

UNIVERSITY OF KWAZULU-NATAL

Fuzzy-based machine learning for predicting narcissistic traits among Twitter users

By

Japheth Kiplang'at Mursi

216068813

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Management, IT and Governance

College of Law and Management Studies

Supervisor: Prof. Prabhakar Rontala Subramaniam

Co-supervisor: Prof. Irene Govender

2022

Declaration

I **Japheth... Kiplang'at ... Mursi**declare that

- (i) The research reported in this dissertation/thesis, except where otherwise indicated, is my original research.
- (ii) This dissertation/thesis has not been submitted for any degree or examination at any other university.
- (iii) This dissertation/thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.
- (iv) This dissertation/thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:
 - a) their words have been re-written but the general information attributed to them has been referenced;
 - b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced.
- (v) Where I have reproduced a publication of which I am an author, co-author or editor, I have indicated in detail which part of the publication was actually written by myself alone and have fully referenced such publications.
- (vi) This dissertation/thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation/thesis and in the References sections.

Signature:



Date: 30/06/2022

Acknowledgements

I would like to express my deepest appreciation to my supervisors, Professor Irene Govender and Professor Prabhakar Rontala, for their guidance and having encouraged and helped me to complete this study.

I'm deeply indebted to my two sets of parents: John Kitur & Jane Kitur, Jackson Yaem & Marion Yaem for their support and prayers throughout this journey.

I'm also extremely grateful to my siblings; Twin Rose Kitur, Carren Kitur, Dancan Mursi, Vibian Chemutai, Angela Chelangat and Bernard Kipngeno for their constant support, encouragement and prayers.

I would also like to extend my deepest gratitude to my wife Jael Kisota for patience, prayers and support during this journey.

I also want thank friends who I met and who contributed in one way or another to this journey. Special mention goes to; Dr. Patrick Wamuyu, Mr & Mrs. Dr. Cherono Kipchumba, Junior Vela, Alex Kasyoki, Odebiri Omasalewa, Winnie Malel, Silas Kipruto, Augustine Mwangi, Abigael Kosgey, Pauline Wambui, Pauline Omollo, Joel Ngetich, Janet Koech and the whole Kenyan Community in Pietermaritzburg.

Dedication

I dedicate this work to

- God for being with me all through.
- My parents, for their continued support and patience.
- My partner, for the support prayers and endless encouragement during this journey.
- My siblings for always believing in me.

Abstract

Social media has provided a platform for people to share views and opinions they identify with or which are significant to them. Similarly, social media enables individuals to express themselves authentically and divulge their personal experiences in a variety of ways. This behaviour, in turn, reflects the user's personality. Social media has in recent times been used to perpetuate various forms of crimes, and a narcissistic personality trait has been linked to violent criminal activities. This negative side effect of social media calls for multiple ways to respond and prevent damage instigated. Eysenck's theory on personality and crime postulated that various forms of crime are caused by a mixture of environmental and neurological causes. This theory suggests certain people are more likely to commit a crime, and personality is the principal factor in criminal behaviour. Twitter is a widely used social media platform for sharing news, opinions, feelings, and emotions by users.

Given that narcissists have an inflated self-view and engage in a variety of strategies aimed at bringing attention to themselves, features unique to Twitter are more appealing to narcissists than those on sites such as Facebook. This study adopted design science research methodology to develop a fuzzy-based machine learning predictive model to identify traces of narcissism from Twitter using data obtained from the activities of a user. Performance evaluation of various classifiers was conducted and an optimal classifier with 95% accuracy was obtained. The research found that the size of the dataset and input variables have an influence on classifier accuracy. In addition, the research developed an updated process model and recommended a research model for narcissism classification.

Keywords: narcissism, personality, social media, design science, fuzzy logic

Journal article submission

1. Mursi, J. K. Subramaniam, P.R. & Govender, I. (2022). Impact of pre-processing techniques in obtaining labelled data using a Twitter dataset. Submitted to International *Journal of Machine Learning and Cybernetics*

Table of Contents

| | |
|---|------|
| Declaration..... | i |
| Acknowledgements..... | ii |
| Dedication..... | iii |
| Abstract..... | iv |
| Journal article submission..... | v |
| Table of Contents..... | vi |
| List of Tables | xv |
| List of Figures | xvi |
| List of acronyms and abbreviations | xvii |
| CHAPTER 1: INTRODUCTION AND BACKGROUND OF THE STUDY | 1 |
| 1.1 Introduction to the study | 1 |
| 1.2 Background | 2 |
| 1.3 Research problem..... | 3 |
| 1.4 Research questions..... | 4 |
| 1.5 Research objectives..... | 5 |
| 1.6 Rationale of the research..... | 5 |
| 1.7 Significance of the study..... | 6 |
| 1.8 Justification of the research | 7 |
| 1.9 Dissertation structure | 8 |
| 1.10 Summary | 9 |
| CHAPTER 2: LITERATURE REVIEW | 10 |
| 2.1 Introduction..... | 10 |
| 2.2 Social media and big data | 10 |

| | |
|--|----|
| 2.2.1 Social network sites | 11 |
| 2.2.2 Categories of social network sites..... | 11 |
| 2.2.3 Impact of social network sites..... | 12 |
| 2.2.4 Twitter as social network site | 13 |
| 2.3 Personality traits..... | 14 |
| 2.3.1 Five-Factor Model (FFM)..... | 15 |
| 2.3.2 The Myers-Briggs Type Indicator (MBTI)..... | 16 |
| 2.3.3 Dark triad personality traits | 17 |
| 2.3.4 Jackson Personality Inventory (JPI)..... | 17 |
| 2.3.5 Eysenck's three-factor model of personality | 17 |
| 2.4 Narcissism and social media..... | 18 |
| 2.4.1 Categories of narcissism | 19 |
| 2.4.2 Narcissism on Facebook | 19 |
| 2.4.3 Narcissism on Twitter | 20 |
| 2.4.4 General impacts of narcissism | 21 |
| 2.4.5 Narcissism and crime | 22 |
| 2.5 Personality assessment approaches..... | 23 |
| 2.5.1 Personality assessment using questionnaires | 23 |
| 2.5.2 Lexicon and open vocabulary approach | 24 |
| 2.5.3 Propensity of word use..... | 25 |
| 2.5.4 Measuring personality processes | 25 |
| 2.6 Data preparation..... | 25 |
| 2.7 Identifying personality traits on social media: A review | 27 |
| 2.7.1 Text-based personality inference | 28 |

| | |
|--|----|
| 2.7.2. Image-based personality inference | 28 |
| 2.8 Text-based annotation for personality identification | 28 |
| 2.9 Levels of sentiment analysis | 29 |
| 2.9.1 Document level | 30 |
| 2.9.2 Sentence Level | 30 |
| 2.9.3 Word/Phrase level | 30 |
| 2.10 Lexicon detection approaches | 31 |
| 2.10.1 Keyword-based detection | 31 |
| 2.10.2 Rule-based construction approach | 32 |
| 2.10.3 Learning-based detection | 32 |
| 2.10.4 Hybrid Method | 33 |
| 2.11 Lexicon detection for narcissism | 34 |
| 2.11.1 Use of swear words | 34 |
| 2.11.2 Social processes and affective processes | 35 |
| 2.11.3: Antisocial Word Use Index | 35 |
| 2.11.4 Self promotion | 35 |
| 2.12 Text classification using artificial intelligence | 36 |
| 2.12.1 Machine learning-based classification | 37 |
| 2.12.2 Fuzzy-based text classification | 38 |
| 2.13 Positioning the research | 40 |
| 2.14 Summary | 41 |
| CHAPTER 3: RESEARCH METHODOLOGY | 43 |
| 3.1 Introduction | 43 |
| 3.2 Types of research methodologies | 43 |

| | |
|---|----|
| 3.3 Design science research | 44 |
| 3.3.1 Choice of design science research methodology | 45 |
| 3.3.2 Relevance of DSRM in this research | 46 |
| 3.3.3 Design research phases | 46 |
| 3.4 Social media methodological frameworks..... | 48 |
| 3.4.1 Social media and forecasting | 48 |
| 3.4.2 CUP framework | 49 |
| 3.4.3 ICUP framework..... | 51 |
| 3.4.4 Iterative CUPP framework..... | 52 |
| 3.5 Process model | 54 |
| 3.5.1 Process model components | 55 |
| 3.6 Data collection | 57 |
| 3.7 Data Analysis | 57 |
| 3.8 Reliability and validity..... | 57 |
| 3.9 Ethical Considerations | 58 |
| 3.10 Summary | 59 |
| CHAPTER 4: RULE-BASED TEXT PRE-PROCESSING | 60 |
| 4.1 Introduction..... | 60 |
| 4.2 Data extraction | 60 |
| 4.2.1 Dataset choice | 61 |
| 4.2.2 Nature and the structure of the dataset..... | 62 |
| 4.3 Data cleaning | 63 |
| 4.3.1 Data cleaning rule and algorithm..... | 64 |
| 4.4 Tokenization and Lemmatization | 66 |

| | |
|--|----|
| 4.4.1 Tokenization rule and procedure | 67 |
| 4.5 Summary | 69 |
| CHAPTER 5: TWEET SENTIMENT ANALYSIS AND DATA LABELLING | 71 |
| 5.1 Introduction..... | 71 |
| 5.2 Sentiment analysis | 71 |
| 5.2.1 TextBlob | 72 |
| 5.2.2 LIWC | 72 |
| 5.2.3 SentiWordNet | 72 |
| 5.2.4 VADER..... | 73 |
| 5.3 Experiment setup: Sentiment analysis | 73 |
| 5.3.1 Sentiment analysis results | 74 |
| 5.4 Experiment setup: Data labelling..... | 76 |
| 5.5 Semi-supervised topic modelling..... | 77 |
| 5.5.1 <i>Latent Dirichlet Allocation</i> (LDA)..... | 77 |
| 5.5.2 Correlation explanation (CorEx)..... | 78 |
| 5.6 Narcissistic lexicon construction | 79 |
| 5.6.1 Seed term collection..... | 80 |
| 5.6.2 Value assignment | 80 |
| 5.6.3 Evaluation of narcissistic lexicon | 81 |
| 5.7 Topic modelling experiment | 81 |
| 5.7.1 LDA topics..... | 81 |
| 5.7.2 Topic coherence | 83 |
| 5.7.3 CorEx topics..... | 83 |
| 5.8 Lexicon detection..... | 86 |

| | |
|--|-----|
| 5.9 Data labelling | 86 |
| 5.10 Summary | 88 |
| CHAPTER 6: MACHINE LEARNING ENSEMBLE CLASSIFIER FOR NARCISSISTIC PERSONALITY PREDICTION..... | 89 |
| 6.1 Introduction..... | 89 |
| 6.2 Text classification tools | 89 |
| 6.3 Classification..... | 90 |
| 6.3.1 Dataset description..... | 90 |
| 6.3.2 Feature extraction..... | 90 |
| 6.3.3 Training and testing | 92 |
| 6.4 Learning classifiers | 93 |
| 6.4.1 Naïve Bayes | 94 |
| 6.4.2 Support-vector machine | 95 |
| 6.4.3 Random Forest | 96 |
| 6.4.4 Voting ensemble classifier | 98 |
| 6.5 Experiment parameters | 99 |
| 6.6 Experiment setup: Classification | 101 |
| 6.6.1 Input and target variables..... | 102 |
| 6.6.2 Vectorization..... | 102 |
| 6.6.3 Data split | 103 |
| 6.6.4 Classifier training: Naïve Bayes | 104 |
| 6.6.5 Support-vector machine | 104 |
| 6.6.6 Random Forest classifier training | 105 |
| 6.6.7 Ensemble 1(En1): SVM and RF | 105 |
| 6.6.8 Ensemble 2 (En2): SVM and Naïve Bayes..... | 106 |

| | |
|--|-----|
| 6.6.9 Ensemble 3 (En3): Naïve Bayes, SVM and RF | 107 |
| 6.6.10 Ensemble 4 (En4): Naïve Bayes and random forest | 107 |
| 6.7 Results and comparison | 108 |
| 6.7.1 Comparison of accuracy and F1-score..... | 108 |
| 6.7.2 Comparison of precision | 109 |
| 6.7.3 Comparison of recall..... | 110 |
| 6.8 Summary | 111 |
| CHAPTER 7: FUZZY-BASED NARCISSISM CLASSIFICATION..... | 112 |
| 7.1 Introduction..... | 112 |
| 7.2 Theory of fuzzy logic system (FLS) | 112 |
| 7.3 Justification of fuzzy-based narcissism classification | 114 |
| 7.4 Application of fuzzy logic classification | 114 |
| 7.4.1 Linguistic variables..... | 115 |
| 7.4.2 Fuzzification | 116 |
| 7.4.3 Membership functions | 117 |
| 7.4.4 Fuzzy logic rules | 117 |
| 7.4.5 Levels of narcissism..... | 118 |
| 7.4.6 Defuzzification..... | 119 |
| 7.5 Experiment setup: Fuzzy logic classification | 120 |
| 7.5.1 Membership functions plot for sentiment polarity..... | 121 |
| 7.5.2 Crisp output levels of narcissism | 122 |
| 7.6 Fuzzy logic system (FLS) experiments..... | 122 |
| 7.7 Enhancing narcissism classification | 128 |
| 7.8 Summary | 130 |

| | |
|---|-----|
| CHAPTER 8: PERFORMANCE EVALUATION OF NARCISSISM CLASSIFICATION..... | 131 |
| 8.1 Introduction..... | 131 |
| 8.2 Performance metrics | 131 |
| 8.2.1 Basic parameters setup..... | 131 |
| 8.2.2 Dataset variables | 132 |
| 8.3 Impact of sentiment analysis on classifier accuracy | 132 |
| 8.4 Classification experiments | 134 |
| 8.4.1 The impact of the number of input variables on classification accuracy..... | 134 |
| 8.4.2 Summary on the impact of input variables on classification accuracy | 139 |
| 8.5 Impact of input variables: Experiment outcome summary | 142 |
| 8.6 Comparison of performance with related research | 144 |
| 8.7 Summary | 146 |
| CHAPTER 9: DISCUSSION AND CONCLUSION | 147 |
| 9.1 Introduction..... | 147 |
| 9.2 Research questions revisited..... | 147 |
| 9.3 Modified process model phases | 148 |
| 9.3.1 Phase1: Rule-based text pre-processing..... | 148 |
| 9.3.2 Phase 2: Data labelling..... | 150 |
| 9.3.3 Phase 3: Machine learning classification | 150 |
| 9.3.4 Phase 4: Fuzzy-based prediction..... | 151 |
| 9.4 Recommended research model | 151 |
| 9.5 Research limitations..... | 151 |
| 9.6 Contribution to knowledge | 152 |
| 9.7 Recommendations..... | 152 |

| | |
|--|-----|
| 9.7.1 Recommendation for Twitter | 152 |
| 9.7.2 Recommendation for users | 153 |
| 9.7.3 Recommendation to the researchers | 153 |
| 9.7.4 Recommendation to governments..... | 153 |
| 9.8 Summary | 154 |
| REFERENCES | 155 |
| APPENDICES | 207 |
| Appendix A: Ethical Clearance letter | 207 |
| Appendix B: Twitter Gatekeeper’s letter (developer account approval) | 209 |

List of Tables

| | |
|--|-----|
| Table 3.1: Research methodologies: Comparison..... | 45 |
| Table 4.1: Dataset attributes | 63 |
| Table 4.2: Pre-processing variables | 65 |
| Table 4.3: Tweet pre-processing for labelling | 66 |
| Table 4.4: Tokenization and lemmatization variables | 67 |
| Table 4.5: Tokenization and Lemmatization output | 69 |
| Table 5.1: Sentiment distribution of the dataset | 75 |
| Table 5.2: LDA topics..... | 82 |
| Table 5.3: Topics and anchor words | 84 |
| Table 5.4: Anchor words and related words | 84 |
| Table 5.5: Topic 1 Keywords..... | 85 |
| Table 5.6: Topic 2 Keywords..... | 85 |
| Table 5.7: Lexicon table | 86 |
| Table 6.1: Dataset characteristics..... | 90 |
| Table 6.2: Classifier variables..... | 93 |
| Table 6.3: Confusion matrix parameters..... | 99 |
| Table 6.4: Classification variables | 102 |
| Table 6.5: Vectorization parameters | 103 |
| Table 6.6: Naïve Bayes confusion matrix | 104 |
| Table 6.7: SVM confusion matrix..... | 105 |
| Table 6.8: Random Forest confusion matrix..... | 105 |
| Table 6.9: Ensemble 1 confusion matrix | 106 |
| Table 6.10: Ensemble 2 confusion matrix | 106 |
| Table 6.11: Ensemble 3 confusion matrix | 107 |
| Table 6.12: Ensemble 4 confusion matrix | 108 |
| Table 7.1: Fuzzy logic system variables | 115 |
| Table 7.2: Input variables | 116 |
| Table 7.3: Output variable | 116 |
| Table 7.4: Fuzzy logic rules variables | 117 |
| Table 7.5: Fuzzy rule table..... | 119 |
| Table 7.6: Fuzzy-based narcissism classification emoticons..... | 129 |
| Table 8.1: Classifier accuracy with respect to sentiment analysis | 133 |
| Table 8.2: Experiment results with four input variables | 135 |
| Table 8.3: Experiment results with three input variables..... | 137 |
| Table 8.4: Experiment results with two (P_t , T_f) input variables | 138 |
| Table 8.5: Experiment results with two (P_t , L_p) input variables..... | 139 |
| Table 8.6: Experimental results of RF and En3 with four input variables..... | 143 |

List of Figures

| | |
|--|-----|
| Figure 2.1: The number of Twitter users worldwide from 2014 to 2020..... | 14 |
| Figure 3.1: Design science research methodology stages..... | 46 |
| Figure 3.2: Social media framework (Schade, 2015)..... | 49 |
| Figure 3.3: CUP framework (Fan & Gordon, 2014)..... | 51 |
| Figure 3.4: ICUP framework (Oh et al., 2015) | 52 |
| Figure 3.5: Iterative CUPP framework (Author) | 54 |
| Figure 3.6: Process model for ICUPP framework (Author) | 56 |
| Figure 4.1: Tokenization and lemmatization flow | 69 |
| Figure 4.2: Rule-base tweet pre-processing | 70 |
| Figure 5.1: Positive polarity word cloud..... | 75 |
| Figure 5.2: Neutral polarity word cloud..... | 75 |
| Figure 5.3: Negative polarity word cloud | 76 |
| Figure 5.4: Topic modelling workflow | 81 |
| Figure 5.5: Topic coherence | 83 |
| Figure 5.6: Data labelling approach..... | 87 |
| Figure 6.1: User categories based on TF..... | 102 |
| Figure 6.2: Vectorization parameters..... | 103 |
| Figure 6.3: Data split (train and test dataset) | 104 |
| Figure 6.4: Accuracy versus F1-score comparison..... | 109 |
| Figure 6.5: Precision comparison..... | 110 |
| Figure 6.6: Recall (sensitivity) performance comparison | 111 |
| Figure 7.1: Block diagram of a general fuzzy logic system..... | 113 |
| Figure 7.2: Fuzzy logic for narcissistic classification..... | 115 |
| Figure 7.3: Sentiment Polarity Membership functions | 121 |
| Figure 7.4: Predicted class membership functions..... | 122 |
| Figure 7.5: Fuzzy-based narcissism classification system | 123 |
| Figure 7.6: Experiment 7.1 output | 124 |
| Figure 7.7: Experiment 7.2 output | 125 |
| Figure 7.8: Experiment 7.3 output | 126 |
| Figure 7.9: Experiment 7.4 output | 127 |
| Figure 7.10: Experiment 7.5 output | 128 |
| Figure 8.1: Classifier accuracy with sentiment as a variable | 134 |
| Figure 8.2: Input variable accuracy comparison | 140 |
| Figure 8.3: Performance accuracy of individual classifiers based on different data sizes | 141 |
| Figure 8.4: Performance accuracy of ensemble classifiers based on different data sizes | 142 |
| Figure 8.5: Performance accuracy comparison with related research..... | 145 |
| Figure 9.1: An updated process model (Author)..... | 149 |
| Figure 9.2: Recommended research model (Author)..... | 151 |

List of acronyms and abbreviations

| | |
|--------|---|
| AR | action research |
| CorEx | correlation explanation |
| CSR | case study research |
| DSR | design science research |
| DSRM | design science research methodology |
| EP | empath personality |
| FFM | Five-Factor Model |
| GN | grandiose narcissism |
| GT | grounded theory |
| JPI | Jackson Personality Inventory |
| LDA | latent Dirichlet allocation |
| LIWC | Linguistic Inquiry and Word Count |
| MBTI | Myers-Briggs Type Indicator |
| ML | machine learning |
| NLP | Natural Language Processing |
| SNSs | social networking sites |
| SR | survey research |
| SVC | support vector classification |
| SVM | support-vector machine |
| TF-IDF | term frequency-inverse document frequency |
| URL | Uniform Resource Locator |
| VEC | voting ensemble classifier |
| VN | vulnerablenarcissism |

CHAPTER 1: INTRODUCTION AND BACKGROUND OF THE STUDY

1.1 Introduction to the study

Social networking sites (SNSs) have become integral channels for communication and self-expression in different people's lives (Alhabash & Ma, 2017). According to Kong, Wang, Zhang, Li, and Sun (2021), SNSs, also referred to as social media, present users with distinct platforms to interact with others to satisfy their self-expression needs. Kuss and Griffiths (2017) asserted that social media has grown in popularity as a medium for obtaining opinions and information about current events. They've become the most frequent place for people to express themselves to others (Ahmed, Ahmad, Ahmad, & Zakaria, 2019). Twitter users post about 500 million messages every day, while Facebook generates 4 petabytes of data (Shu, 2020). The intended purpose of users on social media varies between platforms. These purposes can be to search for knowledge (Tanriverdi & Sağır, 2014), access information (Park & Kim, 2013), maintain communication with friends, and establish professional relationships on platforms like LinkedIn and ResearchGate (Ovadia, 2014).

The increasing use of social network sites has provided unprecedented opportunities for solving problems in various fields with information techniques (Aggarwal, 2011). For example, questionnaires and academic interviews were used for a long time to gather data to predict personality traits. However, most researchers have turned to social media to study personality traits (Wang, Wang, Xu, Wu, & Gia, 2013). The increasing volume of written language on social media provides a huge supply of psychological data with untapped potential (Park et al., 2014). Researchers are currently analysing social media data from the psychological point of view. According to Kosinski, Stillwell, and Graepel (2013), social media platforms like Facebook & Twitter contain a significant amount of autobiographical language and linguistic behaviour correlated to users' psychological traits.

Data mining is the process of extracting raw data to get valuable insights from SNSs (Azeroual, Saake, Abuosba, & Schöpfel, 2018). Text mining is a data mining technique for extracting valuable insights to develop models or identify trends and patterns from unstructured data (Elragal & Haddara, 2014). The primary objective of text mining is to process unstructured (textual) data to obtain meaningful quantitative insights from the text using NLP techniques (Agrawal & Batra,

2013). According to Farnadi et al. (2016), individual personality affects individuals' behaviour and decision-making. This decision-making can be attributed to preferences for websites, products, brands, services, and content such as books and TV shows. By analysing what users post on social media, their personalities can be classified without requiring them to complete lengthy and time-consuming surveys (Lukito, Erwin, Purnama, & Danoekoesoemo, 2016). Even though there are different techniques for analysing personality traits, the most dominant approaches are based on text mining, which implies that critical human personality elements are part of the vocabulary people use to describe themselves (Kulkarni et al., 2018). This research sought to predict personality traits of Twitter users. The research used text mining techniques to detect traces of narcissistic personality from users' behaviour and language-use habits on social network sites.

1.2 Background

The study of personality is regarded as a primary objective of psychology. Personality refers to a set of attributes that make individuals unique (Schwartz, 2020). One of the standard theories used to study personality is the Five Factor Model, which envelopes five essential traits: Agreeableness, Narcissism (Disagreeable Extravert), Conscientiousness, Neuroticism, and Openness to experience. Smith and Canger (2004) stated that the Five-Factor model is essential because it helps classify personality traits. It represents a wide range of personality qualities; each dimension encapsulates a huge number of different and specific personality features (John, Naumann, & Soto, 2008).

Narcissism is characterised by high self-esteem, self-promotion, grandiosity, manipulation, assertiveness, and exhibitionism while at the same time having social withdrawal, low self-esteem, and negative emotionality and rage (Miller et al., 2011). Narcissists are disagreeable extraverts categorised into Vulnerable and Grandiose narcissism. According to Lambe et al. (2018), people with a narcissistic disposition are more likely to be aggressive. Because of the exploitative nature, people with high narcissistic tendencies are also prone to aggressiveness. Furthermore, aggressive, and impulsive people, have a high likelihood of recidivism (Alsheikh & Ahmad, 2020).

Different studies (Hepper, Hart, Meek, Cisek, & Sedikides, 2014; Lowenstein, Purvis, & Rose, 2016) have found a correlation between narcissism and a variety of crimes, showing that understanding or minimising crime requires knowledge of personality traits. Eysenck's theory on personality and crime postulates that different types of crime are triggered by a mixture of

neurological and environmental factors. According to this theory, some people are more likely to commit a criminal act than others, and personality is a key factor in determining criminal behavior (Fakhrzadegan, Gholami-Doon, Shamloo, & Shokouhi Moqhaddam, 2017). According to Kounadi, Ristea, Araujo, and Leitner (2020), little emphasis has been placed on research into and monitoring crime, including incorporating individuals' online behaviour from their digital footprints. Therefore, collection and analysis of crime information remain weak, making it difficult to set a clear crime prevention agenda and establish a clear plan for crime prevention (Jean-Claude, 2014). Traditional media may not be as accurate as social media in describing the crime in a country or city (Curiel, Cresci, Muntean, & Bishop, 2020). Perpetrators, victims, witnesses, or indirect victims may be more willing to share their feelings after witnessing a crime regardless of how minor the incident was (Cresci, Cimino, Avvenuti, Tesconi, & Dell'Orletta, 2018).

Advances in computational technology have assisted researchers in representing personality behaviours digitally and objectively. In addition, it is now possible to collect, store, and analyse data efficiently using computational techniques (Chang, Kauffman, & Kwon, 2014). This use of computational methods helps overcome own human limitations and biases in theory development. The common limitations include biased self-reports and the inability to simultaneously consider the influence of many factors (Jolly & Chang, 2019). Hence, more substantial reliance on computational methods will help advance personality science and create novel, more holistic forms of personality assessment (Stachl et al., 2020). Therefore, this research used computational techniques and methods to develop a predictive model to identify traces of narcissistic behaviours from Twitter. The model would aid law enforcement agencies in integrating personality traits in preventing crime by monitoring narcissists and other prospective perpetrators based on their social media behaviors. Furthermore, this research raises awareness of the negative consequences of a narcissistic personality's possible interaction with innocent social media users.

1.3 Research problem

Social media has transformed the way individuals interact with one another and consequently affected people's ability to empathise negatively and positively (Sankaran, 2019). While social media network sites offer a range of advantages to users, concerns have been expressed about the potential harmful offline implications of exposure to online content and interactions with strangers (Müller et al., 2016). According to Won, Steinert-Threlkeld, and Joo (2017), young people who

read violent content online were more likely to commit severe crimes and had a higher risk of copycat violence. In the development of 'individual' behaviors, personality plays a critical role (Wrzus & Roberts, 2017). Individual characters continue to play a role in deciding how individuals behave on social media. As a vital personality factor correlated with aggression among the youth (Lau & Marsee, 2013), narcissism is significantly associated with various offending behaviours (Fan, Chu, Zhang, & Zhou, 2019). Lobbestael, Baumeister, Fiebig, and Eckel (2014), Maynard, Vaughn, Salas, and Wright (2016), and Li et al. (2015) have highlighted the correlation between narcissism and aggression. Aggression from narcissists stems from the desire to control and manipulate others to attain their desired goals, as well as their exploitative and un-empathetic nature (Lobbestael et al., 2014).

The rapidly evolving nature of digital crime and its influence on physical crimes has prompted law enforcement personnel to add new ways to prevent and curb crime (Norden, 2013, Goodison, Davis, & Jackson, 2015). As they seek as many friends and influence as possible, narcissists thrive in an unmonitored atmosphere provided by social media (Buffardi & Campbell, 2008). As a result, the importance of social media injustice and the criminal justice system can no longer be overlooked. Branley and Covey (2017) assert that users' online behaviour influences their offline decisions and actions. The capacity of law enforcement agencies in dealing with and minimizing harm caused by social media is inadequate. Therefore, understanding or capping crime must incorporate knowledge of personality traits. Whenever there has been use of personality traits in crime prevention, they have not been clearly applied or expressed and thus the need to determine the aspect of personality traits by examining 'users' behaviour on social media (DeVito, Birnholtz, Hancock, French, & Liu, 2018).

1.4 Research questions

The main research question formed to address the problems identified in this research was: *How can traces of narcissistic personality traits be identified among social media users?*

The following research questions contributed to the solution for the main research question:

- i. How can Twitter dataset be prepared for sentiment analysis?

- ii. How can the pre-processed Twitter dataset be labelled into different categories of narcissism?
- iii. How can traces of narcissism be classified using the labelled Twitter dataset?
- iv. How can the classification of narcissism be enhanced?

1.5 Research objectives

The main objective of this research was to identify traces of narcissistic personality traits from social media users using the Twitter dataset. The following were the specific objectives:

- i. To prepare the Twitter dataset for sentiment analysis
- ii. To devise a labelling methodology for the pre-processed Twitter dataset
- iii. To design a classification technique to predict traces of narcissism
- iv. To enhance the classification technique for a better performance.

1.6 Rationale of the research

According to Majid (2012), individuals make themselves visible to potential offenders through their comprehensive and public self-presentation on social network sites. Furthermore, providing personal information (such as information about intimate connections, views and interests) provides potential criminals with a multitude of opportunities to exploit. Because of the two-way nature of communication on social networking sites, individuals become increasingly accessible to those who might wish to victimise them (Majid, 2012).

Given the relationship between narcissism and specific behavioural manifestations on social media, the inadequacy of solely using self-report measures of narcissism, and the value of observing behaviour, research is warranted to support further the efficacy of predicting narcissistic behaviour in everyday settings (Roberts, Woodman, & Sedikides, 2018). Many studies on narcissism rely on the self-report methods, which provide valuable information but do not necessarily predict how a narcissistic individual will act in a given situation (Meagher, Leman, Bias, Latendresse, & Rowatt, 2015). An overreliance on self-report and neglect of behavioural studies creates a gap in research that needs to be filled.

Narcissists' negative emotions can affect the environment one is in (Zajenkowski & Szymaniak, 2021). Thus, it is essential to recognise the psychological components of personality trait-related crime to raise awareness of the harmful consequences of interacting with narcissists. Law enforcement officials can then utilise personality trait prediction to safeguard social media users from narcissists. As a result, knowledge of personality is required to comprehend and handle crime. Therefore, it is necessary to take into account contribution of personality traits to certain criminal behaviors. Rising crime rates globally, coupled with perpetrations of crime from social media platforms, was a primary motivation for this study. Therefore, the main objective of this research was to extract usable, credible information to identify traces of Narcissism from Twitter and assist law enforcement with future crime prevention, thereby contributing to ensuring law enforcement work.

1.7 Significance of the study

People with narcissistic tendencies can use social media to market themselves and satisfy their attention and adoration. As a result, it's critical to identify those who exhibit high levels of narcissism so that strategies and interventions can be developed to mitigate the adverse effects of social media and make it more positive. Personality plays a crucial role in someone's orientation. This work on narcissism prediction is particularly interesting for social network site users who often interact with narcissists. The proposed method effectively identifies traces of narcissism, thus protecting social media users.

Furthermore, according to Reed, Bircek, Osborne, Viganò, and Truzoli (2018), people with high levels of narcissism tend to use Twitter more and more over time. In addition, this study enhances existing literature on personality prediction and its relationship with providing solutions to current societal challenges. Social media instigated crime has been on the rise. Therefore, this work is significant for the following reasons: According to Oltmanns and Widiger (2018), prior research has not been successful in adequately addressing both grandiose and vulnerable narcissism. This research has explored and captured the two facets of narcissism and sought to identify its presence on social media. Lastly, this research will also benefit social media users as it exposes the characteristics of narcissists on social media and how to identify them.

1.8 Justification of the research

For much of its history, personality research psychology has relied on various forms of assessment, including surveys, questionnaires, diagnostic tests, behavioural observation, and interviews which have been time-consuming and expensive (Kosinski, Bachrach, Kohli, Stillwell, & Graepel, 2014). However, these assessments have various limitations. First, according to Van Vaerenbergh and Thomas (2013), the assessments have non-standard response styles and memory limitations (Van Vaerenbergh & Thomas, 2013). However, social media presents a platform with digital footprints that can be used to identify personality traits easily. The platform is cost-effective compared to surveys and reaches a larger population (Azucar, Marengo, & Settanni, 2018). Secondly, social media posts capture communication among friends and acquaintances generated in a natural social setting and thus capture interaction among friends and acquaintances. Thirdly, social media users provide vast amounts of personal information about themselves; it is a common topic of conversation. Finally, social media users usually present themselves as they are, not as idealised representations of themselves (Back et al., 2010).

With widespread social network sites nowadays, Twitter has emerged as among the widely used social media platforms globally. It is a global information network where users make short posts called tweets. A tweet is a short 280-character message (Fearnley & Fyfe, 2018). In recent years, Twitter has proven to be an effective resource for identifying societal interests and general opinions of people. Twitter has been dubbed as electronic word-of-mouth marketing platform (eWOM) (Zhou, Tao, Rahman, & Zhang, 2017). Therefore, Twitter provides an ideal platform for personality research and application.

According to Kaye, Malone, and Wall (2017), emojis display emotional and social meanings and reduce the ambiguity of the message. Emojis also provide contextualisation cues, such as positive or negative attitudes and the organisational role of social relationships. Given the widespread use of emojis in everyday communication, it is vital to consider their adoption to create a safe social media environment. This research has recommended 5 level emoji icons for labelling the levels of narcissism based on digital footprints on Twitter. Through this research, a user can be forewarned on the type of tweet in terms of narcissism thus protecting them from further engagement. Due to the fact that fuzzy rule-based models require identifying sentiments through weighted voting at the

defuzzification stage, judgment bias on both positive and negative tweets can be effectively eliminated in the big data era.

1.9 Dissertation structure

This dissertation consists of nine chapters:

Chapter 1 establishes the problem by presenting the research background, research objectives, research questions, rationale, and the significance of the study. This facilitates to understand the relevance of the research.

Chapter 2 reviews studies on big data, social media, personality traits, crimes in social media and gaps in the literature. The chapter also reviews the existing machine learning algorithms.

Chapter 3 discusses the design science research methodology (DSRM) approach adopted in the study. The DSRM was used to design the artefact (Narcissistic prediction model). In addition, the tools and techniques used in the study are discussed in detail.

Chapter 4 discusses the methodology used in extracting information from social networking sites which form the input for machine learning classifier.

Chapter 5 presents sentiment analysis done on the processed dataset and the process of labelling the dataset as narcissistic or non-narcissistic. The techniques adopted to label the data comprising sentiment analysis, topic modelling, and word detection are discussed in this chapter.

Chapter 6 presents the process of classifying narcissistic personality traits using machine learning techniques. To find the optimal classifier, the researcher examined three machine learning algorithms: random forest (RF), Nave Bayes, and support-vector machine (SVM), as well as four ensemble classifiers.

Chapter 7 presents how the classification of narcissism is improved using fuzzy logic.

Chapter 8 presents the implementations of experiments undertaken to evaluate the model's performance under different types of datasets and attributes. In addition, the results are compared with existing techniques.

Chapter 9 concludes the thesis on this research and summarises the results. A modified process model is also recommended in this chapter. Finally, the chapter recommends possibilities for future work.

1.10 Summary

The use of social media for interaction and communication is highly effective. People can reach a specific target population and influence others through tweets and posts on social media. Millions of people write, post, and share information on social media to express their instant thoughts, emotions, and beliefs on public platforms. Evidence also implies that user generated content on social media reflects the users' true personalities. Scholars and media have asserted that the success of social networking sites is linked to the narcissism of their users and that social networking behaviour reflects narcissistic tendencies (Fishwick, 2016; Buffardi & Campbell, 2008). The next chapter reviews related literature on personality prediction and social media.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

Social media has emerged as an ideal platform for people to easily interact and share opinions with a broad community. Zivanovic, Martinez, & Verplanke (2020) believe that, in contrast to traditional methods of acquiring information about beliefs and behaviour, it is possible to gain insight from people's posts on Twitter. Consequently, Kosinski et al. (2014) posited that it is possible to infer personality traits from social media using machine learning techniques. Technology is rapidly evolving, and everyone, intentionally or otherwise, is producing data in a way. One of the large generators of big data is social media (Gandomi & Haider, 2015). Social media, phone, and server log data are being produced at an unprecedented rate (Rawat & Yadav, 2020). This chapter seeks to give a clear understanding of personality traits identification on social media and how social media plays a crucial role in contributing to big data through appropriate references in literature. It further discusses existing personality traits and expounds on narcissism and its presence on social media.

2.2 Social media and big data

Butler and Matook (2015) defined social media as a broad set of tools and apps that allow individuals to communicate and share information. Furthermore, Kim and Hastak (2018) asserted that social media has rapidly grown as a popular source of information. The popularity of social media continues to soar, resulting in the emergence of new social networks, forums, and blogs (Gundecha & Liu, 2012). The increase in social media use and resultant increase in user generated content has led to the emergence of opportunities to extract and analyse this data (Stieglitz, Mirbabaie, Ross, & Neuberge, 2018). When a user posts on social media, they provide a glimpse into their lives (Blair, Bi, & Mulvenna, 2020). A study of Twitter usage in the London Underground analysed users' tweets at different times of the day. The study cross-referenced the results with their geotagging feature to determine what, where and when users were tweeting (Lansley & Longley, 2016). The findings prompted the researchers to create recommendation algorithms for the types of advertisements displayed on each station's rotating digital billboards at different times of day to enhance their efficiency (Lansley & Longley, 2016).

Big data is characterised by the five dimensions commonly referred to as 5Vs: The first V is “Volume” which relates to the amount of the data generated and collected (Kambatla, Kollias, Kumar, & Grama, 2014). The second V is “Variety” and relates to variations of data (Gandomi & Haider, 2015). The third V is “Velocity” and is the speed at which data is generated and how it should be analysed. The fourth dimension is ‘Variability,’ which refers to the variation in data flow rates. The last dimension is ‘Value,’ which describes big data as having low value relative to its volume in its original form. It is through analysis that high value can be obtained (Gandomi & Haider, 2015).

Social media data are massive, noisy, scattered, unstructured, and constantly changing (Gundecha & Liu, 2012). The vast amounts of user-generated content on social media can be mined and analysed to understand social norms and model user behaviour. One of the approaches to gaining insights from social media big data is data mining. Data mining has three major stages: data pre-processing, pattern discovery (identifying patterns of the targeted data), and pattern evaluation and presentation (Taleb, Dssouli, & Serhani, 2015).

2.2.1 Social network sites

According to Ahmed et al. (2019), social media permits people to share negative or positive thoughts on various topics. Social network sites and social media have been used interchangeably. Facebook, Twitter, Instagram, LinkedIn, Blogs, Wikis, and YouTube are examples of social networking sites that generate vast unstructured data. Because each of these platforms serves a different purpose and has a different audience, it is typical for users to sign up for multiple platforms (Langstedt & Hunt, 2017). These sites allow users to create profiles/online social identities and choose users to share connections. Usage of various SNSs for different purposes has continued to increase globally because of their value in aiding human communication (Yen, Lin, Wang, Shih, & Cheng, 2019). SNS allows the emergence of a different type of social structure that influences its members in one way or another through social relations within the platform (Martínez & De Frutos, 2018).

2.2.2 Categories of social network sites

According to Kemp (2018), there are over three billion social network users. In addition, Aboulhosn (2020) noted that 500 million tweets were sent every day as of 2019, and 695,000 status

updates have been posted on Facebook as of 2020. Although social networking sites generally enable people to connect with each other online, not all of them offer the same services or have the same objectives. (Kwak, Lee, Park, & Moon, 2010). Social network sites can be categorised into three main categories based on usage: social sites, academic networks, and professional sites.

Social sites are platforms used for connection and interaction by different individuals from different spectrums. Facebook, Twitter, and Instagram are the most widely used social networks. Facebook, which has 1.79 billion active members, is a popular social site platform that allows users to create profiles and communicate their emotions through posts, images, and videos. Twitter has 330 million monthly active users, and users (tweeps) can only tweet 280 characters. On the other hand, Instagram has 500 million users and allows users to submit photographs and videos. Users can link their accounts on these three major social media platforms (Lipschultz, 2017).

Academic networks sites connect researchers and academics from various fields to exchange ideas on social media. ResearchGate and Academia.edu are the popular academic network sites. Academics can upload information, read publications from other researchers (Meishar-Tal & Pieterse, 2016) and share their works, thus improving scholarly communication.

Professional sites are sites where professionals meet and discuss their careers and business interests. The main professional site is LinkedIn which seeks to help people network professionally by showing connections to individuals based on their related connections, which can be the same employer, schooling, and same career field (McCabe, 2017).

2.2.3 Impact of social network sites

Social media has increased the rate of collaboration amongst people and as a result, people's behaviours and lifestyles have been transforming (Akram & Kumar, 2017). Social media popularity has also led to the emergence of new careers like influence marketing. This has helped businesses acquire new customers and reach new markets. In addition, according to Aljuboori, Fashakh, and Bayat (2020), social media has provided an opportunity for writers to connect with their clients while also uniting people on a huge platform for the achievement of specific goals.

Although social network sites benefits are well recognised, it's sometimes easier to accept or overlook their drawbacks. One of the limitations is that criminals have taken advantage of these

unregulated, freely accessible communication channels to lure users and perpetrate different forms of crimes through social networks of their own interests (Fox & Moreland, 2015). Despite its significance in enhancing social interactions and growth, social media has created a variety of less suitable impacts, such as online fraud, online child grooming, and online radicalisation (Baccarella, Wagner, & Kietzmann, 2018). Various researchers have focused on multiple risks associated with SNS use. Some of these studies have dwelt on cyberbullying, children grooming, and damage to reputation (Ellison & Boyd, 2013). The increased use of SNS has led to vulnerability to various risks while also providing immense opportunities for users and companies (Metzger, Wilson, & Zhao, 2018). More importantly, Alloway, Runac, Quershi, and Kemp (2014) maintained that social networking sites have significantly changed social relationships. Most social media users frequently accept friends' requests from people unknown to them, which provides a gateway to personal information, putting users' privacy and friends at risk (Ryan, 2008).

2.2.4 Twitter as social network site

Twitter has 335 million monthly active users (Statista, 2018). A microblog entry is called a tweet on Twitter, and it has a maximum character limit of 280. Twitter is a popular platform to broadcast news, current events, beliefs, and user behaviours, and with a character restriction of 280 characters, makes it useful for social monitoring. This rich user-generated data can be used to make sense of public opinions on contemporary issues on social media and on personality studies (Kursuncu et al., 2019). A tweet has various attributes. The first attribute is *retweet* (RT), a tweet reposted by another Twitter user to their followers. The second attribute is *favorite* where a user shows the creator of the tweets that they liked their tweets (Alshehri, 2019). The third attribute is *follower*. Accounts that subscribe to a Twitter user's updates and postings are the user's followers. Twitter users who follow another user show that they wish to keep up with what the user posts (Alshehri, 2019).

Everyone can see how many followers a user has. Twitter users can choose to follow other Twitter users. A user can decide whether they want their tweet to be public (visible for all) or private (visible only to followers) (Marshall, 2018). There are two kinds of data from Twitter. The first is historical data and the second is streaming current data. These kinds of data can also be obtained either by registering as a Twitter developer account (done in this study) and completing the authentication process or purchasing from commercial organisations that have partnered with

Twitter (Hino & Fahey, 2019). Communication on Twitter happens through tweeting, retweeting, and leaving comments. Many users also include hashtags in their tweets (Lowe & Laffey, 2011). A hashtag begins with the # (hash symbol), followed by a word or a merged sentence. One way of finding new users to support is to search for specific hashtags (Shapp, 2014). The popular hashtags globally include #Blacklivesmatter #Bringbackourgirls, #Metoo, #earthday, and #feesmustfall. These hashtags usually refer to the person posting the tweet as a pro hashtag. Figure 2.1 below illustrates the Twitter users between 2014 and 2020.

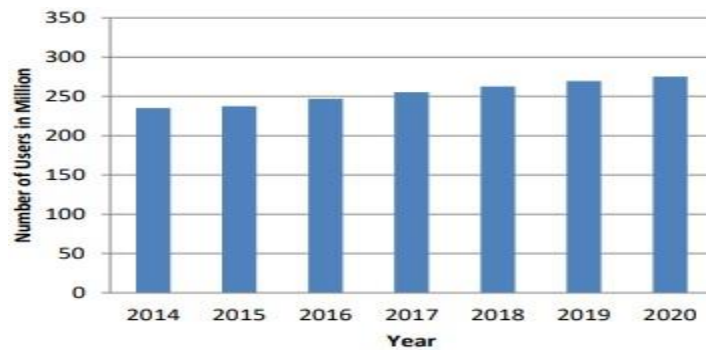


Figure 2.1: The number of Twitter users worldwide from 2014 to 2020

2.3 Personality traits

According to Kulkarni et al. (2018), language remains a fundamental construct in psychology as it enables people to express their inner thoughts and feelings in a way others can comprehend. The development and maintenance of addictive behaviours are all influenced by one's personality (Chung, Morshidi, Yoong, & Thian, 2019). Individual personalities continue to have a role in influencing people's online activities. According to Funder (2012), personality traits are persistent patterns of emotion, thoughts, and behaviours. These comprise all the traits, attributes, and differences that distinguish an individual from all the others. Occasionally, these digital cues are used to shape an impression of someone, especially if they are a stranger or a zero acquaintance (Hinds & Joinson, 2019). Modelling human behaviour has become more feasible owing to recent developments in adopting data-based techniques to social sciences. Because of the quantity of textual data, understanding human behaviour through analysing unstructured social media data has gained much attention (Davahli et al., 2020). Different models of assessing personality traits in psychology exist. These are), Big Five model of personality, the Myers-Briggs Type Indicator, the

dark triad personality traits, the Jackson Personality Inventory (JPI), and Eysenck's three-factor model of personality. There is still no consensus on which model is the best (Davahli et al., 2020).

2.3.1 Five-Factor Model (FFM)

Researchers have employed the Five-Factor Model (FFM) to investigate personality traits on social media (DeYoung, Quilty, & Peterson, 2007). The Big Five model represents the variation in human behaviour and preferences and a framework that integrates research findings in the psychology of individual differences (Kosinski et al., 2014). A growing body of literature indicates that the FFM's five personality variables can effectively capture essential characteristics of various behavioural patterns (Heine & Buchtel, 2009). The Big-Five comprise of: neuroticism, narcissism (disagreeable extravert), openness, conscientiousness, and agreeableness. Gilpin et al. (2018) used supervised learning to predict Big Five personality traits by speech signals. The research examined how individuals look and sound and how they affect an individual's unconscious communication behaviour. The researchers used 640 speech corpora and 11 Big Five assessments to train the classifier. The prediction models were evaluated with 15 speech records, labelled with the same Big Five inventory. It resulted in the accuracy of Agreeableness, 90.78% of Conscientiousness, 77.66% of Emotional Stability, 70.15% for Extraversion, 66.72%, and 78.98% of Intellect/Imagination. Ross et al. (2009) studied university students' big five personalities and their behaviours when using Facebook. According to their findings, there was a partial association between individuals' personalities and their conduct on social media.

Neuroticism is the first trait of FFM, and it is marked by emotional instability and rage. It's also linked to the frequency with which people utilise social media for socialising, emotional disclosure, and expressing personal problems (Seidman, 2013). The second trait in FFM is narcissism and is characterised by having a high degree of extraversion, being self-promoting in nature, and exhibiting extreme selfishness and admiration. In addition, narcissism is also categorised as part of the Dark triad personal, which includes psychopathy and Machiavellianism. In the three subcategories, narcissism is classified as a personality trait characterised by exploitativeness, aggression, entitlement, and self-focus. In addition, psychopathy involves erratic lifestyles, and impulsiveness. Subsequently, Machiavellianism consists of manipulation, a lack of empathy, and multi-dimensional behaviour (Davahli et al., 2020).

Openness to experience is the third trait in FFM and describes sensitive and tolerable people who use social media to share information but not socialisation (McElroy, Hendrickson, Townsend & DeMarie, 2007). They embrace change without resistance and value new ideas tremendously (Hughes, Rowe, Batey, & Lee, 2012). The fourth trait in the Five-Factor Model is conscientiousness and refers to individual ability to control behaviour in pursuit of goals (Seidman, 2013). Conscientiousness refers to the tendency to follow the rules, and resistance to immediate gratification in the interest of longer-term goals. The fifth trait in FFM is agreeableness which relates to people that tend to be friendly, appreciative, and sympathetic. Individuals with a high agreeableness score are sociable, kind, and courteous, and they use Facebook for social connection rather than self-promotion (Seidman, 2013).

2.3.2 The Myers-Briggs Type Indicator (MBTI)

Myers-Briggs Type Indicator (MBTI) is another personality assessment model. The test assigns a human personality type to one of 16 major categories that define personality in terms of its fundamental nature and preferences (Ahmad & Siddique, 2017). There are four dimensions to the MBTI personality theory. These are; sensing vs. intuition, judging vs. perceiving, thinking vs. feeling, and introversion vs. extraversion (Stein & Swan, 2019). Plank and Hovy (2015) used MBTI to identify correlations between personality traits and demographic and linguistic features. The data used in the research was collected from 1200 Twitter user profiles. Each of them had been previously annotated by its owner with an MBTI personality type. The authors tracked posts that mentioned any of the 16 categories associated with Briggs or Myers' words. They gathered between 100 and 2,000 of their most recent tweets from 1,500 different people (Lima & De Castro, 2019). They analysed the attributes in each dimension using logistic regression. The study concluded that the data would provide enough linguistic evidence to accurately predict the dimensions: Feelings/Thinking and Introversion/Extroversion (Lima & de Castro, 2019).

Alsadhan and Skillicorn (2017) devised a method for predicting the Big Five and Myers-Brigg's personality types from text based on word counts. The proposed method did not need special lexicons and has been successfully applied to various languages. Pramodh and Vijayalata (2016) predicted their Big Five personality traits through the authors' articles and writings. They extracted two datasets: one with positive terms and the other with negative phrases corresponding to each of the Big Five personality traits. The authors then conducted pre-processing, which included

tokenization stopwords removal, stemming, scaling, and scoring on the data to predict personality (Pramodh & Vijayalata, 2016).

2.3.3 Dark triad personality traits

Dark triad personality traits are a cluster of three psychological traits at the subclinical level, namely Narcissism, Machiavellianism, and Psychopathy (Kraus, Berchtold, Palmer, & Filser, 2018). Narcissism is associated with exploitative entitlement, dominance, grandiosity, and feelings of superiority (Wright, 2016). Psychopathy comprises individual traits such as aggression, erratic lifestyle, anti-social behaviour, and impulsiveness. The last category is Machiavellianism, which encompasses a lack of compassion, manipulative, and combative and multi-faceted behaviour (Davahli et al., 2020). The Dark Triad personality traits model is widely used in studies which explore negative and undesirable behaviours (Nowak et al., 2020). They've been linked to exposing others in danger for selfish gain (Jones, 2013), as well as a quest for dominance and vanity (Lee et al., 2013), prejudice against outgroups (Hodson, Hogg, & MacInnis, 2009) and antisocial behaviour such as that of bullies (Baughman, Dearing, Giammarco, & Vernon, 2012). Moreover, Carter, Campbell, and Muncer (2014) asserted that the extraverted behaviour of making a good first impression, such as a willingness to socialise and chat with friends, is universal across all three traits.

2.3.4 Jackson Personality Inventory (JPI)

The Jackson Personality Inventory (JPI-R) evaluates personality factors that are significant to an individual's functioning in a range of circumstances (Jackson, 1994). The settings may range from work-related situations to those involving educational/organisational behaviour (Jackson, 1994). The JPI has 15 different scales and contains 300 true-false statements. According to Jackson (1994), the instrument is suitable for career and vocational tests in schools and work settings.

2.3.5 Eysenck's three-factor model of personality

Extraversion (E), Psychoticism (P), and neuroticism (N) are the three key traits in Eysenck's three factor personality model (N). These dimensions represent fundamental, self-contained, and physiologically-based characteristics. They have various degrees of characterising all subjects and allow for the successful description of behavioural, emotional, and individual variations among adults and children (Colledani, Anselmi, & Robusto, 2018). Neuroticism specifies a trait related to

emotional stability and determines how often a person is to experience negative emotions (Eysenck, 1991). People with a high level of neuroticism are known to be worried, moody and irritable (Eysenck & Barrett, 2013). Extraversion is the model's second dimension, and it describes people who are outgoing, carefree, friendly, convivial, easy-going, and spontaneous. Introversion, on the other hand, characterises those who are contemplative, quiet, serious, and restrained (Eysenck & Barrett, 2013). The third dimension is psychoticism. It describes someone who is confrontational, aggressive, untrustworthy, cold, unemotional, unpleasant, devoid of human feelings, and unfriendly (Mor, 2020).

2.4 Narcissism and social media

A narcissistic personality is described by the Diagnostic and Statistical Manual of Mental Disorders-IV as having a pervasive pattern of grandiosity, the need for attention, and a lack of empathy that start early in life and appear in a range of circumstances (Sumner, Byers, Boochever, & Park, 2012). In addition, it is also one of the personality's dark triads are labelled exploitative and display ignorance to the damage they bring to others to achieve their goals (Stone, Segal, & Krus, 2020). Paulhus and Jones (2015) found that people with Dark Triad characteristics are predisposed to violence, but psychopaths appear violent even if they are not provoked. Furthermore, narcissistic aggression was observed to be a more predictable reaction to threats to one's ego and self-image (Goodboy & Martin, 2015). Goodboy and Martin (2015) further noted that Machiavellist's engage in cyberbullying to gain something. Cyberbullying is used by narcissistic people as a kind of retaliation for image restoration, while psychopathic people may cyberbully without provocation (Goodboy & Martin 2015)

In their study, Błachnio and Przepiórka (2018) showed that high levels of Fear of Missing out (FOMO) and narcissism are predictors of intrusive behaviours on Facebook, while a low level of FOMO and high levels of narcissism are related to life satisfaction. According to a study by Ryan and Xenos from 2011, those who use social media are more likely to be extroverted and narcissistic. This study sought to identify traces of narcissism from social media building on these two studies. The study attempted to identify traces of narcissism from Twitter using machine learning techniques. Brailovskaia, Bierhoff, Rohmann, Raeder and Margraf (2020) noted that social media remains appealing to narcissistic individuals because it specifically satisfies the narcissistic person's need to engage in superficial and self-promoting behaviours. This was also

confirmed by the study of Grieve, March, and Watkinson (2020), in which grandiose narcissism predicted a high congruence between the authentic self and that presented on social media, while for vulnerable narcissism, there was a large discrepancy between the authentic self and the one presented online.

2.4.1 Categories of narcissism

Narcissism is divided into two spectrums, namely vulnerable narcissism (VN) and Grandiose Narcissism (GN). Grandiose narcissism refers to traits related to aggression, grandiosity, and dominance. Vulnerable narcissism relates to defensive and insecure grandiosity (Cheiffetz, 2017; Miller, Lynam, Hyatt, & Campbell, 2017). Individuals with high grandiose narcissism promote false beliefs about themselves, repressing knowledge at the same time that is incompatible with an exaggerated self-image (Zajenkowski, Maciantowicz, Szymaniak, & Urban, 2018). Subsequently, vulnerable narcissism is correlated with psychological distress, depression, negative emotions, and feelings of inferiority. Furthermore, vulnerable narcissism is related to distrust, and hostility, which constitutes narcissistic rage (Jauk, 2019).

While almost every user on social media shares information, the type of information exchanged is not identical and is dictated by the different users' personality traits (Hruska & Maresova, 2020). Various studies (Somerville, 2015, Fan, Chu, Zhang, & Zhou, 2019, Jones, Woodman, Barlow, & Roberts, 2017) have demonstrated that social media use contributes to the rising level of narcissism and affects users' self-esteem in detrimental ways. SNSs allow narcissists to participate in exhibitionism and attention-seeking activities that help them maintain their grandiose self-images (Wang, 2017). Since they allow users to benefit from a varied range of loose or "weak tie" relationships, SNSs are a great platform for narcissists to advance their agenda (Blinkhorn, Lyons, & Almond, 2016).

2.4.2 Narcissism on Facebook

One of the most popular elements on social network sites is status updates. Status updates allow users to express their ideas, feelings, and actions on social media contacts. A study of 555 Facebook users by Marshall, Lefringhausen, and Ferenczi's (2015) using the Big Five personality traits model discovered that narcissists' predisposition to update their achievements regularly explains why their posts receive so many likes and comments (Marshall et al., 2015). Narcissists

are more likely to be 'active' SNS users and their activities are oriented at presenting oneself favourably to others (Marshall, Lefringhausen, & Ferenczi, 2015). McCain and Campbell (2018) did a meta-analysis to establish the correlation between social media use and grandiose narcissism. They investigated 62 studies and discovered that there exists a relationship between narcissism and social media use. The findings revealed that narcissism was linked to more social media postings, time spent and having more social media followers. The more narcissistic people use social media, the more positive feedback they receive from their online connections, such as nice comments and likes, which boosts their self-esteem and makes them feel famous and respected (Baccarella, et al., 2018). SNSs, on the other hand, are usually just one of many sources of positive feedback for people who have high degrees of grandiose narcissism.

2.4.3 Narcissism on Twitter

Marshall, Ferenczi, Lefringhausen, Hill and Deng (2020) found that narcissism was more closely linked to numerous motivations for using Twitter and tweeting more topics than any other personality. The usage of Twitter for attention-seeking, social connection, career promotion, and the frequency of tweeting about diet/exercise were all positively associated with narcissists. According to Garcia and Sikström (2014), narcissists may use social media to build social capital if it enables them to take advantage of others. If the number of likes and retweets one's tweets generally receive is a measure of social capital and reward, then narcissists' self-promotion method pays off (Marshall et al., 2020).

Furthermore, Williams, Burnap and Sloan (2017) found that tweets containing keywords related to damaged windows were linked to reported crime rates. According to this research and the expanding usage of social media, there is a correlation between drug-related tweets and crime, which suggests that social media could be used to predict and monitor crime (Wang, Yu, Liu, & Young, 2019). Beiji, Mohammed, Chengzhang, and Rongchang (2016) predicted crime hot spots by analysing Twitter and observed that crime occurs in clusters. They proposed an approach that consisted of three phases: analysis of videos, prediction of crime, and crime mapping. Wang, Gerber and Brown (2012) used social media to predict hit-and-run incidents. The authors used semantic analysis, latent Dirichlet allocation (LDA), and a dimensionality reduction technique for model construction. The test is carried out using a real-world dataset. The proposed model

outperformed the baseline classifier and produced a more accurate receiver operating characteristic (ROC) curve than the baseline method.

2.4.4 General impacts of narcissism

Narcissism as a personality is beneficial when it is connected with leadership traits of dominance, extraversion, confidence, and power which all have a fundamental connection to narcissism. Because of these similarities, narcissistic people may be more likely to lead (Fatfouta, 2019). Furthermore, Nevicka, De Hoogh, Van Vianen, Beersma, and McIlwain (2011) observed that people higher on narcissistic traits emerged as leaders in team tasks. This is owing to their overwhelming desire for recognition, power, and the opportunity to demonstrate how capable they believe they are. Narcissists are driven to and thrive in high-profile leadership roles (Leary & Ashman, 2018).

Whereas narcissism's ability to emerge as a leader can be a positive trait, it can also be detrimental to society (Hudson, 2012). This is due to the harmful elements of narcissism, such as exploitative tendencies, inability to tolerate criticism, and arrogance (Braun, 2017). According to Hudson (2012), followers, on the other hand, tend to see narcissistic people as good leaders, even though they may not be outstanding performers due to incompetence. Furthermore, a narcissistic personality's great self-confidence might be problematic since narcissistic leaders may believe that their followers have nothing to offer the organisation (Braun, 2017).

In addition to having the propensity to display poor leadership, narcissistic persons are more likely to be guilt-free than non-narcissistic people, which leads to a high rate of immoral behaviour (Brunell, Staats, Barden, & Hupp, 2011). They also have a propensity for taking risks (Lakey, Rose, Campbell, & Goodie, 2008). These elements contribute to the emergence of addictive behaviors including drug use, excessive gambling, obsessive shopping, and other negative behaviors like criminal behavior. Studies examining the relationship between narcissism and addictive behaviours have discovered that narcissism is related to increased alcohol use in first-year college students and is significantly associated with binge drinking (Luhtanen & Crocker, 2005). The following section discusses negative impacts with a focus on crime.

2.4.5 Narcissism and crime

According to Hipp, Bates, Lichman, and Smyth, (2019), the use of social media to track criminal activity is a rapidly developing area of personality research. To date, authorities have had to depend significantly on historical data to identify crime-prone areas, such as crime rates and arrest patterns (Schoen et al., 2013). To combat crime, law enforcement agencies ought to know not only where crime hotspots are but also where they will be in the future. Because of the high level of user interaction on social media, offenders now have an advantage that has never been seen before. Before the widespread adoption of the Internet, criminals' activities were limited by territorial boundaries. Many people are now surrounded by technology, social networks, and, as a result, illegal online activity (Casale & Fioravanti, 2018). Analysis of social network data can assist to reveal or predict illegal acts, accounts, relationships, and messages. This enables for the identification of crucial information regarding criminal operations, as well as their location and involvement. Criminals can conceal their communications using a number of methods, including photographs, videos, encryption, and steganography (Grijalva & Zhang, 2016).

While narcissism is often ignored as a significant risk factor for mass shooters, it has been recognised as a common denominator among school shooters by government experts and scholars (Keatley, McGurk, & Allely, 2019). Several lists of risk factors for school shooters, for example, identify narcissism as a significant influence (Bondü & Scheithauer, 2015). Another list categorizes narcissism as one of the seven primary risk factors for the potential of school shooters (Bondü & Scheithauer, 2015). In Blinkhorn et al. (2016), participants were permitted to aggress anyone who threatened or praised them or an innocent third party. The highest aggression levels were shown by narcissists who aggressed directly against the person who offended them (Hyatt et al., 2018). These findings thus support the notion that human personality plays a direct or indirect role in the source of social problems and issues that lead to crime. Kalemi et al. (2019) asserted that it is the presence of narcissistic traits that indicated aggression rather than criminality. The study found that the scores of narcissisms among inmates were significantly greater than those of non-inmates, while the levels of self-esteem were comparable to those of the general population (Kalemi et al., 2019).

Previous research has shown that people with an inflated ego, which is a symptom of narcissism, are more prone to engage in violent behaviour (Kalemi et al., 2019). In Blinkhorn et al. (2016),

their hypothesis further backed the notion that narcissists will have more positive attitudes towards aggression as they believe it is more appropriate. The perpetration of psychological abuse has been connected to narcissism (Gormley & Lopez, 2010), and sexual and physical abuse (Blinkhorn et al., 2016, Carton & Egan, 2017). The grandiose component has dominated most studies on narcissism and domestic violence as the central assessment of narcissism. According to Lambe et al. (2018), a strong association exists between narcissism and aggression following an ego threat. Therefore, to understand aggression, violence, and other criminal activities, a thorough understanding of narcissism is relevant.

2.5 Personality assessment approaches

The growing availability of high-dimensional, fine-grained data about human behaviour on social media has changed how personality psychologists conduct research and personality assessments (Stachl et al., 2020). Personality impacts the formation of social relations, friends, personality traits, thoughts, emotions, and judgments. Self-report surveys have been widely acknowledged as the most accurate method for determining personality for many years (Štajner & Yenikent, 2020). Self-reports, on the other hand, only reflect one facet of personality: people's perceptions of themselves (Boyd & Pennebaker, 2017). This has then led to different approaches emerging with an attempt to overcome self-reported weaknesses. These current approaches include the lexicon and Open Vocabulary strategy, the propensity of word use, and personality processes measurement.

2.5.1 Personality assessment using questionnaires

In this method, personality is determined using language-based personality models that closely resemble the data found in widely used self-report surveys (Hall & Matz, 2020). It entails estimating how people respond to personality surveys using linguistic measures. Researchers use lexical features to maximise their account of variance in questionnaire scores when utilising estimated self-reports using language. The significance of this approach to personality evaluation is that the instruments created using these approaches go through numerous validation processes, ensuring that they are based on strong empirical evidence (Štajner & Yenikent, 2020). However, self-reports are likely to suffer from various limitations. The first limitation is that they require human assessment and training of the assessors. Secondly, self-reports are subject to response

biases, most commonly in the form of social desirability bias. This is where participants react to questions in a way that makes them appear more attractive to others (Krumpal, 2013). Thirdly, the extent to which people's self-reported traits accurately reflect who they are has been questioned in self-reporting questionnaires. The last limitation is the “self-knowledge constraints and response biases”. Questionnaire scores, for example, are viewed as a "real thing" that can be evaluated rather than a collection of supporting psychological processes when calculating people's self-reported neuroticism from language. This paradigm treats self-reported personality as a "gold standard", ignoring the defects that develop during the data collection processes (Boyd & Pennebaker, 2017). Personality researchers have sought innovative methodologies to overcome these limitations by identifying implicit assessments through digital tools rather than explicit self-reports (Stachl et al., 2020).

2.5.2 Lexicon and open vocabulary approach

Besides self-reports, new studies have shown that people's words in ordinary life can reveal a lot about their personalities. As growing number of studies suggest that people's use of words is reliable over time. The use and choice of words are also consistent, predictive of a wide range of behaviours from person to person. According to Kern et al. (2014), language is an important element personality, and unlike other common personality indicators, it does not require people to fill out questionnaires. Open vocabulary refers to linguistic features such as words or phrases that can be identified from the texts (Schwartz et al., 2013). The lexicon method suggests that basic personality traits are expressed in natural languages as words (Das & Das, 2017).

To predict Big 5 self-report measures from Facebook status posts, Schwartz et al. (2013) used an open vocabulary method. The recent expansion of online social media has resulted in a plethora of new avenues for personal expression. Aside from the benefits of a large amount of data, the text is often unique and highlights an individual's daily concerns. Furthermore, recent research has shown that the populations used in online studies are very representative. According to Hezarjaribi, Ashari, Frenzel, Ghasemzadeh, and Hemati (2020), the most universal way of expressing human feelings and thoughts is via language. The lexical hypothesis argues that any language can be an unbiased source of various personality types in psychology. As a result, it suggests that the feasibility of extracting personality types and psychometric features is by examining the linguistic representation (Neuman & Cohen, 2014).

2.5.3 Propensity of word use

This method involves counting the number of times words are used in pre-determined categories of language. A common word used by an individual is built into a lexicon and the personality identified is based on who uses more of such word-category lexica. The most widely used lexicon is Linguistic Inquiry and Word Count (LIWC) (Schwartz et al., 2013). Pennebaker, Chung, Ireland, Gonzales, and Booth (2007) used LIWC to identify personality by examining words in various domains. The researchers discovered that agreeable persons used more articles, introverts and those with low conscientiousness used more words signalling distinctions, and neurotic people used more negative emotion phrases. Kumar et al. (2018) looked at how to classify personality through the Big 5 personality traits and Schwartz's values model, which describes human values (Schwartz, 1992). Before the experiments, the data was pre-processed by stemming and tokenizing it, doing LIWC analysis, and normalising the feature vectors. The features were significant for which personality trait and value type were pre-analysed only to use the notable features in the final classifier and thereby save computation time and power (Kumar et al., 2018). N-grams were also added to the LIWC baseline. The classifier's performance was found to drop by nearly 10% on SVM with uni-grams, while with bi-grams, there were no significant changes in performance.

2.5.4 Measuring personality processes

Another approach to measuring personality is through measuring personality processes. According to Wrzus and Mehl (2015), personality processes relate to thoughts and feelings, assessed using physiological assessment, self-report, and behavioural observation. Physiological assessment entails assessments of physical activity or eye movement which could serve as indicators of attentional focus during conversations (Boyd & Pennebaker, 2017). According to Wrzus and Mehl (2015), behavioural observation entails assessment methodologies that allow for the direct, observation of behaviours in naturalistic settings.

2.6 Data preparation

Data cleaning involves identifying and eliminating anomalies found in a dataset that may negatively impact prediction (Kalra & Aggarwal, 2017). Data cleaning is necessary because real-world data often contains missing values, noisy and inconsistent data preparation (Panda, 2018). Data cleaning in text pre-processing involves transforming the raw data into an understandable

structure (Nhlabano & Lutu, 2018). There are various techniques of data pre-processing. These include data normalisation, stemming, tokenization, lemmatization, and stopword removal (Haryanto & Mawardi, 2018). The order in which these techniques are applied is of utmost importance in some cases. Stemming techniques work by removing the suffix of the word. Stopwords removal eliminates the frequent words that do not influence text classification (Ayedh, Tan, Alwesabi, & Rajeh, 2016). Stopwords are words that usually occur most frequently in a text. Words like a 'the', 'and' and 'for' are stopwords (Kaur & Buttar, 2018). In tokenization, sentences are broken down into words/terms, and word vector known as bag-of-words is constructed (Kadhim, 2018). Lemmatization determines the part of speech of the word. The purpose of lemmatization is to find the lemmas of the terms by removing the prefixes and suffixes based on morphological analysis (Yüksel, Türkmen, Özgür, & Altinel, 2019).

Saif, Fernández, He, and Alani (2014) examined whether removing stopwords improves or hinders the accuracy of Twitter sentiment classification techniques. The study used Twitter data from six separate datasets to test six different stopword identification methods. The study observed that using pre-compiled lists of stopwords has a negative impact on the performance of sentiment classification approaches. According to Deniz and Kiziloz (2017), stemming had little effect on most classification instances, including n-gram models at the character and word levels, as well as author and gender classification.

Srividhya and Anitha (2010) examined Stopwords removal, stemming, and document frequency on the Reuters dataset. The research found that removing stopwords can help to expand words, increase document discrimination, and improve classification performance. Haddi, Liu, and Shi (2013) used two data sets of movie reviews to study the role of text pre-processing in sentiment analysis. To reduce the dataset noise, they applied various pre-processing techniques like alphanumeric characters removal, Uniform Resource Locator (URL) removal, white space removal, negation handling, stemming and stopwords removal (Haddi et al., 2013). The study indicated that by utilising proper pre-processing methods and, sentiment analysis could be significantly improved.

Uysal and Gunal (2014) studied the impact of pre-processing on TC using four pre-processing methods: stopwords removal, tokens, lowercase conversion, and stemming. They investigated four

datasets: Turkish e-mails, English e-mails, Turkish news, and English news, utilising all conceivable combinations of the pre-processing methods. They only used the SVM machine learning (ML) approach with 10, 20, 50, 100, 200-, 500-, 1,000-, and 2,000-word unigrams as feature sizes. Their primary insight was that the proper combination of pre-processing tasks can significantly increase classification accuracy, depending on the domain and language. Inappropriate combinations, on the other hand, can reduce accuracy. According to Uysal and Gunal (2014), regardless of the domain or language, lowercase conversion enhances classification success in terms of accuracy and dimension reduction. However, for each domain and language analysed, there is no one-size-fits-all combination of pre-processing tasks that yields successful classification results.

Text classification methods have been used extensively to solve a variety of natural language processing (NLP) challenges. These classifiers are highly reliant on the size and quality of the training dataset. Poor performance will result from insufficient and unbalanced datasets (Sharifirad, Jafarpour, & Matwin, 2018). The percentage of decreased noise in the datasets determines the efficiency of pre-processing. Also, according to Mehanna and Mahmuddin (2021), some studies consider pre-processing as a standalone process, while others consider it at data preparation and the filtering stage. Depending on the sentiment extraction methodologies used, pre-processing approaches may differ greatly. Text representation systems also have an important role in determining pre-processing approaches (Naseem, Razzak, & Eklund, 2020). In NLP, the impact of the optimal mix of pre-processing techniques is not established in the literature. Experiments by Uysal and Gunal (2014) revealed that selecting the right combination of text pre-processing techniques can enhance classification accuracy significantly.

2.7 Identifying personality traits on social media: A review

Personality psychologists have long attempted to define essential features that differentiate people while often clustering through groups of people to uncover consistent behavioural trends (Kulkarni et al., 2018). Although there are a variety of approaches to analysing individual differences, the most common is lexical analysis, which proposes that key characteristics of human personality would become part of the vocabulary used to describe ourselves (Uher, 2013; Kulkarni et al., 2018). Several studies have looked at Facebook posts, text messages, and Twitter, to establish

correlation between the Big 5 and other individual characteristics with language (Alsadhan & Skillicorn, 2017). There are two main approaches that exist regarding the identification of personality from social media. These main spectrums are a text-based and image-passed personality inference approach.

2.7.1 Text-based personality inference

Yarkoni and Westfall (2017) asserted a relationship between people's personalities and the language they use. This means that the language people use to disclose their personality qualities and the words people use online to coincide with their personalities. Gitari, Zuping, Damien, and Long (2015) employed a lexicon-based technique in online conversation, to 'detect' race, nationality, and religious speech. Existing lexical resources and corpora supported this technique. After combining the retrieved data and the annotation results, a hate speech lexicon was created (Esra'M, 2019).

2.7.2. Image-based personality inference

The study of the personality expressed by photos has been used to infer user personalities, as well as to examine how brands express and shape their identities through social media (Rodriguez, Gonzàlez, Gonfaus, & Roca, 2019). According to Cristani, Vinciarelli, Segalin, and Perina (2013), used aesthetics and content features of 300 Flickr inferred their personality. Ginsberg (2015) used Instagram photos to define each brand's identity along five personality dimensions of honesty, excitement, competence, sophistication, and roughness. The study examined at the photo elements utilized in the top five food businesses' Instagram feeds. Quantitative content analysis was used to look at the frequency of certain qualities and elements in the images.

2.8 Text-based annotation for personality identification

Text annotation is the process of tagging and labelling a dataset to specific categories. The annotation's aim is to create new attributes and expand the corpora. One of the crucial approaches to text annotation is lexicon detection. According to Kolchyna, Souza, Treleaven, and Aste (2015), lexicon-detection comprises two steps. i) creating a list (bag-of-words) ii) check which words of the list are also included in the lexicon. The quality of the lexicon determines the correct classification of text to sentiment categories (Jurek, Mulvenna, & Bi, 2015).

Xiang, Fan, Wang, Hong, and Rose (2012) used an unsupervised approach to detect tweets related to offensive topics. Several supervised learning methods were used. LR achieved slightly better results than other supervised algorithms by 5.4% improvement over the keyword matching baseline method. Research in Putri, Sitepu, Sihombing, and Silvi (2019) was done on Indonesian news sites to identify hoax content. Different machine learning techniques were used to compare the efficacy of the findings after several pre-processing. The stochastic gradient descent (SGD) classifier got the highest accuracy, with an accuracy of 86% over 100 hoaxes and 100 non-hoax websites.

Various attempts have been made to detect themes and identify various personalities using lexicon-based approaches. Warner and Hirschberg (2012) utilised web-based datasets to create a hate speech detection classifier. To identify linguistic hate speech features, they manually analysed chosen paragraphs including anti-feminism, anti-black, anti-Semitism, and anti-Muslim rhetoric. The results were then analysed in order to develop a hate speech model that captured the common characteristics of seven hate speech categories. The model was used to train a parser tool, which analyses text and automatically recognises hate speech (Warner & Hirschberg 2012). The prevalence of more hate words than those recorded by the model was revealed in tests of the model's validity in a corpus. As a result, the study suggested that the model's coverage be increased.

Data labelling is the categorisation and annotation of data before subjected to machine learning classifiers. Cluster-then-label methods are a series of methods that use clustering and classification to label data (Jan, 2020). They use an unsupervised or semi-supervised clustering method on all data available to generate clusters and guide the classification process (Ji, Henriques, & Vedaldi, 2019). With the labelled data contained in each cluster, a classifier is subsequently trained individually for each cluster (Ji et al.,2019). Clusters refer to groups of data with common characteristics which differentiate them from other groups (Ibrahim, Zeebaree, & Jacksi, 2019). Therefore, utilising the classifiers for their respective clusters, the unlabelled data points are categorised.

2.9 Levels of sentiment analysis

Sentiment analysis is the process of categorising people's opinions expressed in a text as positive, negative, or neutral. To find topics with diverse feelings, various sentiment models have been

developed. According to Sharma, Hara and Hirayama (2017), sentiment analysis allows predicting opinions and attitudes or groups of people or individuals. According to Jain (2017), opinion mining will become a necessary and crucial part of big companies and organisations. A sentiment analysis result plays a big role in any significant institutional change, product updates, reviews, launches, strategic planning, and investments (Jain, 2017). There are three primary classification levels of sentiment analysis. These approaches are sentence level, document level and word phrase-level sentiment analysis (Almatarneh & Gamallo, 2018).

2.9.1 Document level

A document is taken into account in document level sentiment analysis. It is classified depending on the paper's overall tone. As a result, it is categorised based on the overall sentiment of the documents (Sabeti & Saad, 2017). The main limitation when using document level is that all the sentences that express opinions are subjective (Al-Shabi, 2020). Khan et al. (2016) noted that it is more precise to look at each sentence individually to obtain more accurate results after sentence level analysis. Only the objective sentences are eliminated, whilst the subjective sentences are extracted for sentiment analysis. The entire document, including the text, is regarded as a basic unit of data. It is considered that the document has a single opinion about a single object (Shirsat, Jagdale, & Deshmukh, 2017). This method is ineffective if the document contains opinions on a variety of topics, such as Twitter or Facebook.

2.9.2 Sentence Level

Sentiment analysis at the sentence level entails assessing if each sentence represents a positive, negative, or neutral sentiment. This format approach is used for one-sentence reviews and comments made by the user (Kharde & Sonawane, 2016). Each sentence is treated as a separate unit in the sentence-level classification, which posits that each sentence should only include one point of view. Because each sentence is considered an independent entity, and each sentence can have a different perspective, polarity is determined for each sentence (Kolkur, Dantal, & Mahe, 2015).

2.9.3 Word/Phrase level

This is a sentiment analysis approach that considers words or phrases at the word or phrase level. This type of sentiment analysis is where a word is the smallest unit with meaning available in a

text. Word lexicons of sentiment analysis include adjectives (happy, cruel, amazing, awesome, sad, annoying), adverbs (poorly, horribly), and some verbs (hate, love, despise) (Saberri & Saad, 2017)

2.10 Lexicon detection approaches

The lexicon-based technique, also known as the dictionary approach, is based on a predefined polarity lexicon or dictionary. This is sometimes described as an unsupervised machine learning approach. The lexicon-based approach's classification quality is exclusively determined by the lexicon's quality. There are four approaches to text lexicon. These are keyword-based detection, rule construction approach, learning-based approach, and hybrid-based approach (Akram & Tahir, 2018).

For lexicon-based approaches, dictionaries are built using human or automatic processes, with seed words serving as the basis for creating more words in the list of terms (Kaity & Balakrishnan, 2020). Most studies that used a lexicon-based method used adjectives to indicate the text's semantic orientation (Ramanathan & Meyyappan, 2019). Adjectives are stored in a dictionary with their polarity and weight values in such circumstances. The dictionary is then utilised to annotate the adjectives collected from the input text data. Enhancements to statistical approaches and the genetic algorithm utilised for dynamic lexicon creation were proposed by Mowlaei, Abadeh, and Keshavarz (2020). The generated vocabulary is combined with static lexicons like SentiWordNet, and the resulting lexicon is used in sentiment analysis. Das and Das (2017) extracted the unique words and phrases associated with each of the Big Five personality types using a frequency-based N-gram approach. In addition to the terms, they introduced a new element to the lexicon: corpus-based probability of occurrence. A small YouTube personality dataset was used to test the lexicon. Their classification experiment achieved 78.52% accuracy with logistic regression and 62.26% with SVM (Das & Das, 2017).

2.10.1 Keyword-based detection

In this approach, lexicon detection is done by extracting keywords related to the study from the text. These keywords are matched with the knowledge base or the dictionary such as Thesaurus to find their relation. The most used and most essential words and expressions are extracted from a text. Through this, various themes and topics are identified from the data. According to Hasan and Ng (2014), a semantic generalization of a paragraph or document, as well as an accurate

representation of the document's content, is a topic keyword retrieved from a document. Documents can be given a certain label based on the topic keywords. Natural language processing (NLP) tasks are influenced by the extraction accuracy of topic keywords (Liu, Huang, Huang, & Duan, 2020).

2.10.2 Rule-based construction approach

Rules-based techniques use rules for knowledge representation. A rule-based system is made up of rules, memory for storing states, a schema to match rules, and a conflict resolution schema if there is more than one rule (Nguyen, Do, Tran, & Pham, 2020). By using a rule-based method, rules needed to extract information from the dataset are developed. A rule-based system's primary goal is to collect and encapsulate the knowledge of a human expert in a specialised domain into a computer system (Golshan, Dashti, Azizi, & Safari, 2018). The rule-based method entails creating and using grammatical and logical rules to find themes in texts. Keyword recognition and lexical affinity methods are used in the rule construction process. The keyword recognition technique is concerned with the creation and application of emotion lexicons using 'IF' rules (antecedent) THEN effect approach (Aubaid & Mishra, 2020).

Seal, Roy, and Basak (2020) sought to detect emotions using emotion keywords and rules with particular focus on phrases. They gathered information from the ISEAR database, pre-processed it, then evaluated it for phrasal verbs. They discovered certain phrasal verbs that may have been associated with emotion terms but weren't. As a result, they compiled a list of phrasal verbs and built a database with the verbs and their synonyms. They identified keywords and phrasal verbs linked with specific emotions and categorised them using the WordNet emotion lexicon (Seal et al., 2020). Although they reported a 65 percent accuracy rate, they acknowledged that their approach did not solve problems with current systems, such as a lack of emotive keywords and a disregard for word semantics based on context (Seal et al., 2020).

2.10.3 Learning-based detection

In this approach, machine learning models are trained to identify the lexicon. Learning-based techniques are probabilistic and use statistical models rather than rules to make decisions. Although the machine learning approach involves less manual work, it does necessitate prelabelled

data and adaption retraining (Rana & Cheah, 2017). The majority of machine learning algorithms rely on supervised categorisation, which detects sentiment in binary terms.

The key disadvantage is the scarcity of labelled data in many domains, limiting the method's applicability to new data and domains (Alessia, Ferri, Grifoni, & Guzzo, 2015). Unsupervised learning is frequently utilized since it doesn't need a lot of human-annotated training data to provide usable results. Khang and Zhou (2017) proposed an unsupervised rule-based method that extracts subjective and objective features from online consumer reviews. They identified accurate features by incorporating relation and review-specific patterns (Khang & Zhou, 2017). Yo and Sasahara (2017) used machine learning algorithms to predict personal attributes from the text of tweets, such as gender, occupation, and age groups. The findings revealed that the machine learning algorithms could accurately predict the three personal qualities of interest by 60– 70%.

2.10.4 Hybrid Method

In a hybrid method, rule-building and machine-learning approaches are combined into a single model. As a result, this technique has a better possibility of outperforming the other two options individually by combining the strengths of both while masking their related shortcomings. According to El Alaoui, Gahi, Messoussi, & Chaabi (2018), a hybrid approach uses both knowledge based and statistical methods for detection. Hung and Chen (2016) adopted a hybrid approach to develop movie-based SWN lexicon from the general-purpose SWN. They used an SVM, Naive Bayes and Decision tree to combine unigrams and bigrams with vector space modeling, and the findings showed that a WSD-based SWN lexicon improved performance (Hung & Chen, 2016).

Priyanta, Hartati, Harjoko, and Wardoyo (2016) used rule-based and machine-learning models to compare the classification of sentence subjects. For rule generation, the authors used opinion patterns. They evaluated sentence subjectivity in Indonesian news to determine if a sentence was a subject or an objective. Two machine-learning models were used to obtain this classification. The results of the evaluation and analysis revealed that the rule-based classifier outperformed SVM (74%) and NBC (71%) with an accuracy of 80.36%.

2.11 Lexicon detection for narcissism

According to Darwich, Mohd, Omar, and Osman (2019), lexicons can be created using either a dictionary-based approach or a corpus-based approach. The dictionary-based approach is constructed by developing a list of seeds-related words manually, while in a corpus-based approach, seed words are collected using statistical and semantic methods. The quality of the dictionaries and lexicons used in the classification process is one of the most important variables in the lexicon-based method. For instance, Kundi, Khan, Ahmad, and Asghar (2014) developed a scoring system for sentiment prediction and created a slang dictionary containing a set of slangs annotated with scores and orientations using a weighted threshold value obtained based on the SWN lexicon.

According Liu, Burns, and Hou (2017), the linguistic content of social media posts has been demonstrated to be valuable since it displays a user's topics of interest and gives information about their lexical usage that may be predictive of specific user types, and the correlation between linguistic clues and personality traits has been identified to discover how to conduct research in the area of automatic personality classification (Kaushal & Patwardhan, 2018). To develop the dictionary, different criteria and measures used in psychology literature were used. The frequency-based N-gram approach is used to extract the unique words as well as phrases concerning narcissistic personality classes. These approaches are discussed in the next sections.

2.11.1 Use of swear words

According to a study by Golbeck (2016), respondents with high levels of narcissism used swear words much more frequently than those with lower levels of narcissism. This was consistent with prior findings, which showed that narcissists curse more (Holtzman, Vazire, & Mehl, 2010; DeWall, Buffardi, Bonser, & Campbell, 2011). According to Kern et al. (2014), a narcissist gets irritated easily and one linguistic marker of their disagreeableness is swearing. This is also motivated by the fact that narcissistic people tend to draw attention to themselves by all means. Furthermore, to create a sexualised environment, narcissistic persons often utilise sexual words (Holtzman et al., 2010). Therefore, the developed narcissistic dictionary included all the swear/curse words that were found in the tweets. These include hate, f*ck, hell, etc.

2.11.2 Social processes and affective processes

Social processes contain words that are pronouns and verbs that reflect social interaction (Pennebaker, Francis., & Booth, 2001). According to Golbeck (2016), people with high scores of narcissisms use many more words about social processes than people with low scores of narcissism. LIWC consists of sub-categories of words that describe 'Family', 'Friends', and 'Humans'. Narcissists pursue social status through the same self-regulatory processes (Grapsas, Brummelman, Back, & Denissen, 2020).

Affective processes contain words that describe positive and negative emotions. Emotional ability relates to people's ability to adapt to life's changes through rational and emotional coping skills. According to Lopes et al. (2011), the ability to deliberately regulate emotions in oneself and others is a critical component of emotional intelligence hostility, social avoidance, and a lack of empathy characterises vulnerable narcissists' social behaviour. Furthermore, Czarna, Zajenkowski, and Dufner (2018) stated that in the event of provocation, vulnerable narcissism exacerbated reactive and misdirected hostility. According to Golbeck (2016), individuals with high narcissistic traits tend to use significantly fewer positive emotion words and more negative emotion words than those with low narcissistic traits. Therefore, the dictionary also included words that represented negative emotion in the text. These include *hate*, *evil*, *kill*, *murder*.

2.11.3: Antisocial Word Use Index

DeWall et al. (2011) created the Antisocial Word Use Index, which combines the frequency counts for swear words and angry terms. After analysing the data, Golbeck (2016) discovered that the high narcissism group used many more antisocial phrases. Some of the tweets in the dataset contained antisocial words, and these were also included in the dictionary. Examples of antisocial words were *evil*, *brat*. DeWall et al. (2011) noted that narcissists tend to use either offensive language or first-person singular pronouns as a means of grabbing attention. Narcissistic persons adopt antisocial word use and offensive words with a goal to get attention (Adams et al., 2014)

2.11.4 Self promotion

Hart, Adams, and Burton (2016) contended that narcissists are more likely to speak in the first person singular while delivering impromptu monologues and to talk about their accomplishments

in conversation. Because pronouns provide extensive information about how people relate to others, first-person pronoun use is one instrumental variable to consider when considering narcissism. First-person singular pronouns are therefore used by narcissists to draw attention to themselves. In cases where narcissists used a less first-person singular pronouns, they call more attention to themselves by using antisocial words (Rathner et al., 2018).

In research, I-talk has been regularly utilized to implement narcissistic self-focus (Ireland & Mehl, 2014). DeWall et al. (2011) looked into how narcissists use social media to communicate information about themselves. First-person singular pronouns, which are thought to be an implicit indicator of narcissistic self-focus, are thought to be used in the two studies. Narcissistic people who didn't use them made up for it by using other attention-getting self-presentation techniques, such as posting provocative photos or using more profanity and verbal aggression in their online self-descriptions.

In addition, according to Carey et al. (2015), narcissists change their linguistic style based on how many first-person singular pronouns they use in an online self-description activity. When narcissists utilise a limited quantity of first-person singular pronouns in tales about themselves, they draw attention to themselves by using more profane and harsh language. However, narcissists did not use more profane and hostile language in their narratives when they had already brought attention to themselves by using an increased frequency of first-person singular pronouns. As a result, narcissistic participants showed symptoms of implicit compensation as a way of attracting attention to themselves when it wasn't available (DeWall et al., 2011). Therefore, if there are more pronouns or negative words, then that would show traces of narcissism (Van der Linden & Rosenthal, 2016). From the literature review, the use of pronouns was found to be synonymous with a grandiose narcissist. Narcissists use pronouns to self-promote themselves and therefore the dictionary included pronouns like; *I, my, we, etc.*

2.12 Text classification using artificial intelligence

Text classification is a text mining task of which the primary goal is to discriminate or characterise a piece of text into a specific format value (Billal, Fonseca, Sadat, & Lounis 2017). Such value can vary from number to labels and classes. Supervised, unsupervised, and semi-supervised algorithms are commonly used to classify text (Jiang, Wang, Wang, & Ding, 2018). The supervised methods

use training data to build a model that gives the required prediction for an unknown instance (i.e., using labelled/annotated data). As a result, unsupervised algorithms are frequently utilised for clustering problems because they do not require labelled data. Personality identification is deemed a classification problem (i.e., categorising an instance as narcissistic or non-narcissistic). Studies on personality detection have mainly been based on supervised algorithms often with performance comparisons among them (Chatzakou et al., 2017; Huang, Ryan, Zabel, & Palmer, 2014).

Text classification relies on labelled data for training. However, labelling a subset of these data is a time-consuming and tedious process that usually has to be done manually (Chen, Yang, & Yang, 2020). The time and effort spent manually labelling an increased amount of data X versus the increase in classification accuracy Y is a trade-off – given the ease with which enormous amounts of unlabelled data are available in most fields. There are large quantities of unlabelled data that are easily accessible in most domains, and the gain in accuracy is low compared to the costs of labelling. The question arises of how to use unlabelled data to improve learning algorithms (Bekker & Davis, 2020). There are two approaches to text classification using artificial intelligence. These are, machine learning-based classification and fuzzy-based classification

2.12.1 Machine learning-based classification

Machine learning is a branch of artificial intelligence based on the premise that computers can learn from data, recognize patterns, and make conclusions with little to no human involvement (AbdulHussien, 2017). Arjaria, Shrivastav, Rathore, & Tiwari (2019) questioned a person's uniqueness from written text in the Big Five model. The study applied to stem in the pre-processing stage, stopword removal, and normalisation to the written datasets. Frequency values were assigned to the datasets, and the vector space model was used to represent them. Finally, they used the Multi-label Nave Bayes algorithm to estimate the outcomes and identified the personality traits derived from the human writing assignment.

Gonçalves, Araújo, Benevenuto, and Cha (2013) averaged the sentiment analysis methods based on their classification performance, demonstrating that combining them results in a higher coverage of correctly classified messages. Existing lexical resources and sentiment analysis approaches were similarly incorporated as meta-level characteristics for supervised learning. The results of the experiments showed that mixing different resources improves polarity and

subjectivity classification accuracy significantly. To categorize diseases based on clinical texts, Yao, Mao, and Luo (2019) classified clinical literature by finding trigger phrases, utilising trigger phrases to predict classes with few samples, then training a knowledge-guided convolutional neural network for classes with more examples (Yao et al., 2019). A related study by Ranganathan, Hedge, Irudayaraj, and Tzacheva (2018) sought to detect emotions on Twitter using decision tree classification. They built a corpus of tweets and related fields where each tweet is classified with respective emotions based on lexicon and emoticons.

Furthermore, Carducci, Rizzo, Monti, Palumbo, and Morisio (2018) suggested a supervised learning strategy for computing personality traits based solely on public tweets. In this study, they segmented tweets in tokens to feed the machine learning classifier. Their dataset had a sample of 24 Twitter users with total tweets of 18473 and 250 Facebook users from the *myPersonality* dataset. After tokenization, they utilized a supervised learning classifier to train word vector representations as embeddings (Carducci et al, 2018).

Tadesse, Lin, Xu, and Yang (2017) used the Big 5 model to predict users' personality based on the *myPersonality* dataset. Using the LIWC and SPLICE dictionaries, textual features were retrieved from comments that represent language usage and have expression and subject count. Second, they used aspects of social interaction such as connectedness, network size, and so on. Pearson's correlation assesses the “strength of the relationship between the variables” and extracts key characteristics. XGBoost is used as a classifier along with three baseline algorithms as Gradient Boosting, and SVM. The study applied Label Distribution learning. The feature extraction was divided into three categories: static features with minor change over time, such as gender and name; dynamic features with major changes over time, such as gender and name; and dynamic features with major changes over time, such as followers. Last but not least was content features like blogs, linguistic and psychological features (Tadesse et al., 2017).

2.12.2 Fuzzy-based text classification

The third approach to text classification is through fuzzy logic. Fuzzy logic is one of the artificial intelligence approaches/techniques in which intelligent behavior is obtained by constructing fuzzy classes of certain parameters (Shailaja, Seetharamulu, & Jabbar, 2018). Fuzzy logic-based classification includes reasoning that is approximate rather than fixed and exact (Manivannan &

Ramakanth, 2018). Variables in fuzzy logic may have a truth value that ranges in degree between 0 and 1, and it has been extended to handle the concept of partial truth. The truth value may range between completely true and completely false (Vashishtha & Susan, 2019).

According to Lee, Choi, and Kim (2017), fuzzy inference systems include rule output aggregation, input variable fuzzification, rule evaluation, and defuzzification. To forecast the level of polarity of reviews, rules are examined after crisp sentiment and subjectivity values obtained from all text reviews are converted into fuzzy values using a fuzzifier. Ekong, Ekong, Uwadiae, Abasiubong, and Onibere (2013) did a study on an effectiveness of fuzzy logic in forecasting risk levels of depression. The authors proposed a Fuzzy Inference System (FIS) model for medical diagnostics in psychology and psychiatry. Based on expert knowledge encoded in fuzzy rules and provided physiological and psychological characteristics, the system predicts depression risk severity levels. The study noted that fuzzy logic is an ideal technique for medical diagnosis and disease management. Kavuri, Kumar, and Rao (2012) used a fuzzy similarity approach to classify text documents where the terms in the feature set were grouped into clusters based on membership function. Then by deriving the membership functions for each extracted feature of each cluster, close matching is done to describe the real distribution of the training dataset. Wilges, Mateus, Nassar, Cislighi, and Bastos (2016) developed a fuzzy classification approach in which the variables demonstrate the ability to analyse the similarity and accuracy of a text document. This was accomplished with the aid of a database compiled from a selection of text documents that had been previously grouped into different categories.

Using fuzzy membership degrees, Jefferson, Liu, and Cocea (2017) developed a fuzzy rule-based approach for sentiment analysis that provide more precise results. The experimental results showed that the fuzzy-based technique outperforms the other algorithms by a slightly better margin. Furthermore, the fuzzy approach eliminates the need for a large number of classes to define distinct degrees of sentiment. Furthermore, the fuzzy technique allowed for the formulation of various degrees of sentiment without the need of a large number of classes (Jefferson et al., 2017). The approach adopted for this research is different from all the above studies in the sense that it proposes the use of fuzzy sets to estimate the level of narcissism given the polarity of a tweet and predicted class.

2.13 Positioning the research

Research has shown that individuals inadvertently leave their ‘behavioural residue’ in their physical and virtual environments. The potential benefits of social media for personality researchers go beyond using large sample sizes. A user's behaviour on social media is reflected in their tweets, status updates, comments, and interests, which reveal traits of their personality. As shown in the above literature on personality detection with machine learning techniques, utilising data from social network sites, it is possible to predict personality automatically, which is simpler, more cost-effective, and more efficient than conventional methods (Dandannavar, Mangalwede, & Kulkarni, 2018).

A growing body of research has investigated the link between of narcissism and motives for using Twitter (Hughes et al., 2012) and the content of tweets. However, few studies have studied the identification of traces of narcissism from social media and specifically Twitter using a fuzzy-based machine learning approach. It is this gap that motivated this research to identify traces of narcissism amongst users on Twitter. In social network sites, the users’ post on social network does not always contain complete information. Social media text is characterised by informal language, short context and noisy sparse contents (Virmani, Pillai, & Juneja, 2017). Extraction of this information involves identifying relevant, precise and useful data from the informal and noisy textual data. Users are faced with the difficulty of extracting precise information from social media without effective techniques that can extract only the facts that match their interest. Processing social media data is often challenging as data structure is informal, noisy and short (Derczynski, Maynard, Aswani, & Bontcheva, 2013).

Therefore, this research adopted a rule-based text pre-processing approach to perform data cleaning. Rules-based techniques use rules as the knowledge representation. Rule-based systems solve problems by taking inputs and then combining together with a set of rules from the rule base to arrive at a solution. When given the same exact problem situation, the system will follow the same approach to come up with the solution (Ahmed, Alfonse, Aref, & Salem, 2015). Text data is split into segmental blocks after pre-processing in preparation for feature selection and analysis through tokenization. Tokenization is a process of splitting sentences into small parts called “tokens”. Lemmatization is performed on tokens after tokenization (Khan & Ratha, 2016). Lemmatization seeks to eliminate inflectional endings from words using vocabulary and

morphological analysis, reverting words to their dictionary form. The text analysis begins with the extraction of relevant sentences, which are then utilised in the text analysis.

Several efforts of personality detection have been made using well-known lexicons such as WordNet, SentiWordNet, and SenticNet, etc. However, a lexicon solely devoted to classifying different personalities is rare (Das & Das, 2017). In this research, a set of seed words were created manually and extended using psychology description of narcissism. Synonyms and antonyms of narcissistic-related terms that are likely to appear on social media posts were identified and added to the seed list. A hybrid approach was adopted to construct the lexicon. This was done by applying unsupervised machine learning to the data where top keywords that describe each topic were generated.

A lexical dictionary was created that contains top words representing the appearance of narcissism in the text. Because words usually have semantic orientations that are the same as their synonyms and the opposite of their antonyms, new words can be added to the lexicon in stages. In literature, few works have attempted to use fuzzy aspects in systems of personality classification. According to Liu, Burnap, Alorainy, and Williams (2019), there is a small percentage of studies in NLP that have focused on fuzzy classification. Social media datasets are naturally unstructured, thus requiring the capability of fuzzy logic in dealing with fuzziness and uncertainty in opinion mining.

2.14 Summary

The use of social media has increased at a faster rate than ever before. More than half of the globe already utilizes social media, with the overall number of users increasing by more than 10% from 2019 to 2020, bringing the total to 3.96 billion by the beginning of July 2020. This indicates that the social media user base rose by more than one million people on average every day in 2020, or over 12 new users every second. Twitter use has soared over the past decade. With this literature review, existing gaps in studying Twitter as a social media platform for personality identification were identified. As evident from the existing literature, there is much need and scope for exploring the potential of the currently unmonitored public stream of Tweets. Twitter users express distinct personality qualities through their tweets and the tweets they like, making it the perfect platform for studying personalities. Besides, existing techniques related to personality prediction have not been studied comprehensively on narcissism. While previous studies emphasize the use of social

network sites data to monitor people's thoughts and feelings in order to identify personality, the approaches in Chapters 3-6 of this thesis look into various psychometric and linguistic features that have not been examined in previous studies in relation to narcissism personality trait classification.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

Research methodology refers to the procedures and methods used in the research study to collect and analyse data (Kassu, 2019). It also examines the tools the researcher must use to complete the study. This chapter presents the methodology that was used to assist in answering the research questions. A summary of existing methods and their limitations are presented and justification is provided for the chosen approach. The chapter also discusses the existing social media methodological frameworks. The frameworks are reviewed, and the modified methodological framework (Iterative CUPP) suitable for the research problem is developed. A process model based on the design science research methodology and CUPP framework are also discussed. Finally, the methods of data collection, analysis, reliability, and validity are discussed. The final section summarises the ethical considerations underpinning the study.

3.2 Types of research methodologies

According to Nabukenya (2012), there are five research methodologies. These are: case study research, action research methodology, grounded theory research methodology, survey research methodology, and design science research methodology.

Case study research (CSR) is an empirical investigation into current phenomena in their real-life setting, particularly when the distinction between phenomenon and context is unclear. It is qualitative and observational, with predetermined research questions (Cronin, 2014). CSR is important when investigators wish to gain a more in-depth look at a topic of interest. On the other hand, CSR has some drawbacks: First, a CSR study must be designed and scoped to guarantee that the research question(s) can be correctly and adequately addressed (Deigh, Farquhar, Palazzo, & Siano, 2016). Secondly, because corporations and other organisations are not always willing to participate in CSR, the availability of appropriate case study locations may be limited.

Action research (AR) inquires how people design and implement action concerning one another. Action research is described as research that results from an investigator's interaction with members of an organisation on a topic of genuine concern to them (Ivankova, 2014). AR aims to

increase understanding of the social situation, emphasising the complex and multivariate nature in the information system domain (Dresch, Lacerda, & Miguel, 2015).

Grounded theory (GT) is a methodology that allows researchers to establish a theoretical description of a topic's general characteristics while being grounded in empirical evidence (El Hussein, Hirst, Salyers, & Osuji, 2014). GT uses qualitative methodologies to collect data on a topic, and a theory develops from the facts rather than collecting data to test a theory or hypothesis. The theory is founded on the data's representation of reality (Flick, 2017).

Survey research (SR) relates to surveys carried out to increase scientific understanding (Nabukenya, 2012). The basic concept underneath survey methodology is to measure variables by asking people questions and looking at their relationships (Ghazi, Petersen, Reddy, & Nekkanti, 2018). According to Nabukenya (2012), SR suffers from two main limitations. The first disadvantage of SR is reactivity, which occurs when respondents give socially acceptable responses that make them look good or appear to be what the researcher wants. The second drawback of SR is that of obtaining the exact number and type of persons required for a representative sample of the target population.

Design science research (DSR) is a problem-solving approach that aims to develop new ideas, practices, technical skills, and products. It entails analysing the usage and output of constructed objects to comprehend, clarify, and enhance information system behaviour (Baskerville, Baiyere, Gregor & Rossi 2018). The design science research methodology (DSRM) seeks to solve problems by introducing into the environment new artefacts. Construction and evaluation are the two main activities in DSRM: Construction is a problem-solving, creative process in which artefacts are developed for a specific objective, whereas evaluation involves determining the usefulness of developed artefacts (Zafar, Harle, Andonovic, & Ashraf, 2007). According to Vom Brocke, Winter, Hevner, and Maedche (2020), DSR establishes a systematic strategy to designing and developing artefacts that solve problems, making it extremely useful in the real world.

3.3 Design science research

According to Baskerville, Kaul, and Storey (2018), DSR has two primary activities. The first task entails the creation of novel artefacts to generate new knowledge. The second activity relates to

the study of the artefact's usage and output through reflection and abstraction. Although DSR is a problem-solving approach, it does not seek an ideal solution but rather a satisfactory one to the issues under consideration (Dresch et al., 2015). In addition, solutions obtained from DSRM's conduct must be generalised to a particular class of problems (Lacerda, Dresch, Proença, & Antunes Júnior, 2013).

3.3.1 Choice of design science research methodology

Because of the explorative nature of this study, DSRM was adopted. DSRM focuses on developing artefacts for business and its environment in information systems (IS). Artefacts are considered research outputs or the end objectives of DSR projects (Hevner & Chatterjee, 2010). Artefacts can be instantiations, methods, models, constructs, or instances.

Table 3.1: Research methodologies: Comparison

| Methodology | Phases | Research output | Problem Solving | Framework/Process model development |
|--------------------|--|--|------------------------|--|
| Action Research | <ul style="list-style-type: none"> • Diagnosing • Action Planning • Action taking • Evaluation | <ul style="list-style-type: none"> • Specific Organizational Solution | Yes | No |
| Ground Theory | <ul style="list-style-type: none"> • Theory • Generating • Theory Evaluation | <ul style="list-style-type: none"> • Abstract Knowledge • Theories | No | No |
| Case Study | <ul style="list-style-type: none"> • Environment analysis • Observation • Evaluation | <ul style="list-style-type: none"> • Phenomenon • Investigation • Generalizations • Tests | No | No |
| Design Science | <ul style="list-style-type: none"> • Analysis • Design • Evaluation • Communication | <ul style="list-style-type: none"> • Construct • Models • Methods (Process) • Instantiations | Yes | Yes |

The fundamental objective of this research was to develop a prediction model. Because of its ability to support the development and evaluation of the model, DSR was chosen as most appropriate methodology for this research. As shown in Table 3.1, the DSR methodology provides defined research outputs in the form of IT artefact. The artefact is evaluated with different datasets and attributes in Chapter 8, and a modified process model based on the developed artefact is presented in Chapter 9. The research approach was problem-driven, meaning that the study's outcome would

solve the research problem and fulfil the business needs through a recommended research model for predicting narcissistic personality traits.

3.3.2 Relevance of DSRM in this research

DSRM ensures that the development of the narcissistic prediction model is done through a rigorous process. Feedback and communication form an integral part, thereby providing an application of high quality and standard. The development of the prediction model, in the form of an artefact, is critical to assess its effectiveness in solving an existing problem. The objective this research was to develop an artefact that can identify of traces narcissism from social network sites.

Figure 3.1 shows the DSR framework, which serves as a data collection and analysis guideline. The framework outlines the design process in response to an identified problem and is conceptualised based on evidence from existing DSR publications. Problem-solving is now widely recognised as a process in which the problem solver conducts a solution search to find a path to a solution (Peffer, Tuunanen, & Niehaves, 2018). The result of the design is a form of design contribution that can range from abstract artefacts (e.g., theories) to material artefacts. The nominal process models for design science consider six standard processes. Figure 3.1 shows the DSRM proposed by Peffer et al. (2018).

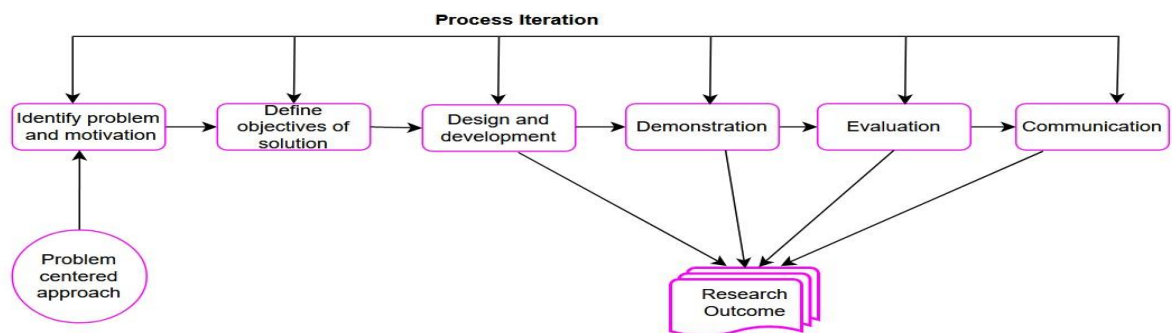


Figure 3.1: Design science research methodology stages

3.3.3 Design research phases

The first stage in DSRM is the identification of the problem which was determined in section 1.3. Design science research must provide solutions to problems faced by people, organisations, and technology (Peffer et al., 2018). The relevance of the problem serves as both a requirement and an evaluation criterion for research and research deliverables. The insights into the problem are

discussed and highlighted in sections of Chapter 1. The evaluation provides feedback on the quality and efficacy of the product (Hevner & Chatterjee, 2010). The artefact is evaluated using various metrics in this research, namely F1-score, precision, accuracy, and recall.

The second stage in DSRM is the formulation of objectives of a solution as determined in section 1.5. The objectives could be quantitative, such that the proposed solution is better than present ones, or qualitative and the new artefact provides solutions to problems that have not been addressed previously (Peffer et al., 2020). The objectives ought to follow logically from the problem description. Knowledge of the current state of issues, existing solutions, and their efficacy, if any, are resources needed for all of this.

The third stage in DSRM is the actual design and the development of the artefact as determined in section 6.7. The real artefact is the result of this phase. Through well-executed evaluation techniques, the artefact's reliability, efficacy, and quality must be carefully evaluated (Dresch et al., 2015).

The fourth stage in DSRM is the “demonstration” of the efficacy of the artefact to solve a problem. These include case studies, proofs, experiments, simulations, and other appropriate activities (Braun, 2017). Effective understanding of how to use the artefact to address the problem is one of the resources required for the demonstration (Peffer et al., 2018).

The fifth stage of the DSRM process is evaluation, which involves comparing a solution's goals to real results gained by the demonstration's use of the artefact (see section 8.4). It involves understanding of key measurements and analysis techniques (Venable, Pries-Heje, & Baskerville, 2012). A functional comparison of the artefact with the solution goals could be included. It also includes performance indicators like response time or availability (Venable et al., 2012). Once evaluation is completed, the researcher can choose whether to return to the design and development stage to increase the artefact's effectiveness, or to go on to communication and leave additional improvements to future studies. (Deng & Ji, 2018).

The last stage in DSRM is the communication of the research results (section 9.6, 9.7). Research in DSRM ought to be effectively presented to both technical and non-technical audiences (Peffer et al., 2018). This communication also includes information on how the organisation can obtain,

build, and use the artefact and a description of the artefact's construction (Hevner, March, Park, & Ram, 2004). This research implemented this last phase of DSRM by means of a written thesis as the primary communication target towards the academic audience. In addition, a journal and learning materials are also used to communicate this research.

3.4 Social media methodological frameworks

According to Zeng, Chen, Lusch, and Li (2010), social media analytics involves applying frameworks and tools to collect, analyse, summarise, and visualise social media data. There are various methodological frameworks in social media analytics. These are discussed in the next section, and then the methodological framework that was adapted for this study from existing frameworks is discussed.

3.4.1 Social media and forecasting

The first framework reviewed in social media was the Forecasting framework by Schade (2015). The framework provides an overview and a guideline for analysing consumer-generated social media data to forecast customers' needs, market trends and changes. The generated framework consists of three phases, namely, *organizational background*, *preparation*, and *data analysis*. In the framework, the initial step of the process is concerned with the *organizational background*. The study noted that organisations must position themselves and become aware of necessary change so exploit social media data (Schade, 2015). These preparations can include acquiring new technology to either store or analyse the data. It also entails hiring specialised human resources with experience processing complicated social media data, building a new management framework, and enacting new legislation, such as privacy laws to prevent personal information from becoming public (Schade, 2015).

The second phase of *preparation* begins after the organisational processes have been put in place. In the preparation phase, organisations have to decide what they would like to know, i.e., the goal of the process. Organisations may, for example, concentrate solely on Twitter data or employ social media data from a certain sector. This step also involves data pre-processing to eliminate noisy from the core analytical process (Schade, 2015). The last phase in the framework is the *data analysis* phase which also contains two activities. The first activity is data pre-processing and choosing a suitable data analysis approach. Organisations can use one or more data analytics

approaches to make sense of the vast data to uncover trends and patterns concerning their customers. The second activity in data analysis involves with interpreting and evaluating analytics performances.

The purpose of the evaluation is to assess the results of the data analytics activity. Following evaluation, interpretation is carried out to gain understanding of the findings and put the organisation's acquired knowledge to good use. In regard to this research, the social media forecasting framework could not be adopted because of two limitations. The first is concerned with organisational background. As per the framework, this step requires the acquisition of new technology for data handling, recruitment specialised human resources to handle complex data and introducing new regulations. These requirements could be costly to establish at once for large firms and small firms may not also afford them. The second limitation is that the framework not only lacks the iteration within the steps but also the framework ends with data analysis. The communication of social media analysis findings is not captured. The framework is shown in Figure 3.2.

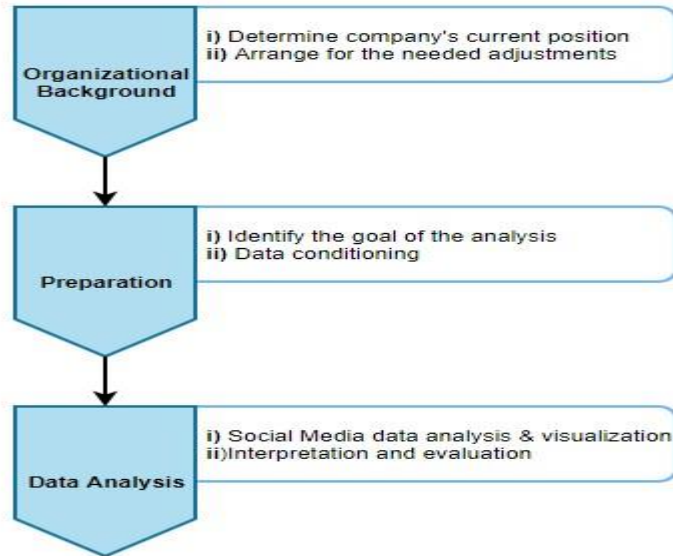


Figure 3.2: Social media framework (Schade, 2015)

3.4.2 CUP framework

Fan and Gordon (2014) developed a social media analytics framework that comprises three stages: *capture*, *understand*, and *present* (see Figure 3.3), the CUP framework. *Capture* refers to the process of extracting social media data from various social media channels and relevant data. It

can be carried out in-house or by a third-party vendor. The capture stage aids a in identifying interactions on social media sites relevant to an organization's activities and interests. Capturing is accomplished by extracting data from social media sites using various tools and techniques (Holsapple, Hsiao, & Pakath, 2018).

After '*capture*', the next stage is to understand where relevant data for analysis is selected while removing noise from the data, using various text processing methods, and gaining insight from them. When a company gathers discussions about its goods and processes, it must determine what they mean (Fan & Gordon, 2014). According to Fan and Gordon (2014), the *understand stage* is the integral process in social media analytics. Its outcomes can directly impact current data and indicators and the progress of future business decisions and actions. Depending on the methodologies employed and the research objective, specific analyses can be pre-processed offline (Fan & Gordon, 2014).

The *present* stage is concerned with presenting the insights from the understand stage in a comprehensive manner. The findings of various analytics are compiled, analysed, and presented to users in understandable format (Fan & Gordon, 2014). Information can be presented using visualisation techniques (Fan & Gordon, 2014).

Overlap exists between some stages; for example, the understanding stage generates models that aid the capture stage. In addition, visualisation supports human judgments that complement the understand stage and enhances the present stage. The stages are carried out sequentially. If the models developed in understand stage fail to uncover meaningful patterns, additional data is captured to increase their predictive power (Fan & Gordon, 2014). If the results lack predictive power, parameters can be modified on *understand* or *capture* stages.

For this research, the CUP framework was further modified because, as it is, the aspect of '*prediction*' is not focused in the framework. While the framework is iterative within the three steps, this research noted the need for the inclusion of prediction just before summarising and presenting the findings.

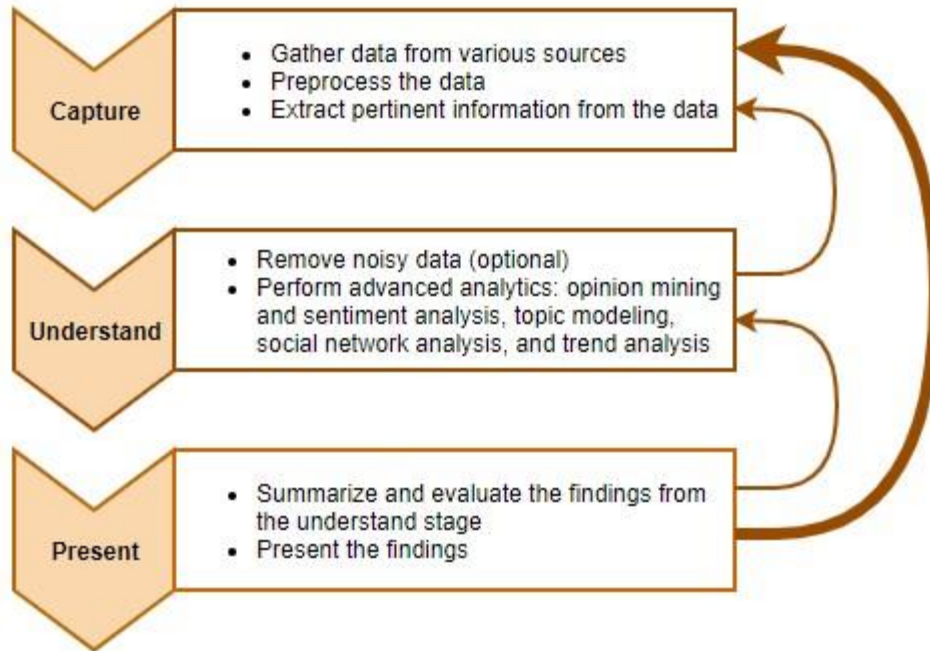


Figure 3.3: CUP framework (Fan & Gordon, 2014)

3.4.3 ICUP framework

Oh, Sasser and Almahmoud (2015) modified the CUP framework proposed by Fan and Gordon (2015) to include the identity stage to identify tweets before the capture stage. The researchers noted that the proposed ICUP framework would be beneficial for both scholars and practitioners in the social media advertising context. The framework could be used specifically in systematic analysis and measurement of ad performance on Twitter (Oh et al., 2015). Thus, the four stages of the modified CUP framework are as follows: *identify*, *capture*, *understand*, and *present* and are illustrated in Figure 3.4 below.

The first stage is to *identify* tweets by determining associated keywords. Whereas some keywords are specific, others are implicitly gathered from other entities. The keywords are then classified into four categories: 1) ad-specific keywords, 2) ad-generic keywords, 3) brand keywords and 4) event keywords (Oh et al., 2015).

The second stage is *capture* which involves two tasks, i.e., downloading and pre-processing of tweets. Keywords extracted in the first stage are used to search for relevant tweet messages. Filtering and removing irrelevant tweets are also done in this stage. Tweets are filtered in three levels. The first level (1) used ad-specific keywords. The second level (2) used ad-generic keywords. The third level (3) filtered brand messages by using brand keywords (Oh et al., 2015).

The third stage is the *understanding* stage and it involves extracting of relevant measures for each tweet and data analysis. These measures consist of tweet likeability characteristics and social media measures. In addition, the sentiment of the tweets is extracted using LIWC. The second task in the understand stage involve classification using machine learning classifiers. The last phase of the framework is *present*. It involves the process of summarisation and reporting of the findings in the SMA framework.

The major drawback of the ICUP framework is the lack of iteration within the steps. Social media analysis is a continuous and iterative process because of the nature of the dataset. Techniques adopted do not always necessarily result in cleaned data or better classifier accuracy. Thus, there is a need for iteration to ensure better classification results.

| Framework Descriptions | | Implementation Descriptions |
|------------------------|---|--|
| 1. Identify | <ul style="list-style-type: none"> Identify relevant keywords to use in collecting social media data | <ul style="list-style-type: none"> Identify relevant keywords from viewing Super Bowl ads to use in collecting tweets about ads and brands |
| 2. Capture | <ul style="list-style-type: none"> Download social media from social media sources using keywords from 'identify' stage Preprocessing | <ul style="list-style-type: none"> Download tweets from twitter API using keywords from the 'identify' stage Preprocessing, removing non-relevant tweets |
| 3. Understand | <ul style="list-style-type: none"> Remove irrelevant tweets Extract relevant metrics Perform analysis | <ul style="list-style-type: none"> Remove irrelevant tweets Extract relevant metrics Perform analysis |
| 4. Present | Summarize and evaluate findings Present the findings | <ul style="list-style-type: none"> Summarize and evaluate findings Present the findings |

Figure 3.4: ICUP framework (Oh et al., 2015)

3.4.4 Iterative CUPP framework

This research formulated an Iterative CUPP framework for social media analysis. This framework was developed based on the shortcomings of the reviewed frameworks and the nature of social media data. Social media data is informal and noisy, which implies that a one-way approach to pre-processing and classification of data is not ideal. This means the need for iteration to further refine the data until optimal results are obtained from the data.

After reviewing the existing social media analytics methodological framework, the Iterative CUPP framework was adapted from the CUP framework. The acronym CUP stands for capture, understand, and present (Fan & Gordon, 2014). *The capture stage* gathers applicable data from social network sites. After collection, the data is preliminarily processed to obtain useful

information using feature extraction (Fan & Gordon, 2014). Through various pre-processing steps, the data is then delivered to the understanding stage.

After *capture*, the next stage is to *understand* the collected data through data cleaning and network analysis. Social media data often consists noise which need to be removed before conducting any analysis. Therefore, NLP techniques are applied to pre-process the data (Fan & Gordon, 2014). In addition, interesting trends and insights concerning users are discovered, covering users' backgrounds, interests, concerns, and relationship networks. According to Fan and Gordon (2014), the results of understand stage have substantial influence on the insights and metrics in the present stage. This stage is critical to the whole social media analysis process since it determines if the *present stage* will be successful.

The last stage of the CUP framework is the *present* stage. In this stage the results of the different analyses are summarised, appraised, and shown to the users (Oh et al., 2015). Various visualisation techniques can be used to present useful information. This stage marks the ends at social network analysis and further analytical process and predictions can be made.

Therefore, the modified CUP framework has four stages – *capture*, *understand*, *predict*, and *present* (*CUPP*). In addition, there is iteration at every stage as shown in Figure 3.5. Data from Twitter that fits within the research objective are collected extracted and pre-processed (*capture*). After pre-processing, sentiment analysis topic modelling and data labelling are done in the understand stage. This stage prepares the data for classification, and after that, prediction happens in the predict stage. After the *understand* stage is completed, classification is done. Classification tasks culminate into a prediction model represented in the predict stage. The predict stage relies on labelled data from the understand stage. Once classifiers are trained, the model can now be used to perform predictions (*predict*). The findings are then summarised and presented in the *present* stage.

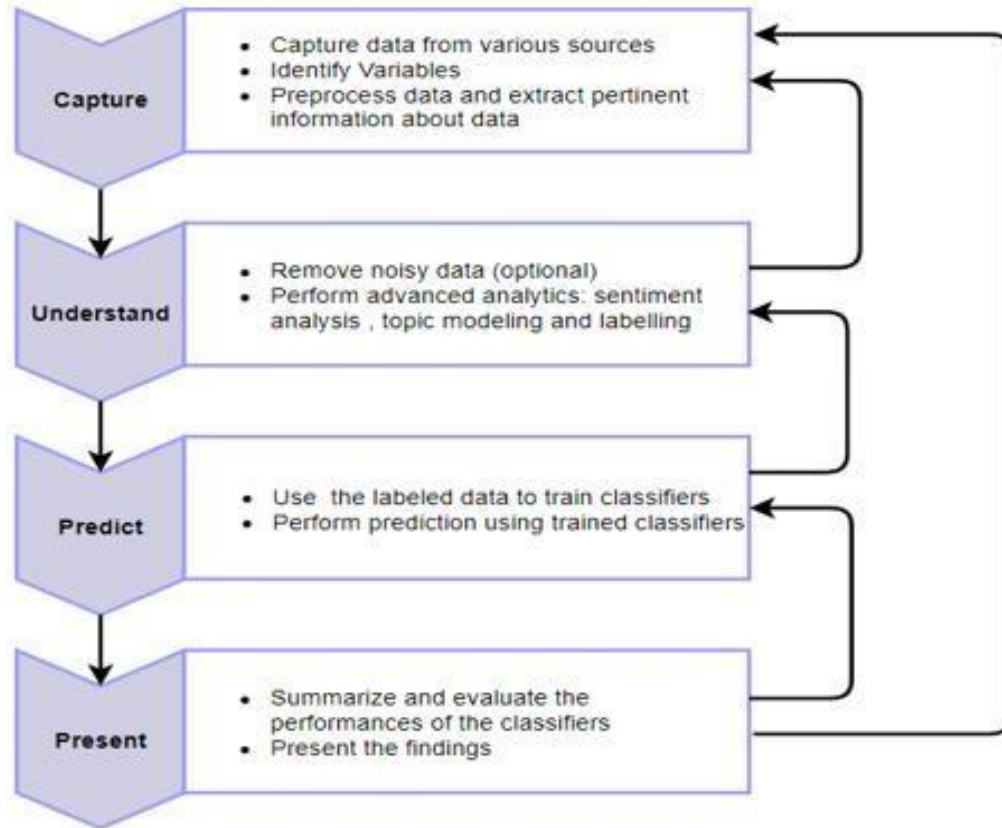


Figure 3.5: Iterative CUPP framework (Author)

3.5 Process model

The main objective of the research was to identify traces of narcissism from social media using a machine learning prediction model. The Iterative CUPP framework was selected as a methodological framework to be able to achieve the research objective. Design science research was presented in Section 3.3, was chosen as the research methodology. Figure 3.6 illustrates the research process model used to implement the four stages of the CUPP framework. The first block indicates a raw dataset that relates to the capture stage of the Iterative CUPP framework, where data from Twitter was extracted. The ‘understand’ stage is implemented through data cleaning, tokenization, labelling, and sentiment analysis in the process model. This process prepares the data for the prediction stage in the Iterative CUPP framework. In Figure 3.6, the application of classifiers and fuzzy logic represent the prediction stage of the Iterative CUPP framework. The output of predictions is presented, which aligns with the last set of the Iterative CUPP framework.

3.5.1 Process model components

The process model consists of five main steps that seek to implement the Iterative CUPP framework. The processes are data cleaning, tokenization and lemmatization, sentiment analysis and labelling, classification and lastly fuzzy logic classification.

Tokenization and lemmatization are the second phase in the model. Tokenization involves the splitting of tweets into words symbols, or some other meaningful element called tokens using a tokenizer. These tokens are form inputs for stemming and lemmatization process. The third phase is data labelling involves adding classes to raw data. This is done through sentiment analysis, topic modelling and lexicon detection. These classes/labels form a representation of what class of objects the data belongs to and is used by machine learning classifiers to identify that particular class of tweets when given new unseen data.

The fourth phase is machine learning classification. This is the training of the labelled data by different classifiers. The predefined categories in this study were grandiose, empath and vulnerable narcissism. The results are a trained classifier that can be used for prediction.

The fifth phase is the prediction of narcissistic personality traits using fuzzy logic. Fuzzy logic is a computation approach based on degrees of truth rather than machine learning prediction, which is based on true or false (1 or 0) logic. Fuzzy logic is incorporated to enhance the prediction of narcissistic personality traits.

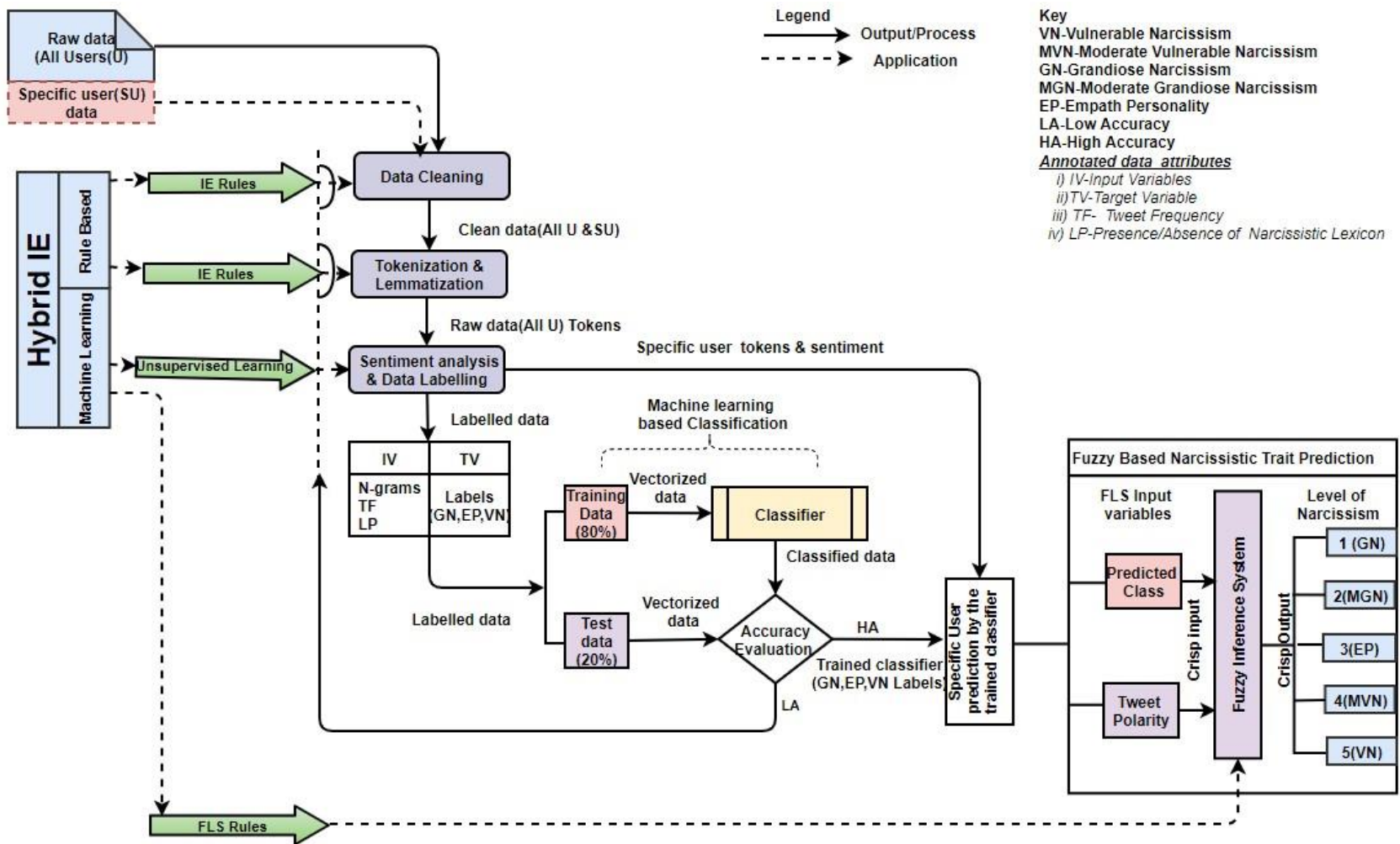


Figure 3.6: Process model for ICUPP framework (Author)

3.6 Data collection

Data collection is an essential component of research (Maxwell, 2012). The effectiveness of data collection is decided by whether or not the purpose of data collection has been identified. The development of the data collection process and the mode in which the data can be evaluated can be aided by defining and understanding this purpose. In the DSR context, the term artefact implies a construction that applies information technology (IT) to organisational tasks. For this study, Twitter was used to gather the primary data. The data was purposively sampled Twitter users. Users had to have at least 20 followers and more than 1000 tweets in their profile. *Rtweet package* in R studio was used to extract tweets from 250 users on Twitter. The researcher used a Twitter developer account which was granted as described in Section 3.9. The access allowed the researcher to access tweets, retweets, likes, favourites, and tweet counts. The prediction model was developed using tweets collected from 15 September 2018 to 15 November 2018, and again retested using tweets collected in April 2021. The experiments conducted in this research study were carried out on a PC, running Windows 10 Enterprise operating system with an Intel(R) Core (TM) i7-470 CPY @ 3.60GHz and 16 GB RAM. A total of 238,317 tweets belonging to 250 users were extracted.

3.7 Data Analysis

In design science research, data analysis entails the use of a set of analytical techniques and methods. However, how the data is analysed and the graphical approaches used to portray the data for better and simpler understanding are all influenced by the researcher's interpretation of the data (Vehkalahti & Everitt, 2018). Exploratory data analysis is done to get an overview of the data before performing data pre-processing. Data analysis in this research was implemented using various Python libraries. Scikit learn was used for classification tasks in the study. Matplotlib library was used to visualise the results. To establish the degree of narcissism, the Fuzzy Logic tool in MATLAB was used.

3.8 Reliability and validity

According to Mohajan (2017), reliability is the degree to which data analysis results can be duplicated with consistency, stability and repeatability. Validity is the accuracy with which a method measures what it is designed to measure (Mohajan, 2017). Cross-validation was used in this study to confirm the validity and reliability of the data. From the literature

review, stratified -k-fold cross-validation is a reliable method of estimating the accuracy of datasets (Hosseini et al., 2020).

3.9 Ethical Considerations

Ethics refers to a set rules of conduct that a researcher should follow in the research process to avoid any misconduct (Resnik, 2015). This research used Twitter datasets that were anonymised by omitting usernames, and Twitter handles during the data cleaning phase to ensure the privacy and confidentiality of the users. Permission and access to the dataset were obtained from Twitter, and ethical clearance was obtained from the University of KwaZulu-Natal's Ethics Committee (See Appendix A). The developer account allows researchers (developers) to get consumer keys and search keys to extract Twitter data (public tweets).

The researcher applied for the permission to use Twitter dataset through a Twitter developer account and permission was granted. To gain access to Twitter data, a researcher has to apply and register a developer account. Application requires a verified phone number and email address, as well as a detailed description of how the API will be used. Use of the Twitter API requires agreeing to Twitter's Developer Agreement and Policy, as well as their related policies. Once a developer account has been established, a developer's Twitter application needs to be registered by providing a name, description and domain. This application authenticates the end user for the application of the Twitter API and gives the user the necessary access key and access token through the app management dashboard. The following script shows the authentication process used in R to register the developer account.

Code listing 3.1: Twitter API authentication process

```
## load rtweet package
library(rtweet)
## create authentication token
create_token(
  app = < APP NAME>,
  consumer_key = < CONSUMER API KEY>
  consumer_secret = <CONSUMER API SECRET KEY>,
  access_token = < ACCESS TOKEN>,
  access_secret = < ACCESS TOKEN SECRET>)
```

3.10 Summary

Since the rise of social media usage, the need to extract information from the social media platforms as an additional source to traditional media has been increasing. Social media networks offer a massive public database of textual data from which vital information can be obtained. This chapter has presented the research design and research methodology used in this research to extract social media data. Social media analytics frameworks were reviewed and an Iterative CUPP framework was developed. While existing frameworks are useful to social media analytics research, they do not offer guidance on how to iterate the processes when evaluation and review of previous steps must be done. Therefore, an Iterative CUPP framework was developed which formed the basis of developing the process model.

CHAPTER 4: RULE-BASED TEXT PRE-PROCESSING

4.1 Introduction

This chapter describes the data processing task, a phase in extracting information from the Twitter dataset. According to Hickman, Thapa, Tay, Cao, and Srinivasan (2020), pre-processing involves transforming text by determining which units (e.g., words and phrases) to use (i.e., tokenize), removing content that is irrelevant for some tasks (i.e., remove nonalphabetic characters and stopwords), and agglomerating semantically related terms to reduce data sparsity and increase predictive power. Therefore, appropriate text pre-processing techniques have to be considered based on the research objective. In this research, text pre-processing was modelled as a rule-based information extraction process to prepare tweets for sentiment analysis and classification. Therefore, the information extraction process comprised rules that would aid in achieving the objective of robust and accurate data ready for a machine learning classifier. Data pre-processing approach adopts an unsupervised rule-based technique approach for information extraction. Tweets were cleaned using a rule-based approach and tokenized using TweetTokenizer. The pre-processed data formed the input for the machine learning classifiers.

4.2 Data extraction

There are three possible ways to extract Tweets for research from Twitter. The first approach is by getting data from open-source repositories like UCI Machine Learning Repository (Nair, Shetty, & Shetty, 2018). The second approach is through APIs (Nair, Shetty, & Shetty, 2018). The search API and the stream API are two types of APIs offered by Twitter and are used to acquire Twitter data based on hashtags and stream Twitter data in real-time. Lastly, data can also be obtained from companies that sell Twitter data for commercial purposes like Audience, Risetag, or Tweetdeck. Kasture (2015) extracted 1,313 posts from Twitter to conduct a study on cyberbullying detection. Squicciarini, Rajtmajer, Liu, and Griffin (2015) extracted 16,000 posts for use to detect cyberbully on Twitter and identify cyberbullying interactions. For this research, approximately 240,000 tweets were extracted from 250 users from Twitter. This research chose Twitter API through a Twitter developer account (see Section 3.6) to extract data. The study used the *rtweet* package on R statistical software to extract the data into CSV format. During extraction, a set of users were randomly selected based on three conditions. The first

condition was that a user must have had at least 1000 tweets by the time of extraction. The second condition was that a user account must have been in existence for at least one year. The last condition was that a user must have at least 20 followers.

4.2.1 Dataset choice

Psychologists postulate that social media provides a simple way to fit in with others (AbdelKhalek, 2016). According to Dandannavar et al. (2018), a user's behaviour on social media is reflected in their tweets, status updates, comments, and interests. Therefore, it is feasible to predict personality automatically using social media data, which is cost-effective and more efficient than traditional methods (Dandannavar et al., 2018). On social media, practically everyone shares content, however the nature of content shared varies depending on the user's personality (Hruska & Maresova, 2020). Therefore, Twitter was selected as the primary data source, as it has been proven to be suitable for analysing personality traits on social media. In addition, Twitter was chosen because Twitter users generate short messages, limited to 280 characters. Secondly, Twitter is a convenient platform to access and collect data. Lastly, Twitter is an instant day-to-day micro-blogging platform and is widely considered to have an advantage over other social network sites like Facebook which not only consists of text but also has videos and photos which make it challenging to study personality with such different variables of data (Park, Park, & Chong, 2020). On the other hand, Twitter provides chronologically ordered posts from each user, making it easy to analyse the data referring to narcissism.

Given the prevalence of social networking sites today, Twitter has emerged as one of the most widely used social networking services worldwide (Fearnley & Fyfe, 2018). Twitter emerged to be an effective instrument for gauging societal interests and popular sentiment in recent years. Twitter has been labelled as a type electronic word-of-mouth marketing by organisations (Zhou et al., 2017). It has also been utilised as an online monitoring platform to gauge public opinion on public health concerns such as immunizations. Twitter has also served as an online surveillance platform about public health issues like vaccines and track public health trends on influenza outbreaks (Zaldumbide & Sinnott, 2015). Therefore, Twitter is an excellent online forum for personality research and application. Personality is essential in various situations; it can predict job happiness, professional and personal relationship success, and even interface preferences.

Since narcissists have inflated self-esteem and use a number of techniques to draw attention to themselves; elements unique to Twitter are more attractive to them than other social media channels like Facebook (Davenport, Bergman, Bergman, & Fearington, 2014). Furthermore, McKinney, Kelly, and Duran (2012) found a link between user volume and narcissism, leading them to conclude that Twitter might be the favoured social media site for narcissists and called for more research. Davenport et al. (2014) noted the association between active Twitter usages is expected, and a closer analysis of the reasons for social media usage and motives users should reveal other strong connections with narcissism. Furthermore, according to Reed et al. (2018), people with high degrees of narcissism tend to use Twitter frequently. However, those who used Facebook grew more narcissistic over time.

4.2.2 Nature and the structure of the dataset

According to Kwon and Sim (2013), previous research classified dataset attributes into three categories. The first category is concerned with the features of the dataset's structure. The second category relates to the attributes of the data set's value of features and target class. The last category is contextual, and it refers to the data set's domain characteristics where the classifier effectively distinguishes one class from the others (Kwon & Sim, 2013). Dataset characteristics are vital in determining the classifier performance. The effectiveness of the classifier may also be affected by interactions between the features in the data set. The purpose of the data extraction task was to extract variables from the Twitter dataset and store extracted dataset in a defined template. The downloaded dataset is in CSV format and contains user_id, created at, screen_ name, text, hashtags, description, user followers, friends count, status count, favourites count, and account_ created at and date when the tweet was created (Table 4.1). The text attribute is pre-processed and used in classifiers. In addition, the status count which shows the frequency of tweeting by individual is also used in training the classifiers.

Table 4.1: Dataset attributes

| Dataset attribute | Description and sample data |
|--------------------|---|
| User id | 958993373536358401 Unique id assigned to the user by Twitter |
| created at | 9/16/2018 Time stamp when a tweet/retweet/Reply post occurred. |
| Screen name | @crypto_guru This refers to the name a user calls themselves on Twitter |
| Text | These murders are happening (to all races and farm workers) and the details are horrific. https://t.co/L0i2g0ngLF |
| Hashtags | #StimulusPackage A hashtag is identified by the symbol (#) in front of any term and they classifying tweets about a specific topic. |
| Description | I care about people This relates to how a user describes themselves in person. Description may also refer to causes a user subscribes to. |
| user followers | 94 A Twitter user's "followers are different accounts that subscribe to the user's posts and updates". "When someone follows an account, it will appear in that account's followers list". |
| friends count | 136 This refers to the number of users a tweep is following. |
| statuses count | 6800 This refers to the number of tweets a user has made since he or she created a Twitter account. |
| favourites count | 5627 This relates to the number of times a tweet has been liked (favourited) by other users. |
| account created at | 2/1/2018 Date when the user created a Twitter account. |

4.3 Data cleaning

Data cleaning is an essential process in natural language processing tasks. Extracted raw tweets are naturally noisy and contain numbers, slangs, missing values, noise, or redundancies (Desai & Mehta, 2016). Data cleaning is a crucial step in the prediction process. By its very nature, Twitter data is noisy, full of typos, informal language, slang, and unwanted content such as URLs and idioms (Pereira, 2017). As a result, effective data cleaning is essential to gain deeper insight and develop great models (Ridzuan & Zainon, 2019). In this step, data cleaning was conducted to convert the tweets to lowercase, remove URLs, hashtags, @, and alphanumeric characters in the tweets that do not provide much semantic meaning to the study.

Regex library in Python was used to eliminate special characters in tweets, such as retweet (RT), users (@), URLs (<http://url>), and punctuation. According to Elbagir and Yang (2019),

hashtags (#) are frequently used to explain a tweet's topic and provide relevant information about a topic. As a result, they are included in the tweet, but the "#" symbol has been omitted. The tweets were then converted to lowercase, stopwords were removed, and the tweets were tokenized into tokens using NLTK methods (See Procedure 4.2).

4.3.1 Data cleaning rule and algorithm

Data pre-processing is critical in generating insights and knowledge discovery from unstructured data (Srividhya & Anitha, 2010). Therefore, extracting useful information from the data required comprehensive data cleaning. By nature, social media text is informal language, short context, and noisy, sparse contents (Virmani et al., 2017). Furthermore, Twitter has its metadata, such as "RT," which denotes Tweets that other users have retweeted, "#," which denotes hashtags, and "@," which denotes other Twitter users. (Liza, 2020). In this research, Key Tweet characteristics were considered when pre-processing the Tweets, and rulebased text pre-processing approach is adopted. A rule for data cleaning was created, and a procedure for implementing the rule is presented. Rule 4.1 was developed to pre-process the data as shown in Procedure 4.1.

Rule 4.1

A clean tweet can only be lowercase and should not have URL (http) links, user handles (@user), special characters, numbers, and punctuations (^a-zA-Z).

Sub rules under 4.1

Rule 4.1: A Clean Tweet is a Tweet with

Rule 4.1.1: Only lowercase

Rule 4.1.2: No URL

Rule 4.1.3: No @user handles

Rule 4.1.4: No special characters

Rule 4.1.5: No numbers or white spaces

Pre-processing is a critical step that has an impact on accuracy of learning models. Rule 4.1 sought to remove usernames, link punctuations, and numeric values from tweets. Punctuation like hashtag symbols, affect classifier interpretation and are thus eliminated. Consequently, all words are transformed to lowercase. In addition, URLs links are removed since the content

of the links is not analysed as it does not provide any helpful information. Rule 4.1 is implemented using Procedure 4.1.

Table 4.2: Pre-processing variables

| Variables | Description |
|---------------------|---|
| Raw Tweet | This relates to tweet extracted from twitter users which has not been cleaned and is still at its informal and noisy form |
| Pre-processed Tweet | This relates to cleaned tweet which has been pre-processed as per Rule 4.1 |

The above two variables in Table 4.2 are used in Procedure 4.1. Raw tweet is the input variable while pre-processed tweet is the output variable.

Procedure 4.1: Data cleaning

Input: Raw Tweet

Output: CleanTweet (Pre-processed Tweet)

Process:

1. For each Raw Tweet
 - 1.1 Initialize temporary column CleanTweet to store the output
 - 1.2 If a tweet is uppercase Then
 - 1.2.1 Convert the tweet to lowercase
 - 1.3 If a tweet has URLs (for example <https://ukzn.ac.za>) Then
 - 1.3.1 Replace all URLs or https:// links with the word 'URL' using regular expression methods and store the result in CleanTweet
 - 1.4 If a tweet has @username Then
 - 1.4.1 Replace all with the word 'AT_USER' and store the result in CleanTweet.
 - 1.5 If a tweet has #hashtags and RT Then
 - 1.5.1 Filter all and store the result in CleanTweet
 - 1.6 If a tweet has any additional special characters ((: \ | [] ; { } - + () < > ? ! @ # % *,) Then
 - 1.6.1 Remove the special characters
 - 1.7 Remove the word 'URL,' which was replaced in step 1, and store the result in CleanTweet.
 - 1.8 Remove the word 'AT_USER,' which was replaced in step 1, and store the result in CleanTweet.
 - 1.9 Return CleanTweet
 - Endif
 - EndFor
-

Table 4.3: Tweet pre-processing for labelling

| Dataset attribute | Sample data |
|--|--|
| Raw Tweet | If you're calling those South African farm murders a conspiracy theory you're an actual rape apologist and child murder lover. These murders are happening (to all races and farm workers) and the details are horrific. https://t.co/L0i2g0ngLF |
| Lowercased Tweet | If you're calling those south african farm murders a conspiracy theory you're an actual rape apologist and child murder lover. these murders are happening (to all races and farm workers) and the details are horrific. https://t.co/l0i2g0nglf |
| URL removed from tweet | https://t.co/l0i2g0nglf |
| Pre-processed tweet (#, @, URL, special characters removed) | if you're calling those south African farm murders a conspiracy theory you re an actual rape apologist and child murder lover these murders are happening to all races and farm workers and the details are horrific. |
| Recognised tokens | ['if', 'you', 're', 'calling', 'those', 'south', 'African', 'farm', 'murders', 'a', 'conspiracy', 'theory', 'you', 're', 'an', 'actual', 'rape', 'apologist', 'and', 'child', 'murder', 'lover', 'these', 'murders', 'are', 'happening', 'to', 'all', 'races', 'and', 'farm', 'workers', 'and', 'the', 'details', 'are', 'horrific']. |

The output of Procedure 4.2 is shown in Table 4.3 above which shows how raw tweet is pre-processed till its tokenized.

4.4 Tokenization and Lemmatization

After pre-processing, the tweets are broken down into tokens in preparation for feature extraction and classification (Allahyari et al., 2017). Lemmatization is another significant NPL task. It is the process of replacing a given token with its corresponding lemma. The main idea behind lemmatization is to reduce sparsity, as different inflected forms of the same lemma may infrequently occur during training (Balakrishnan & Lloyd-Yemoh, 2014).

4.4.1 Tokenization rule and procedure

The tokenization process divides the text data into pieces (or tokens) and often removes special characters such as apostrophes, commas, and periods (Procedure 4.2). The tokens are typically made up of a single word or an N-gram (Purnamasari & Suwardi, 2018). N-grams are sequences of adjacent items, which in this case are words (Mullen, Benoit, Keyes, & Selivanov, 2018). Tokenization helps the classifiers to learn more accurately and efficiently from the dataset. Tokenization aims to split sentences into words to represent each tweet as a feature vector (Elbagir & Yang, 2019). In this study, TweetTokenizer, a python library was used to split pre-processed tweets into tokens and is implemented using Rule 4.2.

The tokens aid in interpreting the context or developing an NLP model (Hapke, Howard, & Lane, 2019). For this research, individuals' words/terms were considered tokens. White space, line breaks, and any other token that is not a word are discarded and not recognized as tokens (Procedure 4.2). The next stage after tokenization is sentiment analysis which relies on words to get a sentiment score. Thus, white spaces could not be considered tokens. Recognised tokens were then lemmatized as shown in Procedure 4.3.

Rule 4.2

For each clean tweet, recognise tokens

Table 4.4: Tokenization and lemmatization variables

| Variables | Description |
|---------------------|--|
| Pre-processed tweet | This relates to cleaned tweet which has been pre-processed as per Rule 4.1 |
| Recognised tokens | This relates to valid tokens recognised as per Rule 4.2 |
| Recognised lemmas | This relates to valid lemmas recognised as per Rule 4.2 |

The output of Procedure 4.1 (pre-processed tweet) serves as the input variable for Procedure 4.2. The output of Procedure 4.2 becomes the input for Procedure 4.3. The variables are described in Table 4.4.

Procedure 4.2: Tokenization

Input: Pre-processedTweet

Output: RecognizedTokens

```
1. For each Pre-processedTweet;  
    1.1 Initialize an empty string RecognizedTokens to store the  
        result of output.  
    1.2 Check if pre-processedTweet has complete words  
    1.3 If the Pre-processedTweet has complete words  
    1.4 Split the Pre-processedTweet into tokens  
  
    Else Remove white spaces  
  
    1.5 Return RecognizedTokens  
  
endif  
End For
```

Procedure 4.3: Lemmatization

Input: RecognizedTokens

Output: RecognizedLemmas

```
1. For each Token;  
    1.1 Initialize an empty string RecognizedLemmas to store the  
        result of output.  
  
    1.2 Perform Lemmatization on RecognizedTokens  
  
    Else Remove white spaces  
  
    1.3 Return RecognizedLemmas  
  
End For
```

Figure 4.1 shows the process of tokenization to obtain valid tokens and transforming the tokens into lemmas. Lemmatization takes the recognised tokens and reduces them to their base form through morphological analysis (Maisto, Pelosi, Polito, & Stingo, 2019).

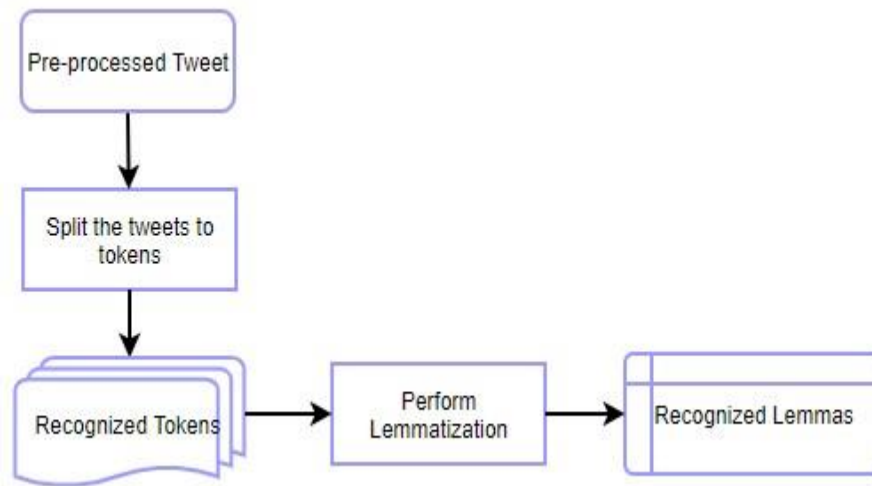


Figure 4.1: Tokenization and lemmatization flow

The output of tokenization and lemmatization is shown in Table 4.5 below.

Table 4.5: Tokenization and Lemmatization output

| Output | Sample data |
|---------------------|--|
| Pre-processed Tweet | if you re calling those south african farm murders a conspiracy theory you re an actual rape apologist and child murder lover these murders are happening to all races and farm workers and the details are horrific |
| Recognised tokens | ['if', 'you', 're', 'calling', 'those', 'south', 'african', 'farm', 'murders', 'a', 'conspiracy', 'theory', 'you', 're', 'an', 'actual', 'rape', 'apologist', 'and', 'child', 'murder', 'lover', 'these', 'murders', 'are', 'happening', 'to', 'all', 'races', 'and', 'farm', 'workers', 'and', 'the', 'details', 'are', 'horrific'] |
| Recognised lemmas | ['if', 'you', 're', 'call', 'those', 'south', 'african', 'farm', 'murder', 'a', 'conspiracy', 'theory', 'you', 're', 'an', 'actual', 'rape', 'apologist', 'and', 'child', 'murder', 'lover', 'these', 'murder', 'be', 'happen', 'to', 'all', 'race', 'and', 'farm', 'worker', 'and', 'the', 'detail', 'be', 'horrific'] |

4.5 Summary

Data pre-processing relates to cleaning methods applied to raw data to prepare it for analysis (Paulauskas & Auskalnis, 2017). Data pre-processing methods transform the data into a format ready for classification. Figure 4.2 shows the summary of text-pre-processing steps.

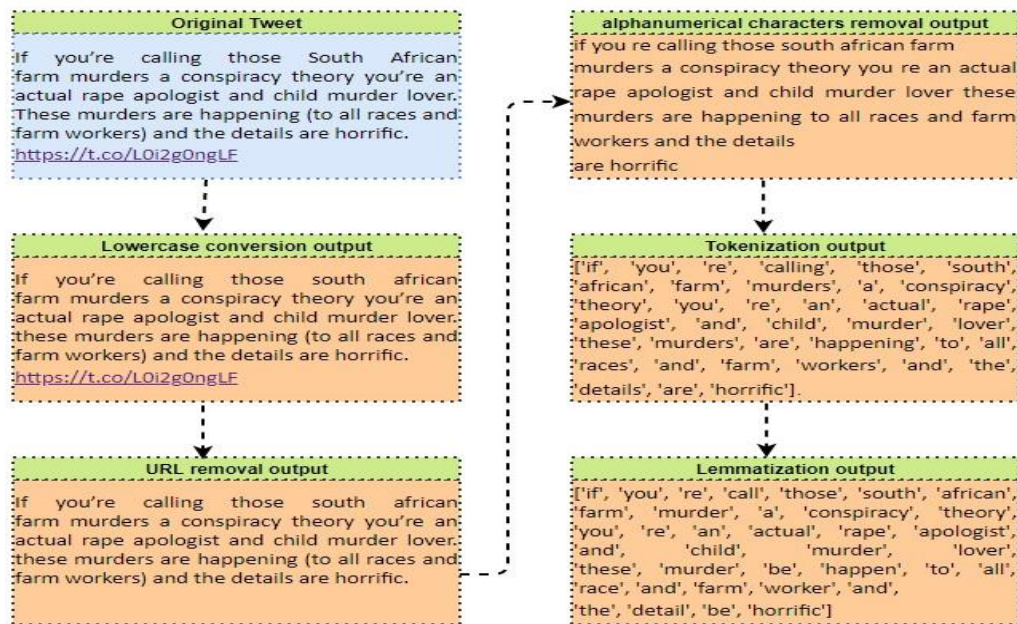


Figure 4.2: Rule-base tweet pre-processing

Machine learning models are only as good as their data, and no matter how good the trained model is, the ultimate problem may lie in the data itself (Tae, Roh, Oh, Kim, & Whang, 2019). According to HaCohen-Kerner, Miller, and Yigal (2020), applying different pre-processing methods can improve text classification results. Knowledge discovery may yield incorrect results if the data is noisy or contains irrelevant and redundant information. Consequently, not all pre-processing techniques are suitable for all text classification problems as some might influence the classification results. In this chapter, three primary data pre-processing methods were discussed, and justification for the choice of each technique was highlighted. The output of this chapter becomes the input of Chapter 5, where the pre-processed data undergoes sentiment analysis and data labelling.

CHAPTER 5: TWEET SENTIMENT ANALYSIS AND DATA LABELLING

5.1 Introduction

Sentiment analysis or opinion mining is the process by which text is analysed to extract opinion and assign a relevant sentiment. The sentiment can be positive, negative or neutral (Bonta & Janardhan, 2019). On Twitter, sentiment analysis involves identifying and categorising the polarity of a tweet made by a user where the goal is to establish whether the tweet is either positive, neutral, or negative. Sentiment analysis of text presumes that lexical items found in the text carry attitudinal loading or the affective state of the author (Diamantini, Mircoli, DiaPotena, & Storti, 2019). Subsequently, data labelling, or data annotation, is the process of segmenting and assigning labels to a dataset before training using supervised machine learning techniques (Cruz-Sandoval et al., 2019). This chapter presents the sentiment analysis that was undertaken to detect the emotions and sentiment of users from their tweets. The data was then labelled for classification using a combination of topic modelling and lexicon detection. Lexicon detection involves utilising the created dictionary (see Section 5.6). In the dictionary, lexicons related to narcissism were selected based on psychology literature, as discussed in Chapter 2, to identify the relationship between narcissism and sentiment in the virtual environment. VADER sentiment analysis tool was chosen over other existing tools because of its suitability to analyse social media data and short text compared to other sentiment analysis tools (Sim, Miller, & Swarup, 2020).

5.2 Sentiment analysis

As discussed in Section 5.1, sentiment analysis relates to a computational way of identifying and classifying the views expressed in a line of text to determine if the writer's attitude toward a specific subject is positive, negative, or neutral (Raghuwanshi & Pawar, 2017). Attitude is usually based on two aspects. First, the individual's judgment and the evaluation of events. Secondly the individual's affective state. Affective state relates to the author's emotional state when writing a review. Sentiments are the thoughts triggered by the feelings associated with events and they are often classified as positive, neutral, or negative (Zhao, Quin, Liu, & Tang, 2016). There are two standard techniques for identifying sentiments from texts: lexicon-based technique and machine learning

techniques. According to Almatarneh and Gamallo (2018), the lexicon-based technique involves extracting opinion-based lexicons from the text and then determining the polarity of those lexicons. According to Desai and Mehta (2016), lexicons are a collection set of predefined sentiment terms. The lexicon-based approach assigns specific weights to each text based on its polarity of association (negative, positive or neutral). This methodology can be performed using SentiWordNet or VADER, among others (AlShabi, 2020).

Machine learning approach involves using trained data to find sentiments in new data. With the use of tagged corpora, supervised learning techniques construct a sentiment classifier utilising feature from text data (Pavan & Prabhu, 2018). The lexicon-based sentiment analysis approach was chosen for this research because the dataset is not yet labelled at this stage. The pre-processed dataset in Section 4.5 is not labelled; thus, lexicon-based sentiment analysis was chosen. There are different types of open-source lexicon tools. These include, Textblob, LIWC, SentiWordNet and VADER

5.2.1 TextBlob

TextBlob is a python library that consists of modules for text mining , sentiment analysis and processing (Desai & Mehta, 2016). TextBlob's polarity ranges from [-1.0, 1.0], where -1.0 is negative polarity, and 1.0 is positive polarity. This score can potentially be zero, which denotes a statement's neutral evaluation because it does not contain any terms from the training set.

5.2.2 LIWC

LIWC has over 4,500 words divided into 76 categories, with 905 terms in two types (Gilbert & Hutto, 2014). Though LIWC has been widely used to find sentiment analysis in social media text, it excludes acronyms, initialises, emoticons, and slang, all of which are important components in sentiment analysis of social media text.

5.2.3 SentiWordNet

SentiWordNet is a sentiment analysis tool based on a WordNet Lexicon (Hardeniya & Borikar, 2016). To identify the text's sentiment, SentiWordNet combines three scores with a WordNet dictionary synset: positive, neutral, and negative. A semi-supervised machine learning technique is used to generate the scores in the range [0, 1] that sum up to 1. SentiWordNet has 117,374 synsets with positive, neutral, and negative scores (Kou & Peng, 2015).

5.2.4 VADER

VADER is an open-source application for sentiment analysis in the English language. VADER contains a systematically built sentiment lexicon and some syntactic rules to improve further sentiment analysis (Bonta & Janardhan, 2019). VADER was constructed especially for tweets and contains both abbreviations and emojis. Emojis are emotional tokens often used on the Internet (Pano & Kashef, 2020). VADER uses four rules to handle sentiments in a social media text. The first rule is the exclamation and interrogation marks. According to Gilbert and Hutto (2014) rules, these marks can increase or decrease the sentiment intensity. The second rule is capitalisation rule. As per VADER, if a word is capitalised while others are not, the sentiment intensity increases for this word. The third rule is the rule on negators. This rule states that, if a word is preceded by a negator such as “not”, this reverts the sentiment. A positive sentiment turns negative and vice versa. The last rule is ‘Booster Words’, such as “extremely”, which if positioned before the word “good” will increase the sentiment of this word.

According to Gilbert and Hutto (2014), VADER performs well in the social media domain. VADER retains the benefits of traditional sentiment lexicons like Linguistic Inquiry and Word Count (LIWC) (Bonta & Janardhan, 2019). VADER’s sentiment lexicon was built for Twitter, and the VADER application also includes an emoji lexicon with 3570 emojis with their textual description. Therefore, VADER was chosen over TextBlob because of the above rules and approach to sentiment analysis. The VADER sentiment lexicon consists of 7,517 entries in the English language, including abbreviations and slang words. The researchers used methods of monitoring the labelling of words to heighten the lexicon's quality. Using Vader, the pre-processed tweets were classified as positive, negative and neutral. VADER puts into consideration slang, capitalisation, and the way words are written, as well as their context in sentiment analysis (Newman & Joyner, 2018). It considers up to three exclamation marks that add additional positive or negative intensity (Gilbert & Hutto, 2014). VADER also supports emoji sentiments. As a result, VADER was chosen since its better option for analysing tweets and their sentiments.

5.3 Experiment setup: Sentiment analysis

The first step of the sentiment analysis was to determine each tweet's sentiment. Sentiment analysis was done using the VADER library in Python. This research aimed at extracting a dataset for the prediction of narcissistic personality traits. This experiment consisted of

input variables and output variables. The *input variable* is clean tweet (Procedure 4.1). The sentiment analysis tool of VADER was used. VADER recommends an aggregated sentiment polarity ≥ 0.5 as high positive, neutral between $[-0.5, +0.5]$, and high negative for sentiment polarity < 0.5 (Saldaña, 2018; Sarkar, 2016). A threshold of ≥ 0.5 for positive and < -0.5 for negative was used in the dataset. The *output variable* for this experiment was TweetPolarity. Three polarities were considered, namely: Positive Polarity -*PosP*. This relates to tweets that have a sentiment polarity of between $+0.5$ and $+1$. The second polarity is Neutral Polarity denoted as *NeuP*. These are tweets that have a sentiments core of -0.5 to $+0.49$. The last polarity considered was negative polarity denoted as *NegP*. These are tweets that have a sentiment polarity of -0.5 to -1 . Sentiment analysis was implemented following as Procedure 5.1.

Procedure 5.1: Sentiment analysis

Input: CleanTweet

Output: Polarity

Process:

1. For each CleanTweet;
 - 1.1 Initialize temporary column *TweetPolarity* to store the tweet sentiments
 - 1.2 Calculate the polarity of each Tweet at sentence level
 - 1.3 If sentiment polarity is between 0.5 and 1 Then
 - 1.3.1 Label the sentiment as *Positive Polarity* (PosP)
 - Else If sentiment score is between $+0.5$ and -0.5 Then
 - 1.3.1 Label the sentiment as *Neutral Polarity* (NeuP)
 - Else
 - 1.3.1 Label the sentiment as *Negative Polarity* (NegP)
 - Endif
 - EndFor
-

5.3.1 Sentiment analysis results

In Table 5.1, most of the tweets were positive, followed by neutral and negative tweets respectively. Only 25% of the tweets seem to reflect some form of negativity. The whole dataset had 238,317 tweets from approximately 250 users.

Table 5.1: Sentiment distribution of the dataset

| Sentiment | Value count |
|-----------------|-------------|
| Positive (PosP) | 90321 |
| Neutral (NeuP) | 88869 |
| Negative (NegP) | 59127 |

The distribution of various words in the respective sentiment category is shown in Figures 5.1, 5.2, and 5.3 in the form of word clouds. A group of words presented in various sizes is called a word cloud. The more frequent a word occurs in the dataset, the larger and bolder it appears (Heimerl, Lohmann, Lange, & Ertl, 2014).



Figure 5.1: Positive polarity word cloud



Figure 5.2: Neutral polarity word cloud

data was then used in classification. Grandiose narcissism refers to words that relate to aggression, grandiosity, and dominance. Empath personalities refers to words that describe more empathy than the average person (Lannin, Guyll, Krizan, Madon, & Cornish, 2014). These words used by empaths are usually more accurate in recognising emotions. They are also more likely to recognise emotions earlier than grandiose and vulnerable individuals. Vulnerable narcissism relates to words that describe defensiveness, psychological distress, anxiety, depression, negative emotions, and feelings of inferiority (Krizan, 2018).

5.5 Semi-supervised topic modelling

To build topics related to narcissism, tokens were grouped into similar relations based on their semantic relatedness (Allahyari et al., 2017). Relation refers to words or groups of words that convey information deemed to be closely related. To cluster the relationships between tokens, the unsupervised learning algorithm of Latent Dirichlet Allocation (LDA) (Procedure 5.2) and CorEx Topic algorithm (Procedure 5.3) are used. After clustering, top words per cluster are generated. If the clusters generate incoherent relations, a further cluster refinement is performed by manipulating the topic word representations. This is done through choosing anchor words, i.e., words that have the highest mutual information with the cluster (Gallagher, Reing, Kale, & Ver Steeg, 2017). Topic modelling approaches are used to find groups in a set of data that are more similar to one another than to those in other groups (Gupta, Banerjee, & Rubin, 2018).

5.5.1 Latent Dirichlet Allocation (LDA)

LDA is an unsupervised machine-learning technique that takes whole documents (tweets of users) as input and finds topics as output (Blei, Ng, & Jordan 2003). It extracts themes from a corpus based on word frequency (Procedure 5.2) (Liu, Preotiuc-Pietro, Samani, Moghaddam, & Ungar, 2016). The model also generates the percentage of what each document talks about each topic (Onan, Bulut, & Korukoglu, 2017). LDA posits that texts are made up of a variety of topics of which generates words based on their probability distribution (Wood, Tan, Wang, & Arnold, 2017). The three basic parameters of LDA are the number of words per theme, the number of themes, and the number of subject matters per document (Maier et al., 2018). Topic modelling using LDA had two main variables. The input variable was '*Documents/Corpus*'. This relates to a dataset consisting of different tweets that as a whole form a document. The parameters used were $k=7$. This relates to the

number of topics to be generated by the algorithm. The output variable was '*Topics*'. This relates to the distribution of related words in a document.

Procedure 5.2: LDA Algorithm

Input: CleanTweet []

Where CleanTweet is an array of Tweets which is referred to as Corpus

Output: Topics & Topic Keywords

Process:

1. For each tweet in corpus
 - 1.1 Initialize the number of topics to be generated (k)
 - 1.2 Count number of occurrences of each word
 - 1.3 Find the word has maximum count.
 - 1.4 Generate the top words per each topic (k)
 - 1.5 Choose the most probable topic from the top words generated
 - 1.6 Given a topic, choose a likely word that belong to a topic
 - 1.7 Generate Topics and Topic Keywords
- EndFor
-

5.5.2 Correlation explanation (CorEx)

The CorEx model allows the incorporation of domain knowledge through user-specific anchor words which guide the model towards the topics of interest. This enables the model to represent topics that do not naturally emerge and provides the ability to separate keywords allowing distinct topics to be identified (Gallagher et al., 2017). The anchor keywords are sets of keywords assigned to each topic. The CorEx model also has a strength parameter that defines the bias of the topics generated towards the anchor keywords. This value should always be above 1 and higher values indicate a stronger bias towards the anchor keywords.

5.5.2.1 Anchoring in CorEx

CorEx has an extension called 'anchoring', where words are anchored to specific topics. CorEx seeks to optimise the mutual information between a word and the anchored topic while anchoring it to a topic (Ver Steeg, 2017). Anchoring provides a way to guide the topic model towards specific subsets of words that the user wants to explore. Anchored

CorEx permits a user to associate certain words to topics in a semi-supervised manner, therefore allowing identification of otherwise elusive topics (Gao, Brekelmans, Ver Steeg, & Galstyan, 2018). For this research Corex was implemented using selected parameters. The first parameter was “*n_hidden*”. This relates to number of hidden latent topics. This was set to three. The second parameter was “*seed*” which refers to the random number seed to use for model initialisation.

Procedure 5.3: CorEx Algorithm

Input: CleanTweet []
 Where CleanTweet is an array of Tweets which is
 referred to as Corpus
Output: Topics & Topic Keywords

Process:

1. For each tweet in corpus
 - 1.1 Initialize the number of latent topics to be generated (*n_hidden*) and a random seed number
 - 1.2 Initialize the anchor words that represent each topic
 - 1.3 Count number of occurrences of each word
 - 1.4 Find the word has maximum count.
 - 1.5 Generate top words that represent topics as per the anchor words in 1.2
 - 1.6 Return Topics and Topic Keywords
- EndFor
-

5.6 Narcissistic lexicon construction

According to Darwich, Mohd, Omar, and Osman (2019), lexicons can be created using either a dictionary-based approach or a corpus-based approach. The dictionary-based approach is constructed by developing a list of seed-related words manually, while in a corpus-based approach, seed words are collected using statistical and semantic methods. One of the most crucial aspects in the lexicon-based method is the quality of the dictionaries and lexicons utilized in the categorization process (Bonta & Janardhan, 2019). Some research focuses on the creation of lexicons to address concerns such as domain adaptiveness (Wang & Xia, 2017), while some focus identifying universal sentiment lexicons and constructing a proper scoring system based on the lexicons. Kundi, Khan, Ahmad, and Asghar (2014) developed a scoring system for sentiment prediction and

created a slang dictionary containing a set of slangs annotated with scores and orientations using a weighted threshold value obtained based on the SWN lexicon.

According to Liu, Burns and Hou (2017), linguistic content of social media posts reveals a user's topic of interest and provides information about their lexical usage which may be predictive of specific user types. The link between linguistic clues and personality traits has been established as a crucial component of automatic personality classification (Kaushal & Patwardhan, 2018). To develop the dictionary, different criteria and measures used in psychology literature were used. The unique words and phrases describing narcissistic personality types were extracted using a frequency-based N-gram technique. The narcissistic lexicon is the key to identifying traces of narcissism through the lexicon-based approach. Previous research has shown that unsupervised approaches to constructing a dictionary rely on the context of words. The construction of the dictionary was done in two steps. The first step was *seed term collection* and entailed collecting seed terms related to narcissism from literature. The second step was *value assignment*. This involved assigning the narcissism orientation value to the lexicon. The last step was *evaluation* which involved evaluating and refining the narcissistic lexicon.

5.6.1 Seed term collection

In this research, narcissistic seed terms refer to words conveying information related to grandiose and vulnerable narcissism, as discussed in Section 2.11. The narcissistic lexicon consists of both universal and domain-specific terms. Prior studies suggest that the domain-specific lexicon contributes to improved classification rather than sentiment since opinions differ greatly across different domains (Qin & Petrounias, 2017). The use of domain-specific words can make it easier to spot explicit and implicit opinions in a writing. In a subjective sentence, a text with an explicit opinion expresses a positive or negative opinion, whereas a text with an implicit opinion implies objective sentence (Liu, Christiansen, Baumgartner, & Verspoor, 2012).

5.6.2 Value assignment

In this stage of lexicon construction value is assigned to the terms based on the two categories of narcissism. This is because each individual word's semantic orientation can be expressed as a numerical value (Qin & Petrounias, 2017). Khan et al. (2016) adopted this approach in their sentiment lexicon construction. A numerical value is added to a

lexicon to indicate the strength of the word and the orientation of the word category in the dictionary, which is vulnerable to grandiose narcissism. Thus, in this research, grandiose lexicons were assigned value 1 and vulnerable lexicons were assigned value 2.

5.6.3 Evaluation of narcissistic lexicon

The evaluation stage in the design science research is crucial because feedback from the evaluation process helps in refinement of the design artefact. In this research, the performance of the classifiers largely depended on lexicon development. Therefore, additional domain corpora were analysed by updated terms during extraction and were added into the lexicon.

5.7 Topic modelling experiment

This section discusses the practical implementation of the topic modelling. Each user tweet was converted into vectors. Thereafter, k number of topics were declared as part of LDA parameters. After the LDA model had been run, top keywords for each topic were generated, as shown in Table 5.2. LDA was used to get an overview of possible words for different topics from the dataset. Thereafter tweets were categorised into two topics by supervising/guiding topic generation using the CorEx topic model. A tweet is considered narcissistic if the dominant words are deemed to be narcissistic. Topic model techniques were used to categorise tweets based on their semantic relatedness into narcissistic (1) or non-narcissistic based on keywords.

Figure 5.4 shows how topic modelling is carried out.

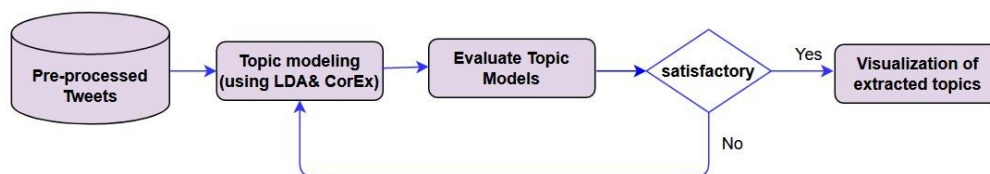


Figure 5.4: Topic modelling workflow

5.7.1 LDA topics

The LDA topic modelling technique was implemented using Gensim package in Python. The number of topics k was set to 7 to be able to give an overview of topics present in the dataset. According to Zhou, Zhao, Rizvi, Bian, Haynos, and Zhang (2019), the number of topics is critical in uncovering the hidden themes from the tweets. A high number of topics may lead to irrelevant and redundant topics. To select the optimal number of topics, a

coherence score for each topic was calculated. The model with optimal number of topics was further examined, and topics generated related semantically were grouped together. Table 5.2 shows the distribution of LDA topics. Top keywords are shown in each topic based on semantic similarity.

Table 5.2: LDA topics

| No. of Topics | LDA top words per topic |
|---------------|---|
| 1 | (0, '0.035*"think" + 0.023*"must" + 0.019*"death" + 0.019*"wrong" + 0.018*"rape" '+ 0.017*"guy" + 0.017*"person" + 0.017*"good" + 0.016*"pay" + 0.015*"right" + 0.014*"police" + 0.013*"family" + 0.012*"watch" + 0.012*"last" + ' '0.012*"student" + 0.011*"friend" + 0.011*"video" + 0.010*"victim" + ' '0.010*"hear" + 0.009*"twitter" + 0.009*"become" + 0.008*"week" + ' '0.007*"next" + 0.007*"long" + 0.007*"care" + 0.007*"place" + 0.007*"enough" + 0.006*"away" + 0.006*"protect" + 0.006*"follow"'), |
| 2 | (1, '0.051*"black" + 0.034*"white" + 0.020*"would" + 0.018*"look" + ' '0.018*"fight" + 0.017*"try" + 0.017*"steal" + 0.015*"attack" + ' '0.014*"african" + 0.014*"ever" + 0.014*"world" + 0.013*"fail" + ' '0.012*"start" + 0.012*"girl" + 0.008*"end" + 0.008*"job" + 0.008*"power" + '0.008*"point" + 0.008*"new" + 0.007*"real" + 0.007*"news" + 0.006*"part" + '0.006*"business" + 0.006*"read" + 0.006*"history" + 0.006*"send" + ' '0.005*"listen" + 0.005*"suck" + 0.005*"matter" + 0.005*"better"'), |
| 3 | (2, '0.051*"make" + 0.032*"stop" + 0.031*"never" + 0.030*"thing" + 0.022*"find" + 0.021*"leave" + 0.015*"land" + 0.014*"blame" + 0.012*"hard" + ' '0.010*"home" + 0.010*"run" + 0.009*"hurt" + 0.009*"fake" + 0.008*"war" + ' '0.008*"young" + 0.008*"continue" + 0.008*"learn" + 0.008*"farm" + ' '0.007*"alone" + 0.007*"relationship" + 0.007*"mind" + 0.007*"lead" + ' '0.007*"seem" + 0.007*"claim" + 0.007*"angry" + 0.007*"full" + ' '0.007*"stupid" + 0.006*"force" + 0.006*"system" + 0.006*"little"'), |
| 4 | (3, '0.049*"kill" + 0.037*"man" + 0.031*"woman" + 0.025*"murder" + 0.023*"sad" + '0.021*"day" + 0.020*"today" + 0.018*"much" + 0.017*"many" + 0.016*"crime" + '0.015*"arrest" + 0.013*"always" + 0.012*"may" + 0.012*"mean" + ' '0.011*"violence" + 0.011*"could" + 0.010*"abuse" + 0.009*"destroy" + ' '0.009*"high" + 0.008*"leader" + 0.037*"man" + 0.007*"kid" + 0.007*"suspect" + '0.007*"love" + 0.007*"member" + 0.007*"idea" + 0.007*"great" + ' '0.006*"poverty" + 0.006*"accuse" + 0.006*"beat" + 0.005*"weak"'), |
| 5 | (4, '0.033*"know" + 0.032*"take" + 0.032*"bad" + 0.028*"even" + 0.027*"want" + ' '0.023*"life" + 0.022*"need" + 0.019*"shit" + + 0.019*"tell" + 0.018*"lose" + ' '0.016*"live" + 0.014*"child" + 0.014*"problem" + 0.013*"way" + 0.012*"keep" '+ 0.011*"break" + 0.010*"help" + 0.010*"money" + 0.009*"fuck" + '0.008*"show" + 0.008*"change" + 0.008*"big" + 0.007*"medium" + '0.006*"struggle" + 0.006*"actually" + 0.006*"report" + 0.006*"forget" + '0.006*"understand" + 0.006*"whole" + 0.006*"ass"'), |
| 6 | (5, '0.089*"people" + 0.028*"die" + 0.022*"give" + 0.020*"racist" + 0.017*"work" '+ 0.015*"government" + 0.015*"use" + 0.015*"really" + 0.014*"feel" + '0.014*"criminal" + 0.014*"back" + 0.013*"racism" + 0.017*"ask" + 0.012*"talk" + '0.012*"lie" + 0.009*"well" + 0.009*"dead" + 0.008*"case" + '0.008*"support" + 0.007*"vote" + 0.007*"speak" + 0.007*"name" + '0.007*"farmer" + 0.006*"word" + 0.006*"already" + 0.006*"protest" + '0.006*"hand" + 0.006*"fear" + 0.005*"threaten" + 0.005*"political"'), |
| 7 | (6, '0.049*"go" + 0.040*"say" + 0.037*"get" + 0.026*"see" + 0.024*"call" + '0.021*"time" + 0.020*"still" + 0.019*"year" + 0.019*"shit" + 0.018*"come" + '0.013*"poor" + 0.013*"happen" + 0.011*"damn" + 0.010*"fire" + 0.010*"also" '+ 0.009*"put" + 0.008*"face" + 0.008*"state" + 0.007*"school" '+ 0.007*"believe" + 0.006*"first" + 0.006*"economy" + 0.006*"fact" + '0.006*"can" + 0.006*"remember" + 0.006*"cry" + 0.005*"suffer" + '0.005*"turn" + 0.005*"old"')] |

5.7.2 Topic coherence

After the topics had been generated, topic coherence was calculated to identify the optimal topics from the model. The top keywords were then used as anchors in CorEx topic modelling. Topic coherence is the degree of semantic similarity between terms in a single topic (Yazdavar et al., 2017). The higher the topic coherence score, the higher the quality of the topics. Given that the purpose the experiment was to extract meaningful themes associated with narcissism, topics with a high coherence score were chosen. As shown in Figure 5.5, the optimal number of topics as per LDA was 5 with a coherence score of 0.25.

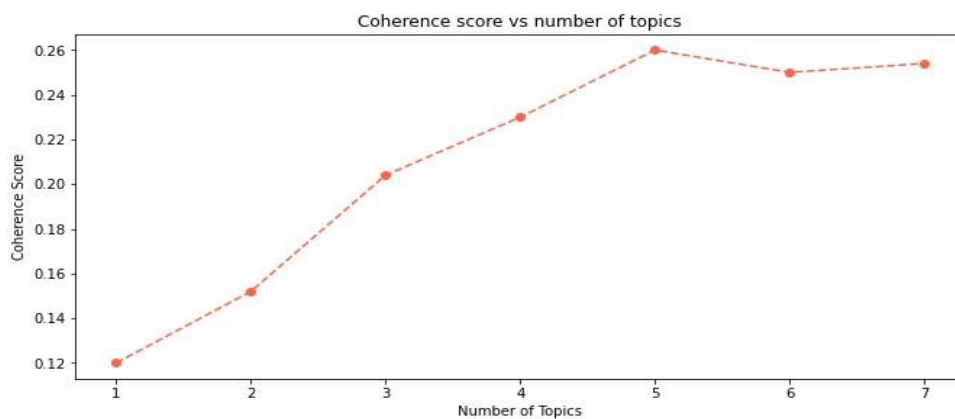


Figure 5.5: Topic coherence

5.7.3 CorEx topics

Using the CorEx technique, three topics were derived in the first instance. This first instance of CorEx modelling without anchoring (i.e., completely unsupervised) was done to discover topics spontaneously emerging from the data. Thereafter, selected keywords from each topic were used as anchors to fine-tune the topics regarding their semantic relationship. The study used CorEx's anchoring technique to improve topic separability and lexicon detection since the study was exploring for narcissistic-related words.

CorEx top words

Topic 1#Vulnerable narcissistic words: black, damn, kill, court, capture, dagga, weed, smoke, killshot, shit, violent, victim, criminal, attack, lie, die, fuck, hate, weak, beat, suck

Topic 2 #Empath-related words: retweets, follower, gain, school, player, score, turn notification, like, notification, gain follower, fast, retweets, follower, turn, like, take, grow account,

Topic 3#Grandiose narcissist words: *life, love, great, watch, best, happy, song, brilliant, great, daily, happy birthday, life, really, birthday, new song, leader, relationship, live, history*

After generating the top keywords, selected anchor words were chosen to further narrow down the topic distributions of the dataset into two, i.e., those with narcissistic-related words and those not related. The anchor words used for the two topics are shown below. In defining the anchor stage, anchors (set of words) were used in one topic and to help extract a topic that did not come naturally initially. Table 5.3 shows the defined anchors.

Table 5.3: Topics and anchor words

| Topic | Anchor words |
|-------|--|
| 1 | 'my', 'mine', 'crazy', 'happy', 'kill', 'bullshit', 'fuck', 'racist', 'hate', 'brilliant' |
| 2 | 'follow', 'government', 'cape', 'thank', 'retweets', 'gain', 'education', 'dear' |

Table 5.4 lists the top keywords based on the anchor words listed in Table 5.3. As shown, the topics are categorised into narcissistic and non-narcissistic words based on semantic similarity. After the application of the anchor words, two main topics were generated as shown in Tables 5.5 and 5.6. The topics consist of keywords that relate to all the words in the topic. The dataset is then labelled based on these two topics.

Table 5.4: Anchor words and related words

| Anchor words | Related words |
|---|---|
| 'my', 'mine', 'crazy', 'happy', 'kill', 'bullshit', 'fuck', 'racist', 'hate', 'brilliant', 'best' | my', 'happy', 'racist', 'hate', 'kill', 'fuck', 'crazy', 'mine', 'brilliant', 'bullshit', 'on my', 'in my', 'me', 'violent', 'my life', 'with my', 'damn' |
| 'follow', 'government', 'thank', 'retweets', 'gain', 'education', 'dear' | follow', 'government', 'thank', 'cape', 'dear', 'retweets', 'gain', 'education', 'thank you', 'follow back', 'follow everyone', 'the', 'follow everyone', 'the', 'cape town', 'followers', 'retweet', 'likes' |

Tables 5.5 and 5.6 show the top words and associated mutual information. Mutual information is a metric indicating how comparable two labels on the same data are (Onan, Bulut, & Korukoglu, 2017).

Table 5.5: Topic 1 Keywords

| Top words | Mutual information |
|-------------|-----------------------|
| ‘my’ | 0.23543547311349738 |
| ‘happy’ | 0.029746948940105968 |
| ‘racist’ | 0.0226683468272996 |
| ‘hate’ | 0.017492366089102557 |
| ‘kill’ | 0.011114543387142803 |
| ‘fuck’ | 0.006966174009169974 |
| ‘crazy’ | 0.005648598503413456. |
| ‘brilliant’ | 0.00392533525681853 |
| ‘bullshit’ | 0.012441019772974583 |
| ‘suck’ | 0.011663166733752365 |
| ‘damn’ | 0.011663166733752365 |
| ‘me’ | 0.011506874412556694 |
| ‘shit’ | 0.010840370910136156 |
| ‘my life’ | 0.006238972418612966 |
| ‘with my’ | 0.005149384911680893 |
| ‘is my’ | 0.005376262314217887 |

Table 5.6: Topic 2 Keywords

| Top words | Mutual information |
|-------------------|----------------------|
| ‘follow’ | 0.052414605479735144 |
| ‘government’ | 0.044429666207309655 |
| ‘thank’ | 0.040771516644581035 |
| ‘dear’ | 0.024973795804116176 |
| ‘retweets’ | 0.021449259654409768 |
| ‘gain’ | 0.01871348183566067 |
| ‘education’ | 0.016032269062716525 |
| ‘thank you’ | 0.03512996746433948 |
| ‘watch’ | 0.01880452575532225 |
| ‘follow everyone’ | 0.016304531047168943 |
| ‘birthday’ | .015628828306896225 |
| ‘notification’ | 0.016667688870065242 |
| ‘the’ | .015628828306896225 |
| ‘followers’ | 0.013246765993018227 |
| ‘retweet’ | 0.01313798702742559 |
| ‘likes’ | 0.012167943964177955 |

5.8 Lexicon detection

In this experiment, words that relate to narcissism were extracted from the output of the topic model experiments. The dictionary was created as follows: after topic modelling, lexicons about narcissistic discourse from Tables 5.2, 5.4 and 5.5 were extracted from the training data. These were then grouped into three different categories (grandiose, vulnerable & empath) of narcissism. The grandiose category contains mostly pronouns, social and affective words. Furthermore, one of the ways narcissists call attention to themselves is by utilizing first-person singular pronouns. The vulnerable category contains words deemed anger words, swear words and belong to the category of antisocial word use index.

Tweets with any lexicon present in the dictionary were given more weight than the other tweets and were labelled as '1' while the rest were labelled as '0'. A tweet was annotated to have traces of narcissism based on the presence of at least one word from the narcissistic lexicon dictionary. In total, a dictionary of 50 lexicons was developed as shown below in Table 5.7. In total, nine terms were generated related to grandiose narcissism and 41 terms were generated related to vulnerable narcissism. Each dictionary was assigned numerical value to represent narcissistic orientation. Empath personality dictionary was not considered in generating the lexicons as the focus was on narcissism in tweets. Thus, the tweets that did not contain lexicons for grandiose and vulnerable dictionary were categorised as empath.

Table 5.7: Lexicon table

| Category | Lexicons |
|-----------------------|--|
| Grandiose dictionary | ['i','my', 'me', 'myself', 'I'm', 'mine', 'oneself', 'happy', 'brilliant'] |
| Vulnerable dictionary | ['worthless','worst','worse','weird','brat','vulnerable','violently','violent','useless','ugly','terribly','terrible','sucks','sucker','stupidity','fuck','stupid','sickening','shocking','shocked','shit','bullshit','selfish','scary','sadly','sad','rude','ridiculous','retarded','retard','pity','pathetic','outrageous','outraged','hell','crap','suck','heck','mad','damn','crazy']. |

5.9 Data labelling

The data labelling experiment was done with the objective of labelling the dataset into the three main categories of grandiose, empath and vulnerable narcissism. The input variable

for this experiment was the pre-processed tweet that had been cleaned as per Rule 4.1. The second input variable was the polarity score of the tweet, as obtained in Section 5.3. The third input variable was the lexicon presence which was 1 if lexicon was present and 0 if lexicon was absent. The output variable was labelled tweets which is a tweet label based on the sentiment score and presence of tokens in the lexicon dictionary. Figure 5.6 shows how labelling was conducted.

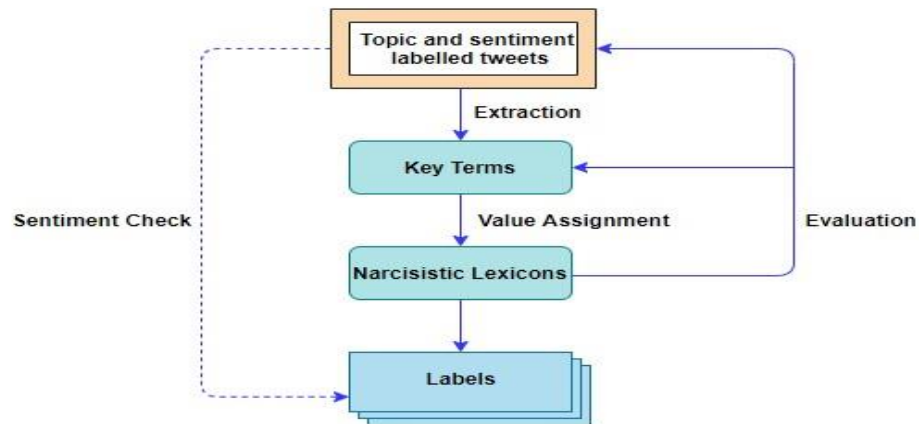


Figure 5.6: Data labelling approach

Procedure 5.4 shows the labelling steps. After the topic modelling experiment had been conducted, data was annotated into three different labels: ‘grandiose narcissist (GN)’, ‘empath (EP)’ and ‘vulnerable narcissist (VN)’. The annotation incorporated the output of sentiment analysis and topic modelling. The GN label relates to tweets containing at least one lexicon/word from the GN dictionary and has a sentiment greater than +0.5. The label VN relates to tweets that contain at least one lexicon/word from the VN dictionary and has a sentiment less than -0.5. The label EP relates to tweets that do not contain at least one lexicon/word from any VN & GN dictionary and has a sentiment between -0.5 & +0.5.

Procedure 5.4: Labelling process

Input: CleanTweet

Output: LabelledTweet

Process:

1. For each Clean Tweet
 - 1.1 Initialize temporary column LabelledTweet to store the label of the tweet
 - 1.2 Calculate the presence of Grandiose and Vulnerable narcissism lexicon in the tweet

```

1.3  If tweet has at least one lexicon from Grandiose
      Dictionary and the Sentiment polarity is greater
      than +0.5;
      1.3.1      Label the tweet as Grandiose
                  Narcissism (GN)
      Else If tweet has at least one lexicon from
      Vulnerable Dictionary and the Sentiment polarity
      is Less than -0.5;
      1.3.1 Label the tweet as Vulnerable Narcissism
(VN)
      Else
      1.3.1 Label the tweet as Empath Personality (EP)
1.4  Return LabelledTweet
      Endif
EndFor

```

5.10 Summary

In this chapter, the process of identifying the labels required to establish if a tweet is positive, negative, or neutral was discussed. VADER sentiment analysis tool was chosen for sentiment analysis over Textblob, LIWC, and SentiWordNet because of its ability to handle informal social media text. In addition, the narcissistic lexicon that was constructed based on the literature review was discussed. The chapter presented how the lexicon construction had started with key terms extraction to construct different dictionaries; how the value had then been assigned to denote the different dictionaries created from the seed terms; and thereafter that seed words related to narcissism had been extracted from the literature. Lastly, it was explained that the lexicons had been put into three dictionaries and evaluated. The output of this task in the process model is annotated data that was used to train selected machine learning classifiers – as discussed in Chapter 6.

CHAPTER 6: MACHINE LEARNING ENSEMBLE CLASSIFIER FOR NARCISSISTIC PERSONALITY PREDICTION

6.1 Introduction

This chapter discusses the techniques and approaches used to classify individuals as either narcissistic or non-narcissistic based on their tweets. The chapter presents the efforts to model narcissism using the features extracted in Chapters 4 and 5. According to Liu, Wang, and Jiang (2016), the combination of various social network usage features can help to predict a narcissistic personality more accurately. Selected supervised classifiers are used. Supervised classifiers techniques depend on training data to perform classification. Classification is done in two phases, namely training and testing. Classifiers are fed training data during the training phase to learn and develop knowledge about data patterns. The trained classifier is subsequently used to predict test data labels during the testing phase. The classifiers were chosen because of their suitability to the research problem and nature of the data. Data was split into training and test data in an 80-20% proportion to perform classification. In training, the classifier is trained to recognise the correct label of the tweets.

6.2 Text classification tools

In this research, libraries and packages were chosen to help achieve the research objective. Potential tools were explored to determine how they would enhance or hinder the research. Python 3.7 was selected because of its widespread use in data analytics and machine learning communities. The data analysis and evaluation processes were the two critical factors considered when selecting a language. NumPy, Panda, and Scikit-learn were used for data analysis. Panda provided the required data representation and allowed for quick attribute modifications. The Panda data frames were converted to arrays using NumPy. The data was divided into training and test datasets using Scikit-learn (Renström, 2018). Scikit-learn is a machine learning toolkit that provides a collection of algorithms for applications such as classification, regression, and clustering (Raschka, Patterson, & Nolet, 2020). Because of its powerful classification of models and pre-processing functionality, this library was chosen for this study (Tohid et al., 2018).

6.3 Classification

Classification is a subset of supervised machine learning tasks that predict an element's class based on data attributes (Nikam, 2015). This research employed multiclass text classification in an attempt to identify traces of narcissism based on their text samples. In this research, supervised learning for classification was used for training. The dataset had three predefined labels. The first label was Grandiose, the second label was Empath, and the last label was Vulnerable Tweets. The training data was fed into the model in batches (size dependent on optimisation method) which learns the weights. The error in predictions was minimised across the training data. The trained model could then be used to make predictions on unseen feature vectors.

6.3.1 Dataset description

The raw dataset that was used comprised 238,317 tweets belonging to 250 users. From Table 6.1 it is clear that the size of the dataset was reduced by 24.3% after various pre-processing techniques compared to the initial raw uncleaned data.

Table 6.1: Dataset characteristics

| Variables | Value count |
|--|--------------------|
| Number of users | 250 |
| Number of tweets before pre-processing | 238,317 |
| Number of tweets after pre-processing | 180,389 |
| Number of tokens before pre-processing | 4,401,672 |
| Number of tokens after pre-processing | 4,031,554 |

6.3.2 Feature extraction

Machine learning classifiers work only with numeric feature vectors that the classifiers can understand; therefore, text data has to be transformed into feature vectors in a process referred to as text vectorization (Singh & Shashi, 2019). Feature vectors are n-dimensional vectors that represent individual objects. Machine learning techniques require numerical data to perform statistical analysis. Two feature extraction techniques were adopted in this research. These are n-grams and Tfidf.

6.3.2.1 N-grams

N-grams are a set of n-consecutive tokens from a given tweet or text (Ahuja, Chug, Kohli, Gupta, & Ahuja, 2019). In text classification, features can be unigrams, bigrams, trigrams and more. Because n-gram features represent more distinct text information than unigrams, n-gram features are extracted from text documents (Wan & Gao, 2015).

6.3.2.2 Term frequency-inverse document frequency (TF-IDF)

TF-IDF is a variant of bag-of-words models and utilises character n-grams. It converts tokens into feature vectors containing (weighted) frequencies of the ngrams present in each token sample (Sidorov, Velasquez, Stamatatos, Gelbukh, & ChanonaHernández, 2014). TF-IDF gives weighted features to a document using the product of Term Frequency (TF) and Inverse Document Frequency (IDF). The frequency of a term in a document is measured in terms of TF, which is proportional to the text's length (Shahmirzadi, Lugowski, & Younge, 2019).

To prepare the training and testing tweets, Scikit-learn's Tfidfvectorizer feature extraction library was used. Tfidfvectorizer converts all of the tweets into a feature matrix weighted by TFIDF term weighting (a weight calculated by term frequency). The use of n-gram frequency for text classification can be challenging. Both prevalent and very rare words are not adopted for classification since they provide little information that can be used to anticipate class labels. The TF-IDF algorithm helps to overcome this problem. TF-IDF provides insight into the true importance of a term, whereas word counting techniques just provide information on its frequency in the present document. It assigns a value to each term's frequency that represents the term's overall importance in the corpus. The TF-IDF algorithm is shown in Equations 6.1, 6.2 and 6.3 shows how TFIDF scores were calculated as shown Procedure 6.1.

$$TF = \frac{Nt}{TD} \dots\dots\dots (Eq. 6.1),$$

where

t-Token

Nt- Number of times a token, t appears in Individual tweets

D-Training dataset

TD- total number of tweets in the training dataset (D)

$$IDF = \log\left(\frac{TD}{TD_t}\right) \dots\dots\dots (Eq. 6.2),$$

where

TD_t -Number of tweets in the training dataset D, a token t appears

$$TFIDF = TF \times IDF \dots\dots\dots (Eq.6.3)$$

The basic idea behind the TF-IDF term is to reduce the weights of the unnecessary terms which occur frequently and do not contribute much to meaning of the text. Thus, the more frequent the term, the higher the denominator of the IDF ratio, and hence the less the value of IDF is, which satisfies the condition of giving less weights to the frequently occurring terms. 1 is added in the denominator to avoid divisions by zero for an unseen word and thus acts as a smoothing parameter.

Procedure 6.1: TFIDF algorithm

Input: Training Dataset (TD)

Output: TFIDFscore

Process:

1. For each tweet in a training dataset (TD)
 - 1.1 Initialize temporary column TFIDFscore to store the term frequency stores
 - 1.2 Count the number of times a token appears in an individual tweet in the Dataset (TD)
 - 1.3 Get $TF(N_t/TD)$ where:

TF is Number of times token t appears in the individual tweets N_t (Eq6.1)
 - 1.4 Get IDF where:

total number of tweets in the dataset)/(Number of tweets in the training dataset a token appears) (Eq. 6.2)
 - 1.5 Get TFIDF where:

TFIDF is $TF \times IDF$ (Eq.6.3)
 - 1.6 Return TFIDFscore

End For

6.3.3 Training and testing

Training and testing are vital parts of the machine learning process. Training is the process of selecting a batch of input data and feeding a classifier to learn and recognise the data

patterns (Pirina & Çöltekin, 2018). During training, the model adjusts its weights to improve upon its output. When training a model, the training data must consist of an example of an object to classify and the correct classification of the object. The first step in building a classification model is to properly prepare the training set of data, and the second step is to set the parameters of the designated machine learning algorithms. In this research, the training sets used in this study involved different combinations of various sets of features (tokens, sentiment, number of tweets, and tweeting frequency). The dataset was divided into training and testing on an 80-20% basis based on the Pareto principle. In the testing phase, a tweet without a class label was provided which means that the tweet was unseen before.

6.4 Learning classifiers

The classifiers that were used to classify the dataset are described in this section. Only supervised learning methods were employed in this study, as previously stated. Training and testing are the two fundamental processes that all classifiers go through. Classifiers are provided training data during the training phase. The model will then forecast test data labels based on the knowledge acquired. Three supervised machine classifiers were considered for this research. These (Random Forest, Support Vector Machines & Naïve Bayes) classifiers were chosen because of their suitability in text classification experiments.

In addition, four ensemble classifiers emanating from the combination of the three core classifiers were also trained. The objective was to obtain a classifier that could achieve optimum performance from the dataset. The variables used in the classifier procedures were N-grams and Labels. When the size of n is 1, it is a unigram, and n-grams when the sizes are greater than 1. Labels refer to the assigned class (C) of the tweets and they were categorised into three, namely grandiose narcissism (GN), empath personality (EP) and vulnerable narcissism (VN).

Table 6.2: Classifier variables

| Variables | Description |
|--------------------|--|
| D | Training dataset consisting of pre-processed Tweets and N-grams |
| n-grams (1,2) | A sequence of N words |
| Labels (Y) | This relates to the target class variable the classifier will learn the data from. GN, EP, VN are the labels in this chapter |
| Trained classifier | This is the trained classifier |

6.4.1 Naïve Bayes

This classification technique is based on Bayes' Theorem with an assumption of independence among predictors (AbdulHussien, 2017). The Naive Bayes classifier assumes that the presence of one feature in a class is unrelated to the presence of any other feature. Naïve Bayes presume that every feature (in this research 'Label' attribute) is considered independent from all other features in the given class (Procedure 6.2). As a result, to compute Bayesian probability, it will multiply all members of the feature vector in the specified class. A dataset with a given class (Label) C (GN, VN and EP), where X is n-gram defined by a feature vector $\{X_1, X_2 \dots X_n\}$ with n being the number of features in the dataset. Therefore, given class C with an instance X can be computed Bayesian probability is $P(C|X)$ as shown in Eq. 6.4.

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \dots\dots\dots (Eq. 6.4)$$

where

$P(C)$ = prior probability of class.

$P(X)$ = prior probability of predictor.

$P(C|X)$ = posterior probability of class (target/label) given predictor (n-gram).

$P(X|C)$ = likelihood which is the probability of the predictor given class.

The input variables are pre-processed tweets split into tokens/n-grams, the tweet frequency, lexicon presence and sentiment of the tweet. The target variables are classification of narcissism which can either be GN, EP or VN. The output of this algorithm is a trained classifier with knowledge to identify the target variables of GN, EP and VN. The parameters are tfidf and additive smoothing.

Procedure 6.2: Naïve Bayes algorithm

Input: DT (n-grams, Labels)-Training Dataset

Output: NB trained classifier

Process:

```
1.For each n-gram in DT
  1.1 Compute TFIDF for n-grams in each label
  1.2 Using Bayesian rule, calculate the probability of
      each n-gram in the tweets against each label as per
      Eq. 6.4
  1.3 If a tweet has maximum likelihood for an n-gram in
      GN
      1.3.1 Label the tweet as GN
      Else If it has maximum likelihood for an n-gram in
      VN
      1.3.1 Label the tweet as VN
      Else
      1.3.1 Label the tweet as EP
  1.4 Return Trained classifier
EndFor
```

6.4.2 Support-vector machine

Support-vector machine (SVM) is a supervised learning classifier that attempts to discover the best hyperplane based on the labelled data (training data) that can be used to categorise new data points (Procedure 6.3). The hyperplane is a simple line in two dimensions. Each text document is represented as a vector in SVM, with the dimension being the number of different keywords. These support vectors represent data points of the classes and define the position and the margin of the hyperplane. SVM does not use all the training data points like RF or NB to make classifications but uses only the support vectors (Ahire, Kolhe, Kirange, Karale, & Bhole, 2015).

The objective of SVM in this research was to find an optimal hyperplane ‘h’ that best distinguishes the three labels in the study, i.e., GN, EP and VN. The hyperplane establishes decision boundaries that aid in data classification. The different classes are represented by data points on either side of the hyperplane (Ahire et al., 2015). The main goal is to identify a hyperplane that provides a greater margin between the hyperplane and the nearest data points/vectors. The support vectors are the document representatives closest to the decision surface (Nalepa & Kawulok, 2019). There are three labels/classes of the dataset. The first

label is grandiose narcissism denoted as GN. The second label is empath personality denoted as EP. The last label is vulnerable narcissism denoted as VN in this research.

Procedure 6.3: SVM algorithm

Input: DT (n-grams, Labels) Training dataset

Output: SVM trained classifier

Process:

```

1.For each n-gram in DT
  1.1 Divide the associated n-grams into three set of data
      items based on the associated labels
  1.2 If a tweet has a corresponding label of 1 Then
      1.2.1 Label the tweet as GN
      Else If it has corresponding label of 2
      1.2.1 Label the tweet as VN
      Else
      1.2.1 Label the tweet as EP
  1.3 Construct a hyperplane to separate the feature
      vectors based on the three labels
  1.4 Add the feature vectors into support vectors set V
  1.5 Classify n-grams as GN, VN, or EP
  1.6 Return trained classifier
EndFor

```

6.4.3 Random Forest

Random forest (RF) is an ensemble classifier that comprises a group of decision trees. RF is a bagging-based ensemble model. It creates numerous trees and votes among them to reach a majority decision (Li, Yan, Liu, & Li, 2018). As the number of trees grows, so does the accuracy of the prediction. In addition, Random Forest was selected as it reduces the problem of over-fitting through the use of bootstrap sampling technique. The sample is picked at random in RF, and a decision tree is constructed for the random sample (Procedure 6.4). The process is repeated, with a different sample picked at random each time (Reis, Baron, & Shahaf, 2018). A result is generated from various tress with a different sample which results in different predictions. Each tree is built on a sample of objects drawn with replacement from the original dataset; thus, each tree has some objects that it hasn't seen (Daho, Settouti, Lazouni, & Chikh, 2014). As pointed out, this research had the three labels GN, EP, and VN. Seven parameters in random forest classifier were used in this research. The first one is *n estimators* which are number of trees in the forest. In a random

forest, using more trees boosts accuracy, but at a certain point, the classifier accuracy reduces.

Various sizes of trees were experimented, and an optimal accuracy was achieved with 180 trees. The second parameter is *max features*. The third parameter is *gini criterion*. This is the function that determines the purity of each decision tree node. Gini was chosen because it is computationally faster than the alternative metric, entropy. The fourth parameter used was *bootstrap*. This relates to the use of trees as samples with replacement. The fifth parameter is the *node* which indicates the beginning of a tree, in this case the beginning of the tweet classification to determine whether it is GN, EP or VN. The sixth parameter is *Arc* which splits the classification further into nodes. The seventh parameter is the leaf node which refers to the tweet classification which can no longer be split further into any category. The last parameter is the branch which links the three classifications of tweets.

Procedure 6.4: Random Forest Classifier

Input: DT (n-grams, Labels)-Training Dataset

DTs - Training Data size

DT-Bootstrap sample, subset of DT

Dt_s-Bootstrap sample size of Dt

Decision Trees (D_i) where i=1 to 180

Output: Random trained classifier(s)

Process:

1. $N \leftarrow DT_s / Dt_s ; n=0$
2. $Dt \leftarrow \text{Subset}(DT)$
 - 2.1 If $n=N$ and $DT \neq \text{Empty}$
 $Dt \leftarrow \text{Balance}(DT)$
 - Else
Break
 - 2.2 Randomly sample the training data DT with replacement to obtain bootstrap sample
 - 2.3 Train the bootstrap sample using first Decision Tree $D_i(D)$ with replacement
 - 2.4 $N = n+1$
 - 2.5 Repeat steps 2.2 $N+1$ times
3. Obtain majority votes of decision trees for each tweet
 - 3.1 Form the ensemble classifier by combining the decision trees results using majority voting
 - 3.2 If the Majority for the tweet is 1,
 - 3.2.1 label the tweet as GN
 - Else If the Majority for the tweet is 2,

```

3.2.1 label the tweet as VN
Else
3.2.1 Label the tweet as EP
3.3 Return Trained classifier
EndFor

```

6.4.4 Voting ensemble classifier

Voting ensemble classifier (VEC) combines different machine learning classifiers and makes predictions based on a voting mechanism. The VEC is divided into hard and soft categories (Shahzad & Lavesson, 2013). The final class prediction in hard voting is determined by a majority vote – the estimator selects the most appearing among the base models. The final class prediction in soft voting is based on the average probability (weighted) generated from all of the base model predictions (Zhu, Moh, & Moh, 2016). Multiple classifiers are used to perform a classification task using an ensemble learning approach. The ensemble approach is based on the majority voting of different classifiers that have learned various features from dataset to perform classification. (Catal & Nangir, 2017). In this ensemble classifier, support vector machine, random forest, and Naive Bayes classifiers were used to create different combination ensemble classifiers (Procedure 6.5).

Procedure 6.5: Voting ensemble classifiers

Input: Dt (n-grams, Labels)-Training Dataset

DTs - Training Data size

DT-Bootstrap sample, subset of DT

Dt_s-Bootstrap sample size of Dt

Base Classifiers (Bc_i) where i=1 to 3 (RF, SVM, NB)

Output: Ensemble trained classifier(s)

Process:

1. $N \leftarrow DT_s / Dt_s ; n=0$
 - a. $Dt \leftarrow \text{Subset}(DT)$
2. If $n=N$ and $DT \neq \text{Empty}$
 - $Dt \leftarrow \text{Balance}(DT)$
 - Else
 - Break
 - 2.1 Train the bootstrap sample using first base classifier Bc_i(Dt)
 - 2.2 Train the bootstrap sample using second base classifier Bc_{ii}(Dt)
 - 2.3 Train the bootstrap sample using third base classifier Bc_{iii}(Dt)

```

2.4  N= n+1
2.5  Repeat steps 1 N+1 times
3. Obtain majority votes of base classifiers for each tweet
3.1  Form the ensemble classifier by combining the base
      learners using majority voting
3.2  If the Majority for the tweet is 1,
      3.2.1 label the tweet as GN
      Else If the Majority for the tweet is 2,
      3.2.1 label the tweet as VN
      Else
      3.2.1 Label the tweet as EP
3.3  Return Trained classifier
EndFor

```

6.5 Experiment parameters

Four parameters were used to evaluate the performance of the classifiers. The parameters are precision score, recall score, accuracy and F1- score. Accuracy refers to the proportion of correct predictions. Accuracy and F1-score are converted into percentage for presentation purposes. Classifiers were trained using 80% of the dataset and evaluated using 20% of the dataset.

All parameters were calculated per class, resulting in three multiclass classification problems using GN, EP, and VN samples. For every class, true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) were calculated, as described in Table 6.3. In addition, Confusion matrix which is table that shows how well a classifier performs was used (Visa, Ramsay, Ralescu, & Van der Knaap, 2011). By comparing the actual and predicted classifications, the confusion matrix shows how accurate a classifier is. Confusion matrix table consists of recall, precision, F1-score and accuracy. The maximum score for Recall and precision is 1 and minimum is 0. The maximum score for F1-score and accuracy is 100% and minimum is 0%.

Table 6.3: Confusion matrix parameters

| Parameter | Description |
|----------------------|--|
| True Positives (TP) | The number of correct positive predictions of a class made by the model. |
| True Negatives (TN) | The number of correct negative predictions of a class made by the model. |
| False Positives (FP) | The number of incorrect positive predictions of a class made by the model. |

| | |
|----------------------|---|
| False Negatives (FN) | The negative outcomes that the model predicted incorrectly. |
|----------------------|---|

Precision is the proportion of successfully predicted positive observations to total expected positive observations. It displays how close the predicted values are to the actual values. It is derived by the number of TPs divided by the sum of TPs and FPs.

$$\text{Precision} = \frac{TP}{TP+FP} \dots\dots\dots (\text{Eq. 6.5})$$

The number of positive outcomes that are accurately labeled as positive is known as **recall**. The recall also reveals how accurate the model is at predicting class membership, so it covers the quantitative part of classification success. In sentiment analysis, if a dataset is annotated in two sentiment polarities: positive and negative, there are two recalls: positive recall and negative recall. The example of positive recall calculation formula is shown below:

$$\text{Recall} = \frac{TP}{TP+FN} \dots\dots\dots (\text{Eq. 6.6})$$

The fraction of correctly identified instances is measured by **accuracy**. It demonstrates how close the predicted value is to the actual or theoretical value (Joshi, 2018).

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \dots\dots\dots (\text{Eq. 6.7})$$

F-measure (F1_score) – F1-score is defined as the measure that combines precision and recall and tries to convey the balance between them.

$$\text{F1 Score} = \frac{2*\text{Precision}* \text{Recall}}{\text{Precision}+\text{Recall}} \dots\dots\dots (\text{Eq. 6.8})$$

Cross-validation

Cross-validation is a machine learning evaluation technique that attempts to approximate the model's effectiveness on data that it has never seen before. It is very beneficial when working with a small dataset and identifying a strong predictor (Ramezan, Warner, & Maxwell, 2019). The number of groups that a dataset should be split into in k-fold cross-validation is determined by k (Ramezan et al.,2019). This method is performed k times, with each iteration evaluating the model using a different fold. Increased generalisability on unknown data and a lesser likelihood of overfitting are two advantages of this method. In comparison to the traditional train-test split technique, it usually produces less biased

estimates. The gold standard for evaluating a model's performance is k-fold cross-validation, which usually outperforms the alternatives (Ramezan et al., 2019).

This research used a variation k-fold cross validation called “stratified k-fold”. Since the distribution of each type of narcissism in this dataset is imbalanced, stratified k-fold was the most suitable validation technique. Stratified k-fold solves this problem by maintaining the same dataset distribution throughout all folds. The parameters of stratified k-folds are: *n splits*. This relates to the number of folds used. In this research, different 5, 10 and 15 folds were experimented with and optimal accuracy was achieved with 10 folds. The second parameter is *shuffle*. It refers to the data prior to creating the folds. The strength of the techniques is the enhanced generalisability on unknown data and a reduced risk of overfitting (Szalma & Weiss, 2020). According to Zhang et al. (2016), a k value of 5 or 10 is appropriate as it reduces the possibility of variance and biasness. The number of samples that can be obtained is limited by a larger k value, whereas bias is increased by a lower k value.

6.6 Experiment setup: Classification

This section presents the experiments undertaken and how the optimal classifier was built and evaluated using the labelled data. As explained in Section 6.1, the objective of the experiment was to train a classifier to predict whether a tweet has traces of narcissism or not. That is whether a tweet is narcissistic or not narcissistic. The prediction model was developed using three input variables and one target variable. As noted in Section 6.4, three classifiers were chosen in this research. The three classifiers were then combined to create four ensemble classifiers using the majority voting approach. Four different combinations of the classifiers were experimented with, and performance comparison was made. The ensemble classifiers experimented with the aim of getting the classifier with optimum performance from the data. An ensemble of classifiers is a group classifier of which the decisions are combined by voting (Liu & Ge, 2018). Each classifier was evaluated using accuracy, precision, F1-score, and confusion matrix and classification report.

Table 6.4: Classification variables

| Variable | Description |
|---|--|
| Clean tweet: (Input variable) | Pre-processed tweet. |
| Tweet frequency (T_f): (Input variable) | Number of tweets (statuses posted) made by the user. |
| Sentiment polarity (S_p) (Input variable) | This refers to the sentiment of each tweet and can be positive, neutral or negative. |
| Lexicon presence (L_p):(Input variable) | This relates to presence or absence of narcissistic-related word/clean tweet based on the narcissistic dictionary. |
| Label (GN, EP, VN): <i>Target variable</i>) | This refers to the target variable which is label/class to be learned and predicted. |

6.6.1 Input and target variables

Tweet frequency (T_f): Number of tweets (statuses posted) made by the user. The rate at which users' tweets are classified into three categories to enhance the training of the classifiers, as shown in Figure 6.1. The first category was *users with Tweets less than 50,000*. The second category was *users with tweets above 50000 but less than 100,000*. The last category was *users with tweets more than 100,000*.

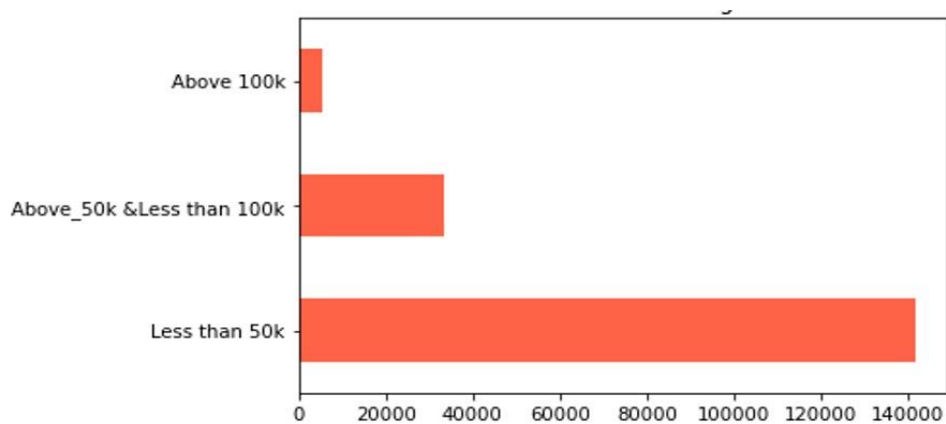


Figure 6.1: User categories based on TF

6.6.2 Vectorization

The data is transformed into feature vectors before being trained with machine learning classifiers. For this research, a data frame mapper library in Python was used. DataFrame Mapper is a python library for mapping DataFrame columns to transformations, then later recombined into features, as defined in Figure 6.2.

Table 6.5: Vectorization parameters

| Parameter | Description |
|-----------------|---|
| ngram_range | (1, 2) represent the lower and upper boundary of the range of n-values for different n-grams to be extracted. In this research, it extracted all one gram and two gram. |
| max_df = 0.8 | This parameter is used to remove terms that appear too frequently in the tweets l. For this study 0.8 was used as Max df. |
| Min_df = 0.0025 | This parameter is used to remove terms that appear too infrequent in the tweets. For this study 0.0025 was used as Min df. |

```
#Add the features of the dataframe that you want to transform and/or combine
mapper = DataFrameMapper([
    ('clean_tweet', TfidfVectorizer(min_df=.0025,
    max_df=0.8, ngram_range=(1,2))),
    ('TweetPolarity', None)])
    ('Tweet Frequency', None),
    ('Lexicon Presence', None)])
```

Figure 6.2: Vectorization parameters

6.6.3 Data split

After transformation, the dataset was split in an 80-20 basis. This was done in accordance with the Pareto principle, which argues that 20% of all causes (or inputs) result in 80% of all outcomes (or outputs) for any particular occurrence (Koch, 2011). According to Liu and Cosea (2017), the classifiers use training dataset set to discover any new patterns, and the test set is then used to validate the degree to which the patterns truly exist and are trustable. In addition, machine learning classifiers could learn what they needed to know from the data in the training set. It then used what it had learned to generate a prediction about the data in the test set. The prediction could then be compared to the actual target variables in the test set to determine how accurate the classifier was. The data was divided into four primary variables. The first variable was X_Train which was the 80% of the training dataset. The second variable was X_Test which was the test dataset. The third variable was Y_Train which was the target variable of the training dataset. The last variable was Y_Test which was the target variable of the test dataset. The variables were transformed into four parameters. The first parameter was feature which was the transformed input variables of X_Train and X_Test. The second parameter was categories. These were transformed to X_Train and Y_Test variables. The third parameter was test size which was 0.2 or 20%. The last parameter was train size which was 0.8 or 80%. As shown in Figure 6.3, the training and test datasets consisted of 144,460 tweets and 36116 tweets respectively split on the 80-20% basis.

```
# Split the data between train and test
X_Train, X_Test, Y_Train, Y_Test = train_test_split(features, categories ,test_size=0.2,train_size=0.8)

#Show the size of the split train and test dataset
X_Train.shape

(144460, 686)

X_Test.shape

(36116, 686)
```

Figure 6.3: Data split (train and test dataset)

6.6.4 Classifier training: Naïve Bayes

Three classifiers and four ensemble classifiers were used in this study. The three classifiers were combined differently to create different ensemble classifiers using the majority voting approach. Four different combinations of the classifiers were experimented with, and a performance comparison was made.

Table 6.6: Naïve Bayes confusion matrix

| Classifier | Metric | Class /Label | | | Accuracy (%) | F1-score (%) |
|-------------|---------------------|--------------|-------------|-------------|--------------|--------------|
| | | GN | EP | VN | | |
| Naïve Bayes | Precision | 0.90 | 0.83 | 0.90 | 83.83 | 81 |
| | Recall | 0.37 | 0.99 | 0.27 | | |
| | F1-score (%) | 0.52 | 0.91 | 0.41 | | |

As shown in Table 6.6 above, Class GN achieved a precision of 0.90 or 90% and a recall of 0.37 or 37%. The precision for Class EP was 0.83 and a recall of 0.99. Class VN achieved a precision of 0.90 and a recall of 0.27. Class EP had a high F1-score of 0.91, followed by class GN and lastly class VN. The low F1-score of VN can be attributed to less size of training and test data for the class. Cumulatively, the classifier achieved an accuracy of 83.83 and an F1score of 81%.

6.6.5 Support-vector machine

Table 6.7 shows the confusion matrix of SVM. Class GN achieved a precision of 0.64 and a recall of 0.01. Class EP achieved a precision of 0.86 and a recall of 1. This implies that SVM identified all proportions of EP instances correctly. Class VN achieved a precision of 0.67 and a recall of 0.01. Class EP had a high F1-score of 0.92, while class GN and class

VN had an F1score of 0.02. Cumulatively, the classifier achieved an accuracy of 85.44 and an F1-score of 79%.

Table 6.7: SVM confusion matrix

| Classifier | Metric | Class /Label | | | Accuracy (%) | F1- score (%) |
|------------------------|---------------------|--------------|------------|----------|--------------|---------------|
| | | GN | EP | VN | | |
| Support Vector Machine | Precision | 0.64 | 0.8 | 0.67 | 85.44 | 79 |
| | Recall | 0.01 | 6 | 0.01 | | |
| | F1-score (%) | 2 | 0.1 | 2 | | |

6.6.6 Random Forest classifier training

In a random forest (RF) classification model, the two main hyper parameters used are the number of estimators and how deep a tree is allowed to grow. Table 6.8 shows the confusion matrix of random forest. Class EP achieved a precision of 0.90 and a recall of 0.96. This implies that RF identified 96% of EP instances correctly. Class VN achieved a precision of 0.83 and a recall of 0.52. Class EP had a high F1-score of 0.93, while class GN had an F1-score of 0.75 and class VN had an F1-score of 0.64. Cumulatively, the classifier achieved an accuracy of 88.19 and an F1-score of 88%.

Table 6.8: Random Forest confusion matrix

| Classifier | Metric | Class /Label | | | Accuracy (%) | F1-score (%) |
|---------------|---------------------|--------------|-----------|-----------|--------------|--------------|
| | | GN | EP | VN | | |
| Random Forest | Precision | 0.79 | 0.90 | 0.83 | 88.19 | 88 |
| | Recall | 0.71 | 0.96 | 0.52 | | |
| | F1-score (%) | 75 | 93 | 64 | | |

6.6.7 Ensemble 1(En1): SVM and RF

Table 6.9 shows the confusion matrix of ensemble 1 (SVM & RF). Ensemble 1 achieved an accuracy of 86.37%. Class EP achieved a precision of 0.86 and a recall of 0.98. This implies that ensemble 1 identified 98% of EP instances correctly. Class VN achieved a precision of 0.98 and a recall of 0.27. Class EP had a high F1-score of 92%, while class

GN had an F1-score of 70% and class VN had an F1score of 42%. Cumulatively, the classifier achieved an accuracy of 86.37 and an F1-score of 84%.

Table 6.9: Ensemble 1 confusion matrix

| Classifier | Metric | Class /Label | | | Accuracy (%) | F1-score (%) |
|------------------------------------|---------------------|--------------|-----------|-----------|--------------|--------------|
| | | GN | EP | VN | | |
| Ensemble 2: SVM and Naïve Bayes | Precision | 0.83 | 0.86 | 0.98 | 86.37 | 84 |
| | Recall | 0.60 | 0.98 | 0.27 | | |
| | F1-score (%) | 70 | 92 | 42 | | |

6.6.8 Ensemble 2 (En2): SVM and Naïve Bayes

Table 6.10 shows the confusion matrix of ensemble 2 (SVM& Naïve Bayes). Class GN achieved a precision of 0.89 and a recall of 0.37. Class EP achieved a precision of 0.83 and a recall of 1. This implies that ensemble 2 identified 100% of EP instances correctly. Class VN achieved a precision of 0.98 and a recall of 0.25. Class EP had a high F1-score of 0.91, while class GN had an F1-score of 0.52 and class VN had an F1-score of 0.40. The second ensemble classifier achieved an accuracy of 84% and an F1-score of 81%.

Table 6.10: Ensemble 2 confusion matrix

| Classifier | Metric | Class /Label | | | Accuracy (%) | F1-score (%) |
|------------------------------------|---------------------|--------------|-----------|-----------|--------------|--------------|
| | | GN | EP | VN | | |
| Ensemble 2: SVM and Naïve Bayes | Precision | 0.89 | 0.83 | 0.98 | 84 | 81 |
| | Recall | 0.37 | 1 | 0.25 | | |
| | F1-score (%) | 52 | 91 | 40 | | |

As shown in Table 6.10, ensemble 2 achieved an accuracy of 84%. It achieved a recall of 1 for class EP and precision of 0.98 for class VN. The classifier did not perform well in recall and F1-score of class VN.

6.6.9 Ensemble 3 (En3): Naïve Bayes, SVM and RF

The third ensemble classifier achieved an accuracy of 86.65%. In this ensemble, the estimators were SVM, RF and NB. Table 6.11 shows the confusion matrix of ensemble 3 (Naïve Bayes SVM & RF). Class GN achieved a precision of 0.81 and a recall of 0.62. Class EP achieved a precision of 0.87 and a recall of 0.97. Class VN achieved a precision of 0.90 and a recall of 0.36. Class EP had a high F1-score of 0.91, while class GN had an F1-score of 0.52 and class VN had an F1-score of 0.40. The third ensemble classifier achieved an accuracy of 86.65% and an F1-score of 85%.

Table 6.11: Ensemble 3 confusion matrix

| Classifier | Metric | Class /Label | | | Accuracy (%) | F1-score (%) |
|------------------------------------|--------------|--------------|------|------|--------------|--------------|
| | | GN | EP | VN | | |
| Ensemble 3: Naïve Bayes SVM and RF | Precision | 0.81 | 0.87 | 0.90 | 86.65 | 85 |
| | Recall | 0.62 | 0.97 | 0.36 | | |
| | F1-score (%) | 71 | 91 | 52 | | |

6.6.10 Ensemble 4 (En4): Naïve Bayes and random forest

Table 6.12 shows the confusion matrix of ensemble 4 (random forest & Naïve Bayes). Class GN achieved a precision of 0.86 and a recall of 0.59. Class EP achieved a precision of 0.88 and a recall of 98. This implies that ensemble 4 identified 98% of EP instances correctly. This is because of the size of the training dataset which was higher than those of class GN and VN. Class VN achieved a precision of 0.98 and a recall of 0.25. Class EP had a high F1-score of 0.91, while class GN had an F1-score of 0.52 and class VN had an F1-score of 0.40. The second ensemble classifier resulted in an accuracy of 87.35% and an F1-score of 86%.

Table 6.12: Ensemble 4 confusion matrix

| Classifier | Metric | Class /Label | | | Accuracy (%) | F1-score (%) |
|---|---------------------|--------------|-----------|-----------|--------------|--------------|
| | | GN | EP | VN | | |
| Ensemble 4: Naïve Bayes and random forest | Precision | 0.86 | 0.88 | 0.85 | 87.35 | 86 |
| | Recall | 0.59 | 0.98 | 0.47 | | |
| | F1-score (%) | 70 | 92 | 61 | | |

6.7 Results and comparison

In this section, the parameters used to evaluate the classifiers are compared. Section 6.7.1 compares the classifier accuracy and F1-score (Fscore). Section 6.7.2 compares the precision score of the classifiers, and Section 6.7.3 compares the recall score of the classifiers.

6.7.1 Comparison of accuracy and F1-score

Figure 6.4 below shows the comparison between the average accuracy and F1-scores for each algorithm and it can be seen that that the accuracy scores are slightly higher than the F1-scores. As can be seen, random forest (RF) had better accuracy than the rest of the classifiers. In addition, RF had a high F1-score followed by an ensemble of the three classifiers. Support vector machine (SVM) had both lowest accuracy and F1-score followed by Naïve Bayes. According to Gustafsson (2020), F1-score can be a good metric for imbalanced data, because it takes into consideration both axis of the confusion matrix.

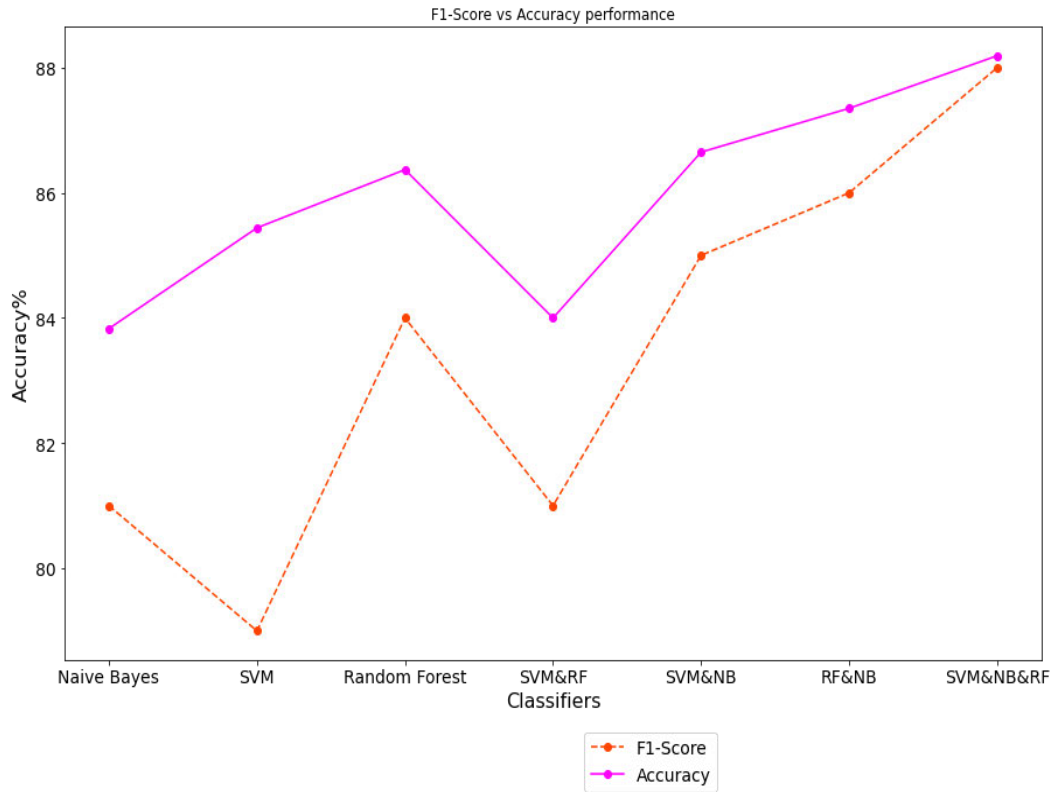


Figure 6.4: Accuracy versus F1-score comparison

Figure 6.4 shows that scores differ greatly depending on the classifier, from average F1-scores of 79 for SVM to 88% for RF. This corresponds to a performance gain of almost +9% and confirms that comparing and selecting the most suitable supervised model is critical to increase the efficiency of the prediction and ranking process. This may be due to the ensemble nature of the two classifiers. Random forest uses a bagging approach form of ensemble learning to perform prediction. Moreover, it can be seen that the three ensemble classifiers remain stable even if retrained and iterated over different random state values. Indeed, the learning process of these models does not imply random sampling from data, unlike NB, and SVM. The results of this experiment indicate that predicting an individual's narcissistic traits from text is possible with reasonable accuracy.

6.7.2 Comparison of precision

Out of the total number of components that the classifier asserts belong to that class, precision shows how many were accurately identified as belonging to that class (Gustafsson, 2020). A high precision shows that the classifier gave more accurate results than approximate results (Kunte & Panicker, 2019). Figure 6.5 shows the Precision of various classifiers for the three classes.

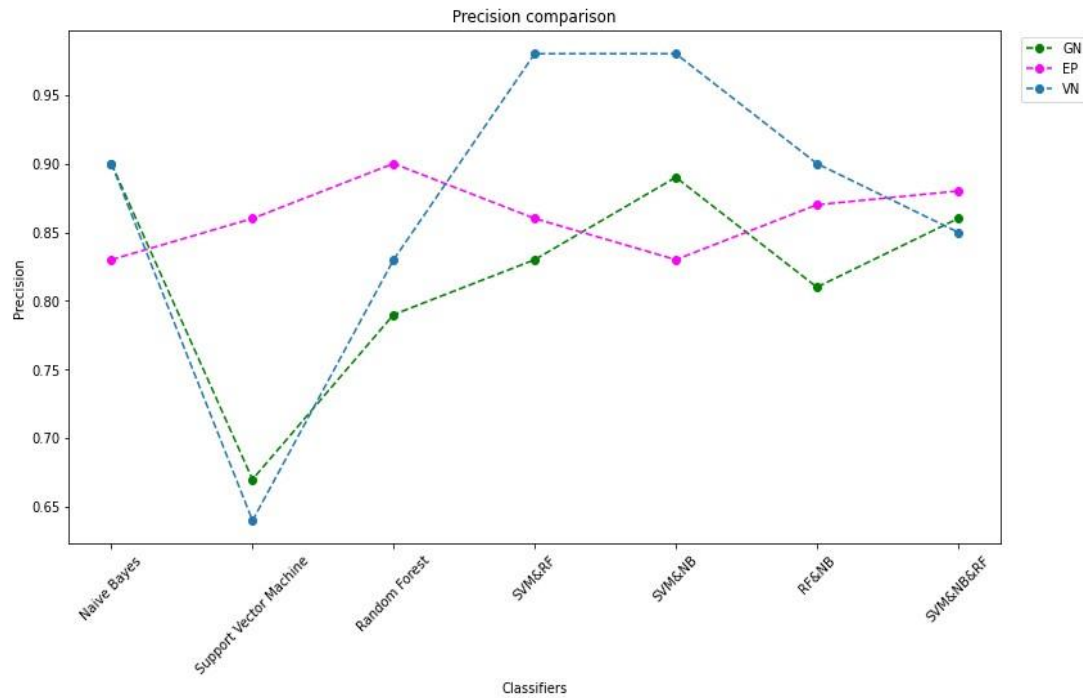


Figure 6.5: Precision comparison

In Figure. 6.5, ensemble 1 (SVM&RF) and ensemble 2 (SVM&NB) have a precision of 0.98 for class ‘VN’ which is highest among other categories. Random forest had the second largest precision after the ensemble classifiers, which is 0.88. F1-score is used to measure the accuracy of the algorithm. F1-score considers both precision and recall in calculating accuracy.

6.7.3 Comparison of recall

Recall is a metric that indicates how well a classifier can recognize relevant data (Kharde & Sonawane, 2016). As shown in Figure 6.6, all the classifiers had a higher recall for class EP and a low recall for class VN. According to Gustafsson (2020), a high recall can mean that most tweets were labelled neutral, in this case class EP. The low recall in class VN may be as a result of the size of the training dataset in the class, VN.

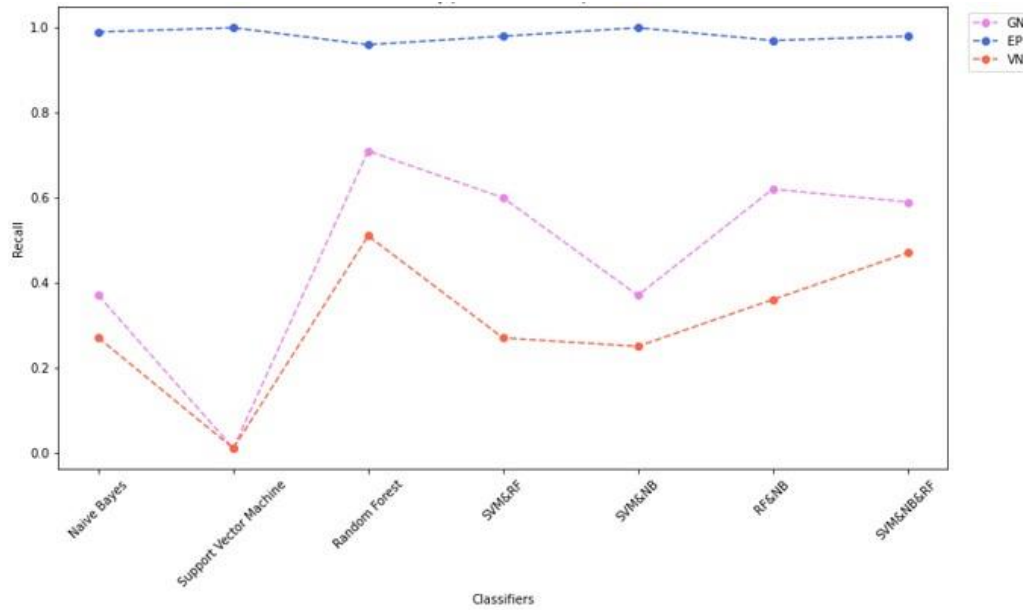


Figure 6.6: Recall (sensitivity) performance comparison

6.8 Summary

Linguistic tweet patterns have been demonstrated to be strong predictors of personality on social media in previous studies. The use of social media is increasing, and the data provided by users can be utilized to better understand their personalities and preferences, as well as recommend services and facilities or predict their behavior. This chapter discussed how to predict narcissism personality traits based on twitter datasets. To determine the best model for classifying a narcissistic personality type, different classifiers were trained and compared using the features described in Section 6.7.1 and based on their suitability for text classification. Before modelling, the data was split into a training set (80%) and a test set (20%), using stratified sampling to preserve the relative ratio of classes across sets. The results from the analyses showed that accuracy was highest (88.19%) for the random forest (RF) classifier and lowest for the Naïve Bayes classifier (83.83%). An ensemble classifier of RF had the highest score of 88% followed by an ensemble 4 of Naïve Bayes and RF. The high accuracy by RF classifiers may be due the ensemble nature of the two classifiers. Random forest uses a bagging approach form of ensemble learning to perform prediction which leads to higher performance when compared to individual classifiers.

CHAPTER 7: FUZZY-BASED NARCISSISM CLASSIFICATION

7.1 Introduction

This research models the uncertainty of narcissism classification using fuzzy-based approach. Machine learning-based classifiers treat text classification in a "black-and-white" approach, whereas text classification is rarely that simple (Madani, Erritali, Bengourram, & Sailha, 2019). Furthermore, Jefferson et al. (2017) asserted that previous works on text classification focused on deterministic algorithms' methods without considering the text's fuzziness. Sentiment keywords in text categorization are ambiguous since even words in the same context might have different sentimental orientations. In addition, human sentiments are often fuzzy as one may use one word to express more than one feeling simultaneously (Jefferson et al., 2017). Fuzzy logic is a technique for obtaining a decision from input data that is unclear, unreliable, noisy, or incomplete (Molina-Gil, Concepción-Sánchez, & Caballero-Gil, 2019). Incorporating fuzzy logic in narcissistic classification helps to handle the imprecision of knowledge coming from gaps and blanks presenting the model classifications. In this research, fuzzy logic was used to summarise the results of a classification considering two inputs (sentiment and class) into an output with a level of relevance. The fuzzy inputs to the model are classifier predicted class and sentiment analysis scores. The model utilises a Triangular membership function for the fuzzification of user data.

7.2 Theory of fuzzy logic system (FLS)

Fuzzy logic is a problem-solving control method that uses a simplistic methodology to reach a definite conclusion based on unclear, messy, or incomplete input data (Madhusudhanan & Moorthi, 2019). FLS consists of three main steps: Fuzzification, and inference engine using rules, and defuzzification. Figure 7.1 illustrates a standard fuzzy logic system.

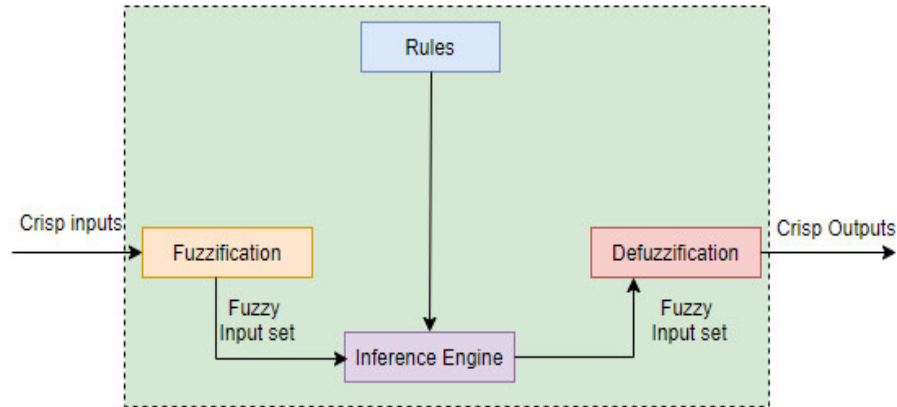


Figure 7.1: Block diagram of a general fuzzy logic system

The first process in FLS is fuzzification and involves the transformation of crisp input, deriving the membership functions for input, and representing them with linguistic variables (Eshuis & Firat, 2018). The membership functions transform each crisp value into a fuzzy set. Membership functions can be triangular, trapezoidal, Gaussian bell-shaped, sigmoidal and Scurve waveform (Shynu, Shayan, & Chowdhary, 2020).

The second component in FLS is the application of the rules on the fuzzy set in the inference engine (Vashishtha & Susan, 2019). Every rule in the knowledge base is compared to facts in the database by the inference engine. A fuzzy rule is a conditional IF-THEN rule with a conclusion. The rule is activated and its (action) part is done when the condition part of the rule matches a fact (Liu & Zhang, 2018). Tsukamoto, Mamdani, and Sugeno are the three Fuzzy logic inference methods. The main difference between the FLS methods is the methodology used to create the crisp output out of the fuzzy input (Wilges et al., 2016).

The third component in FLS is defuzzification and refers to the conversion of a fuzzy quantity to a specific quantity. In the real-world environment, the output fuzzy sets must be defuzzified to crisp values for decision-making reasons (Rahmani, Hosseinzadeh, Rostamy-Malkh, & Allahviranloo, 2016). The goal of the defuzzification is to find the linguistic term with the highest membership degree for the value of the class attribute (Mary & Arockiam, 2018). In this step, the outcome of the previous steps is used to calculate the final crisp output. According to Liu et al. (2019), in defuzzification, new instance is categorized by allocating it to the class with the maximum membership degree.

7.3 Justification of fuzzy-based narcissism classification

Membership functions and their intervals are chosen in fuzzy rule-based techniques to represent the intrinsic fuzziness of the system. Fuzzy logic has the ability to make machine learning model results better by bridging the intelligence gap because it deals with uncertainty, vagueness, or imprecise factors in human language (Howells & Ertugan, 2017). This motivated its incorporation in this research to improve prediction further. In this research, fuzzy logic could deal with linguistic uncertainty that might have present from the clustered data. Fuzzy logic considers the classification of a problem based on degrees of truth, thereby reducing it on both positive and negative sides (Jefferson et al., 2017).

According to Liu and Cocca (2017), rule-based systems are more interpretable than computational models in text classification techniques such as support vector machine learning. Furthermore, in fuzzy logic, the classification result is provided with a certainty component (fuzzy truth value) rather than an absolute truth. This is because rule-based models operate in a white box environment, making the mapping relationship between an input and an output completely transparent. According to Liu and Cocca (2017), when fuzzy logic and rule-based systems are combined, rules can be represented in a way that is similar to natural language, making the information derived from rules more understandable. This will result in higher confidence in the outcome for people who would like to see the thinking process of text classification by classification techniques.

For this research, the Sugeno fuzzy logic method was chosen because of the nature of its output variables. A Takagi–Sugeno-type fuzzy inference system was chosen. As opposed to Mamdani, Sugeno output variables are constant and interpretable. Furthermore, according to Subramaniam and Venugopal (2020), the Sugeno fuzzy logic system works well with the linear design and is efficient in mathematical analysis. In addition, Sugeno fuzzy is ideal for classification problems.

7.4 Application of fuzzy logic classification

The primary objective of fuzzy logic in this research was to further establish the degree of narcissism through representation in terms of levels. Fuzzy logic considers the classification of a problem based on degrees of truth, thereby reducing bias (Jefferson et al., 2017). In FLS (Figure 7.2), each variable in the input or output is called a linguistic variable. Each linguistic

variable has several values that can be referred to as linguistic terms or fuzzy sets (Karami, Gangopadhyay, Zhou, & Kharrazi, 2015). After the classifier had been trained, the best performing classifier, random forest with an accuracy of 88%, was used to predict the new dataset. After prediction, the classified dataset was further classified into different degrees of narcissism using fuzzy logic as per the process model. The predictions made by the trained model and the output of Vader (tweets polarities) were inputted to the fuzzy-based narcissistic classification system. Figure 7.2 shows a fuzzy logic system with two input variables (polarity & predicted class of narcissism) and five output variables (level of narcissism).

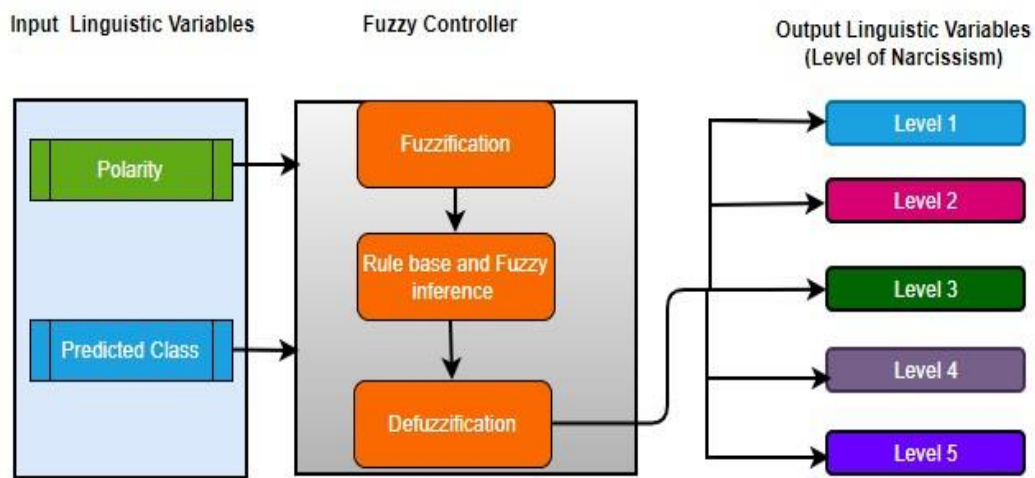


Figure 7.2: Fuzzy logic for narcissistic classification

7.4.1 Linguistic variables

According to Karami et al. (2015), linguistic variables are the input and output variables of the system. This chapter classifies the level of narcissism according to five levels. Therefore, two input variables of polarity and the predicted class of the tweets are defined, and one output variable of degree of narcissism of the tweet is defined. Table 7.1 shows the linguistic variables defined for this study.

Table 7.1: Fuzzy logic system variables

| Type | Linguistic term |
|----------------------------|---------------------|
| Input linguistic variable | Sentiment polarity |
| Input linguistic variable | Predicted labels |
| Output linguistic variable | Level of narcissism |

7.4.2 Fuzzification

Based on the input and output variables, associated fuzzy sets are identified (Table 7.2). For this research, six fuzzy sets were identified for the input variables. The first three fuzzy sets belong to the first input variable of polarity. The polarity can be positive, neutral or negative based on the sentiment score. The second input variable is the predicted class (C) from the trained classifier and is classified into three linguistic terms. The terms are grandiose narcissism (GN), empath (EP) and vulnerable narcissism (VN). The three labels were obtained in Chapter 6.

Table 7.2: Input variables

| Linguistic variable | Fuzzy sets |
|----------------------------|----------------------------|
| Sentiment polarity | Positive (PosP) |
| | Neutral (NeuP) |
| | Negative (NegP) |
| Predicted class | Grandiose narcissism (GN) |
| | Empath personality (EP) |
| | Vulnerable narcissism (VN) |

The output variable had five fuzzy sets. The fuzzy set describes the degree/level of narcissism given two inputs of Tweet polarity and predicted class Level 1 describes a high level of grandiose narcissism. Level 2 relates to a moderate level of grandiose narcissism. Level 3 there relates to neutral personality that is neither grandiose nor vulnerable. Level 4 relates to moderate vulnerable narcissism. Level 5 relates to a higher level of vulnerable narcissism.

Table 7.3: Output variable

| Linguistic variable | Fuzzy sets |
|----------------------------|-------------------|
| Level of narcissism | Level 1 |
| | Level 2 |
| | Level 3 |
| | Level 4 |
| | Level 5 |

7.4.3 Membership functions

Through the membership function, the crisp inputs of the linguistic variables are turned into fuzzy sets. Appropriate membership functions for the partitions of linguistic variables map the inputs into degrees of membership. There are various shapes of membership functions, namely trapezoidal, bell like, triangular, or Gaussian forms (Priyanka & Gupta, 2015; Couso, Borgelt, Hullermeier, & Kruse, 2019). In this research, a triangular membership function was adopted because it retains three variables and establishes a relationship between them (Sheeba & Vivekanandan, 2014). Sentiment analysis score is categorised into three linguistic terms of positive, neutral, and negative. In addition, the second input variable is the predicted class. The classes represent the three categories of narcissism, namely, GN, VN and EP.

7.4.4 Fuzzy logic rules

Fuzzy inference is a collection set of IF–THEN-type rules which convert the fuzzy input to the fuzzy output. Rules are a set of linguistic statements based on IF-THEN statements that follow human expert knowledge (Wu, Zhou, Lu, & Huang, 2017). By using these rules, controlled output, and the conclusion can be derived. In Tsugeno FLS; If Polarity is -1 and predicted class is 1, then $z = f(-1, 1)$. The following nine inference rules (listed after Table 7.4 below) based on Sugeno fuzzy inference mechanism were used in this study.

Table 7.4: Fuzzy logic rules variables

| Linguistic variable | Fuzzy sets |
|---------------------|----------------------------|
| Sentiment polarity | Positive (PosP) |
| | Neutral (NeuP) |
| | Negative (NegP) |
| Predicted class | Grandiose narcissism (GN) |
| | Empath personality (EP) |
| | Vulnerable narcissism (VN) |

Rules

- 1 IF C is GN and S_p is Posp THEN LN is Level 1
- 2 IF C is GN and S_p is NeuP THEN LN is Level 2
- 3 IF C is GN and S_p is NegP THEN LN is Level 3

- 4 IF C is EP and S_p is Posp THEN LN is Level 2
- 5 IF C is EP and S_p is NeuP THEN LN is Level 3
- 6 IF C is EP and S_p is Negp THEN LN is Level 4
- 7 IF C is VN and S_p is Posp THEN LN is Level 3
- 8 IF C is VN and S_p is NeuP THEN LN is Level 4
- 9 IF C is VN and S_p is Negp THEN LN is Level 5

7.4.5 Levels of narcissism

The levels of narcissism are categorised from Level 1 to 5. Level 1 implies a high form of grandiose narcissism, while Level 5 denotes a high level of vulnerable narcissism. Level 3 indicates empath personality, an individual who exhibits neither vulnerable nor grandiose narcissistic traits. After the application of the rules, the next step is the implication of the rules on the output when applied to the two input variables.

Level 1 implies a high form of GN. The seed words for this level are '*i*', '*my*', '*me*', '*myself*', '*I'm*', '*mine*', '*oneself*'. Psychologically, narcissists display a high tendency for self-presentation and self-admiration (Wang, 2017). According to Ozimek, Bierhoff, and Hanke (2018), people with GN frequently communicate positively about themselves in social networks. SNSs are used as channels for self-presentation since they allow users to put self-enhancing posts and status updates (Kauten, Lui, Stary, & Barry, 2015). Casale, Fioravanti, and Rugai (2016) noted that self-representation, and desire for more comments and likes tend to feed into vanity and desire for attention by grandiose narcissists.

Level 2 denotes a moderate level of GN. The seed words for this level are '*happy*', '*brilliant*', '*beautiful*'. While grandiose narcissists are self-centred, they also use affective and social processes words on social media to describe themselves or situations. This category of tweets has been denoted as moderate grandiose and characterises users who tweet to endear themselves to their followers in attempt to increase follower likability and engagement (Bernarte, Festijo, Layaban, & Ortiz, 2015).

Level 3 indicates EP. The seed words identified in this research for this level are *Follow*, '*retweets*', and '*gain*'. It is described as a personality that is neither vulnerable nor grandiose narcissists

Level 4 denotes a moderate level of VN. The seed words for this level are '*worthless*', '*sad*', '*rude*'. Narcissists tend to deflect all their feelings onto others because of their underlying pain and insecurity. Level 4 describes a moderate variation of vulnerable narcissists on social media. Some of their words on social media can at times be positive in nature.

Level 5 denotes a high level of VN. The seed words for this level are '*kill*', '*bullshit*', '*fuck*', '*racist*', '*hate*'. Vulnerable narcissism is marked with hypersensitivity to other people's opinions, desire for approval, defensiveness, introversion, neuroticism, low self-esteem, and insecurity. Oversensitivity by vulnerable narcissists correlates with their negativity as displayed in their negative social media posts. In addition, according to DeWall et al. (2011), vulnerable narcissist uses profane and aggressive language on social media.

7.4.6 Defuzzification

The final process in fuzzy-based narcissism classification is defuzzification and involves determining the degree of narcissism based on the input variables and applicable rules. Defuzzification is the calculation of dependent the variable value based on the resulting fuzzy set after application of rules. There are three defuzzification methods. The first one is the average method: In the output fuzzy set, the average value of the dependent variable. The second method is "average of maximum method". It relates to the average numerical value of the dependent variable with the maximum degree of truth in the output fuzzy set. The last method is the centroid method. It relates to the weighted numerical value of the dependent variable. The degree of truth determines the weight. The Tsugeno inference system is used to compute the output given the inputs. For this research, the Centroid method was used. This method provides a crisp value based on the centre of gravity of the fuzzy set (Vashishtha, & Susan, 2019).

Table 7.5: Fuzzy rule table

| Predicted class (Pc) | F (Pc, Pol) | | |
|----------------------|-----------------|-----------------|-----------------|
| | PosP | NeuP | NegP |
| GN | LN ₁ | LN ₂ | LN ₃ |
| EP | LN ₂ | LN ₃ | LN ₄ |
| VN | LN ₃ | LN ₄ | LN ₅ |

LN₁-Level 1; LN₂-Level 2; LN₃-Level 3; LN₄-Level 4; LN₅-Level 5

PosP-Positive Polarity; NeuP-Neutral Polarity NegP-Negative Polarity

7.5 Experiment setup: Fuzzy logic classification

After the prediction of a new dataset by the classifier, fuzzy logic was incorporated to further break down the degrees of narcissism. The experiment had two input variables and one output variables. The parameters applied were a triangular membership function and nine fuzzy logic rules (Table 7.5).

The input for fuzzy logic is polarity of the tweet and predicted class. As discussed in Section 3.5 – the process model – after application of machine learning classifiers, fuzzy logic is incorporated to improve the identification of traces of narcissism from text. Fuzzy-based narcissism classification has two main input variables and one output variable. The first input variable is the predicted class. This relates to the class/label of a tweet based on prediction by the trained classifier. The label can be GN labelled as 1, EP labelled as 0 and VN labelled as 2. The second input variable is the polarity of the tweet of the new dataset. The polarity can be positive, neutral or negative.

Procedure 7.1: Fuzzy logic application algorithm

Input: SentimentPolarity, Predicted Class (C)

Output: LevelofNarcissism (LN)

Process:

1. For each CleanTweet;
 - 1.1 Initialize temporary empty column LevelofNarcissism to store the result of output
 - 1.2 Identify the narcissistic class label C {GN, EP, VN} of the tweet
 - 1.3 Apply the classifier to the training data to learn the dataset attributes
 - 1.4 Use the trained classifier, to predict the new data into three classes (GN, EP, VN)
 - 1.5 Using the new data create fuzzy profiles with polarity and predicted class of the tweets as input variable
 - 1.6 Create fuzzy inference system (FIS) to classify the tweets
 - 1.7 Using the classifier;
 - 1.7.1 Select the positive tweets and find its Level of Narcissism
 - 1.7.2 Select the negative tweets and find its Level of Narcissism
 - 1.7.3 Select the Neutral tweets and find its Level of Narcissism

```

1.8 Predict the overall level of narcissism of a user the
    basis of Level of Narcissism
1.9 Return LevelofNarcissism
End For

```

7.5.1 Membership functions plot for sentiment polarity

The membership function for the tweet polarity input variable was negative, neutral, and positive. A triangular membership function was adopted. The range for negative polarity was -0.2 to 1; neutral polarity was between -0.2 to + 0.2, while positive polarity had a range between +0.2 to +1, as shown in Figure 7.3. The membership for predicted class were in the range of (0,1) for GN, (1 2) for EP and (2 ,3) for VN. Figures 7.3 and 7.4 are pictorial presentations of the membership functions.

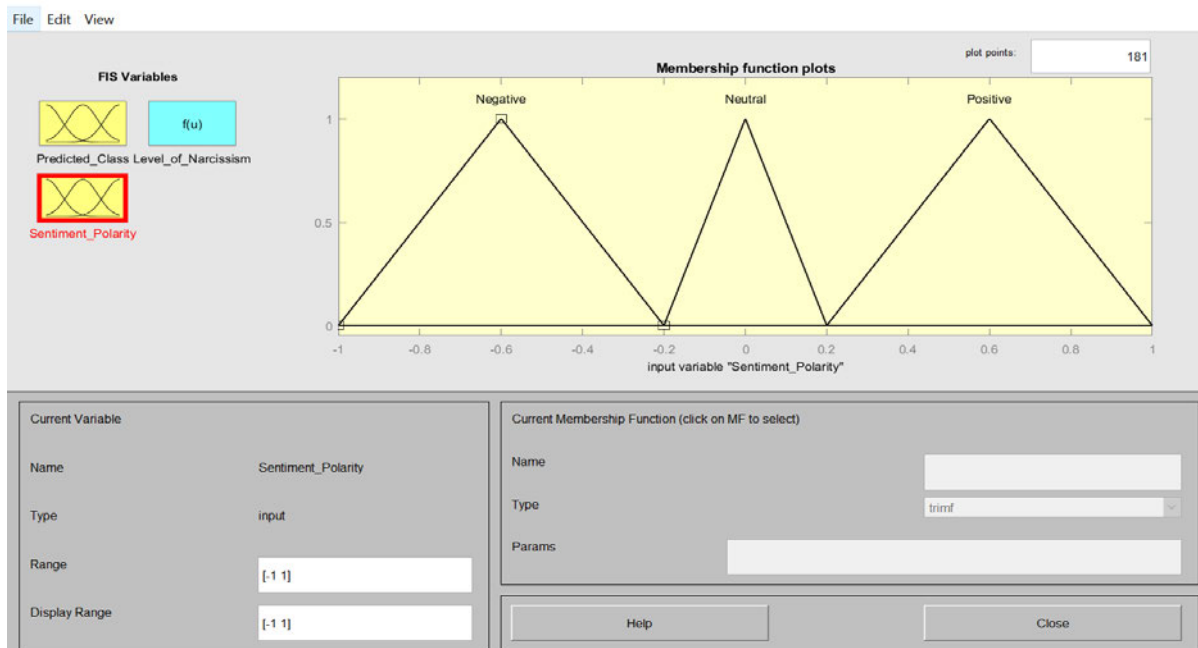


Figure 7.3: Sentiment Polarity Membership functions

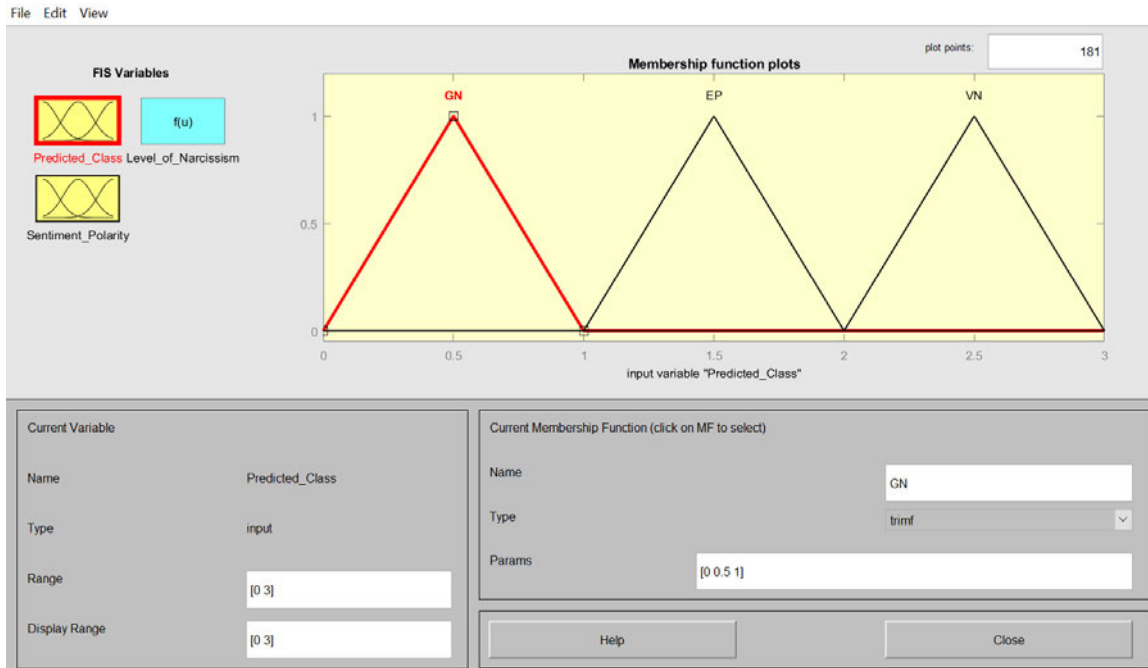


Figure 7.4: Predicted class membership functions

7.5.2 Crisp output levels of narcissism

The output of the FLS is the degrees of narcissism ranging from highly grandiose (represented by GN) which is labelled as level 1 to highly vulnerable (represented by VN) labelled as level 5 of narcissism. The output variable was the level of narcissism which ranged from level 1 to level 5. Level 1 implied low levels of narcissism, while level 5 implied a high degree of narcissism. The output variable was computed based on the rules, and the input variable.

7.6 Fuzzy logic system (FLS) experiments

Five experiments were conducted to establish the degree of narcissism given a tweet polarity and predicted label from the machine learning classifier. A new dataset of 20,000 tweets was classified into three narcissism categories by the trained classifier presented in Chapter 6. The classified dataset together with the respective polarity was fed into the fuzzy logic system in MATLAB. New predictions based on the fuzzy logic rules are discussed in Section 7.4.4 are presented. The value of input variables was varied with the aim of getting the level of narcissism (output variable) given a predicted class score and tweet polarity. The experiments sought to answer the question of “*What is the degree/level of narcissism given Sentiment Polarity and predicted class by trained classifier*”?

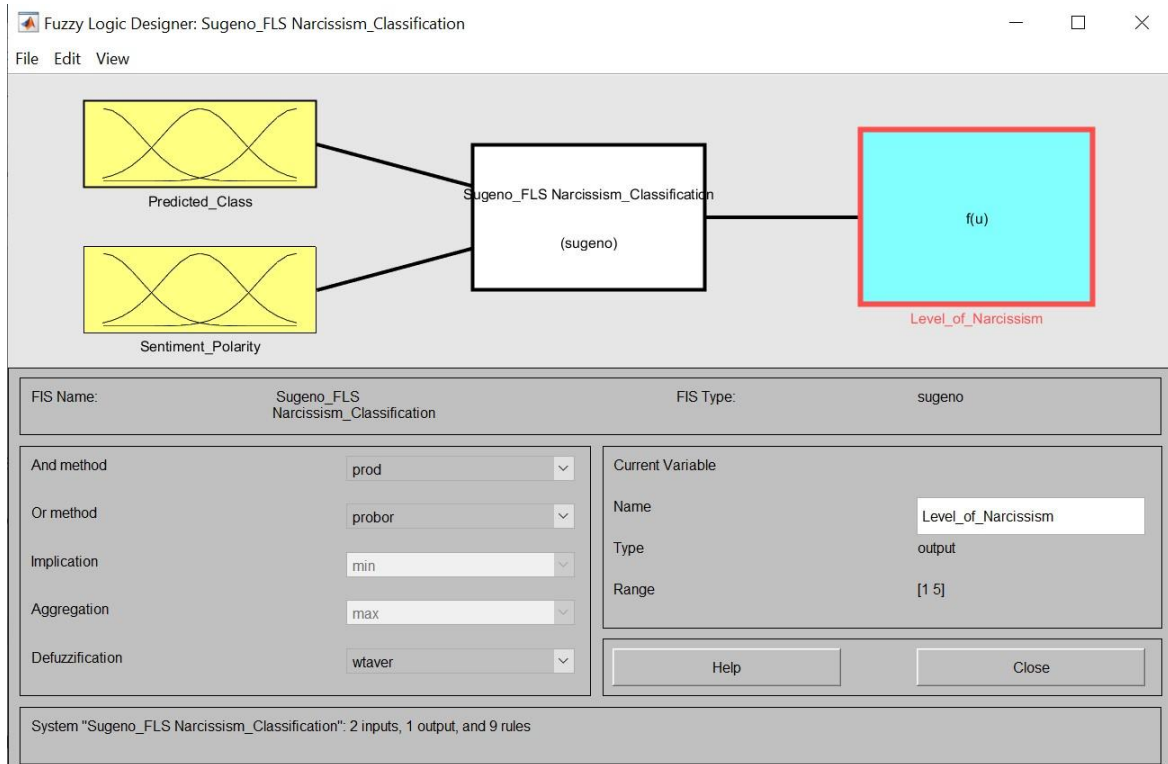


Figure 7.5: Fuzzy-based narcissism classification system

FLS experiment 7.1

The objective of the experiment 7.1 was to determine the level of narcissism given a narcissistic classification and a tweet polarity. Therefore, the input variables were predicted class and sentiment polarity, as discussed in Section 7.4. A new dataset classified by the trained classifier was fed into fuzzy logic inference system in MATLAB. The parameters in the experiment were a triangular membership function and Sugeno inference system. The experiment was subjected to the nine fuzzy logic rules shown in Table 7.5. For each output variable, the aggregation method produces a fuzzy set. The goal of the defuzzification stage is to determine the level of narcissism at which the input vector's membership degree is at its highest. In this experiment, when the membership degree value of the input vector is (05, 0.9) the level of narcissism is '1' as shown in Figure 7.6. This shows that when predicted class falls under class GN as per the rule table and the sentiment polarity is a positive polarity of 0.9, the tweet is then classified as grandiose narcissism.

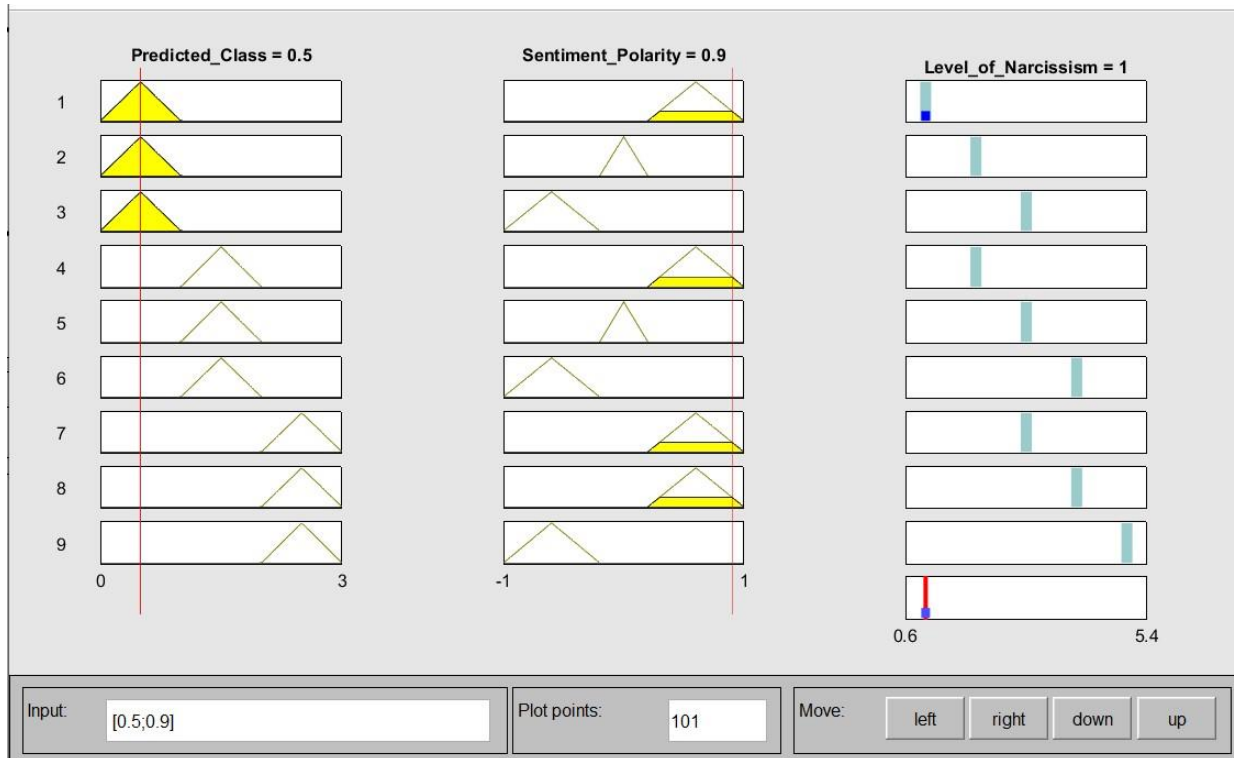


Figure 7.6: Experiment 7.1 output

FLS experiment 7.2

The objective of the experiment 7.2 was to determine the level of narcissism given an input factor of (0.354, 0.125). The two input values represent the lower value of predicted class 1 which is GN and lower class of positive polarity. The parameters in the experiment were a triangular membership function and Sugeno inference system. During defuzzification, FLS attempts to categorise the level of narcissism given a predicted class and a positive polarity. Figure 7.7 shows that when predicted class is 0.354 and the polarity of the tweet is 0.125 the level of narcissism is 2. This implies moderate grandiose narcissism as it has aspects of neutral sentiment.

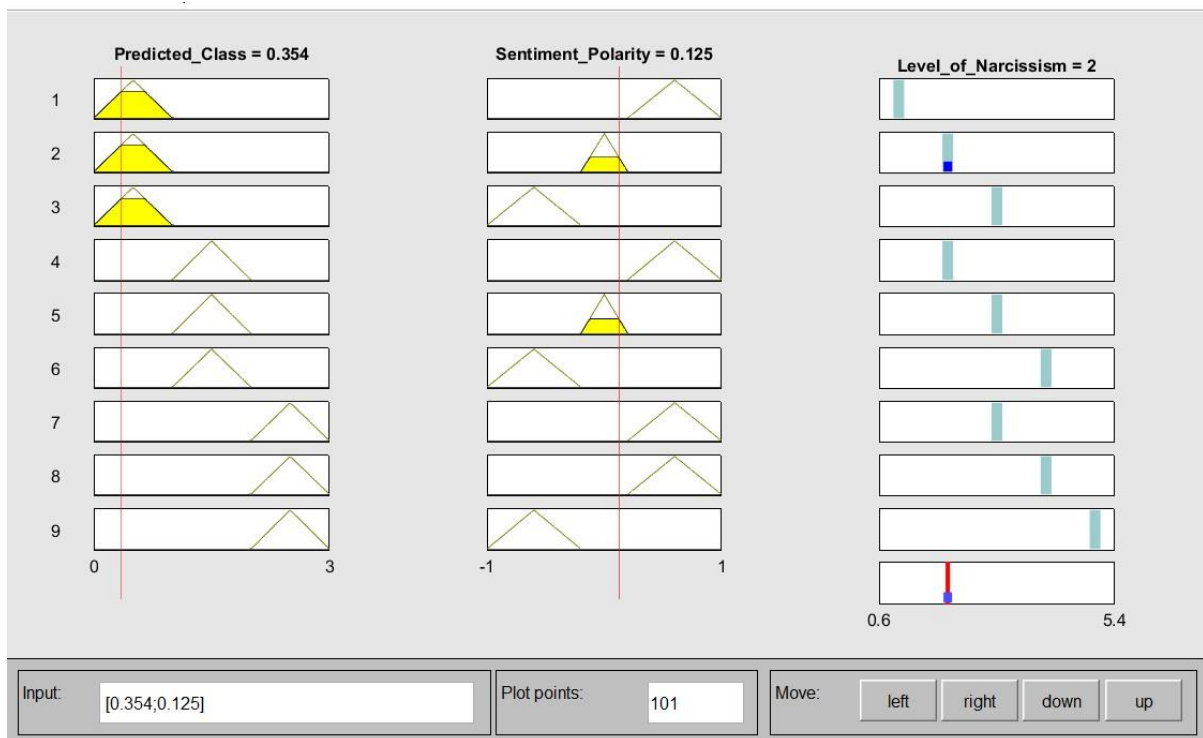


Figure 7.7: Experiment 7.2 output

FLS experiment 7.3

The objective of the experiment 7.3 was to determine the level of narcissism (crisp output) given input factor (1.89, -0.125). In this experiment, when the predicted class of is 1.89 and the sentiment polarity is -0.125 the crisp output is 3 (Figure 7.8). This shows that when the tweet belongs to grandiose narcissism as per classifier prediction and the sentiment polarity is negative, then the level of narcissism is level 3 which is empath personality. This is because while the sentiment belongs to the higher negative category, the lexicons in the tweet are grandiose (positive) in nature. Level 3 implies a neutral level of narcissism that does not fall to either the grandiose category or vulnerable category. This type of narcissism level is discussed in detail in Section 7.4.5.

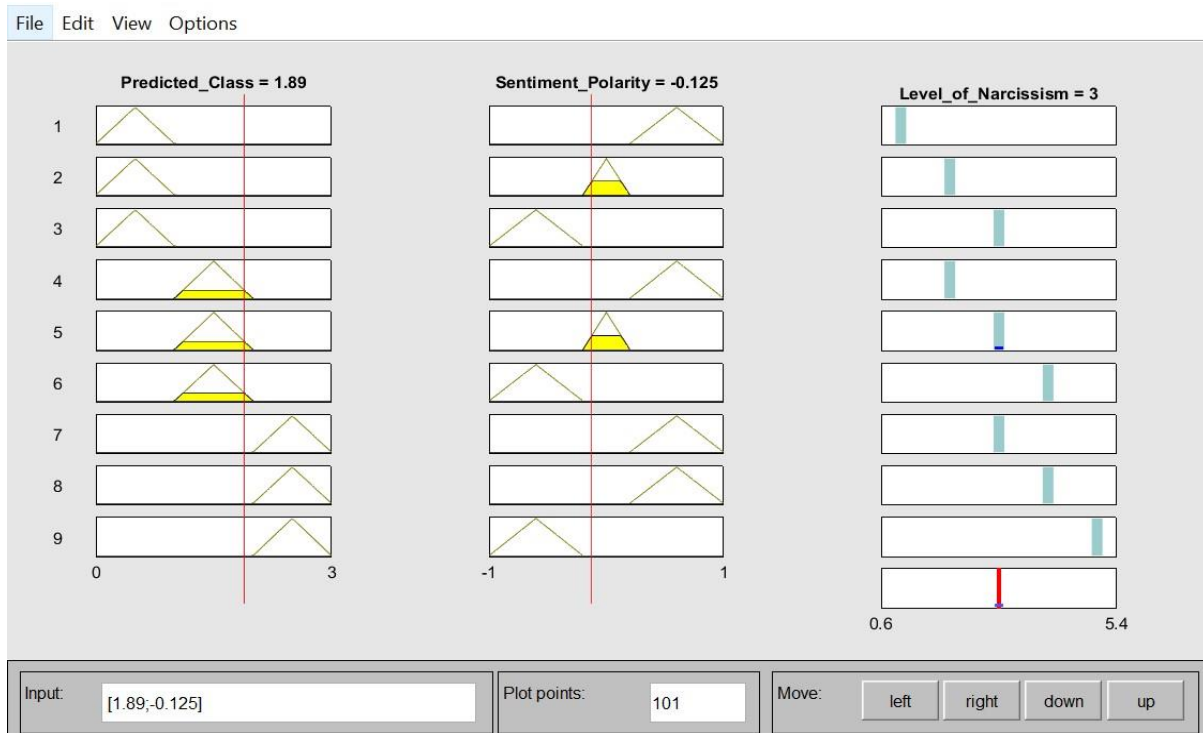


Figure 7.8: Experiment 7.3 output

FLS Experiment 7.4

In experiment 7.4, the research sought to identify the level of narcissism given f (1.12, -0.795). The experiment was implemented in the Tsugenio inference system. The output of this experiment is shown in Figure 7.9 and it can be observed that the level of narcissism is 4 when the sentiment of the tweet is high, and the predicted class is 1 or grandiose narcissism. This represents a moderate level of vulnerable narcissism as described in Section 7.4.5. This is because the first input factor represents strong grandiose narcissism denoted by 1 and a strong negative of -0.795.

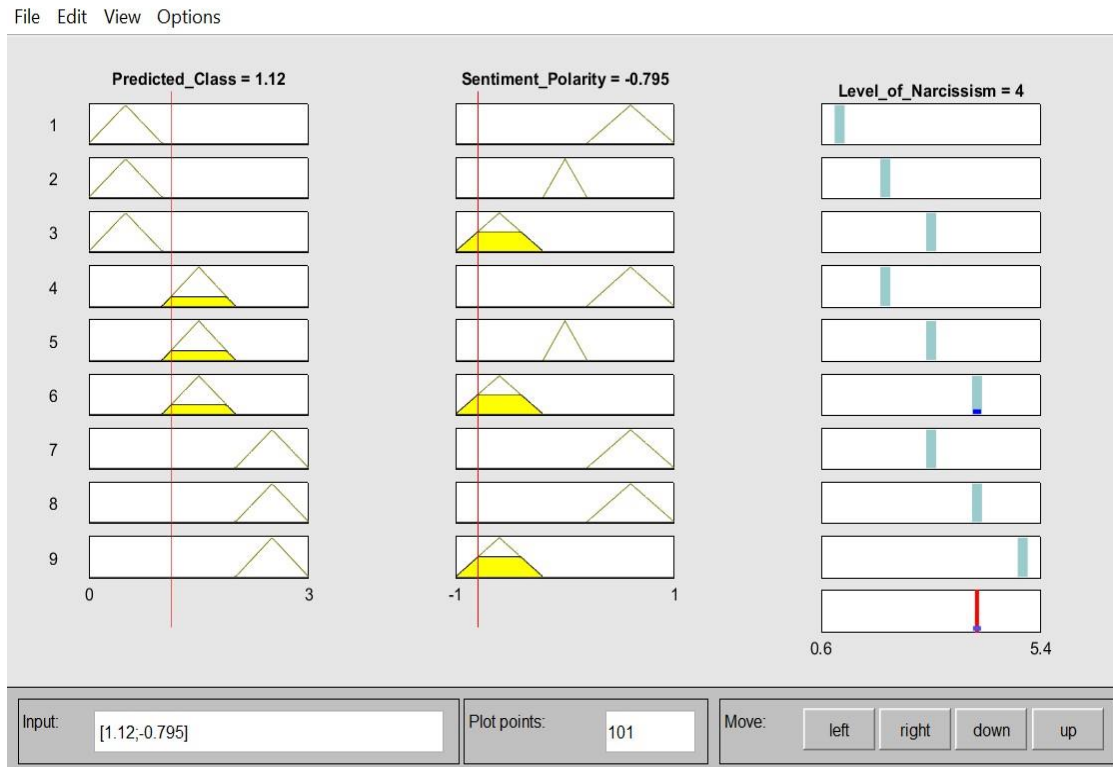


Figure 7.9: Experiment 7.4 output

FLS Experiment 7.5

The objective of the experiment 7.5 was to determine the level of narcissism given an input factor of (2.94, -0.65). The parameters in the experiment were a triangular membership function and the Sugeno inference system. Figure 7.10 shows a crisp output of 5 or level 5 of narcissism. This is because the predicted class is VN and the sentiment polarity is negative. As discussed in Section 7.4.5, the two-input variables represent the higher categories of sentiment polarity and vulnerable narcissism.

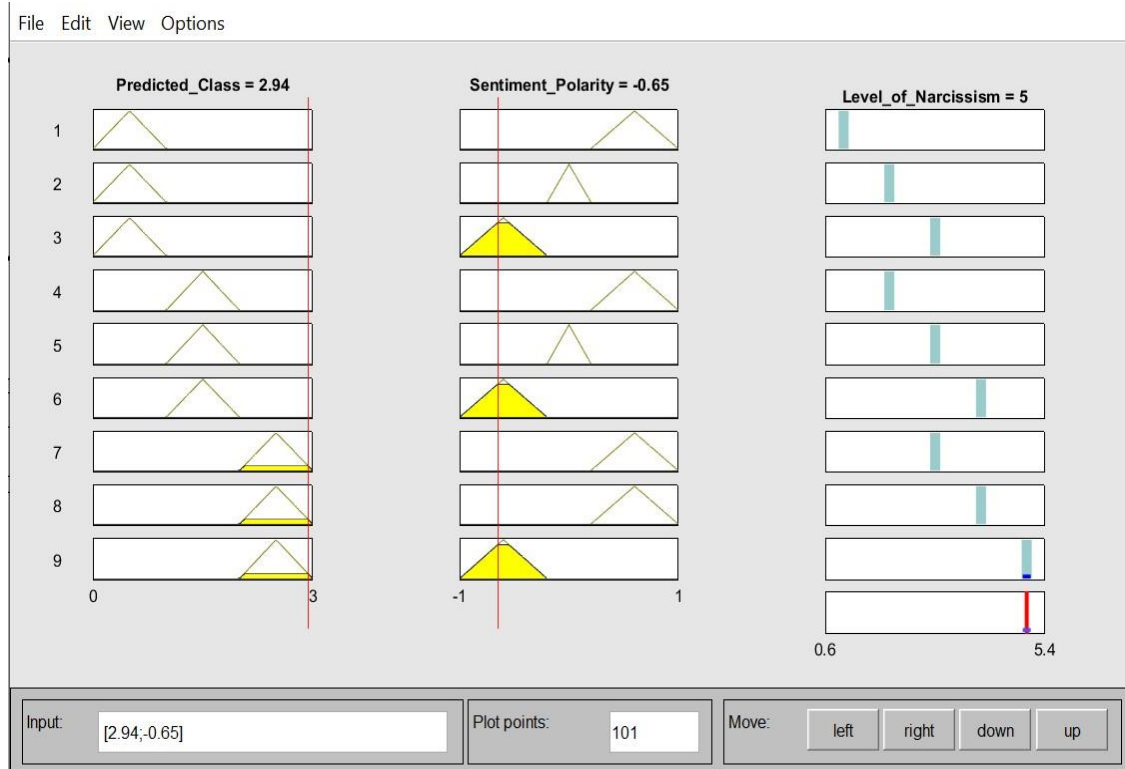





Figure 7.10: Experiment 7.5 output



7.7 Enhancing narcissism classification

Previous works on text classification have often focused on deterministic algorithms' methods without considering the text's fuzziness (Jefferson et al., 2017). Furthermore, different text classification classes are unlikely to be mutually exclusive. According to Liu and Zhang (2018), in emotion recognition, a person can display two or more emotions at the same time. Therefore, this research presents a new approach to classify narcissism into five levels (Level 1 to Level 5) using fuzzy logic and machine learning classifier results and sentiment analysis. Non-fuzzy approaches use techniques that distinguish one class from another in order to categorize the occurrence uniquely, whereas fuzzy logic uses techniques that treat each class equally, by establishing the instance membership degree to each class (Liu & Zhang, 2018). To establish the degree of narcissism, the research used the results generated on the test dataset by the trained classifier and sentiment of the tweets. These two variables become the inputs of the fuzzy logic systems. The level of narcissism (1-5) is determined after the fuzzification and defuzzification process based on FLS rules set in 7.4.4. According to Serrano, Guerrero, Romero, and Olivas (2021), fuzzy logic not only allows for a more realistic representation of real-world data, but it also accomplishes so in a straightforward manner. In comparison to machine learning techniques, fuzzy logic-based techniques typically need fewer rules and variables.

The research recommends emoticons that can be used to indicate the level of narcissism from a high to a low level. Emojis are visual representations of facial expressions, body language, and hand gestures used to indicate emotions and attitudes in text (Gülşen, 2016). According to Yuasa, Saito, and Mukawa (2011), the use of emoticons enriches the communication between the sender and the recipient. Emoticons exist on social media and have also been widely adopted in various channels of communication to express sentiment. Even though emoticons are graphical representations of human emotions with less expressive power than real facial expressions, they are able to convey emotions (Yuasa et al., 2011). Fuzzy logic is applied on the level of narcissism by considering the fuzziness and the vagueness of text classification. Each step in the fuzzy logic system contains a set of treatments. This research argues that interaction with tweets by users and the use of graphical emoticons that shows traces of narcissism will protect users from narcissistic people on social media. Narcissists post on Twitter with the intention to influence others because of the nature of their personality. Thus, if tweets can be flagged using emoticons as a warning to users before interaction on social media, narcissism influence can be reduced. Accordingly, the use of emoticons can strengthen the power of positive interactions on social media interactions. Therefore, this research emphasizes the need to incorporate emoticons on tweets as it will reduce the effect of narcissists by warning and even filtering traces of narcissism from users. Researchers can also adopt the recommended emoticons to complement their personality studies.

Table 7.6: Fuzzy-based narcissism classification emoticons

| Level of narcissism | Emoticon | Description | Sentiment | Seed words |
|---------------------|---|---|----------------|--|
| 1 |  | Level 1 implies high form of grandiose narcissism | +0.5 to +1 | <i>'I', 'my', 'me', 'myself', 'I'm', mine', 'oneself',</i> |
| 2 |  | Level 2 denotes a moderate level of grandiose narcissism. | +0.2 to +0.49 | <i>'happy', 'brilliant', beautiful</i> |
| 3 |  | Level 3 indicates empath personality. | +0.19 to -0.19 | <i>Follow, 'retweets', 'gain'</i> |

| | | | | |
|---|---|--|---------------|--|
| 4 |  | Level 4 denotes a moderate level of vulnerable narcissism. | -0.2 to -0.49 | 'worthless', 'sad', 'rude' |
| 5 |  | Level 5 denotes a high level of vulnerable narcissism. | -0.5 to -1 | 'kill', 'bullshit', 'fuck', 'racist', 'hate' |

7.8 Summary

Fuzzy logic is a simplistic approach to achieve a precise outcome using input data that is unclear, ambiguous, imprecise, noisy, or missing. The degree to which a person exhibits "narcissistic" characteristics varies on several levels. The improved narcissism classification presented in this chapter categorised users based on their level of narcissism using machine learning classifiers and fuzzy approach. Vagueness is associated with the difficulty of making precise and clear distinctions in personality classification. Uncertainty refers to a scenario in which the choice between two or more options is unclear. Through the use of fuzzy sets, uncertainty at various stages can be solved. Both narcissism and the amount of people using social media are on the rise, and research on the relationship between the two is still at infancy stage. The application of fuzzy rules enhances machine learning classification and allows better interpretation on how the final classification was derived from a classifier. This chapter has also recommended emoji icons to distinguish the five levels of narcissism given two input variables.

CHAPTER 8: PERFORMANCE EVALUATION OF NARCISSISM CLASSIFICATION

8.1 Introduction

The methodology adopted for this research was design science research, which places emphasis on generating new knowledge by constructing and evaluating designed artefacts. In design science research, evaluation is critical for determining a model's credibility and any necessary improvements (Peffer et al., 2018). The designed artefacts included the process model, classification model for narcissism prediction, and the recommended research model for this research. In this chapter, after establishing the optimal classification model hyperparameters, the influence of sentiment polarity, input attributes, and dataset size which were evaluated are presented in this chapter. Each classifier was evaluated using different attributes and data sizes with the different parameters presented in Sections 8.3 and 8.4. Section 8.6 presents the comparisons of the classifier performance based on the different parameters and in comparison, with existing literature. The comparison was carried out according to the accuracy and the F1 score value which these methods achieved.

8.2 Performance metrics

The performance evaluation was conducted on the accuracy, the F1-score, precision and recall parameters of the scikit-learn metrics. Accuracy is the proportion of total correct predictions. F-measure (F1-score) consolidates precision and recall and tries to demonstrate the balance between them. The percentage of relevant instances accurately identified is referred to as recall. Precision is the percentage of correctly predicted relevant instances (Kynkäänniemi, Karras, Laine, Lehtinen, & Aila, 2019).

8.2.1 Basic parameters setup

All experiments were implemented in Python 3.7. Before training classifiers, the transformation of the data into feature vectors was done. Data frame mapper library in Python was used to map data frame columns to transformations, then later recombined into features. The datasets were split into training and test data on an 80%-20% basis. All classifiers were implemented in the scikit-learn; a library for Python that offers efficient data mining and data analysis tools such as classifications, regressions, and clustering (Nguyen et al., 2019).

8.2.2 Dataset variables

Four input variables and one target variable were used in the experiments. The first input variable is **Pre-processed tweet** (P_t). This relates to the tweets that had been pre-processed as Rule 4.1 and is denoted as (P_t). The second input variable is **Tweet Frequency** (T_f). It relates to the number of tweets (statuses posted) made by the user. This was categorised into three as discussed in Section 6.7.1. The third input variable is the '**Lexicon presence** (L_p)'. It relates to the presence or absence of a narcissistic-related words based on the narcissistic dictionary as discussed in Section 5.6. The fourth variable used in training is the sentiment polarity (S_p) of the tweet. It relates to the polarity of the pre-processed tweet (P_t) and can be positive, neutral, or negative. The target variable is narcissism class and classified into three categories: *Grandiose Narcissism-GN/1*; *Empath Personality-EP/0*; *Vulnerable Narcissism-VN/2*. The output variable is a trained classifier based on the four input variables and the target variable.

8.3 Impact of sentiment analysis on classifier accuracy

According to Lin, Mao, and Zeng (2017), personality influences user expressions and attitudes but is seldom accounted for in emotional classification. People with the same personality have been found to exhibit similarities in writing and expressions (Wilson, DeRue, Matta, & Howe, 2016). This feature is the basis for introducing sentiment analysis into personality prediction. Sentiment analysis classifies the polarity orientation of tweets based on their sentiment inclination. In this experiment, the effect of sentiment on the classification accuracy of various classifiers is evaluated. The results are presented relative to the classifier accuracy without a sentiment as a training variable to show how sentiment influences classifier performance easily. Therefore, a classifier is trained on three variables, i.e., pre-processed text, number of tweets per user and presence and absence of lexicon. The scores of each classifier with these parameters are then presented.

Table 8.1: Classifier accuracy with respect to sentiment analysis

| Classifier | Metric | Class/Label | | | Accuracy (%) | F1-score (%) |
|---------------|---------------------|-------------|-----------|-----------|--------------|--------------|
| | | GN | EP | VN | | |
| Naïve Bayes | Precision | 0.89 | 0.81 | 0.91 | 82.49 | 80 |
| | Recall | 0.48 | 0.99 | 0.40 | | |
| | F1-score (%) | 62 | 89 | 56 | | |
| SVM | Precision | 0.87 | 0.83 | 0.84 | 81.44 | 81 |
| | Recall | 0.42 | 0.96 | 0.44 | | |
| | F1-score (%) | 85 | 88 | 58 | | |
| Random forest | Precision | 0.86 | 0.86 | 0.89 | 86.46 | 86 |
| | Recall | 0.67 | 0.97 | 0.54 | | |
| | F1-score (%) | 76 | 91 | 67 | | |
| Ensemble 1 | Precision | 0.87 | 0.85 | 0.93 | 85.62 | 84 |
| | Recall | 0.63 | 0.98 | 0.49 | | |
| | F1-score (%) | 73 | 91 | 64 | | |
| Ensemble 2 | Precision | 0.85 | 0.88 | 0.91 | 84.8 | 85 |
| | Recall | 0.61 | 0.95 | 0.52 | | |
| | F1-score (%) | 76 | 89 | 66 | | |
| Ensemble 3 | Precision | 0.87 | 0.84 | 0.92 | 85.08 | 84 |
| | Recall | 0.61 | 0.98 | 0.44 | | |
| | F1-score (%) | 72 | 91 | 60 | | |
| Ensemble 4 | Precision | 0.89 | 0.85 | 0.94 | 85.90 | 85 |
| | Recall | 0.63 | 0.98 | 0.49 | | |
| | F1-score (%) | 74 | 91 | 65 | | |

Experiment 8.1: Classifier accuracy with sentiment as a variable

The objective of experiment 8.1 was to determine the impact of sentiment analysis on classifier accuracy. Supervised learning algorithms were utilised to create models with labelled data. The labelled part of the dataset was utilised here for the creation of the model. The input variables were (P_0) , (L_p) and (T_f) . The classifiers were trained with and without a sentiment as part of the training variables. From the accuracy of the classifiers, random forest and ensemble classifier of Naïve Bayes, SVM and random forest had higher accuracies of 94% more than the other classifiers. In addition, it can be observed in the line plot that using sentiment polarity as a variable has an effect on classifiers accuracies. As

seen in Figure 8.1, there is a clear difference in regards to classifier accuracy when the variable was used and not used. This shows that when classifying personality, the sentiment of texts made by users ought to be taken into consideration

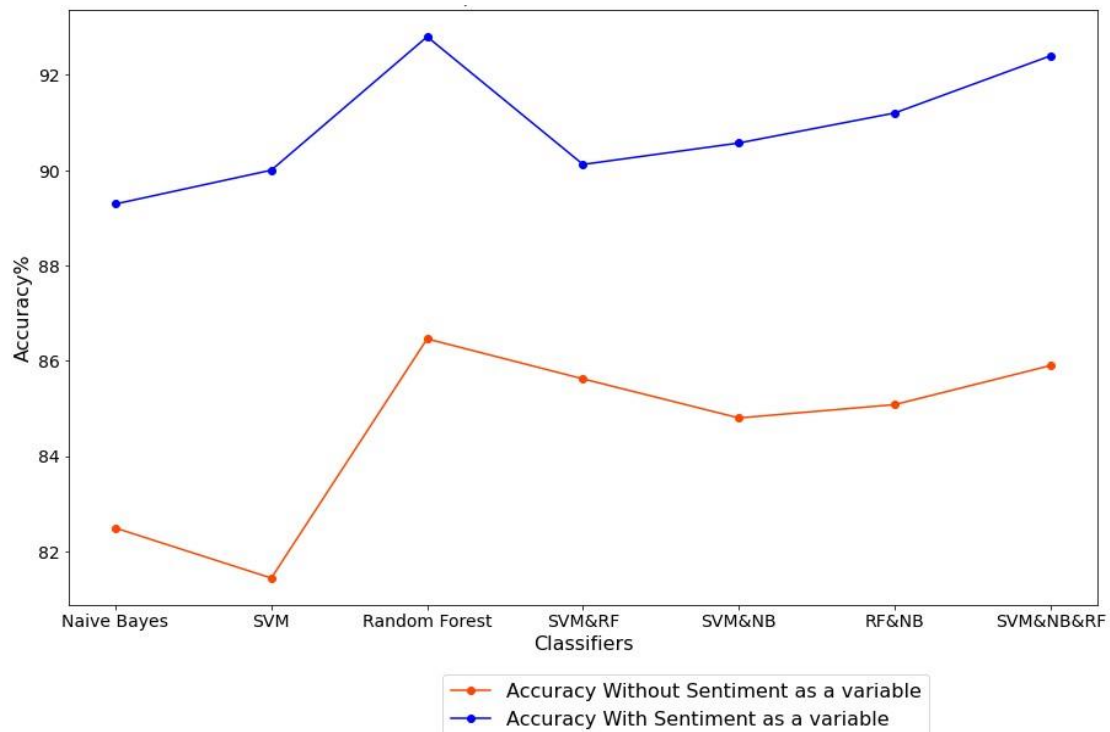


Figure 8.1: Classifier accuracy with sentiment as a variable

8.4 Classification experiments

In these experiments, classifier performances based on different dataset sizes and different input attributes are presented. Three supervised classifiers and four ensemble classifiers models were trained and evaluated in scikit-learn in Python. The data used to construct and evaluate the models was transformed into numeric values to make them suitable for the construction of the model. The dataset has input variables, target variable, and output variable. In experiments 8.2, 8.3, 8.4, and 8.5 the influence of the input variables on the accuracy is presented. Generating a model using the technique of selected classifiers involves mapping all the significant patterns and relationships among specified input attributes to predict the target variable.

8.4.1 The impact of the number of input variables on classification accuracy

The objective of these experiments was to examine how various dataset input variables influence the classifier performance in terms of accuracy. In this experiment, the research was interested in finding out the relation between the performance of classifiers and the

number of variables that the model considers. Thus, three experiments were conducted with a different number of training variables.

Experiment 8.2: Performance of classifiers based on four input attributes

In experiment 8.2, the performance of the classifiers when trained with four input attributes was evaluated. The input variables used were (P_t), (L_p), (T_f) and (S_p). Three individual classifiers and four ensemble classifiers were trained in the experiment.

Experiment testbed:

Input:

Input Variable 1: Pre-processed tweet (P_t)

Input Variable 2: Lexicon presence (L_p)

Input Variable 3: Tweet frequency (T_f)

Input Variable 4: Sentiment polarity (S_p)

Classifiers: SVM, NB, RF, ensemble vote classification

Output: Refer Table 8.2

Table 8.2: Experiment results with four input variables

| Classifier | Metric | Class/Label | | | Accuracy (%) | F1-score (%) |
|---------------|---------------------|-------------|-----------|-----------|--------------|--------------|
| | | GN | EP | VN | | |
| Naïve Bayes | Precision | 0.91 | 0.93 | 0.88 | 91.95 | 90 |
| | Recall | 0.78 | 1 | 0.66 | | |
| | F1-score (%) | 84 | 96 | 76 | | |
| SVM | Precision | 0.87 | 0.93 | 0.84 | 91.23 | 91 |
| | Recall | 0.82 | 0.96 | 0.74 | | |
| | F1-score (%) | 85 | 95 | 78 | | |
| Random forest | Precision | 0.94 | 0.95 | 0.96 | 95.02 | 95 |
| | Recall | 0.89 | 0.99 | 0.80 | | |
| | F1-score (%) | 92 | 97 | 87 | | |

Table 8.2: Experiment results with four input variables cont.

| | | | | | | |
|---|--|---------------------------|------------------------|---------------------------|-------|----|
| Ensemble 1 (Naïve Bayes & SVM) | Precision recall F1-score (%) | 0.94 0.87 90 | 0.94 1 97 | 0.98 0.75 85 | 94.37 | 94 |
| Ensemble 2 (random forest & SVM) | Precision recall F1-score (%) | 0.94 0.77 85 | 0.90 1 95 | 0.96 0.60 74 | 91.05 | 90 |
| Ensemble 3 (Naïve Bayes, SVM and random forest) | Precision recall F1-score (%) | 0.95 0.86 90 | 0.94 1 97 | 0.97 0.76 95 | 94.42 | 94 |
| Ensemble 4 (Naïve Bayes and random forest) | Precision recall F1-score (%) | 0.92 0.83 87 | 0.92 1 96 | 0.98 0.65 78 | 92.75 | 92 |

As shown in Table 8.2, random forest obtained the highest accuracy at 95%, followed by an ensemble of SVM, RF and Naïve Bayes with an accuracy of 94%. The least performing classifier was SVM with an accuracy of 91.9% and Naïve Bayes with F1-score of 90%. In addition, ensemble 3 achieved 94.42% accuracy and 94% F1-score. The classifier performances confirm its effectiveness at classifying narcissistic personality on Twitter with four input variables. These results can help researchers and authorities identify traces of narcissism from Twitter while incorporating various variables in the model for better accuracy.

Experiment 8.3: Performance of classifiers based on three input attributes

The goal of experiment 8.3 was to see how well the classifiers performed after being trained with three input variables. The input variables were (P_t) , (L_p) , and (T_f) . Three individual classifiers and four ensemble classifiers were trained in the experiment. Table 8.3 shows the performances of the classifiers using three input variables.

Experiment testbed:

Input:

Input Variable 1: Pre-processed tweet (P_t)

Input Variable 2: Tweet frequency (T_f)

Input Variable 3: Lexicon presence (L_p)

Classifiers: SVM, NB, RF, ensemble vote classification

Output: Refer Table 8.3

Table 8.3: Experiment results with three input variables

| Classifier | Metric | Class/label | | | Accuracy (%) | F1-score (%) |
|--|---------------------|-------------|------|------|--------------|--------------|
| | | GN | EP | VN | | |
| Naïve Bayes | Precision | 0.86 | 0.91 | 0.82 | 89.29 | 89 |
| | recall | 0.75 | 0.97 | 0.64 | | |
| | F1-score (%) | 80 | 94 | 72 | | |
| SVM | Precision | 0.87 | 0.93 | 0.84 | 90 | 91 |
| | recall | 0.82 | 0.96 | 0.74 | | |
| Random forest | Precision | 0.89 | 0.94 | 0.90 | 92.8 | 93 |
| | recall | 0.88 | 0.97 | 0.78 | | |
| | F1-score (%) | 88 | 95 | 84 | | |
| Ensemble 1 (Naïve Bayes & random forest) | Precision | 0.86 | 0.88 | 0.82 | 90.12 | 89 |
| | recall | 0.72 | 0.94 | 0.66 | | |
| | F1-score (%) | 79 | 91 | 74 | | |
| Ensemble 2 (Naïve Bayes & SVM) | Precision | 0.85 | 0.93 | 0.84 | 90.57 | 89 |
| | recall | 0.81 | 0.96 | 0.74 | | |
| | F1-score (%) | 83 | 94 | 94 | | |
| Ensemble 3 (Naïve Bayes, SVM and random forest) | Precision | 0.88 | 0.94 | 0.89 | 91.2 | 90 |
| | recall | 0.88 | 0.96 | 0.77 | | |
| | F1-score (%) | 85 | 95 | 78 | | |
| Ensemble 4 (Naïve Bayes, and random forest) | Precision | 0.90 | 0.93 | 0.89 | 92.40 | 92 |
| | recall | 0.86 | 0.97 | 0.75 | | |
| | F1-score (%) | 88 | 96 | 79 | | |

When the input variables were reduced from four to three, it was observed that compared to the four variables, the accuracy and F1-score of the classifiers reduced by an average of 2%. Random forest classifier F1-score decreased from 95% to 93% while ensemble 4 achieved 92.4 % as opposed to 94% in the first experiment with four training variables. SVM and Naïve Bayes also reduced in accuracy and F1-score by 2%. This shows that performance of ML classifiers varies with input variables techniques used, thus implying that for any research the optimal number of variables have to be identified and which variables results in optimal performance.

Experiment 8.4: Performance of classifiers based on two attributes of dataset (P_t) & (T_f)

Experiment 8.4 sought to examine how the classifiers will perform with only two input variables. The variables were pre-processed tweet and tweet frequency. The first input variable is the pre-processed tweet and the second input variable was tweet frequency. Tweet frequency relates to the number of times a user had tweeted as shown in the dataset (Section 6.7.1).

Experiment test bed:

Input:

Input variable 1: Pre-processed tweet (P_t)

Input variable 2: Tweet frequency (T_f)

Classifiers: SVM, NB, RF, ensemble vote classifiers

Output: Refer to Table 8.4

Table 8.4: Experiment results with two (P_t, T_f) input variables

| Classifier | Accuracy (%) | F1-score (%) |
|---|--------------|--------------|
| Naïve Bayes | 84.15 | 85 |
| SVM | 83 | 82 |
| Random forest | 86.45 | 86 |
| Ensemble 1(Naïve Bayes & random forest) | 84.15 | 84 |
| Ensemble 2 (Naïve Bayes & SVM) | 82 | 83 |
| Ensemble 3 (Naïve Bayes, SVM & random forest) | 84 | 84 |
| Ensemble 4 (Naïve Bayes & random forest) | 85 | 85 |

As shown in Table 8.4, the highest accuracy with two input variables of tweets and tweet frequency was random forest with 86.45%. In comparison with experiments 8.2 and 8.3 which had a high accuracy of 92.8 and 94.42 respectively, it can be concluded that the more the input training variables the better the classifier model.

Experiment 8.5: Performance of classifiers based on two attributes of dataset -(P_t) & (L_p)

The last classification experiment sought to examine how the classifiers would perform with two input variables. The variables were pre-processed tweet and lexicon presence.

Experiment test bed:

Input:

Input variable 1: Pre-processed tweet (P_t)

Input variable 2: Lexicon presence (L_p)

Classifiers: SVM, NB, RF, ensemble vote classifiers

Output: Refer to Table 8.5

Table 8.5: Experiment results with two (P_t, L_p) input variables

| Classifier | Accuracy (%) | F1-score (%) |
|---|--------------|--------------|
| Naïve Bayes | 83.50 | 81 |
| SVM | 82 | 84 |
| Random forest | 86.97 | 86 |
| Ensemble 1(Naïve Bayes & random forest) | 84.25 | 84 |
| Ensemble 2 (Naïve Bayes & SVM) | 83 | 82 |
| Ensemble 3 (Naïve Bayes, SVM & random forest) | 85 | 84 |
| Ensemble 4 (Naïve Bayes & random forest) | 86 | 85 |

Table 8.5 shows the performance of the classifiers using pre-processed tweet and tweet frequency as the input variables. Random forest had a high accuracy of 86.97 followed by ensemble 4 with accuracy of 86%.

8.4.2 Summary on the impact of input variables on classification accuracy

Figure 8.2 shows the summary of experiments 8.2, 8.3, 8.4 and 8.5 for test. The experiments revealed a significant correlation between the classifier performance and the number of input variables. From Figure 8.2 it can be observed that performance of all classifiers was enhanced

as the number of input variables increased. From the five experiments (8.1, 8.2, 8.3, 8.4, 8.5) conducted, it is evident that the number of input variables influence classifier accuracy. This finding supports Choi and Lee's (2017) conclusion that big data analysis necessitates a greater understanding of data set attributes, data structure, and rate of update frequency.

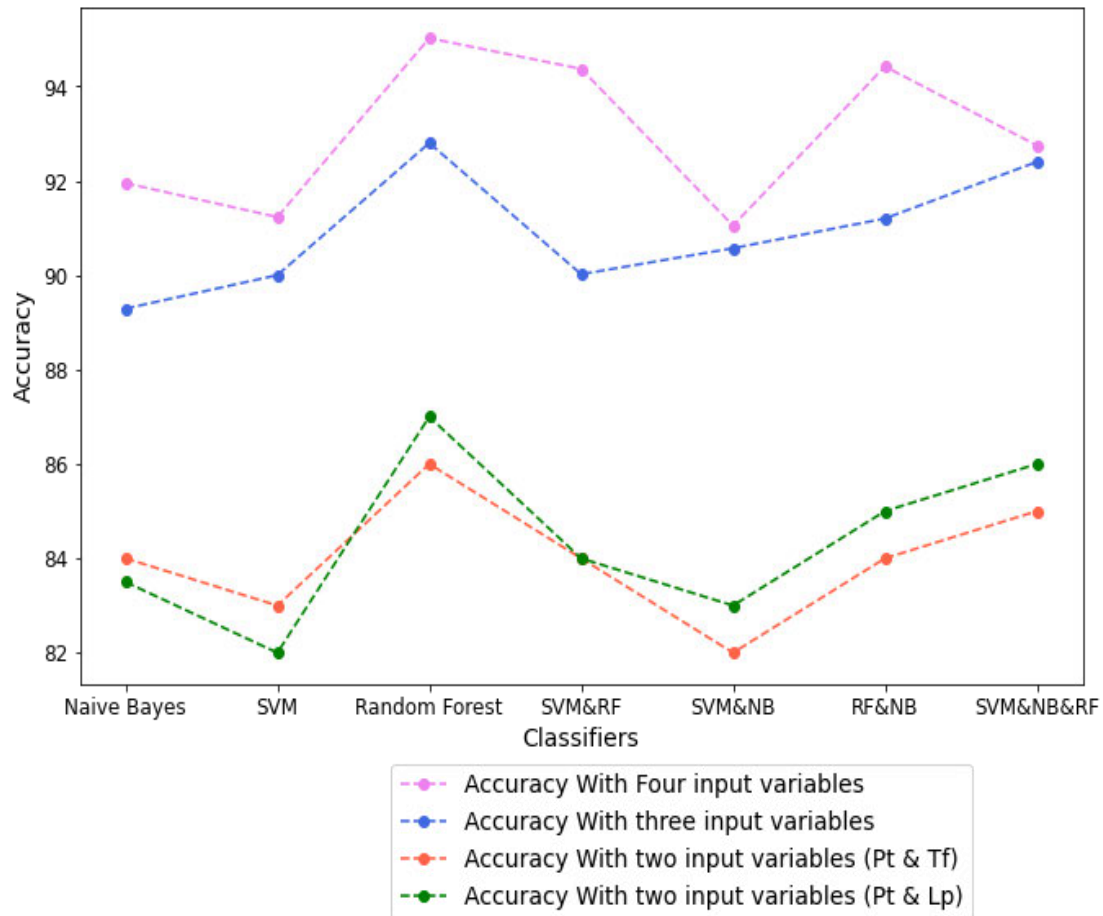


Figure 8.2: Input variable accuracy comparison

Experiment 8.6: The impact of on data size variation on classification accuracy

In this experiment, the research sought to investigate scalability and the impact of data size on classifier performance. To achieve the experiment objective, the properties of the training datasets were controlled. Instead of using the whole dataset used in Section 6.6 experiments, the size of the dataset was split into three categories. The first dataset used was annotated as *consolidated tweet lists 100* and refers to the whole dataset used in Chapter 6 experiments. Three other datasets were split from this (*Consolidated tweet lists 100*). The first dataset is *consolidated tweet lists 80*. This refers to the 80% tweets of the whole dataset extracted randomly. The second dataset comprised 60% of the

80% *consolidated tweet lists* and is labelled *consolidated tweet sublist 60*. The third dataset comprised 40% from *consolidated tweet lists 80 dataset* and 20% from *consolidated tweet lists 100*. The third dataset is labelled as *consolidated tweet sublist 40*.

Experiment test bed:

Input:

Input variable 1: Pre-processed tweet (P_t)

Input variable 2: Tweet frequency (T_f)

Input variable 3: Lexicon presence (L_p)

Input variable 4: Sentiment polarity (S_p)

Classifiers: SVM, NB, RF and ensemble voting classifiers

Output: Refer Figure 8.3

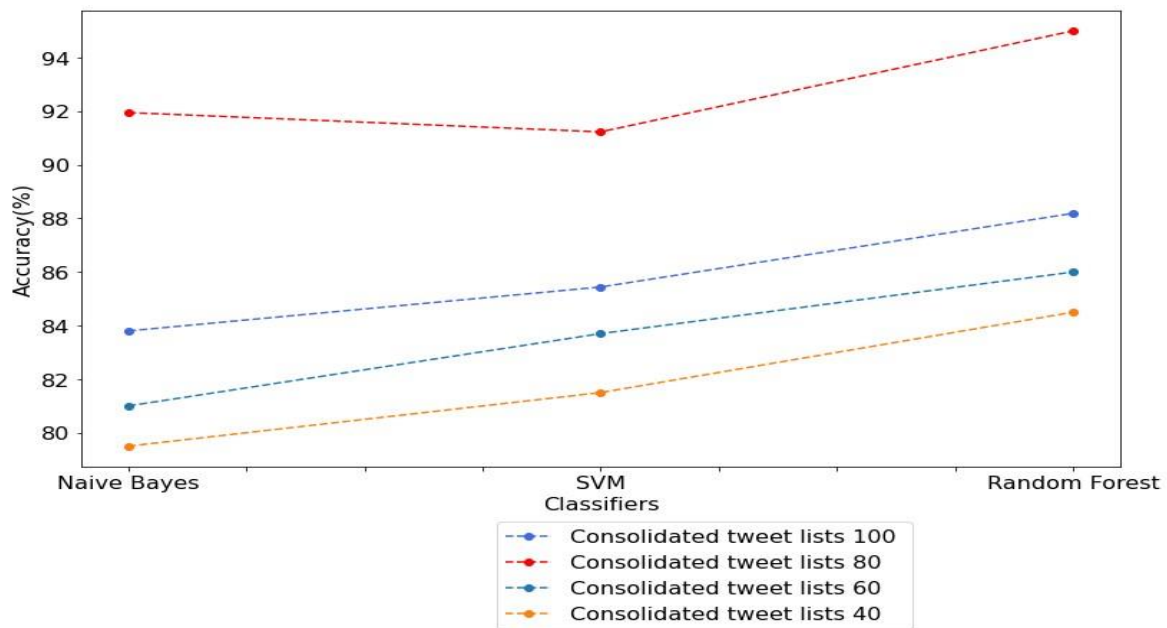


Figure 8.3: Performance accuracy of individual classifiers based on different data sizes

The objective of experiment 8.6 was to evaluate the significance of dataset size on the performance of classifiers. According to Althnian et al. (2021), the size of a dataset is considered as a significant factor in evaluating the performance of the classifier. Small datasets can lead to over-fitting, while large datasets can lead to better classification results. Furthermore, in regards to large dataset, there is an optimal size of dataset whereby any further increase in the dataset size does not result in any improvement in classifier accuracy. Althnian et al. (2021) noted that previous studies have focused on maximizing classifier's

accuracy on limited size datasets, while paying less attention to the impact of the size attributed of the dataset.

As shown Figure 8.3, one large dataset (consolidated tweet lists 100) was split into three small subsets. In terms of accuracy, the change in model performance as a result of the reduction in dataset size was assessed. While there was a general upward trend in accuracy from sublist 40,60,80, the accuracy dropped with list 100. The results highlight significant influence of data size on the classifiers performance and that optimal dataset size exists where any increase of data does not increase the accuracy of any classifier further. The optimal dataset for this experiment was sublist 80 as shown in Fig. 8.3 and Fig. 8.4.

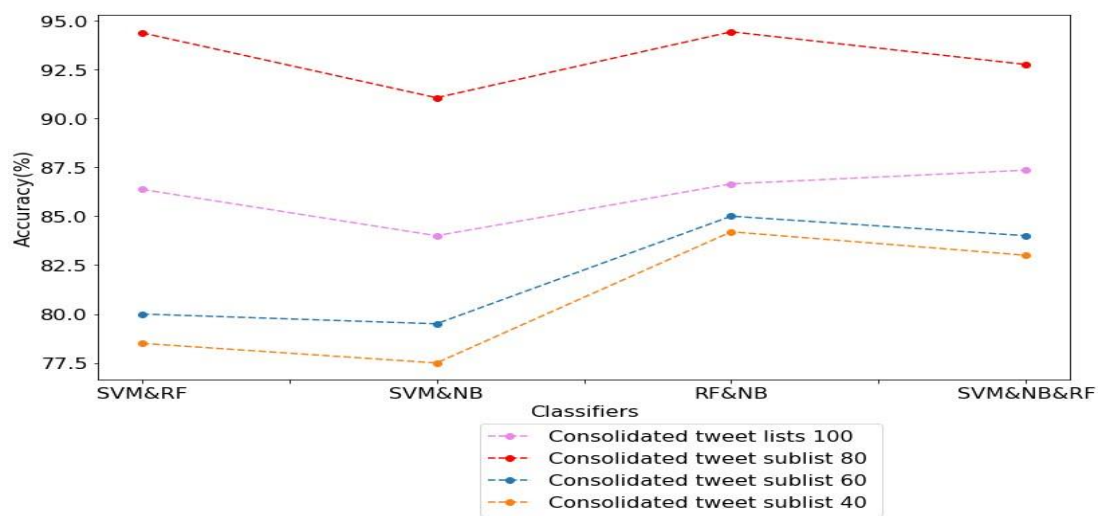


Figure 8.4: Performance accuracy of ensemble classifiers based on different data sizes

Figure 8.4 shows the performance of different combinations to ensemble classifiers based on various data sizes. Consistent with individual classifiers, better accuracy was achieved with consolidated tweet lists 80. It can be observed that for this experiment the optimum data size is 80% of the original dataset. This implies data for any prediction research, researchers have to perform different experiments to identify their optimal dataset, and with which type of attributes.

8.5 Impact of input variables: Experiment outcome summary

Experiments conducted using more input variables (4) resulted in comparatively better performance than those trained with two and three input variables. Different metrics were utilised to determine the model with the best performance. As shown in Figure 8.2, there was a precise distinction classifier accuracy when the training variables increased. This indicates

that models ought to be trained with more than one input variable to perform successful prediction.

In addition, it can be noted that sentiment is a key variable in personality classification. As shown in Figure 8.1, the accuracy of classifiers increased by +5% when the sentiment variable was included in training the classifier. Social media users communicate to their circles of closely known and unknown friendships through text posts (Khowaja, Mahar, Nawaz, Wasi, & Rehman, 2019). Thus, incorporating sentiment in narcissism classification helps provide overall insight opinion of users to different topics on social media. It is also notable to point out that there was a strong correlation between classifier performance and training variables. When fewer variables were used, the accuracy was low, as shown in Figure 8.2, but the accuracy increased when the variables increased.

In addition, ensemble learning is a powerful solution for combining the learning models. From the above discussion, it can be observed that all the ensemble classifiers including random forest, which is an ensemble of decision trees, surpassed other ML models in classifying of narcissism tweets. Table 8.6 shows the performance of the two best performing classifiers (random forest (RF) and ensemble 3). RF outperforms ensemble classifier 3 by 1% in accuracy and F1-score metric. Ensemble three had a better recall of 1 for class EP and better precision of 0.97 compared to RF. In regards to the VN class, RF had a better precision of 0.8 compared to EN3 with 0.76. It was concluded that semi-supervised models outperformed supervised models.

Table 8.6. Experimental results of RF and En3 with four input variables

| Classifier | Accuracy (%) | Class | Precision | Recall | F1-score (%) |
|---------------------------|---------------------|--------------|------------------|---------------|---------------------|
| Random Forest | 95.02 | GN | 0.94 | 0.89 | 95 |
| | | EP | 0.95 | 0.99 | |
| | | VN | 0.96 | 0.80 | |
| Ensemble 3 (RF, SVM & NB) | 94.42 | GN | 0.95 | 0.86 | 94 |
| | | EP | 0.94 | 1 | |
| | | VN | 0.97 | 0.76 | |

8.6 Comparison of performance with related research

According to Swearingen et al. (2017) and Gemert (2017), choosing the suitable classifier for a dataset and configuring and using an optimal classifier and the right dataset attributes is a current research problem in machine learning. This is largely due to the large number of different algorithms available and their difficulty in deployment, fine-tuning and interpretation (Luque, Carrasco, Martín, & De las Heras, 2019). At the same time, more data than ever before has become available to users but the quality of the data has not necessarily improved. These challenges require novel solutions. In addition, Choi and Lee (2017) asserted that the common challenge in machine learning studies is obtaining an optimum data size that can produce a reasonably high level of accuracy. This chapter presented various experiments that had been conducted in order to explore the influence of pre-processing, data set size and the number of input variables on the performance classifiers. High accuracy was achieved with more input variables. The use of four variables and 80% of the dataset achieved a high accuracy of 95%, which was better than existing studies.

Saeidi, Sousa, Milios, Zeh, and Berton (2019) conducted a study with the aim of categorising online harassment in Twitter posts. The study used Random Forest, linear SVM, Gaussian SVM, Polynomial SVM, multilayer perceptron, and AdaBoost methods. The dataset consisted of 10,622 Tweets. They tested different approaches to develop features for the classifiers and, thereafter, conducted classification and applied 10-fold cross-validation setup. Random forest was their best ranked classifier, with an accuracy of 93.5%. The study concluded that the use of TF-IDF vectors presented higher performance (Saeidi et al., 2019). Compared to the experiments conducted in Section 8.4, the random forest classifier trained with a large data size and training variables achieved better accuracy than their study.

Christian, Suhartono, Chowanda, and Zamli (2021) conducted a study to predict personality from Facebook and Twitter. They used my Personality dataset consisting of 9,917 status updates and their second dataset consisted of 46,238 Tweets. The best classifier obtained an 86.2% accuracy and 91.2% F1-score on the Facebook dataset. The Twitter dataset had an accuracy of 88.5% and 88.2% F1-score score. In comparison with the results of this research classification experiments, the research achieved better accuracy of 95% in the first dataset of 142,000 tweets and 92% in the second dataset of 100,000 tweets. On both datasets, Christian et al. (2017) observed that sentiment analysis and the NRC lexicon database had a significant impact on their personality prediction. For this research, it was observed that the number of

attributes together with sentiment analysis and lexicons contributed to the better performance of the classifier. Higher F1-score and accuracy was achieved with four input variables which included sentiment of the tweets, followed by the test with three input variables.

Kunte and Panicker (2019) conducted a study to predict the personality of Twitter users using XGBoost and ensemble classifiers. Their dataset consisted of 9,918 tweets. The classifiers were trained with only one training variable, i.e., the tweets and five labels (Kunte & Panicker, 2019). Their study achieved an accuracy of 82.59% with their ensemble classifier. In comparison, the current research achieved higher accuracy (95%) with three labels and four training variables. In addition, the dataset used in this research was 142,000 tweets which was a higher sample than Kunte and Panicker's (2019) study.

Kumar, Sharma, and Arora (2019) sought to predict depression from social media data. They used a dataset of 100 Twitter users. The study used five input variables to train their classifiers. Anxiety-related lexicon, tweet frequency, time the tweet was posted, sentiment, and the occurrence of at least 25% polarity contrast in postings within 24 hours were all used as input variables. Three classifiers of, gradient boosting, Naïve Bayes and random forest and one ensemble classifiers were trained. Their best classifier achieved an accuracy of 85.09%. The current research further achieved a higher accuracy than this research with four input variables and three labels.

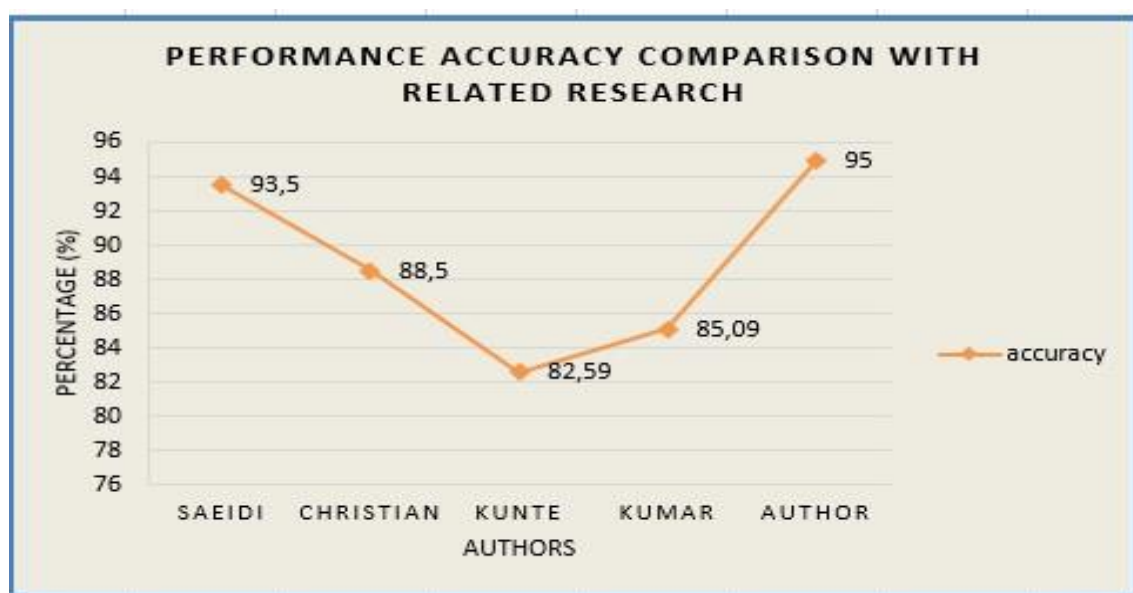


Figure 8.5: Performance accuracy comparison with related research

8.7 Summary

In this chapter, the effects of various variables in the dataset on classifier accuracy were investigated. Many factors play a role in why people spend time on social media. It is critical to conceptualise and constructively break down the factors associated with using major social media platforms. Through experiments using Twitter data, it was noted that sentiment is beneficial in predicting a user personality with higher accuracy as opposed to not incorporating the sentiment. The results from classification experiments showed that attributes of a dataset could affect the performance of machine learning classifiers. The number of training variables have a significant effect on personality prediction accuracy. The adopted approach, i.e., the combination of more input variables (4), the classifiers generated a better result in comparison with the result obtained by applying less input variables (2). Therefore, the findings suggest that researchers and practitioners need to consider the attributes of the datasets before choosing appropriate classifiers. The findings further support the theory that proper control of training datasets is fundamental in text classification. In this regard, the research proposes determining what type of variables are found in the dataset. As observed in the experiments, the more the training variables used, the higher the classifier accuracy. An important finding from the classification tests is that the more the training attributes, the higher the accuracy. The results from this chapter corroborate the findings from previous studies (Choi & Lee, 2017) that input variables and data size influence the machine learning model's performance, and appropriate algorithms ought to be selected based on the dataset. SNSs encourage users to focus on only a few elements of their lives and personalities, hence a person's personality influences how they use SNSs. By recognising relationships between variables such as personality and SNS, psychologists can better understand how online web influences our social lives and social behaviours.

CHAPTER 9: DISCUSSION AND CONCLUSION

9.1 Introduction

Design science research methodology was adopted in this research. The need to predict narcissistic behaviour using Twitter as the social media platform was presented in Chapter 1. Chapter 2 discussed previous studies related to personality prediction, existing personality theories, and gaps in past research. This chapter summarises the research undertaken, presents the significant findings regarding the research questions, and discusses the implications of this research. The chapter also discusses the recommended research model for the study that emanates from the research findings. The process model that underpinned this research had four main tasks. First was data pre-processing, the second task was sentiment analysis and data annotation, the third task was classifying tweets into narcissistic and non-narcissistic tweets. The fourth and final task was the prediction of the level of narcissism of a user using fuzzy logic. Therefore, the end goal was to train two classifiers. The first step approach of classification categorised a tweet as grandiose, empath, or vulnerable narcissism. Subsequently, the second step used fuzzy logic to predict the degree of narcissism, given two input variables.

Through various experiments, three classifiers and four ensemble classifiers (SVM&NB, RF&NB, SVM&RF, SVMRF&NB) were evaluated. The SVM, RF, and RF with the Naïve Bayes ensemble classifier were most effective and provided the best performance based on the experimental results. Secondly, various pre-processing methods were studied, and some interesting observations were evident from the experiments. The experimental results proved that stopwords removal contributed to the meaning of words in the sentiment analysis and eventually affected the accuracy of classifiers. It was also noted that stemming and lemmatization influenced topic models as they tend to alter the meaning of words. The research concludes that appropriate pre-processing techniques and input variables should be considered based on the research objective.

9.2 Research questions revisited

The answers to the research questions identified in Section 1.4 are summarised in this chapter. The main research question for this study was "*How can traces of narcissistic personality traits be identified among social media users?*" This research developed a process model (Section 3.5) used to answer the research questions. After all processes and experiments had

been conducted in the initial process model, a modified process model for predicting personality traits is recommended. The process model comprises four main phases, namely Phase 1 (rulebased text processing), Phase 2 (data labelling), Phase 3 (classification), and finally Phase 4 (prediction). The phases in the modified research model summarise the answers to the four research questions. Figure 9.1 below presents the phases in detail.

9.3 Modified process model phases

As noted in Section 9.2, a modified process model has been developed that shows the different steps undertaken to answer the research questions Section 1.4. The process model is organised into four phases, and the activity in each phase must be conducted before proceeding to the next phase. There is also iteration among phase 1, phase 2, and phase 3 based on the classifier results and the needs of the researcher.

9.3.1 Phase1: Rule-based text pre-processing

The first research question was *How can Twitter dataset be prepared for sentiment analysis?* To answer this research question, different data pre-processing techniques were used to prepare data for sentiment analysis, as discussed in Chapter 5 and shown in phase 1 of the process model. The research noted that appropriate text pre-processing techniques ought to be selected based on the research objective. For this study, text pre-processing was done using a rule-based approach, and selected techniques were adopted.

Pre-processing can potentially eliminate useful information or add errors into the analysis (for example, when stemming alters semantically different phrases) and can affect classifier results (Boyd, 2016). This research noted that despite various text pre-processing techniques, not all of them are appropriate for personality classification. Therefore, proper text pre-processing techniques have to be considered based on the research objective. This research recommends pre-processing approaches for personality classification: conversion to lowercase, alphanumeric characters removal, removal of @ and # symbols, tokenization, and lemmatization. These five approaches are adopted as they do not alter the meaning or order of words in a dataset.

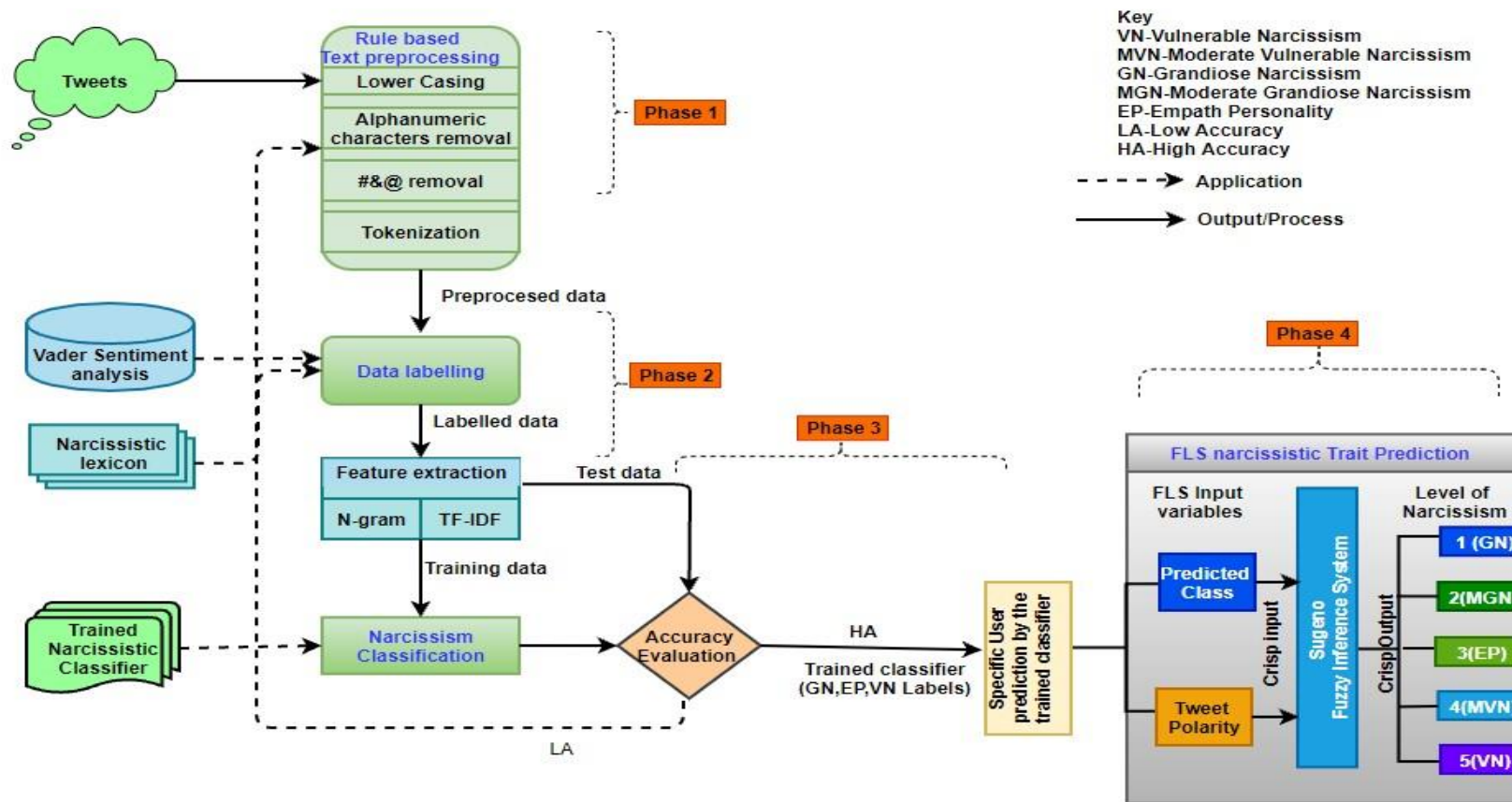


Figure 9.1: An updated process model (Author)

9.3.2 Phase 2: Data labelling

The second research question was "*How can prepared twitter dataset be labelled into different categories of narcissism?*" The dataset was labelled into grandiose narcissism, vulnerable narcissism, and empath personality. This was achieved through simultaneous tasks of sentiment analysis, topic modelling, and lexicon detection. Lexicon detection was achieved in phase 2 of the process model through the dictionary built from literature.

Data labelling refers to the process of annotation of the dataset into certain categories. Raw social media data is unstructured and does not have any labels. Therefore, after pre-processing, this research recommends applying two approaches with the objective of labelling data. In the first approach, the research suggests that sentiment analysis is done on the processed tweets. This is because users use social media to express themselves emotionally. Emotions could be negative, neutral or positive, depending on the attitude and the subject matter the user is reacting to. In addition, sentiment is an essential item in identifying individual personality. The second approach is lexicon detection. This involves categorising tweets into two classes of narcissism based on the presence or absence of keywords in the developed lexicon (dictionary). A tweet can be labelled narcissistic if it has at least one lexicon present in the developed dictionary and non-narcissistic if it does not have any lexicon in the dictionary. To further differentiate grandiose and vulnerable narcissism, Vader sentiment analysis results are adopted. The research suggests that scores between +0.5 and +1 be considered positive for grandiose narcissism; a score of between -0.49 and +0.49 be considered neutral; and a score of -0.5 to -1 be considered for vulnerable narcissism.

9.3.3 Phase 3: Machine learning classification

The third research question was "*How can traces of narcissism be classified using a labelled Twitter dataset?*" This study sought to identify traces of narcissism as a personality on social media. To achieve this objective, existing tools and techniques were reviewed in Chapter 2. Various machine learning classifiers were explored from the literature review based on their suitability to the study and classification.

Once the whole dataset is labelled, this research recommends applying machine learning classifiers to the dataset based on the needs of the researchers. Various classifiers can be evaluated based on accuracy and F1- score as used in this research. The classifier with higher accuracy can then be adopted and be used in the next stage of prediction.

9.3.4 Phase 4: Fuzzy-based prediction

The last research question was, *"How can the classification of narcissism be improved?"* This was answered in phase 4 of the process model, as shown in Figure 9.1. This is the last phase of the process model, where the trained classifier is used to predict new data. In the process model, predictions can be made using both a trained classifier and fuzzy logic or just the trained classifier alone. To increase the effectiveness of the trained model in prediction, fuzzy logic is incorporated. Linguistic vagueness and uncertainty are well-suited to fuzzy logic. The research recommends the adoption of either of the approaches based on the research objective. Fuzzy logic is a computational approach where truth values range from 0 and 1. Fuzzy logic assisted to further establish the degree of narcissism by way of representation in terms of levels.

9.4 Recommended research model

This research concludes by recommending a research model that can identify traces of narcissism on social media. This research suggests the analysis of tweets made by the user to identify narcissistic personality while also considering three other moderating variables. These variables are the polarity of the tweet. Secondly, researchers need to consider the number of tweets made by a user. Finally, researchers ought to consider the lexicons used by a user when they are tweeting. As discussed in the literature review, Twitter users behave differently. Some tweet frequently, while some tweet occasionally. Those who tweet often may have tweeted with positive polarity or negative polarity. Thus, the sentiment, the frequency of tweeting, and the type of lexicon used ought to be included in identifying traces of narcissism

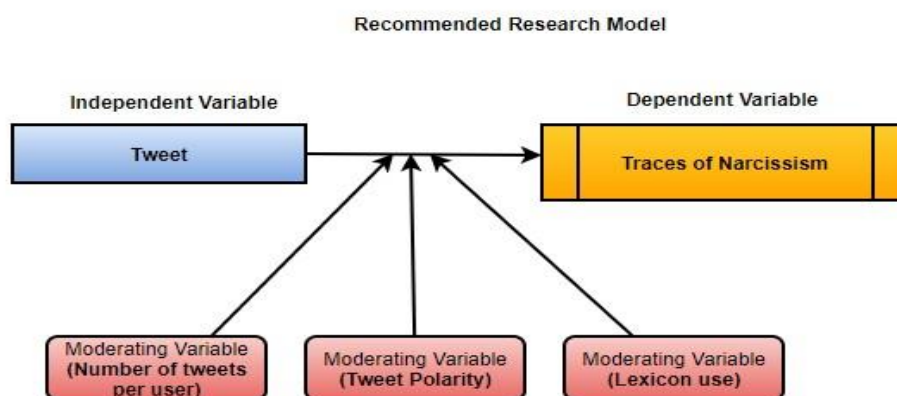


Figure 9.2: Recommended research model (Author)

9.5 Research limitations

The focus of this research was to develop a model that can identify traces of narcissism from social media. While this was a success based on the different experiments done, the research

had a few limitations. The patterns of user activity within Twitter may differ from social media platforms, which limits the ability to generalise the results outside of Twitter. While the dataset provides a strong starting point for solving the problem, however, due structural changes, a model constructed with this training set may not translate well to other social media platforms like reddit or Instagram. Expanding the research to other social network sites would significantly increase the diversity of the variables in the dataset as information those platforms are more diverse than the dataset used in this study. The three main classifiers used in this study, proved to be fairly effective at solving the problem. As previously stated, the implementations in this study produced excellent results on a small dataset.

9.6 Contribution to knowledge

This dissertation advances personality trait prediction by proposing a process model that can be adopted by research in social media and personality research. Moreover, the findings of predicting personality traits could be used to assist the law enforcement sector in assessing criminal behaviour.

This research also expanded Fan and Gordon's (2014) CUPP framework by incorporating iteration and prediction. The original framework has only three stages of *capture*, *understand* and *present*. This research has modified the framework and included the '*predict*' stage before the '*present*' stage. Once classification is done in the 'understanding' stage in the CUP framework, it is necessary to use the trained classifiers to perform new predictions. This is handled by the 'predict' stage in the modified Iterative CUPP framework.

9.7 Recommendations

In general, this research makes various recommendations to different social media stakeholders, as discussed in the next sections.

9.7.1 Recommendation for Twitter

This research recommends the use of emoticons to Twitter to flag tweets that have traces of narcissism. With the popularity of Twitter rising on daily basis there is need to protect the users by providing them with a safe space for tweeting. Thus, flagging narcissistic tweets will help to prevent users' exposure to such individuals and provide a safe space for tweeting and interacting amongst users. In addition, Twitter should put in place measures to handle

complaints about narcissistic posts on their platform, which should then hide further sharing of such posts.

9.7.2 Recommendation for users

Individuals use social media sites for various reasons, but the most popular motivation is to build and sustain interpersonal relationships. According to research, certain personality factors can lead to increased social media usage. The use of social networking sites for self-promotion is linked to narcissism. Individual with narcissistic traits is likely to claim their desire to gain favourable attention from others and grow their social network. Therefore, this research also recommends users be cautious when establishing such interpersonal relations on Twitter.

This research also provides a way for users to discern potential threats and a sense of susceptibility within SNSs. Examining personalities and behaviours as they occur in the actual world and applying the results to the virtual is one way to capture the essence of SNS cybercrime. Many people spend an unusual amount of time connecting with social media sites and posting a large volume of personal data. As a result, users and their audiences intentionally release personal data into the public sphere. Cybercriminals may take advantage of this by creating fake identities and obtaining private information from users on social media sites.

9.7.3 Recommendation to the researchers

This study only used the Twitter dataset. Future research can include other social media platforms like Instagram, Reddit, and Facebook. Secondly, future research can also incorporate traditional questionnaires together with the digital footprints of the different users. In addition, this application of deep learning techniques to predict narcissism from social media can also be another direction of future work.

Researchers can also employ a "mixed study methodology" by formulating questionnaires to users and using their tweets to conduct a study on personality traits. In addition, whereas the dataset used in this research was limited in terms of producing a generalisable classifier, with a more robust and versatile dataset, the model's capabilities may be used for more reasons.

9.7.4 Recommendation to governments

This research also recommends the incorporation of personality traits, specifically narcissism, in law enforcement. While the existing approaches seem to work, incorporation of online personality traits can help deter and speed up identification and prevention of crime. The link

between narcissism and aggression corroborates previous research findings that show people with inflated egos, which are a sign of narcissism, are more likely to engage in violent behaviour. This research has discussed narcissism as a personality trait and its potential association with crime and violent behaviours. This finding was further supported by Barry et al. (2007), who observed that teens with criminal history are more narcissistic than general population. Hepper, Hart, Meek, Cisek, and Sedikides (2014) observed that narcissistic traits are linked to higher risk of criminal activities. Therefore, law enforcement authorities in government may incorporate personality traits to prevent and investigate crime.

9.8 Summary

A model for detecting and identifying traces of narcissism based on tweets was developed in this work. In this research, publicly available Tweets were collected to develop a narcissistic prediction model to identify traces of narcissism from social media. The experiments used various features and techniques to find a combination that performs well for the tasks. The best classifier obtained provides an accuracy of 88.19% and an f1-score of 88%. In Chapter 4, prediction workflow and outline of the steps of achieving the research objective were discussed. The proposed solution achieves high accuracy and is scalable with large datasets. The experiments show that the fuzzy-based machine learning narcissistic classification is precise in identifying traces of narcissism from social media and achieved a better accuracy than previous studies. In conclusion, this thesis expands the existing body of knowledge by incorporating new ways of identifying narcissistic personality traits from Twitter. Social network sites have become an integral part of people's lives, and have increased in popularity. Therefore, future studies could explore how existing and future social media platforms can be enhanced to safeguard individuals' well-being, particularly those who are highly vulnerable in the society.

REFERENCES

- Abdel-Khalek, A. M. (2016). Introduction to the psychology of self-esteem. In F. Holloway (Ed.), *Self-esteem: perspectives, influences and improvement strategies* (pp. 1-23). United Kingdom: Nova Science Publishers.
- AbdulHussien, A. A. (2017). Comparison of machine learning algorithms to classify web pages. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 8(11), 205-209. Retrieved from <http://dx.doi.org/10.14569/IJACSA.2017.081127>
- Aboulhosen, S. (2020, January 21). 18 Facebook statistics every marketer should know in 2020. *SproutSocial*. Retrieved from <https://sproutsocial.com/insights/facebook-stats-for-marketers/>
- Adams, J. M., Florell, D., Burton, K. A., & Hart, W. (2014). Why do narcissists disregard social-etiquette norms? A test of two explanations for why narcissism relates to offensive-language use. *Personality and Individual Differences*, 58, 26-30.
- Aggarwal, C. C. (2011). *Social network data analytics*. New York: Springer.
- Agrawal, R., & Batra, M. (2013). A detailed study on text mining techniques. *International Journal of Soft Computing and Engineering (IJSCE)*, 2(6), 2231-2307.
- Ahire, P. G., Kolhe, S., Kirange, K., Karale, H., & Bhole, A. (2015). Implementation of improved ID3 algorithm to obtain more optimal decision tree. *International Journal of Engineering Research and Development*, 11(2), 44-47.
- Ahmad, N., & Siddique, J. (2017). Personality assessment using Twitter tweets. *Procedia Computer Science*, 112, 1964-1973.
- Ahmed, I. M., Alfonse, M., Aref, M., & Salem, A. B. M. (2015). Reasoning techniques for diabetics expert systems. *Procedia Computer Science*, 65, 813-820. Retrieved from <https://doi.org/10.1016/j.procs.2015.09.030>
- Ahmed, Y. A., Ahmad, M. N., Ahmad, N., & Zakaria, N. H. (2019). Social media for knowledge-sharing: A systematic literature review. *Telematics and Informatics*, 37, 72-112.

- Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, 152, 341-348.
- Akram, J., & Tahir, A. (2018). Lexicon and heuristics based approach for identification of emotion in text. *In 2018 International conference on frontiers of information technology (FIT)* (pp. 283-297). US/Canada: IEEE.
- Akram, W., & Kumar, R. (2017). A study on positive and negative effects of social media on society. *International Journal of Computer Sciences and Engineering*, 5(10), 351-354.
- Alessia, D., Ferri, F., Grifoni, P., & Guzzo, T. (2015). Approaches, tools and applications for sentiment analysis implementation. *International Journal of Computer Applications*, 125(3), 26-33.
- Alhabash, S., & Ma, M. (2017). A tale of four platforms: Motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students? *Social Media+ Society*, 3(1), 2056305117691544. Retrieved from <https://journals.sagepub.com/doi/pdf/10.1177/2056305117691544>
- Aljuboori, A. F., Fashakh, A. M., & Bayat, O. (2020). The impacts of social media on university students in Iraq. *Egyptian Informatics Journal*, 21(3), 139-144. Retrieved from <https://doi.org/10.1016/j.eij.2019.12.003>
- Allahyari, M., Pouriye, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). *A brief survey of text mining: classification, clustering and extraction techniques*. Retrieved from <https://arxiv.org/pdf/1707.02919.pdf>
- Alloway, T., Runac, R., Quershi, M., & Kemp, G. (2014). Is Facebook linked to selfishness? Investigating the relationships among social media use, empathy, and narcissism. *Social Networking*. Retrieved from <https://doi.org/10.4236/sn.2014.33020>
- Almatarneh, S., & Gamallo, P. (2018). A lexicon based method to search for extreme opinions. *PloS one*, 13(5), e0197816. Retrieved from <https://doi.org/10.1371/journal.pone.0197816>

- Alsadhan, N., & Skillicorn, D. (2017, November). Estimating personality from social media posts. In *2017 IEEE international conference on data mining workshops (ICDMW)* (pp. 350-356). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/icdmw.2017.51>
- Al-Shabi, M. A. (2020). Evaluating the performance of the most important Lexicons used to Sentiment analysis and opinions Mining. *International Journal of Computer Science and Network Security (IJCSNS)*, 20(1), 51-57.
- Alshehri, A. (2019). *A machine learning approach to predicting community engagement on social media during disasters* (Unpublished PhD dissertation). Tampa, Florida: University of South Florida.
- Alsheikh, A., & Ahmad, S. (2020). Delinquency as predicted by dark triad factors and demographic variables. *International Journal of Adolescence and Youth*, 25(1), 661-675.
- Althnian, A., AlSaeed, D., Al-Baity, H., Samha, A., Dris, A. B., Alzakari, N., ... & Kurdi, H. (2021). Impact of dataset size on classification performance: an empirical evaluation in the medical domain. *Applied Sciences*, 11(2), 796. Retrieved from <https://doi.org/10.3390/app11020796>
- Amichai-Hamburger, Y., & Vinitzky, G. (2010). Social network use and personality. *Computers in Human Behavior*, 26(6), 1289-1295.
- Arjaria, S., Shrivastav, A., Rathore, A. S., & Tiwari, V. (2019). Personality trait identification for written texts using MLNB. In *Data, Engineering and Applications* (pp. 131-137). Singapore: Springer. Retrieved from https://doi.org/10.1007/978-981-13-6347-4_12
- Aubaid, A. M., & Mishra, A. (2020). A rule-based approach to embedding techniques for text document classification. *Applied Sciences*, 10(11), 4009. Retrieved from <https://doi.org/10.3390/app10114009>
- Ayedh, A., Tan, G., Alwesabi, K., & Rajeh, H. (2016). The effect of preprocessing on arabic document categorization. *Algorithms*, 9(2), 27. Retrieved from <http://dx.doi.org/10.3390/a9020027>

- Azeroual, O., Saake, G., Abuosba, M., & Schöpfel, J. (2018). Text data mining and data quality management for research information systems in the context of open data and open science. In *3rd International Colloquium Open Access – Open Access to Science Foundations, Issues and Dynamics* (pp. 29-46). Rabat-Morocco.
- Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences*, 124, 150-159.
- Baccarella, C. V., Wagner, T. F., & Kietzmann, J. H. (2018). Social media? It's serious! Understanding the dark side of social media. *European Management Journal*, 36(4), 431-438.
- Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, 21(3), 372-374. Retrieved from <https://doi.org/10.1177/0956797609360756>
- Balakrishnan, V., & Lloyd-Yemoh, E. (2014). Stemming and lemmatization: A comparison of retrieval performances. *Lecture Notes on Software Engineering*, 2(3), 262-267.
- Barry, T. D., Thompson, A., Barry, C. T., Lochman, J. E., Adler, K., & Hill, K. (2007). The importance of narcissism in predicting proactive and reactive aggression in moderately to highly aggressive children. *International Society for Research on Aggression*, 33(3), 185-197.
- Baskerville, R. L., Kaul, M., & Storey, V. C. (2018). Aesthetics in design science research. *European Journal of Information Systems*, 27(2), 140-153.
- Baskerville, R., Baiyere, A., Gregor, S., & Rossi, M. (2018). Design science research contributions: finding a balance between artifact and theory. *Journal of the Association for Information Systems*, 19(5). Retrieved from <https://aisel.aisnet.org/jais/vol19/iss5/3>
- Baughman, H. M., Dearing, S., Giammarco, E., & Vernon, P. A. (2012). Relationships between bullying behaviours and the Dark Triad: A study with adults. *Personality and*

- Individual Differences*, 52(5), 571-575. Retrieved from <https://doi.org/10.1016/j.paid.2011.11.020>
- Beiji, Z., Mohammed, N., Chengzhang, Z., & Rongchang, Z. (2017). Crime hotspot detection and monitoring using video based event modeling and mapping techniques. *International Journal of Computational Intelligence Systems*, 10(1), 962. Retrieved from <https://doi.org/10.2991/ijcis.2017.10.1.64>
- Bekker, J., & Davis, J. (2020). Learning from positive and unlabeled data: A survey. *Machine Learning*, 109(4), 719-760.
- Bernarte, R. P., Festijo, A. I., Layaban, M. D., & Ortiz, S. U. (2015). Me, myself and i: what makes Filipino millennials narcissist? *Asia Pacific Higher Education Research Journal (APHERJ)*, 2(1). Retrieved from <https://po.pnuresearchportal.org/ejournal/index.php/apherj/article/view/90/80>
- Billal, B., Fonseca, A., Sadat, F., & Lounis, H. (2017). Semi-supervised learning and social media text analysis towards multi-labeling categorization. In *2017 IEEE International Conference on Big Data* (pp. 1907-1916). US/Canada: IEEE.
- Błachnio, A., & Przepiórka, A. (2018). Facebook intrusion, fear of missing out, narcissism, and life satisfaction: a cross-sectional study. *Psychiatry Research*, 259, 514-519. Retrieved from <https://doi.org/10.1016/j.psychres.2017.11.012>
- Blair, S. J., Bi, Y., & Mulvenna, M. D. (2020). Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*, 50(1), 138-156. Retrieved from <https://doi.org/10.1007/s10489-019-01438-z>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Blinkhorn, V., Lyons, M., & Almond, L. (2016). Drop the bad attitude! Narcissism predicts acceptance of violent behaviour. *Personality and Individual Differences*, 98, 157-161.
- Bondü, R., & Scheithauer, H. (2015). Narcissistic symptoms in German school shooters. *International Journal of Offender Therapy and Comparative Criminology*, 59(14), 1520-1535.

- Bonta, V., & Janardhan, N. K. (2019). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, 8(S2), 1-6.
- Boyd, R. (2016). *General use: RIOT scan*. Retrieved from <https://riot.ryanb.cc/general-use/>
- Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: a new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, 18, 63-68.
- Brailovskaia, J., Bierhoff, H. W., Rohmann, E., Raeder, F., & Margraf, J. (2020). The relationship between narcissism, intensity of Facebook use, Facebook flow and Facebook addiction. *Addictive Behaviors Reports*, 11, 100265. Retrieved from <https://doi.org/10.1016/j.abrep.2020.100265>
- Branley, D. B., & Covey, J. (2017). Pro-ana versus pro-recovery: a content analytic comparison of social media users' communication about eating disorders on Twitter and Tumblr. *Frontiers in Psychology*, 8, 1356. Retrieved from <https://doi.org/10.3389/fpsyg.2017.01356>
- Braun, S. (2017). Leader narcissism and outcomes in organizations: a review at multiple levels of analysis and implications for future research. *Frontiers in Psychology*, 8, 773. Retrieved from <https://doi.org/10.3389/fpsyg.2017.00773>
- Brunell, A. B., Staats, S., Barden, J., & Hupp, J. M. (2011). Narcissism and academic dishonesty: the exhibitionism dimension and the lack of guilt. *Personality and Individual Differences*, 50(3), 323-328. Retrieved from <https://doi.org/10.1016/j.paid.2010.10.006>
- Bryan, C. J., Butner, J. E., Sinclair, S., Bryan, A. B. O., Hesse, C. M., & Rose, A. E. (2018). Predictors of emerging suicide death among military personnel on social media networks. *Suicide and Life-Threatening Behavior*, 48(4), 413-430. Retrieved from <https://doi.org/10.1111/sltb.12370>
- Buffardi, L. E., & Campbell, W. K. (2008). Narcissism and social networking web sites. *Personality and Social Psychology Bulletin*, 34(10), 1303-1314. Retrieved from <https://doi.org/10.1177/0146167208320061>

- Butler, B. S., & Matook, S. (2015). Social media and relationships. *The International Encyclopedia of Digital Communication and Society*, 1-12. Retrieved from <https://doi.org/10.1002/9781118767771.wbiedcs097>
- Carducci, G., Rizzo, G., Monti, D., Palumbo, E., & Morisio, M. (2018). Twitpersonality: computing personality traits from tweets using word embeddings and supervised learning. *Information*, 9(5), 127. Retrieved from <https://doi.org/10.3390/info9050127>
- Carey, A. L., Brucks, M. S., Küfner, A. C., Holtzman, N. S., Back, M. D., Donnellan, M. B., & Mehl, M. R. (2015). Narcissism and the use of personal pronouns revisited. *Journal of Personality and Social Psychology*, 109(3), e1-e5.
- Carter, G. L., Campbell, A. C., & Muncer, S. (2014). The Dark Triad: beyond a 'male' mating strategy. *Personality and Individual Differences*, 56, 159-164. Retrieved from <https://doi.org/10.1016/j.paid.2013.09.001>
- Carton, H., & Egan, V. (2017). The dark triad and intimate partner violence. *Personality and Individual Differences*, 105, 84-88. Retrieved from <https://doi.org/10.1016/j.paid.2016.09.040>
- Casale, S., & Fioravanti, G. (2018). Why narcissists are at risk for developing Facebook addiction: The need to be admired and the need to belong. *Addictive Behaviors*, 76, 312-318.
- Casale, S., Fioravanti, G., & Rugai, L. (2016). Grandiose and vulnerable narcissists: who is at higher risk for social networking addiction?. *Cyberpsychology, Behavior, and Social Networking*, 19(8), 510-515. Retrieved from <https://doi.org/10.1089/cyber.2016.0189>
- Catal, C., & Nangir, M. (2017). A sentiment classification model based on multiple classifiers. *Applied Soft Computing*, 50, 135-141.
- Chang, R. M., Kauffman, R. J., & Kwon, Y. (2014). Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems*, 63, 67-80. Retrieved from <https://doi.org/10.1016/j.dss.2013.08.008>
- Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., & Vakali, A. (2017, June). Mean birds: detecting aggression and bullying on twitter. In

Proceedings of the 2017 ACM on Web Science Conference (pp. 13-22). Retrieved from <https://doi.org/10.1145/3091478.3091487>

- Cheiffetz, R. T. (2017). *Examination of the Dark Triad and its association with antisocial behavior and cheating in undergraduates*. City University of New York, John Jay College of Criminal Justice. Retrieved from <https://doi.org/10.4324/9781315721606-73>
- Chen, J., Yang, Z., & Yang, D. (2020). Mixtext: Linguistically-informed interpolation of hidden space for semi-supervised text classification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 2147-2157). Retrieved from <https://aclanthology.org/2020.acl-main.194>
- Choi, Y., & Lee, H. (2017). Data properties and the performance of sentiment classification for electronic commerce applications. *Information Systems Frontiers*, 19(5), 993-1012.
- Christian, H., Suhartono, D., Chowanda, A., & Zamli, K. Z. (2021). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *Journal of Big Data*, 8(1), 1-20. Retrieved from <https://doi.org/10.1186/s40537-021-00459-1>
- Chung, K. L., Morshidi, I., Yoong, L. C., & Thian, K. N. (2019). The role of the dark tetrad and impulsivity in social media addiction: findings from Malaysia. *Personality and Individual Differences*, 143, 62-67. Retrieved from <https://doi.org/10.1016/j.paid.2019.02.016>
- Colledani, D., Anselmi, P., & Robusto, E. (2018). Using item response theory for the development of a new short form of the Eysenck Personality Questionnaire – revised. *Frontiers in Psychology*, 9, 1834. Retrieved from <https://doi.org/10.3389/fpsyg.2018.01834>
- Couso, I., Borgelt, C., Hullermeier, E., & Kruse, R. (2019). Fuzzy sets in data analysis: from statistical foundations to machine learning. *IEEE Computational Intelligence Magazine*, 14(1), 31-44. Retrieved from <https://doi.org/10.1109/mci.2018.2881642>

- Cresci, S., Cimino, A., Avvenuti, M., Tesconi, M., & Dell'Orletta, F. (2018, June). Real-world witness detection in social media via hybrid crowdsensing. In *Twelfth international AAAI conference on web and social media*, 12(1). Retrieved from <https://ojs.aaai.org/index.php/ICWSM/article/view/15072>
- Cristani, M., Vinciarelli, A., Segalin, C., & Perina, A. (2013, October). Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis. In *Proceedings of the 21st ACM international conference on Multimedia* (pp. 213-222). Retrieved from <https://doi.org/10.1145/2502081.2502280>
- Cronin, C. (2014). Using case study research as a rigorous form of inquiry. *Nurse Researcher*, 21(5), 19-27.
- Cruz-Sandoval, D., Beltran-Marquez, J., Garcia-Constantino, M., Gonzalez-Jasso, L. A., Favela, J., Lopez-Nava, I. H., & Nugent, C. (2019). Semi-automated data labeling for activity recognition in pervasive healthcare. *Sensors*, 19(14), 1-19.
- Curiel, R. P., Cresci, S., Muntean, C. I., & Bishop, S. R. (2020). Crime and its fear in social media. *Palgrave Communications*, 6(1), 1-12.
- Czarna, A. Z., Zajenkowski, M., & Dufner, M. (2018). How does it feel to be a narcissist? Narcissism and emotions. In *Handbook of trait narcissism* (pp. 255-263). Cham.: Springer.
- Daho, M. E. H., Settouti, N., Lazouni, M. E. A., & Chikh, M. E. A. (2014, April). Weighted vote for trees aggregation in random forest. In *2014 International Conference on Multimedia Computing and Systems (ICMCS)* (pp. 438-443). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/icmcs.2014.6911187>
- Dandannavar, P. S., Mangalwede, S. R., & Kulkarni, P. M. (2018). Social media text – A source for personality prediction. In *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)* (pp. 62-65). US/Canada: IEEE.
- Darwich, M., Mohd, S. A., Omar, N., & Osman, N. A. (2019). Corpus-based techniques for sentiment Lexicon generation: a review. *Journal of Digital Information Management*, 17(5), 296-305. Retrieved from <https://doi.org/10.6025/jdim/2019/17/5/296-305>

- Das, K. G., & Das, D. (2017). Developing lexicon and classifier for personality identification in texts. In *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)* (pp. 362-372). ICON.
- Davahli, M. R., Karwowski, W., Gutierrez, E., Fiok, K., Wróbel, G., Taiar, R., & Ahram, T. (2020). Identification and prediction of human behavior through mining of unstructured textual data. *Symmetry*, 12(11), 1902. Retrieved from <https://doi.org/10.3390/sym12111902>
- Davenport, S. W., Bergman, S. M., Bergman, J. Z., & Fearing, M. E. (2014). Twitter versus Facebook: Exploring the role of narcissism in the motives and usage of different social media platforms. *Computers in Human Behavior*, 32, 212-220. Retrieved from <https://doi.org/10.1016/j.chb.2013.12.011>
- Davidow, B. (2013, March 26). The Internet “narcissism epidemic”. *The Atlantic*. Retrieved from <https://www.theatlantic.com/health/archive/2013/03/the-internet-narcissism-epidemic/274336/>
- Deigh, L., Farquhar, J., Palazzo, M., & Siano, A. (2016). Corporate social responsibility: Engaging the community. *Qualitative Market Research: An International Journal*, 19(2), 225-240.
- Deng, Q., & Ji, S. (2018). A review of design science research in information systems: concept, process, outcome, and evaluation. *Pacific Asia Journal of the Association for Information Systems*, 10(1), 2. Retrieved from <https://aisel.aisnet.org/pajais/vol10/iss1/2>
- Deniz, A., & Kiziloğlu, H. E. (2017, October). Effects of various preprocessing techniques to Turkish text categorization using n-gram features. In *2017 International Conference on Computer Science and Engineering (UBMK)* (pp. 655-660). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/ubmk.2017.8093491>
- Derczynski, L., Maynard, D., Aswani, N., & Bontcheva, K. (2013, May). Microblog-genre noise and impact on semantic annotation accuracy. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media* (pp. 21-30). Retrieved from <https://doi.org/10.1145/2481492.2481495>

- Desai, M., & Mehta, M. A. (2016). Techniques for sentiment analysis of Twitter data: A comprehensive survey. In *2016 International Conference on Computing, Communication and Automation (ICCCA)* (pp. 149-154). US/Canada: IEEE.
- DeVito, M. A., Birnholtz, J., Hancock, J. T., French, M., & Liu, S. (2018, April). How people form folk theories of social media feeds and what it means for how we study self-presentation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-12). Association for Computing Machinery. Retrieved from <https://doi.org/10.1145/3173574.3173694>
- DeWall, C. N., Buffardi, L. E., Bonser, I., & Campbell, W. K. (2011). Narcissism and implicit attention seeking: evidence from linguistic analyses of social networking and online presentation. *Personality and Individual Differences*, 51(1), 57-62. Retrieved from <https://doi.org/10.1016/j.paid.2011.03.011>
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of Personality and Social Psychology*, 93(5), 880. Retrieved from <https://doi.org/10.1037/0022-3514.93.5.880>
- Diamantini, C., Mircoli, A., DiaPotena, D., & Storti, E. (2019). Social information discovery enhanced by sentiment analysis techniques. *Future Generation Computer Systems*, 95, 816-828.
- Dilmegani, C. (2021, June 21). *Data labeling in 2021: How to choose a data labeling partner*. Retrieved from <https://research.aimultiple.com/data-labeling/>
- Dresch, A., Lacerda, D. P., & Miguel, P. A. (2015). A distinctive analysis of case study, action research and design science research. *RBGN*, 17(56), 1116-1133.
- Ekong, V. E., Ekong, U. O., Uwadiae, E. E., Abasiubong, F., & Onibere, E. A. (2013). A fuzzy inference system for predicting depression risk levels. *African Journal of Mathematics and Computer Science Research*, 6(10), 197-204.
- El Alaoui, I., Gahi, Y., Messoussi, R., Chaabi, Y., Todoskoff, A., & Kobi, A. (2018). A novel adaptable approach for sentiment analysis on big social data. *Journal of Big Data*, 5(1), 1-18. Retrieved from <https://doi.org/10.1186/s40537-018-0120-0>

- El Hussein, M., Hirst, S., Salyers, V., & Osuji, J. (2014). Using grounded theory as a method of inquiry: Advantages and disadvantages. *Qualitative Report*, 19, 1-15.
- Elbagir, S., & Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and VADER sentiment. In *Proceedings of the International MultiConference of Engineers and Computer Scientists* (p. 122). Hong Kong: IMECS.
- Ellison, N. B., & Boyd, D. (2013). Sociality through social network sites. In *The Oxford Handbook of Internet Studies* (pp. 151-172). Retrieved from <https://doi.org/10.1093/oxfordhb/9780199589074.013.0008>
- Elragal, A., & Haddara, M. (2014). Big data analytics: A text mining-based literature analysis. In *Norsk Konferanse for Organisasjoners Bruk Av It*, 22(1). Retrieved from https://www.academia.edu/47294109/Big_Data_Analytics_A_Text_Mining_Based_Literature_Analysis
- Eshuis, R., & Firat, M. (2018, September). Modeling uncertainty in declarative artifact-centric process models. In *International Conference on Business Process Management* (pp. 281-293). Cham.: Springer. Retrieved from https://doi.org/10.1007/978-3-030-11641-5_22
- Esra'M, A. (2019). Lexicon-based detection of violence on social media. *Cognitive Semantics*, 5(1), 32-69.
- Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3?—Criteria for a taxonomic paradigm. *Personality and Individual Differences*, 12(8), 773-790. Retrieved from [https://doi.org/10.1016/0191-8869\(91\)90144-z](https://doi.org/10.1016/0191-8869(91)90144-z)
- Eysenck, S., & Barrett, P. (2013). Re-introduction to cross-cultural studies of the EPQ. *Personality and Individual Differences*, 54(4), 485-489.
- Fakhrzadegan, S., Gholami-Doon, H., Shamloo, B., & Shokouhi-Moghaddam, S. (2017). The relationship between personality disorders and the type of crime committed and substance used among prisoners. *Addiction & Health*, 9(2), 64-71.
- Fan, C. Y., Chu, X. W., Zhang, M., & Zhou, Z. K. (2019). Are narcissists more likely to be involved in cyberbullying? Examining the mediating role of self-esteem. *Journal of*

- Interpersonal Violence*, 34(15), 3127-3150. Retrieved from <https://doi.org/10.1177/0886260516666531>
- Fan, W., & Gordon, M. D. (2014). The power of social media analytics. *Communications of the ACM*, 57(6), 74-81.
- Farnadi, G., Sitaraman, G., Sushmita, S., Celli, F., Kosinski, M., & Stillwell, D. (2016). Computational personality recognition in social media. *User Modeling and User-Adapted Interaction*, 26(2), 109-142.
- Fatfouta, R. (2019). Facets of narcissism and leadership: A tale of Dr. Jekyll and Mr. Hyde? *Human Resource Management Review*, 29(4), 100669. Retrieved from <https://doi.org/10.1016/j.hrmr.2018.10.002>
- Fearnley, F., & Fyfe, R. (2018). Twitter: an emerging source for geographical study. *Geography*, 103(2), 97-101.
- Fishwick, C. (2016). *I, Narcissist—vanity, social media, and the human condition*. Retrieved from <http://www.guardian.com/world/2016/mar/17/i-narcissist-vanity-social-media-and-the-human-condition>
- Flick, U. (2017). *The Sage handbook of qualitative data collection*. Thousand Oaks, CA: Sage.
- Fox, J., & Moreland, J. J. (2015). The dark side of social networking sites: an exploration of the relational and psychological stressors associated with Facebook use and affordances. *Computers in Human Behavior*, 45, 168-176. Retrieved from <https://doi.org/10.1016/j.chb.2014.11.083>
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177-182. Retrieved from <https://doi.org/10.1177/0963721412445309>
- Furnham, A., Richards, S. C., & Paulhus, D. L. (2013). The Dark Triad of personality: A 10-year review. *Social and Personality Psychology Compass*, 7(3), 199-216. Retrieved from <https://doi.org/10.1111/spc3.12018>

- Gallagher, R. J., Reing, K., Kale, D., & Ver Steeg, G. (2017). Anchored correlation explanation: Topic modeling with minimal domain knowledge. *Transactions of the Association for Computational Linguistics*, 5, 529-542.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.
- Gao, S., Brekelmans, R., Ver Steeg, G., & Galstyan, A. (2018). Auto-encoding total correlation explanation. In *The 22nd International Conference on Artificial Intelligence and Statistics* (pp. 1157-1166). PMLR.
- Garcia, D., & Sikström, S. (2014). The dark side of Facebook: semantic representations of status updates predict the Dark Triad of personality. *Personality and Individual Differences*, 67, 92-96. Retrieved from <https://doi.org/10.1016/j.paid.2013.10.001>
- Gemert, T. V. (2017). *On the influence of dataset characteristics on classifier performance* (Bachelor's thesis). Utrecht: Utrecht University.
- Ghazi, A. N., Petersen, K., Reddy, S. S., & Nekkanti, H. (2018). Survey research in software engineering: Problems and mitigation strategies. *IEEE Access*, 7, 24703-24718.
- Gilbert, C. H., & Hutto, E. (2014). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *8th International Conference on Weblogs and Social Media (ICWSM-14)*, 8(1), 216-225. Retrieved from <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>
- Gilpin, L. H., Bau, D., Yuan, B. Z., Bajwa, A., Specter, M., & Kagal, L. (2018, October). Explaining explanations: an overview of interpretability of machine learning. In *2018 IEEE 5th International Conference on data science and advanced analytics (DSAA)* (pp. 80-89). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/dsaa.2018.00018>
- Ginsberg, K. (2015). Instabranding: shaping the personalities of the top food brands on instagram. *Elon Journal of Undergraduate Research in Communications*, 6(1), 78-91.

- Gitari, N. D., Zuping, Z., Damien, H., & Long, J. (2015). A lexicon-based approach for hate speech detection. *International Journal of Multimedia and Ubiquitous Engineering*, 10(4), 215-230. Retrieved from <https://doi.org/10.14257/ijmue.2015.10.4.21>
- Golbeck, J. A. (2016). Predicting personality from social media text. *AIS Transactions on Replication Research*, 2(1), 2. Retrieved from <https://doi.org/10.17705/1attr.00009>
- Golshan, P. N., Dashti, H. R., Azizi, S., & Safari, L. (2018). *A study of recent contributions on information extraction*. Retrieved from <https://arxiv.org/ftp/arxiv/papers/1803/1803.05667.pdf>
- Gonçalves, P., Araújo, M., Benevenuto, F., & Cha, M. (2013, October). Comparing and combining sentiment analysis methods. In *Proceedings of the 1st ACM Conference on Online Social Networks* (pp. 27-38). Retrieved from <https://doi.org/10.1145/2512938.2512951>
- Goodboy, A. K., & Martin, M. M. (2015). The personality profile of a cyberbully: Examining the Dark Triad. *Computers in Human Behavior*, 49, 1-4.
- Goodison, S. E., Davis, R. C., & Jackson, B. A. (2015). Digital evidence and the US criminal justice system – Identifying technology and other needs to more effectively acquire and utilize digital evidence. *Priority Criminal Justice Needs Initiative*, 1-31. Retrieved from <https://www.ojp.gov/pdffiles1/nij/grants/248770.pdf>
- Gormley, B., & Lopez, F. G. (2010). Psychological abuse perpetration in college dating relationships: contributions of gender, stress, and adult attachment orientations. *Journal of Interpersonal Violence*, 25(2), 204-218. Retrieved from <https://doi.org/10.1177/0886260509334404>
- Grapsas, S., Brummelman, E., Back, M. D., & Denissen, J. J. (2020). The “why” and “how” of narcissism: a process model of narcissistic status pursuit. *Perspectives on Psychological Science*, 15(1), 150-172. Retrieved from <https://doi.org/10.1177/1745691619873350>
- Grieve, R., March, E., & Watkinson, J. (2020). Inauthentic self-presentation on facebook as a function of vulnerable narcissism and lower self-esteem. *Computers in Human Behavior*, 102, 144-150. Retrieved from <https://doi.org/10.1016/j.chb.2019.08.020>

- Grijalva, E., & Zhang, L. (2016). Narcissism and self-insight: A review and meta-analysis of narcissists' self-enhancement tendencies. *Personality and Social Psychology Bulletin*, 42(1), 3-24.
- Gülşen, T. T. (2016). You tell me in emojis. In *Computational and Cognitive Approaches to Narratology* (pp. 354-375). IGI Global.
- Gundecha, P., & Liu, H. (2012). Mining social media: a brief introduction. *INFORMS Tutorials in Operations Research*, 1-17. Retrieved from <https://doi.org/10.1287/educ.1120.0105>
- Gupta, A., Banerjee, I., & Rubin, D. L. (2018). Automatic information extraction from unstructured mammography reports using distributed semantics. *Journal of Biomedical Informatics*, 78, 78-86. Retrieved from <https://doi.org/10.1016/j.jbi.2017.12.016>
- Gustafsson, M. (2020). *Sentiment analysis for tweets in Swedish: Using a sentiment lexicon with syntactic rules* (Unpublished dissertation). Sweden: Linnaeus University.
- HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PloS One*, 15(5), e0232525. Retrieved from <https://doi.org/10.1371/journal.pone.0232525>
- Haddi, E., Liu, X., & Shi, Y. (2013). The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, 17, 26-32. Retrieved from <https://doi.org/10.1016/j.procs.2013.05.005>
- Hall, A. N., & Matz, S. C. (2020). Targeting item-level nuances leads to small but robust improvements in personality prediction from digital footprints. *European Journal of Personality*, 34(5), 873-884.
- Hapke, H., Howard, C., & Lane, H. (2019). *Natural language processing in action: Understanding, analyzing, and generating text with Python*. Simon and Schuster Digital.

- Hardeniya, T., & Borikar, D. A. (2016). Dictionary based approach to sentiment analysis – a review. *International Journal of Advanced Engineering, Management and Science*, 2(5), 317-322.
- Hart, W., Adams, J. M., & Burton, K. A. (2016). Narcissistic for the people: narcissists and non-narcissists disagree about how to make a good impression. *Personality and Individual Differences*, 91, 69-73. Retrieved from <https://doi.org/10.1016/j.paid.2015.11.045>
- Haryanto, A. W., & Mawardi, E. K. (2018, September). Influence of word normalization and chi-squared feature selection on support vector machine (SVM) text classification. In *2018 International Seminar on Application for Technology of Information and Communication* (pp. 229-233). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/isemantic.2018.8549748>
- Hasan, K. S., & Ng, V. (2014, June). Automatic keyphrase extraction: a survey of the state of the art. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 1262-1273). Retrieved from <https://doi.org/10.3115/v1/p14-1119>
- Hassanein, M., Hussein, W., Rady, S., & Gharib, T. F. (2018, December). Predicting personality traits from social media using text semantics. In *2018 13th International Conference on Computer Engineering and Systems (ICCES)* (pp. 184-189). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/icces.2018.8639408>
- Hawk, S. T., Van den Eijnden, R. J., Van Lissa, C. J., & Ter Bogt, T. F. (2019). Narcissistic adolescents' attention-seeking following social rejection: Links with social media disclosure, problematic social media use, and smartphone stress. *Computers in Human Behavior*, 92, 65-75. Retrieved from <https://doi.org/10.1016/j.chb.2018.10.032>
- Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word cloud explorer: Text analytics based on word clouds. In *47th Hawaii International Conference on System Sciences* (pp. 1833-1842). US/Canada: IEEE.

- Heine, S. J., & Buchtel, E. E. (2009). Personality: the universal and the culturally specific. *Annual Review of Psychology*, 60, 369-394. Retrieved from <https://doi.org/10.1146/annurev.psych.60.110707.163655>
- Hepper, E. G., Hart, C. M., Meek, R., Cisek, S., & Sedikides, C. (2014). Narcissism and empathy in young offenders and non-offenders. *European Journal of Personality*, 28(2), 201-210. Retrieved from <https://doi.org/10.1002/per.1939>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 75-105. Retrieved from <https://doi.org/10.2307/25148625>
- Hevner, A., & Chatterjee, S. (2010). Design science research in information systems. In *Design research in information systems* (pp. 9-22). Boston, MA: Springer. Retrieved from https://doi.org/10.1007/978-1-4419-5653-8_2
- Hezarjaribi, N., Ashari, Z. E., Frenzel, J. F., Ghasemzadeh, H., & Hemati, S. (2020). Personality assessment from text for machine commonsense reasoning. Retrieved from <https://doi.org/10.1063/pt.5.028530>
- Hickman, L., Thapa, S., Tay, L., Cao, M., & Srinivasan, P. (2022). Text preprocessing for text mining in organizational research: review and recommendations. *Organizational Research Methods*, 25(1), 114-146. Retrieved from <https://doi.org/10.1177/1094428120971683>
- Hinds, J., & Joinson, A. (2019). Human and computer personality prediction from digital footprints. *Current Directions in Psychological Science*, 28(2), 204-211.
- Hino, A., & Fahey, R. A. (2019). Representing the Twittersphere: Archiving a representative sample of Twitter data under resource constraints. *International Journal of Information Management*, 48, 175-184.
- Hipp, J. R., Bates, C., Lichman, M., & Smyth, P. (2019). Using social media to measure temporal ambient population: Does it help explain local crime rates? *Justice Quarterly*, 36(4), 718-748.

- Hodson, G., Hogg, S. M., & MacInnis, C. C. (2009). The role of “dark personalities” (narcissism, Machiavellianism, psychopathy), Big Five personality factors, and ideology in explaining prejudice. *Journal of Research in Personality*, 43(4), 686-690.
- Holsapple, C. W., Hsiao, S. H., & Pakath, R. (2018). Business social media analytics: Characterization and conceptual framework. *Decision Support Systems*, 110, 32-45.
- Holtzman, N. S., Vazire, S., & Mehl, M. R. (2010). Sounds like a narcissist: behavioral manifestations of narcissism in everyday life. *Journal of Research in Personality*, 44(4), 478-484. Retrieved from <https://doi.org/10.1016/j.jrp.2010.06.001>
- Hosseini, M., Powell, M., Collins, J., Callahan-Flintoft, C., Jones, W., Bowman, H., & Wyble, B. (2020). I tried a bunch of things: the dangers of unexpected overfitting in classification of brain data. *Neuroscience & Biobehavioral Reviews*, 119, 456-467. Retrieved from <https://doi.org/10.1016/j.neubiorev.2020.09.036>
- Howells, K., & Ertugan, A. (2017). Applying fuzzy logic for sentiment analysis of social media network data in marketing. *Procedia Computer Science*, 120, 664-670.
- Hruska, J., & Maresova, P. (2020). Use of social media platforms among adults in the United States – behavior on social media. *Societies*, 10(1), 27. Retrieved from <http://dx.doi.org/10.3390/soc10010027>
- Huang, J. L., Ryan, A. M., Zabel, K. L., & Palmer, A. (2014). Personality and adaptive performance at work: a meta-analytic investigation. *Journal of Applied Psychology*, 99(1), 162. Retrieved from <https://doi.org/10.1037/a0034285>
- Hudson, E. J. (2012). Understanding and exploring narcissism: impact on students and college campuses. *CMC Senior Theses*, paper 381. Retrieved from http://scholarship.claremont.edu/cmc_theses/381
- Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior*, 28(2), 561-569. Retrieved from <https://doi.org/10.1016/j.chb.2011.11.001>
- Hung, C., & Chen, S. J. (2016). Word sense disambiguation based sentiment lexicons for sentiment classification. *Knowledge-Based Systems*, 110, 224-232.

- Hyatt, C. S., Sleep, C. E., Lynam, D., Widiger, T. A., Campbell, W. K., & Miller, J. D. (2018). Ratings of affective and interpersonal tendencies differ for grandiose and vulnerable narcissism: a replication and extension of Gore and Widiger (2016). *Journal of Personality*, 86(3), 422-434.
- Ibrahim, R., Zeebaree, S., & Jacksi, K. (2019). Survey on semantic similarity based on document clustering. *Advances in Science, Technology and Engineering Systems Journal*, 4(5), 115-122.
- Ireland, M. E., & Mehl, M. R. (2014). Natural language use as a marker. In *The Oxford Handbook of Language and Social Psychology* (pp. 201-237). Retrieved from <https://doi.org/10.1093/oxfordhb/9780199838639.013.034>
- Ivankova, N. V. (2014). *Mixed methods applications in action research*. Thousand Oaks, CA: Sage.
- Jackson, D. N. (1994). *Jackson personality inventory – revised manual*. Port Heron: Sigma Assessment Systems, Research Psychologists Press Division.
- Jain, H. (2017). A web-based application for sentiment analysis. *International Journal of Education and Management Engineering*, 1, 25-35. Retrieved from <https://doi.org/10.5815/ijeme.2017.01.03>
- James, S., Kavanagh, P. S., Jonason, P. K., Chonody, J. M., & Scrutton, H. E. (2014). The Dark Triad, schadenfreude, and sensational interests: dark personalities, dark emotions, and dark behaviors. *Personality and Individual Differences*, 68, 211-216. Retrieved from <https://doi.org/10.1016/j.paid.2014.04.020>
- Jan, T. G. (2020). Clustering of tweets: a novel approach to label the unlabelled tweets. In *Proceedings of ICRIC 2019* (pp. 671-685). Cham.: Springer.
- Jauk, E. (2019). A bio-psycho-behavioral model of creativity. *Current Opinion in Behavioral Sciences*, 27, 1-6. Retrieved from <https://doi.org/10.1016/j.cobeha.2018.08.012>
- Jean-Claude, M. (2014). Townships as crime ‘hot-spot’ areas in Cape Town: perceived root causes of crime in site B, Khayelitsha. *Mediterranean Journal of Social Sciences*, 5(8), 596-603.

- Jefferson, C., Liu, H., & Cocea, M. (2017). Fuzzy approach for sentiment analysis. In *2017 IEEE international conference on fuzzy systems (FUZZ-IEEE)* (pp. 1-6). US/Canada: IEEE.
- Ji, X., Henriques, J. F., & Vedaldi, A. (2019). Invariant information clustering for unsupervised image classification and segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 9865-9874). US/Canada: IEEE.
- Jiang, C., Wang, Z., Wang, R., & Ding, Y. (2018). Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending. *Annals of Operations Research*, 266(1), 511-529. Retrieved from <https://doi.org/10.1007/s10479-017-2668-z>
- John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In R. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). New York: The Guilford Press.
- Jolly, E., & Chang, L. J. (2019). The flatland fallacy: Moving beyond low-dimensional thinking. *Topics in Cognitive Science*, 11(2), 433-454. Retrieved from <https://doi.org/10.1111/tops.12404>
- Jones, B. D., Woodman, T., Barlow, M., & Roberts, R. (2017). The darker side of personality: narcissism predicts moral disengagement and antisocial behavior in sport. *The Sport Psychologist*, 31(2), 109-116. Retrieved from <https://doi.org/10.1123/tsp.2016-0007>
- Jones, D. N. (2013). What's mine is mine and what's yours is mine: The Dark Triad and gambling with your neighbor's money. *Journal of Research in Personality*, 47(5), 563-571. Retrieved from <https://doi.org/10.1016/j.jrp.2013.04.005>
- Jurek, A., Mulvenna, M. D., & Bi, Y. (2015). Improved lexicon-based sentiment analysis for social media analytics. *Security Informatics*, 4(1), 1-13.
- Kadhim, A. I. (2018). An evaluation of preprocessing techniques for text classification. *International Journal of Computer Science and Information Security (IJCSIS)*, 16(6), 22-32.

- Kaity, M., & Balakrishnan, V. (2020). Sentiment lexicons and non-English languages: a survey. In *Knowledge and Information Systems* (pp. 1-36). Springer.
- Kalemi, G., Michopoulos, I., Efstathiou, V., Konstantopoulou, F., Tsaklakidou, D., Gournellis, R., & Douzenis, A. (2019). Narcissism but not criminality is associated with aggression in women: A study among female prisoners and women without a criminal record. *Frontiers in Psychiatry*, 10, 21. Retrieved from <https://doi.org/10.3389/fpsyt.2019.00021>
- Kalra, V., & Aggarwal, R. (2017). Importance of text data preprocessing & implementation in RapidMiner. In *Proceedings of the First International Conference on Information Technology and Knowledge Management* (pp. 71-75). Retrieved from https://annals-csis.org/Volume_14/drp/pdf/46.pdf
- Kambatla, K., Kollias, G., Kumar, V., & Grama, A. (2014). Trends in big data analytics. *Journal of Parallel and Distributed Computing*, 74(7), 2561-2573.
- Kang, Y., & Zhou, L. (2017). RubE: rule-based methods for extracting product features from online consumer reviews. *Information & Management*, 54(2), 166-176. Retrieved from <https://doi.org/10.1016/j.im.2016.05.007>
- Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: past, present and future. *Information Systems Frontiers*, 20(3), 531-558.
- Karami, A., Gangopadhyay, A., Zhou, B., & Kharrazi, H. (2015, August). FLATM: a Fuzzy Logic Approach Topic Model for medical documents. In *2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC)* (pp. 1-6). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/nafips-wconsc.2015.7284190>
- Kassu, J. S. (2019). Research design and methodology. In A. Evon, E. M. Abdelkrim, & H. Issam (Eds.), *Cyberspace*. Retrieved from <https://doi.org/10.5772/intechopen.85731>
- Kasture, A. S. (2015). *A predictive model to detect online cyberbullying* (Doctoral dissertation). Auckland: Auckland University of Technology.

- Kaufman, S. B., Weiss, B., Miller, J. D., & Campbell, W. K. (2020). Clinical correlates of vulnerable and grandiose narcissism: a personality perspective. *Journal of Personality Disorders*, 34(1), 107-130. Retrieved from https://doi.org/10.1521/pedi_2018_32_384
- Kaur, J., & Buttar, P. K. (2018). A systematic review on stopword removal algorithms. *International Journal on Future Revolution in Computer Science & Communication Engineering*, 4, 207-210.
- Kaushal, V., & Patwardhan, M. (2018). Emerging trends in personality identification using online social networks—a literature survey. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 12(2), 1-30.
- Kauten, R. L., Lui, J. H., Stry, A. K., & Barry, C. T. (2015). “Purging my friends list. Good luck making the cut”: perceptions of narcissism on Facebook. *Computers in Human Behavior*, 51, 244-254. Retrieved from <https://doi.org/10.1016/j.chb.2015.05.010>
- Kavuri, D., Kumar, P. A., & Rao, D. V. S. (2012). Text and image classification using fuzzy similarity based self constructing algorithm. *International Journal of Engineering Science & Advanced Technology*, 2(6), 1572-1576.
- Kaye, L. K., Malone, S. A., & Wall, H. J. (2017). Emojis: insights, affordances, and possibilities for psychological science. *Trends in Cognitive Sciences*, 21(2), 66-68. Retrieved from <https://doi.org/10.1016/j.tics.2016.10.007>
- Keatley, D. A., McGurk, S., & Allely, C. S. (2019). Understanding school shootings with crime script analysis. *Deviant Behavior*, 1-13.
- Kemp, S. (2018). Global digital 2018: World’s internet users pass 4 billion. We Are Social. Retrieved from <https://wearesocial.com/blog/2018/01/global-digital-report-2018>
- Kern, M. L., Eichstaedt, J. C., Schwartz, H. A., Dziurzynski, L., Ungar, L. H., Stillwell, D. J., ... & Seligman, M. E. (2014). The online social self: an open vocabulary approach to personality. *Assessment*, 21(2), 158-169. Retrieved from <https://doi.org/10.1177/1073191113514104>
- Khan, A. U., & Ratha, B. K. (2016, March). Business data extraction from social networking. In *2016 3rd International Conference on Recent Advances in Information Technology*

- (RAIT) (pp. 651-656). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/rait.2016.7507976>
- Khan, M. T., Durrani, M., Ali, A., Inayat, I., Khalid, S., & Khan, K. H. (2016). Sentiment analysis and the complex natural language. *Complex Adaptive Systems Modeling*, 4(1), 1-19.
- Kharde, V., & Sonawane, P. (2016). Sentiment analysis of twitter data: a survey of techniques. *International Journal of Computer Applications*, 139(11), 5-15.
- Khowaja, A., Mahar, M. H., Nawaz, H., Wasi, S., & Rehman, S. (2019). Personality evaluation of student community using sentiment analysis. *International Journal of Computer Science and Network Security*, 19(3), 167-180.
- Kim, J., & Hastak, M. (2018). Social network analysis: characteristics of online social networks after a disaster. *International Journal of Information Management*, 38(1), 86-96. Retrieved from <https://doi.org/10.1016/j.ijinfomgt.2017.08.003>
- Koch, R. (2011). *The 80/20 principle: The secret of achieving more with less: Updated 20th anniversary edition of the productivity and business classic*. United Kingdom: Nicholas Brealey Publishing Limited.
- Kolchyna, O., Souza, T. T., Treleaven, P., & Aste, T. (2015). Twitter sentiment analysis: Lexicon method, machine learning method and their combination. *Handbook of Sentiment Analysis in Finance*. Retrieved from <https://doi.org/10.48550/arXiv.1507.00955>
- Kolkur, S., Dantal, G., & Mahe, R. (2015). Study of different levels for sentiment analysis. *International Journal of Current Engineering and Technology*, 5(2), 768-770.
- Kong, F., Wang, M., Zhang, X., Li, X., & Sun, X. (2021). Vulnerable narcissism in social networking sites: The role of upward and downward social comparisons. *Frontiers in Psychology*, 3921. Retrieved from <https://www.frontiersin.org/articles/10.3389/fpsyg.2021.711909/full>

- Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, 95(3), 357-380. Retrieved from <https://doi.org/10.1007/s10994-013-5415-y>
- Kosinski, M., Matz, S. C., Gosling, S. D., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, 70(6), 543. Retrieved from <https://doi.org/10.1037/a0039210>
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences (PNAS)*, 110(15), 5802-5805. Retrieved from <https://doi.org/10.1073/pnas.1218772110>
- Kou, G., & Peng, Y. (2015). An application of latent semantic analysis for text categorization. *International Journal of Computers Communications & Control*, 10(3), 357-369.
- Kounadi, O., Ristea, A., Araujo, A., & Leitner, M. (2020). A systematic review on spatial crime forecasting. *Crime Science*, 9(1), 1-22.
- Kraus, S., Berchtold, J., Palmer, C., & Filser, M. (2018). Entrepreneurial orientation: the dark triad of executive personality. *Journal of Promotion Management*, 24(5), 715-735. Retrieved from <https://doi.org/10.1080/10496491.2018.1405524>
- Krizan, Z. (2018). The narcissism spectrum model: A spectrum perspective on narcissistic personality. In *Handbook of trait narcissism* (pp. 15-25). Cham.: Springer.
- Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity*, 47(4), 2025-2047.
- Kulkarni, V., Kern, M. L., Stillwell, D., Kosinski, M., Matz, S., Ungar, L., & Schwartz, H. A. (2018). Latent human traits in the language of social media: An open-vocabulary approach. *PloS One*, 13(11), e0201703. Retrieved from <https://doi.org/10.1371/journal.pone.0201703>

- Kumar, A., Sharma, A., & Arora, A. (2019, March). Anxious depression prediction in real-time social data. In *International Conference on Advances in Engineering Science Management & Technology (ICAESMT)-2019*. Dehradun, India: Uttaraanchal University. Retrieved from <https://doi.org/10.2139/ssrn.3383359>
- Kumar, U., Reganti, A. N., Maheshwari, T., Chakroborty, T., Gambäck, B., & Das, A. (2018). Inducing personalities and values from language use in social network communities. *Information Systems Frontiers*, 20(6), 1219-1240.
- Kundi, F. M., Khan, A., Ahmad, S., & Asghar, M. Z. (2014). Lexicon-based sentiment analysis in the social web. *Journal of Basic and Applied Scientific Research*, 4(6), 238-48.
- Kunte, A. V., & Panicker, S. (2019). Using textual data for personality prediction: a machine learning approach. In *4th International Conference on Information Systems and Computer Networks (ISCON)* (pp. 529-533). US/Canada: IEEE.
- Kursuncu, U., Gaur, M., Lokala, U., Thirunarayan, K., Sheth, A., & Arpinar, I. B. (2019). Predictive analysis on Twitter: Techniques and applications. In *Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining* (pp. 67-104). Cham.: Springer.
- Kuss, D. J., & Griffiths, M. D. (2017). Social networking sites and addiction: Ten lessons learned. *International Journal of Environmental Research and Public Health*, 14(3), 311. Retrieved from <https://doi.org/10.3390/ijerph14030311>
- Kwak, H., Lee, C., Park, H., & Moon, S. (2010, April). What is Twitter, a social network or a news media?. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 591-600). Retrieved from <https://doi.org/10.1145/1772690.1772751>
- Kwon, O., & Sim, J. M. (2013). Effects of data set features on the performances of classification algorithms. *Expert Systems with Applications*, 40(5), 1847-1857.
- Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., & Aila, T. (2019). Improved precision and recall metric for assessing generative models. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)* (pp. 1-10). Vancouver, Canada.

- Lacerda, D. P., Dresch, A., Proença, A., & Antunes Júnior, J. A. V. (2013). Design science research: A research method to production engineering. *Gestão & Produção*, 20, 741-761. Retrieved from <https://doi.org/10.1590/s0104-530x2013005000014>
- Lakey, C. E., Rose, P., Campbell, W. K., & Goodie, A. S. (2008). Probing the link between narcissism and gambling: the mediating role of judgment and decision-making biases. *Journal of Behavioral Decision Making*, 21(2), 113-137. Retrieved from <https://doi.org/10.1002/bdm.582>
- Lambe, S., Hamilton-Giachritsis, C., Garner, E., & Walker, J. (2018). The role of narcissism in aggression and violence: a systematic review. *Trauma, Violence, & Abuse*, 19(2), 209-230.
- Langstedt, E., & Hunt, D. S. (2017). An exploration into the brand personality traits of social media sites. *The Journal of Social Media in Society*, 6(2), 315-342.
- Lannin, D. G., Gyll, M. K., Krizan, Z., Madon, S., & Cornish, M. (2014). When are grandiose and vulnerable narcissists least helpful? *Personality and Individual Differences*, 56, 127-132.
- Lansley, G., & Longley, P. A. (2016). The geography of Twitter topics in London. *Computers, Environment and Urban Systems*, 58, 85-96.
- Lau, K. S., & Marsee, M. A. (2013). Exploring narcissism, psychopathy, and Machiavellianism in youth: Examination of associations with antisocial behavior and aggression. *Journal of Child and Family Studies*, 22(3), 355-367. Retrieved from <https://doi.org/10.1007/s10826-012-9586-0>
- Leary, T., & Ashman, J. (2018). Narcissistic leadership: important considerations and practical implications. *International Leadership Journal*, 10(2), 62-74.
- Lee, E., Choi, C., & Kim, P. (2017). Intelligent handover scheme for drone using fuzzy inference systems. *IEEE Access*, 5, 13712-13719. Retrieved from <https://doi.org/10.1109/access.2017.2724067>
- Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, power, and money: prediction from the Dark Triad and Honesty–Humility.

- European Journal of Personality*, 27(2), 169-184. Retrieved from <https://doi.org/10.1002/per.1860>
- Lee, V. L. S., Gan, K. H., Tan, T. P., & Abdullah, R. (2019). Semi-supervised learning for sentiment classification using small number of labeled data. *Procedia Computer Science*, 161, 577-584. Retrieved from <https://doi.org/10.1016/j.procs.2019.11.159>
- Li, C., Sun, Y., Ho, M. Y., You, J., Shaver, P. R., & Wang, Z. (2016). State narcissism and aggression: The mediating roles of anger and hostile attributional bias. *Aggressive Behavior*, 42(4), 333-345. Retrieved from <https://doi.org/10.1002/ab.21629>
- Li, Y., Yan, C., Liu, W., & Li, M. (2018). A principle component analysis-based random forest with the potential nearest neighbor method for automobile insurance fraud identification. *Applied Soft Computing Journal*. Retrieved from <http://dx.doi.org/10.1016/j.asoc.2017.07.027>
- Lima, A. C., & De Castro, L. N. (2019). TECLA: a temperament and psychological type prediction framework from Twitter data. *Plos One*, 14(3), e0212844. Retrieved from <http://dx.doi.org/10.1371/journal.pone.0212844>
- Lin, J., Mao, W., & Zeng, D. D. (2017). Personality-based refinement for sentiment classification in microblog. *Knowledge-Based Systems*, 132, 204-214. Retrieved from <https://doi.org/10.1016/j.knosys.2017.06.031>
- Lipschultz, J. H. (2017). *Social media communication: concepts, practices, data, law and ethics*. New York, NY: Routledge.
- Liu, F., Huang, X., Huang, W., & Duan, S. X. (2020). Performance evaluation of keyword extraction methods and visualization for student online comments. *Symmetry*, 12(11), 1923. Retrieved from <https://doi.org/10.3390/sym12111923>
- Liu, H., & Cocea, M. (2017). Fuzzy rule based systems for interpretable sentiment analysis. In *9th International Conference on Advanced Computational Intelligence* (pp. 129-136). US/Canada: IEEE.
- Liu, H., & Zhang, L. (2018). Fuzzy rule-based systems for recognition-intensive classification in granular computing context. *Granular Computing*, 3, 355-365.

- Liu, H., Burnap, P., Alorainy, W., & Williams, M. (2019). A fuzzy approach to text classification with two-stage training for ambiguous instances. *IEEE Transactions on Computational Social Systems*, 6(2), 227-240.
- Liu, H., Christiansen, T., Baumgartner, W. A., & Verspoor, K. (2012). BioLemmatizer: a lemmatization tool for morphological processing of biomedical text. *Journal of biomedical semantics*, 3(1), 1-29.
- Liu, L., Preotiuc-Pietro, D., Samani, Z. R., Moghaddam, M. E., & Ungar, L. (2016). Analyzing personality through social media profile picture choice. In *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1), 211-220.
- Liu, L., Tang, L., Dong, W., Yao, S., & Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, 5(1), 1608. Retrieved from <https://doi.org/10.1186/s40064-016-3252-8>
- Liu, X., Burns, A. C., & Hou, Y. (2017). An investigation of brand-related user-generated content on Twitter. *Journal of Advertising*, 46(2), 236-247.
- Liu, Y., & Ge, Z. (2018). Weighted random forests for fault classification in industrial processes with hierarchical clustering model selection. *Journal of Process Control*, 64, 62-70. Retrieved from <https://doi.org/10.1016/j.jprocont.2018.02.005>
- Liu, Y., Wang, J., & Jiang, Y. (2016). PT-LDA: A latent variable model to predict personality traits of social network users. *Neurocomputing*, 210, 155-163. Retrieved from <https://doi.org/10.1016/j.neucom.2015.10.144>
- Liza, F. F. (2020). Sentence classification with imbalanced data for health applications. In *Proceedings of the 5th Social Media Mining for Health Applications Workshop & Shared Task* (pp. 138-145). Barcelona, Spain.
- Lobbestael, J., Baumeister, R. F., Fiebig, T., & Eckel, L. A. (2014). The role of grandiose and vulnerable narcissism in self-reported and laboratory aggression and testosterone reactivity. *Personality and Individual Differences*, 69, 22-27. Retrieved from <https://doi.org/10.1016/j.paid.2014.05.007>

- Lopes, P. N., Nezlek, J. B., Extremera, N., Hertel, J., Fernández-Berrocal, P., Schütz, A., & Salovey, P. (2011). Emotion regulation and the quality of social interaction: does the ability to evaluate emotional situations and identify effective responses matter? *Journal of Personality*, 79(2), 429-467. Retrieved from <https://doi.org/10.1111/j.1467-6494.2010.00689.x>
- Lowe, B., & Laffey, D. (2011). Is Twitter for the birds? Using Twitter to enhance student learning in a marketing course. *Journal of Marketing Education*, 33(2), 182-192.
- Lowenstein, J., Purvis, C., & Rose, K. (2016). A systematic review on the relationship between antisocial, borderline and narcissistic personality disorder diagnostic traits and risk of violence to others in a clinical and forensic sample. *Borderline Personality Disorder and Emotion Dysregulation*, 3(1), 1-12. Retrieved from <https://doi.org/10.1186/s40479-016-0046-0>
- Luhtanen, R. K., & Crocker, J. (2005). Alcohol use in college students: effects of level of self-esteem, narcissism, and contingencies of self-worth. *Psychology of Addictive Behaviors*, 19(1), 99. Retrieved from <https://doi.org/10.1037/0893-164x.19.1.99>
- Lukito, L. C., Erwin, A., Purnama, J., & Danoekoesoemo, W. (2016). Social media user personality classification using computational linguistic. In *8th International Conference on Information Technology and Electrical Engineering (ICITEE)* (pp. 1-6). US/Canada: IEEE.
- Luque, A., Carrasco, A., Martín, A., & De las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216-231.
- Madani, Y., Erritali, M., Bengourram, J., & Sailhan, F. (2019). Social collaborative filtering approach for recommending courses in an E-learning platform. *Procedia Computer Science*, 151, 1164-1169. Retrieved from <https://doi.org/10.1016/j.procs.2019.04.166>
- Madhusudhanan, S., & Moorthi, M. (2019). Optimized fuzzy technique for enhancing sentiment analysis. *Cluster Computing*, 22(5), 11929-11939. Retrieved from <https://doi.org/10.1007/s10586-017-1514-z>

- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., ... Schmid-Petri, H. (2018). Applying LDA topic modeling in communication research: toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2-3), 93-118.
- Maisto, A., Pelosi, S., Polito, M., & Stingo, M. (2019). Automatic text preprocessing for intelligent dialog agents. In *Workshops of the International Conference on Advanced Information Networking and Applications* (pp. 805-814). Cham.: Springer.
- Majid, Y. (2012). E-Crime 2.0: the criminological landscape of new social media. *Information & Communications Technology Law*, 21(3), 207-219.
- Manivannan, P. V., & Ramakanth, P. (2018). Vision based intelligent vehicle steering control using single camera for automated highway system. *Procedia Computer Science*, 133, 839-846.
- Marshall, B. H. (2018). *The complete guide to personal digital archiving*. Chicago: American Library Association.
- Marshall, T. C., Ferenczi, N., Lefringhausen, K., Hill, S., & Deng, J. (2020). Intellectual, narcissistic, or Machiavellian? How Twitter users differ from Facebook-only users, why they use Twitter, and what they tweet about. *Psychology of Popular Media*, 9(1), 14. Retrieved from <https://doi.org/10.1037/ppm0000209>
- Marshall, T. C., Lefringhausen, K., & Ferenczi, N. (2015). The Big Five, self-esteem, and narcissism as predictors of the topics people write about in Facebook status updates. *Personality and Individual Differences*, 85, 35-40. Retrieved from <https://doi.org/10.1016/j.paid.2015.04.039>
- Martínez, J. A. D., & de Frutos, T. H. (2018). Connectivism in the network society. the coming of social capital knowledge. *Tendencias Sociales. Revista de Sociología*, (1), 21-37. Retrieved from <https://doi.org/10.5944/ts.1.2018.21358>
- Mary, A., & Arockiam, L. (2018, April). FDSS: Fuzzy Based Decision Support System for aspect based sentiment analysis in big data. In *International Conference on Advances in Computing and Data Sciences* (pp. 77-87). Singapore: Springer. Retrieved from https://doi.org/10.1007/978-981-13-1813-9_8

- Maxwell, J. A. (2012). *Qualitative research design: An interactive approach*. Thousand Oaks, California: Sage Publications. Retrieved from <https://doi.org/10.4135/9781529770278.n4>
- Maynard, B. R., Vaughn, M. G., Salas-Wright, C. P., & Vaughn, S. (2016). Bullying victimization among school-aged immigrant youth in the United States. *Journal of Adolescent Health, 58*(3), 337-344. Retrieved from <https://doi.org/10.1016/j.jadohealth.2015.11.013>
- McCabe, M. B. (2017). Social media marketing strategies for career advancement: an analysis of LinkedIn. *Journal of Business and Behavioral Sciences, 29*(1), 85-99.
- McCain, J. L., & Campbell, W. K. (2018). Narcissism and social media use: a meta-analytic review. *Psychology of Popular Media Culture, 7*(3), 308-327. Retrieved from <https://doi.org/10.1037/ppm0000137>
- McClellan, C., Ali, M. M., Mutter, R., Kroutil, L., & Landwehr, J. (2017). Using social media to monitor mental health discussions— evidence from Twitter. *Journal of the American Medical Informatics Association, 24*(3), 496-502. Retrieved from <https://doi.org/10.1093/jamia/ocw133>
- McElroy, J. C., Hendrickson, A. R., Townsend, A. M., & DeMarie, S. M. (2007). Dispositional factors in internet use: personality versus cognitive style. *MIS Quarterly, 809-820*. Retrieved from <https://doi.org/10.2307/25148821>
- McKinney, B. C., Kelly, L., & Duran, R. L. (2012). Narcissism or openness?: college students' use of Facebook and Twitter. *Communication Research Reports, 29*(2), 108-118. Retrieved from <https://doi.org/10.1080/08824096.2012.666919>
- Meagher, B. R., Leman, J. C., Bias, J. P., Latendresse, S. J., & Rowatt, W. C. (2015). Contrasting self-report and consensus ratings of intellectual humility and arrogance. *Journal of Research in Personality, 58*, 35-45. Retrieved from <https://doi.org/10.1016/j.jrp.2015.07.002>
- Mehanna, Y. S., & Mahmuddin, M. (2021). The effect of pre-processing techniques on the accuracy of sentiment analysis using bag-of-concepts text representation. *SN*

Computer Science, 2(4), 1-13. Retrieved from <https://doi.org/10.1007/s42979-021-00453-7>

- Meishar-Tal, H., & Pieterse, E. (2016, June). Academics' use of academic social networking sites: The case of Research Gate and Academia.Edu. In *Proceedings of the European Distance and E-Learning Network 2016 Annual Conference*, Budapest, 14-17 June. Retrieved from https://www.researchgate.net/publication/309732348_ACADEMICS'_USE_OF_ACADEMIC_SOCIAL_NETWORKING_SITES_THE_CASE_OF_RESEARCHGATE_AND_ACADEMIAEDU
- Metzger, M. J., Wilson, C., & Zhao, B. Y. (2018). Benefits of browsing? The prevalence, nature, and effects of profile consumption behavior in social network sites. *Journal of Computer-Mediated Communication*, 23(2), 72-89.
- Miller, J. D., Hoffman, B. J., Gaughan, E. T., Gentile, B., Maples, J., & Campbell, W. K. (2011). Grandiose and vulnerable narcissism: A nomological network analysis. *Journal of Personality*, 79(5), 1013-1042.
- Miller, J. D., Lynam, D. R., Hyatt, C. S., & Campbell, W. K. (2017). Controversies in narcissism. *Annual Review of Clinical Psychology*, 13, 291-315. Retrieved from <https://doi.org/10.1146/annurev-clinpsy-032816-045244>
- Mohajan, H. K. (2017). Two criteria for good measurements in research: validity and reliability. *Annals of Spiru Haret University: Economic Series*, 17(4), 59-82.
- Molina-Gil, J., Concepción-Sánchez, J. A., & Caballero-Gil, P. (2019). Harassment detection using machine learning and fuzzy logic techniques. In *Multidisciplinary Digital Publishing Institute Proceedings*, 31(1), 27. Retrieved from <https://doi.org/10.3390/proceedings2019031027>
- Mor, N. (2020). Eysenck personality questionnaire. In *The Corsini Encyclopedia of Psychology* (pp. 1-2). John Wiley & Sons.
- Mowlaei, M. E., Abadeh, M. S., & Keshavarz, H. (2020). Aspect-based sentiment analysis using adaptive aspect-based lexicons. *Expert Systems with Applications*, 148, 113234. Retrieved from <https://doi.org/10.1016/j.eswa.2020.113234>

- Mullen, L., Benoit, K., Keyes, O., & Selivanov, D. (2018). Fast, consistent tokenization of natural language text. *Journal of Open Source Software*, 3(23), 655. Retrieved from <http://dx.doi.org/10.21105/joss.00655>
- Müller, K. W., Dreier, M., Beutel, M. E., Duven, E., Giralt, S., & Wölfling, K. (2016). A hidden type of internet addiction? Intense and addictive use of social networking sites in adolescents. *Computers in Human Behavior*, 55, 172-177. Retrieved from <https://doi.org/10.1016/j.chb.2015.09.007>
- Nabukenya, J. (2012). Combining case study, design science and action research methods for effective collaboration engineering research efforts. In *45th Hawaii International Conference on System Sciences* (pp. 343-352). US/Canada: IEEE.
- Nair, L. R., Shetty, S. D., & Shetty, S. D. (2018). Applying spark based machine learning model on streaming big data for health status prediction. *Computers & Electrical Engineering*, 65(C), 393-399.
- Nalepa, J., & Kawulok, M. (2019). Selecting training sets for support vector machines: a review. *Artificial Intelligence Review*, 52(2), 857-900.
- Namatevs, I., Sudars, K., & Polaka, I. (2019). Automatic data labeling by neural networks for the counting of objects in videos. *Procedia Computer Science*, 149, 151-158.
- Naseem, U., Razzak, I., & Eklund, P. W. (2020). A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter. *Multimedia Tools and Applications*, 80(4), 1-28.
- Neuman, Y., & Cohen, Y. (2014). A vectorial semantics approach to personality assessment. *Scientific Reports*, 4(1), 1-6.
- Nevicka, B., De Hoogh, A. H., Van Vianen, A. E., Beersma, B., & McIlwain, D. (2011). All I need is a stage to shine: narcissists' leader emergence and performance. *The Leadership Quarterly*, 22(5), 910-925. Retrieved from <https://doi.org/10.1016/j.leaqua.2011.07.011>

- Newman, H., & Joyner, D. (2018). Sentiment analysis of student evaluations of teaching. In *International Conference on Artificial Intelligence in Education* (pp. 246-250). Cham: Springer.
- Nguyen, H. D., Do, N. V., Tran, N. P., & Pham, X. H. (2020). Some criteria of the knowledge representation method for an intelligent problem solver in STEM education. *Applied Computational Intelligence and Soft Computing*, 2020, 9834218. Retrieved from <https://doi.org/10.1155/2020/9834218>
- Nhlabano, V. V., & Lutu, P. E. (2018). Impact of text pre-processing on the performance of sentiment analysis models for social media data. In *International Conference on Advances in Big Data, Computing and Data Communication Systems* (pp. 1-6). US/Canada: IEEE.
- Nikam, S. S. (2015). A comparative study of classification techniques in data mining algorithms. *Oriental Journal of Computer Science & Technology*, 8(1), 13-19.
- Norden, S. (2013). *How the internet has changed the face of crime* (Unpublished Doctoral dissertation). Florida: Florida Gulf Coast University.
- Nowak, B., Brzóska, P., Piotrowski, J., Sedikides, C., Żemojtel-Piotrowska, M., & Jonason, P. K. (2020). Adaptive and maladaptive behavior during the COVID-19 pandemic: the roles of dark triad traits, collective narcissism, and health beliefs. *Personality and Individual Differences*, 167, 110232. Retrieved from <https://dx.doi.org/10.1016%2Fj.paid.2020.110232>
- Oh, C., Sasser, S., & Almahmoud, S. (2015). Social media analytics framework: the case of Twitter and Super Bowl ads. *Journal of Information Technology Management*, 26(1), 1-18.
- Oltmanns, J. R., & Widiger, T. A. (2018). Assessment of fluctuation between grandiose and vulnerable narcissism: development and initial validation of the FLUX scales. *Psychological Assessment*, 30(12), 1612. Retrieved from <https://doi.org/10.1037/pas0000616>

- Onan, A., Bulut, H., & Korukoglu, S. (2017). An improved ant algorithm with LDA-based representation for text document clustering. *Journal of Information Science*, 43(2), 275-292.
- Ovadia, S. (2014). ResearchGate and Academia.edu: Academic social networks. *Behavioral & Social Sciences Librarian*, 33(3), 165-169. Retrieved from <https://doi.org/10.1080/01639269.2014.934093>
- Ozimek, P., Bierhoff, H. W., & Hanke, S. (2018). Do vulnerable narcissists profit more from Facebook use than grandiose narcissists? An examination of narcissistic Facebook use in the light of self-regulation and social comparison theory. *Personality and Individual Differences*, 124, 168-177. Retrieved from <https://doi.org/10.1016/j.paid.2017.12.016>
- Panda, M. (2018). Developing an efficient text pre-processing method with sparse generative Naive Bayes for text mining. *International Journal of Modern Education and Computer Science*, 10(9), 11-19.
- Pano, T., & Kashef, R. (2020). A complete VADER-based sentiment analysis of Bitcoin (BTC) tweets during the era of COVID-19. *Big Data and Cognitive Computing*, 4(4), 33. Retrieved from <https://doi.org/10.3390/bdcc4040033>
- Park, C. H., & Kim, Y. J. (2013). Intensity of social network use by involvement: A study of young Chinese users. *International Journal of Business and Management*, 8(6), 22-33. Retrieved from <http://dx.doi.org/10.5539/ijbm.v8n6p22>
- Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M. S., Stillwell, D. J., ... Seligman, M. E. (2014). Automatic personality assessment through social media language. *Journal of Personality and Social Psychology*, 108(6), 934-952. Retrieved from <https://doi.org/10.1037/pspp0000020>
- Park, H. W., Park, S., & Chong, M. (2020). Conversations and medical news frames on twitter: infodemiological study on covid-19 in South Korea. *Journal of Medical Internet Research*, 22(5), e18897. Retrieved from <https://doi.org/10.2196/18897>

- Paulauskas, N., & Auskalnis, J. (2017). Analysis of data pre-processing influence on intrusion detection using NSL-KDD dataset. In *Open Conference of Electrical, Electronic and Information Sciences (eStream)* (pp. 1-5). US/Canada: IEEE.
- Paulhus, D. L., & Jones, D. N. (2015). Measures of dark personalities. In *Measures of Personality and Social Psychological Constructs* (pp. 562-594). Academic Press. Retrieved from <https://doi.org/10.1016/b978-0-12-386915-9.00020-6>
- Pavan, K. M., & Prabhu, J. (2018). Role of sentiment classification in sentiment analysis: a survey. *Annals of Library and Information Studies (ALIS)*, 65(3), 196-209.
- Peffer, K., Tuunanen, T., & Niehaves, B. (2018). Design science research genres: introduction to the special issue on exemplars and criteria for applicable design science research. *European Journal of Information Systems*, 27(2), 129-139.
- Peffer, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2020). Design science research process: a model for producing and presenting information systems research (Unpublished thesis). New York: Cornell University.
- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007 [LIWC manual]*. Austin, TX: LIWC.net.
- Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic inquiry and word count: LIWC 2001*. Mahway: Lawrence Erlbaum Associates.
- Pereira, J. F. F. (2017). *Social media text processing and semantic analysis for smart cities*. Retrieved from <https://arxiv.org/pdf/1709.03406.pdf>
- Pirina, I., & Çöltekin, C. (2018). Identifying depression on reddit: The effect of training data. In *Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task* (pp. 9-12). Brussels, Belgium: Association for Computational Linguistics.
- Plank, B., & Hovy, D. (2015). Personality traits on Twitter or how to get 1500 personality tests in a week. *Proceedings of the 6th Workshop on Computational Approaches to*

- Subjectivity, Sentiment and Social Media Analysis* (pp. 92-98). Retrieved from <https://aclanthology.org/W15-2913>
- Pramodh, K. C., & Vijayalata, Y. (2016). Automatic personality recognition of authors using big five factor model. In *IEEE International Conference on Advances in Computer Applications (ICACA)* (pp. 32-37). US/Canada: IEEE.
- Preotiuc-Pietro, D., Carpenter, J., Giorgi, S., & Ungar, L. (2016, October). Studying the Dark Triad of personality through Twitter behavior. In *Proceedings of the 25th ACM international on conference on information and knowledge management* (pp. 761-770). Retrieved from <https://doi.org/10.1145/2983323.2983822>
- Priyanka, C., & Gupta, D. (2015). Fine grained sentiment classification of customer reviews using computational intelligent technique. *International Journal of Engineering and Technology*, 7(4), 1453-1468. Retrieved from citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1067.2107&rep=rep1&type=pdf
- Priyanta, S., Hartati, S., Harjoko, A., & Wardoyo, R. (2016). Comparison of sentence subjectivity classification methods in Indonesian News. *International Journal of Computer Science and Information Security*, 14(5), 407-414.
- Purnamasari, K. K., & Suwardi, I. S. (2018). Rule-based part of speech tagger for Indonesian language. In *IOP Conference Series: Materials Science and Engineering* (p. 407). Rajpura, India: IOP Publishing.
- Putri, T. T., Sitepu, I. Y., Sihombing, M., & Silvi, S. (2019). Analysis and detection of hoax contents in Indonesian News based on machine learning. *Journal of Informatic Pelita Nusantara*, 4(1). Retrieved from <https://doi.org/10.30534/ijatcse/2020/297942020>
- Qin, Z., & Petrounias, I. (2017). A semantic-based framework for fine grained sentiment analysis. In *19th IEEE Conference on Business Informatics* (pp. 295-301). US/Canada: IEEE.
- Raghuwanshi, A. S., & Pawar, S. K. (2017). Polarity classification of Twitter data using sentiment analysis. *International Journal on Recent and Innovation Trends in Computing and Communication*, 5(6), 434-439.

- Rahmani, A., Hosseinzadeh, L., Rostamy-Malkh, M., & Allahviranloo, T. (2016). A new method for defuzzification and ranking of fuzzy numbers based on the statistical beta distribution. *Advances in Fuzzy Systems*, 2016, 6945184. Retrieved from <https://doi.org/10.1155/2016/6945184>
- Ramanathan, V., & Meyyappan, T. (2019). Twitter text mining for sentiment analysis on people's feedback about Oman tourism. In *4th MEC International Conference on Big Data and Smart City (ICBDSC)* (pp. 1-5). US/Canada: IEEE.
- Ramezan, A., Warner, A. T., & Maxwell, E. A. (2019). Evaluation of sampling and cross-validation tuning strategies for regional-scale machine learning classification. *Remote Sensing*, 11(2), 185. Retrieved from <https://doi.org/10.3390/rs11020185>
- Rana, T. A., & Cheah, Y. N. (2017). A two-fold rule-based model for aspect extraction. *Expert Systems with Applications*, 89, 273-285. Retrieved from <https://doi.org/10.1016/j.eswa.2017.07.047>
- Ranganathan, J., Hedge, N., Irudayaraj, A. S., & Tzacheva, A. A. (2018, July). Automatic detection of emotions in Twitter data: a scalable decision tree classification method. In *Proceedings of the Workshop on Opinion Mining, Summarization and Diversification* (pp. 1-10). Retrieved from <https://doi.org/10.1145/3301020.3303751>
- Raschka, S., Patterson, J., & Nolet, C. (2020). Machine learning in python: main developments and technology trends in data science, machine learning, and artificial intelligence. *Information*, 11(4), 193. Retrieved from <https://doi.org/10.3390/info11040193>
- Rathner, E. M., Djamali, J., Terhorst, Y., Schuller, B., Cummins, N., Salamon, G., & Baumeister, H. (2018). How did you like 2017? Detection of language markers of depression and narcissism in personal narratives. *Proc. Interspeech* (pp. 3388-3392). Retrieved from <https://doi.org/10.21437/Interspeech.2018-2040>
- Rawat, R., & Yadav, R. (2020). Big data: big data analysis, issues and challenges and technologies. In *IOP Conference Series: Materials Science and Engineering*, 1022 (pp. 1-7). Rajpura, India: IOP Publishing.

- Reed, P., Bircek, N. I., Osborne, L. A., Viganò, C., & Truzoli, R. (2018). Visual social media use moderates the relationship between initial problematic internet use and later narcissism. *The Open Psychology Journal*, 11(1), 163-170.
- Reis, I., Baron, D., & Shahaf, S. (2018). Probabilistic random forest: a machine learning algorithm for noisy data sets. *The Astronomical Journal*, 157(1), 16.
- Renström, M. (2018). *Fraud detection on unlabeled data with unsupervised machine learning* (Doctoral dissertation). Stockholm: KTH. Retrieved from <https://kth.diva-portal.org/smash/get/diva2:1217521/FULLTEXT01.pdf>
- Resnik, D. B. (2015). *What is ethics in research & why is it important?* Retrieved from <https://www.niehs.nih.gov/research/resources/bioethics/whatis/index.cfm>
- Ridzuan, F., & Zainon, W. M. N. W. (2019). A review on data cleansing methods for big data. *Procedia Computer Science*, 161, 731-738. Retrieved from <https://doi.org/10.1016/j.procs.2019.11.177>
- Roberts, R., Woodman, T., & Sedikides, C. (2018). Pass me the ball: narcissism in performance settings. *International Review of Sport and Exercise Psychology*, 11(1), 190-213. Retrieved from <https://doi.org/10.1080/1750984x.2017.1290815>
- Robinson, J., Bailey, E., Hetrick, S., Paix, S., O'Donnell, M., Cox, G., ... & Skehan, J. (2017). Developing social media-based suicide prevention messages in partnership with young people: exploratory study. *JMIR Mental Health*, 4(4), e7847. Retrieved from <https://doi.org/10.2196/mental.7847>
- Rodriguez, P., González, J., Gonfaus, J. M., & Roca, F. X. (2019). Integrating vision and language in social networks for identifying visual patterns of personality traits. *International Journal of Social Science and Humanity*, 9(1), 6-12.
- Ross, C., Orr, E. S., Sisic, M., Arseneault, J. M., Simmering, M. G., & Orr, R. R. (2009). Personality and motivations associated with Facebook use. *Computers in Human Behavior*, 25(2), 578-586. Retrieved from <https://doi.org/10.1016/j.chb.2008.12.024>

- Ryan, L. (2008). 'I had a sister in England': family-led migration, social networks and Irish nurses. *Journal of Ethnic and Migration Studies*, 34(3), 453-470. Retrieved from <https://doi.org/10.1080/13691830701880293>
- Ryan, T., & Xenos, S. (2011). Who uses Facebook? An investigation into the relationship between the Big Five, shyness, narcissism, loneliness, and Facebook usage. *Computers in Human Behavior*, 27(5), 1658-1664. Retrieved from <https://doi.org/10.1016/j.chb.2011.02.004>
- Saberi, B., & Saad, S. (2017). Sentiment analysis or opinion mining: a review. *International Journal of Advanced Science Engineering Information Technology*, 7, 1660-1667.
- Saeidi, M., Sousa, S. B. D., Milios, E., Zeh, N., & Berton, L. (2019, September). Categorizing online harassment on Twitter. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 283-297). Cham.: Springer. Retrieved from https://doi.org/10.1007/978-3-030-43887-6_22
- Saif, H., Fernandez, M., He, Y., & Alani, H. (2014). *On stopwords, filtering and data sparsity for sentiment analysis of twitter*. Retrieved from <https://doi.org/10.1016/j.ipm.2015.01.005>
- Saldaña, Z. W. (2018). Sentiment analysis for exploratory data analysis. *Programming Historian*. Retrieved from <https://programminghistorian.org/en/lessons/sentiment-analysis>
- Sankaran, A. (2019). *BullyNet: unmasking cyberbullies on social networks* (Unpublished MSc thesis). United States: Boise State University. Retrieved from <https://scholarworks.boisestate.edu/cgi/viewcontent.cgi?article=2758&context=td>
- Sarkar, D. (2016). *Text analytics with python*. New York, NY, USA: Apress. Retrieved from <https://doi.org/10.1007/978-1-4842-2388-8>
- Schade, L. (2015). *Social media and forecasting: what is the predictive power of social media?* (Unpublished thesis). Netherlands: University of Twente.
- Schmidt, A., & Wiegand, M. 2017. A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language*

- Processing for Social Media* (pp. 1-10). New York: ACM. Retrieved from <https://doi.org/10.18653/v1/w17-1101>
- Schoen, H., Gayo-Avello, D., Metaxas, P. T., Mustafaraj, E., Strohmaier, M., & Gloor, P. (2013). The power of prediction with social media. *Internet Research*. Retrieved from <https://doi.org/10.1108/intr-06-2013-0115>
- Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., & Agrawal, M. (2013). Personality, gender, and age in the language of social media: the open-vocabulary approach. *PloS One*, 8(9), e73791. Retrieved from <https://doi.org/10.1371/journal.pone.0073791>
- Schwartz, S. H. (1992). Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In *Advances in Experimental Social Psychology* (vol. 25, pp. 1-65). Academic Press. Retrieved from [https://doi.org/10.1016/s0065-2601\(08\)60281-6](https://doi.org/10.1016/s0065-2601(08)60281-6)
- Schwartz, T. (2020). Where is the culture? Personality as the distributive locus of culture. In *The Making of Psychological Anthropology* (pp. 419-441). California: University of California Press.
- Seal, D., Roy, U. K., & Basak, R. (2020). Sentence-level emotion detection from text based on semantic rules. In *Proceedings of the Information and Communication Technology for Sustainable Development* (pp. 423-430). Switzerland: Springer.
- Seidman, G. (2013). Self-presentation and belonging on Facebook: how personality influences social media use and motivations. *Personality and individual differences*, 54(3), 402-407. Retrieved from <https://doi.org/10.1016/j.paid.2012.10.009>
- Serrano-Guerrero, J., Romero, F. P., & Olivas, J. A. (2021). Fuzzy logic applied to opinion mining: a review. *Knowledge-Based Systems*, 222, 107018. Retrieved from <https://doi.org/10.1016/j.knosys.2021.107018>
- Shahmirzadi, O., Lugowski, A., & Younge, K. (2019, December). Text similarity in vector space models: a comparative study. In *2019 18th IEEE International Conference on Machine Learning and Applications (ICMLA)* (pp. 659-666). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/icmla.2019.00120>

- Shahzad, R. K., & Lavesson, N. (2013). Comparative analysis of voting schemes for ensemble-based malware detection. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 4(1), 98-117.
- Shailaja, K., Seetharamulu, B., & Jabbar, M. A. (2018). Machine learning in healthcare: A review. In *2nd International Conference on Electronics, Communication and Aerospace Technology (ICECA)* (pp. 910-914). US/Canada: IEEE.
- Shapp, A. (2014). *Variation in the use of Twitter hashtags* (Unpublished thesis). New York: New York University.
- Sharifirad, S., Jafarpour, B., & Matwin, S. (2018). Boosting text classification performance on sexist tweets by text augmentation and text generation using a combination of knowledge graphs. In *Proceedings of the 2nd workshop on abusive language online (ALW2)* (pp. 107-114). Brussels, Belgium: Association for Computational Linguistics.
- Sharma, R. C., Hara, K., & Hirayama, H. (2017). A machine learning and cross-validation approach for the discrimination of vegetation physiognomic types using satellite based multispectral and multitemporal data. *Scientifica*, 2017, 9806479. Retrieved from <https://doi.org/10.1155/2017/9806479>
- Sheeba, J. I., & Vivekanandan, K. (2014). A fuzzy logic based on sentiment classification. *International Journal of Data Mining & Knowledge Management Process*, 4(4), 27. Retrieved from <https://doi.org/10.5121/ijdkp.2014.4403>
- Shirsat, V. S., Jagdale, R. S., & Deshmukh, S. N. (2017). Document level sentiment analysis from news articles. In *2017 international conference on computing, Communication, Control and Automation (ICCUBE)* (pp. 1-4). US/Canada: IEEE.
- Shu, X. (2020). *Knowledge discovery in the social sciences: a data mining approach*. California: University of California Press.
- Shynu, P. G., Shayan, H. M., & Chowdhary, C. L. (2020, February). A fuzzy based data perturbation technique for privacy preserved data mining. In *2020 International Conference on Emerging Trends in Information Technology and Engineering (IC-ETITE)* (pp. 1-4). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/ic-etite47903.2020.244>

- Sidorov, G., Velasquez, F., Stamatatos, E., Gelbukh, A., & Chanona-Hernández, L. (2014). Syntactic n-grams as machine learning features for natural language processing. *Expert Systems with Applications*, 41(3), 853-860. Retrieved from <https://doi.org/10.1016/j.eswa.2013.08.015>
- Sim, J., Miller, P., & Swarup, S. (2020). Tweeting the High line life: a social media lens on urban green spaces. *Sustainability*, 12(21), 8895. Retrieved from <https://doi.org/10.3390/su12218895>
- Singh, A. K., & Shashi, M. (2019). Vectorization of text documents for identifying unifiable news articles. *International Journal of Advanced Computer Science and Applications*, 10(7), 305-310.
- Smith, M. A., & Canger, J. M. (2004). Effects of supervisor “Big Five” personality on subordinate attitudes. *Journal of Business and Psychology*, 18(4), 465-481.
- Somerville, T. A. (2015). The effect of social media use on narcissistic behavior. *Journal of Undergraduate Research*, 25. Retrieved from <http://www.mckendree.edu/academics/scholars/somerville-issue-25.pdf>
- Souri, A., Hosseinpour, S., & Rahmani, A. M. (2018). Personality classification based on profiles of social networks’ users and the five-factor model of personality. *Human-Centric Computing and Information Sciences*, 8(1), 1-15. Retrieved from <https://doi.org/10.1186/s13673-018-0147-4>
- Squicciarini, A., Rajtmajer, S., Liu, Y., & Griffin, C. (2015). Identification and characterization of cyberbullying dynamics in an online social network. In *ASONAM* (pp. 280-285). Retrieved from <https://doi.org/10.1145/2808797.2809398>
- Srividhya, V., & Anitha, R. (2010). Evaluating preprocessing techniques in text categorization. *International Journal of Computer Science and Application*, 47(11), 49-51.
- Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Harari, G. M., Schoedel, R., ... Bühner, M. (2020). Personality research and assessment in the era of machine learning. *European Journal of Personality*, 34(5), 613-631.

- Štajner, S., & Yenikent, S. (2020, December). A survey of automatic personality detection from texts. In *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 6284-6295). Retrieved from <https://doi.org/10.18653/v1/2020.coling-main.553>
- Statista. (2018). *Number of monthly active Twitter users worldwide from 1st quarter 2010 to 3rd quarter 2018 (in millions)*. Retrieved from <https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/#0>
- Stein, R., & Swan, A. B. (2019). Evaluating the validity of Myers-Briggs Type Indicator theory: a teaching tool and window into intuitive psychology. *Social and Personality Psychology Compass*, 13(2), e12434. Retrieved from <https://doi.org/10.1111/spc3.12434>
- Stieglitz, S., Mirbabaie, M., Ross, B., & Neuberger, C. (2018). Social media analytics – challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39, 156-168.
- Stone, L. E., Segal, D. L., & Krus, G. C. (2020). Relationships between pathological narcissism and maladaptive personality traits among older adults. *Aging & Mental Health*, 25(5), 930-935.
- Subramaniam, P. R., & Venugopal, C. (2020). Comparison of Mamdani and Sugeno inference methods in calorie burn calculation for activity using treadmill. *Journal of Computational and Theoretical Nanoscience*, 17(4), 1703-1709.
- Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012, December). Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *2012 11th international conference on machine learning and applications* (vol. 2, pp. 386-393). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/icmla.2012.218>
- Swearingen, T., Drevo, W., Cyphers, B., Cuesta-Infante, A., Ross, A., & Veeramachaneni, K. (2017, December). ATM: a distributed, collaborative, scalable system for automated machine learning. In *2017 IEEE international conference on big data (big data)* (pp. 151-162). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/bigdata.2017.8257923>

- Szalma, J., & Weiss, B. (2020). Data-driven classification of dyslexia using eye-movement correlates of natural reading. In *ETRA '20 Short Papers: ACM Symposium on Eye Tracking Research and Applications* (pp. 1-4). Retrieved from <https://doi.org/10.1145/3379156.3391379>
- Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in reddit social media forum. *IEEE Access*, 7, 44883-44893. Retrieved from <https://doi.org/10.1109/access.2019.2909180>
- Tae, K. H., Roh, Y., Oh, Y. H., Kim, H., & Whang, S. E. (2019, June). Data cleaning for accurate, fair, and robust models: a big data-AI integration approach. In *Proceedings of the 3rd International Workshop on Data Management for End-to-End Machine Learning* (pp. 1-4). Retrieved from <https://doi.org/10.1145/3329486.3329493>
- Taleb, I., Dssouli, R., & Serhani, M. A. (2015). Big data pre-processing: a quality framework. In *IEEE International Congress on Big Data* (pp. 191-198). US/Canada: IEEE.
- Tanriverdi, H., & Sağır, S. (2014). Lise Öğrencilerinin Sosyal Ağ Kullanım Amaçlarının Ve Sosyal Ağları Benimseme Düzeylerinin Öğrenci Başarisina Etkisi. *Adıyaman Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 18, 775-822.
- Tohid, R., Wagle, B., Shirzad, S., Diehl, P., Serio, A., Kheirhahan, A., & Kaiser, H. (2018). Asynchronous execution of python code on task-based runtime systems. In *IEEE/ACM 4th International Workshop on Extreme Scale Programming Mode*. IS/Canada: IEEE.
- Uher, J. (2013). Personality psychology: lexical approaches, assessment methods, and trait concepts reveal only half of the story—why it is time for a paradigm shift. *Integrative Psychological and Behavioral Science*, 47(1), 1-55. Retrieved from <https://doi.org/10.1007/s12124-013-9230-6>
- Uysal, A. K., & Gunal, S. (2014). The impact of preprocessing on text classification. *Information Processing & Management*, 50(1), 104-112. Retrieved from <https://doi.org/10.1016/j.ipm.2013.08.006>

- Van der Linden, S., & Rosentha, S. A. (2016). Measuring narcissism with a single question? A replication and extension of the Single-Item Narcissism Scale (SINS). *Personality and Individual Differences*, 90, 238-241.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: a literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25(2), 195-217. Retrieved from <https://doi.org/10.1093/ijpor/eds021>
- Vashishtha, S., & Susan, S. (2019). Fuzzy rule based unsupervised sentiment analysis from social media posts. *Expert Systems with Applications*, 138(12), 112834. Retrieved from <http://dx.doi.org/10.1016/j.eswa.2019.112834>
- Vehkalahti, K., & Everitt, B. S. (2018). *Multivariate analysis for the behavioral sciences*. Boca Raton: CRC Press.
- Venable, J., Pries-Heje, J., & Baskerville, R. (2012). A comprehensive framework for evaluation in design science research. In *International Conference on Design Science Research in Information Systems* (pp. 423-438). Berlin, Heidelberg: Springer.
- Ver Steeg, G. (2017). Unsupervised learning via total correlation explanation. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17)* (pp. 5151-5155). Retrieved from <https://www.ijcai.org/proceedings/2017/0740.pdf>
- Virmani, C., Pillai, A., & Juneja, D. (2017). Clustering in aggregated user profiles across multiple social networks. *International Journal of Electrical and Computer Engineering*, 7(6), 3692. Retrieved from <https://doi.org/10.11591/ijece.v7i6.pp3692-3699>
- Visa, S., Ramsay, B., Ralescu, A. L., & Van Der Knaap, E. (2011). Confusion matrix-based feature selection. *MAICS*, 710, 120-127.
- Vom Brocke, J., Winter, R., Hevner, A., & Maedche, A. (2020). Special issue editorial—accumulation and evolution of design knowledge in design science research: a journey through time and space. *Journal of the Association for Information Systems*, 21(3), 9. Retrieved from <https://doi.org/10.17705/1jais.00611>

- Wan, Y., & Gao, Q. (2015, November). An ensemble sentiment classification system of twitter data for airline services analysis. In *2015 IEEE international conference on data mining workshop (ICDMW)* (pp. 1318-1325). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/icdmw.2015.7>
- Wang, D. (2017). A study of the relationship between narcissism, extraversion, drive for entertainment, and narcissistic behavior on social networking sites. *Computers in Human Behavior*, 66, 138-148. Retrieved from <https://doi.org/10.1016/j.chb.2016.09.036>
- Wang, F., Wang, H., Xu, K., Wu, J., & Jia, X. (2013, July). Characterizing information diffusion in online social networks with linear diffusive model. In *2013 IEEE 33rd international conference on distributed computing systems* (pp. 307-316). US/Canada: IEEE. Retrieved from doi.org/10.1109/ICDCS.2013.14
- Wang, L., & Xia, R. (2017, September). Sentiment lexicon construction with representation learning based on hierarchical sentiment supervision. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 502-510). Retrieved from <https://doi.org/10.18653/v1/d17-1052>
- Wang, X., Gerber, M. S., & Brown, D. E. (2012, April). Automatic crime prediction using events extracted from twitter posts. In *International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction* (pp. 231-238). Berlin, Heidelberg: Springer. Retrieved from https://doi.org/10.1007/978-3-642-29047-3_28
- Wang, Y., Yu, W., Liu, S., & Young, S. D. (2019). The relationship between social media data and crime rates in the United States. *Social Media+ Society*, 5(1). Retrieved from <https://doi.org/10.1177/2F2056305119834585>
- Warner, W., & Hirschberg, J. (2012, June). Detecting hate speech on the world wide web. In *Proceedings of the Second Workshop on Language in Social Media* (pp. 19-26). Retrieved from <https://doi.org/10.5220/0010070912421247>
- Wilges, B., Mateus, G. P., Nassar, S. M., Cislighi, R., & Bastos, R. C. (2016). Fuzzy modeling for multi-label text classification supported by classification algorithms. *Journal of Computer Sciences*, 12(7), 341-349.

- Williams, M. L., Burnap, P., & Sloan, L. (2017). Crime sensing with big data: the affordances and limitations of using open-source communications to estimate crime patterns. *The British Journal of Criminology*, 57, 320-340. Retrieved from <https://doi.org/10.1093/bjc/azw031>
- Wilson, K. S., DeRue, D. S., Matta, F. K., & Howe, M. (2016). Personality similarity in negotiations: Testing the dyadic effects of similarity in interpersonal traits and the use of emotional displays on negotiation outcomes. *Journal of Applied Psychology*, 101(10), 1405-1421.
- Won, D., Steinert-Threlkeld, Z. C., & Joo, J. (2017). Protest activity detection and perceived violence estimation from social media images. In *Proceedings of the 25th ACM International Conference on Multimedia* (pp. 786-794). Retrieved from <https://arxiv.org/pdf/1709.06204.pdf>
- Wood, J., Tan, P., Wang, W., & Arnold, C. (2017). Source-LDA: enhancing probabilistic topic models using prior knowledge sources. In *International Conference on Data Engineering (ICDE)* (pp. 411-422). US/Canada: IEEE.
- Wright, A. G. (2016). On the measure and mismeasure of narcissism: A response to “Measures of narcissism and their relations to DSM-5 pathological traits: a critical reappraisal”. *Assessment*, 23(1), 10-17.
- Wright, W. R., & Chin, D. N. (2014, July). Personality profiling from text: introducing part-of-speech N-grams. In *International Conference on User Modeling, Adaptation, and Personalization* (pp. 243-253). Cham.: Springer. Retrieved from https://doi.org/10.1007/978-3-319-08786-3_21
- Wrzus, C., & Mehl, M. R. (2015). Lab and/or field? Measuring personality processes and their social consequences. *European Journal of Personality*, 29(2), 250-271. Retrieved from <https://doi.org/10.1002/per.1986>
- Wrzus, C., & Roberts, B. W. (2017). Processes of personality development in adulthood: the TESSERA framework. *Personality and Social Psychology Review*, 21(3), 253-277.
- Wu, K., Zhou, M., Lu, X. S., & Huang, L. (2017, October). A fuzzy logic-based text classification method for social media data. In *2017 IEEE International Conference*

on Systems, Man, and Cybernetics (SMC) (pp. 1942-1947). USA/Canada: IEEE.

Retrieved from <https://doi.org/10.1109/smc.2017.8122902>

Xiang, G., Fan, B., Wang, L., Hong, J., Rose, C., 2012. Detecting offensive tweets via topical feature discovery over a large scale twitter corpus, *21st ACM International Conference on Information and Knowledge Management-CIKM '12*. Maui, Hawaii, USA: ACM Press. Retrieved from <https://doi.org/10.1145/2396761.2398556>

Yao, L., Mao, C., & Luo, Y. (2019). Clinical text classification with rule-based features and knowledge-guided convolutional neural networks. *BMC Medical Informatics and Decision Making*, 19(3), 71. Retrieved from <https://doi.org/10.1186/s12911-019-0781-4>

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: lessons from machine learning. *Perspectives on Psychological Science*, 12(6), 1100-1122.

Yazdavar, A. H., Al-Olimat, H. S., Ebrahimi, M. B., Bajaj, G., Banerjee, T., Thirunarayan, T., & Sheth, A. (2017). Semi-supervised approach to monitoring clinical depressive symptoms in social media. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. US/Canada: IEEE.

Yen, W. C., Lin, H. H., Wang, Y. S., Shih, Y. W., & Cheng, K. H. (2019). Factors affecting users' continuance intention of mobile social network service. *The Service Industries Journal*, 39(13-14), 983-1003.

Yo, T., & Sasahara, K. (2017, December). Inference of personal attributes from tweets using machine learning. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 3168-3174). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/bigdata.2017.8258295>

Yuasa, M., Saito, K., & Mukawa, N. (2011). Brain activity when reading sentences and emoticons: an fMRI study of verbal and nonverbal communication. *Electronics and Communications in Japan*, 94(5), 17-24. Retrieved from <https://doi.org/10.1002/ecj.10311>

- Yue, L., Chen, W., Li, X., Zuo, W., & Yin, M. (2019). A survey of sentiment analysis in social media. *Knowledge and Information Systems*, 60(2), 617-663. Retrieved from <https://doi.org/10.1007/s10115-018-1236-4>
- Yüksel, A., Türkmen, Y., Özgür, A., & Altinel, B. (2019). Turkish tweet classification with transformer encoder. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)* (pp. 1380-1387). Retrieved from https://doi.org/10.26615/978-954-452-056-4_158
- Zafar, H., Harle, D., Andonovic, I., & Ashraf, M. (2007). Partial-disjoint multipath routing for wireless ad-hoc networks. In *32nd IEEE Conference on Local Computer Networks (LCN 2007)*. US/Canada: IEEE.
- Zajenkowski, M., & Szymaniak, K. (2021). Narcissism between facets and domains: the relationships between two types of narcissism and aspects of the Big Five. *Current Psychology*, 40(5), 2112-2121. Retrieved from <https://doi.org/10.1007/s12144-019-0147-1>
- Zajenkowski, M., Maciantowicz, O., Szymaniak, K., & Urban, P. (2018). Vulnerable and grandiose narcissism are differentially associated with ability and trait emotional intelligence. *Frontiers in Psychology*, 9, 1606. Retrieved from <https://doi.org/10.3389/fpsyg.2018.01606>
- Zaldumbide, J., & Sinnott, R. O. (2015). Identification and validation of real-time health events through social media. In *IEEE International Conference on Data Science and Data Intensive Systems* (pp. 9-16). US/Canada: IEEE.
- Zeng, D., Chen, H., Lusch, R., & Li, S. H. (2010). Social media analytics and intelligence. *IEEE Intelligent Systems*, 25(6), 13-16. Retrieved from <https://doi.org/10.1109/mis.2010.151>
- Zhang, Y. D., Yang, Z. J., Lu, H. M., Zhou, X. X., Phillips, P., Liu, Q. M., & Wang, S. H. (2016). Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation. *IEEE Access*, 4, 8375-8385.
- Zhao, Y., Qin, B., Liu, T., & Tang, D. (2016). Social sentiment sensor: a visualization system for topic detection and topic sentiment analysis on microblog. *Multimedia Tools and*

Applications, 75(15), 8843-8860. Retrieved from <https://doi.org/10.1007/s11042-014-2184-y>

Zhou, S., Zhao, Y., Rizvi, R., Bian, J., Haynos, A. F., & Zhang, R. (2019, June). Analysis of Twitter to identify topics related to eating disorder symptoms. In *2019 IEEE International Conference on Healthcare Informatics (ICHI)* (pp. 1-4). USA/Canada: IEEE. Retrieved from <https://doi.org/10.1109/ichi.2019.8904863>

Zhou, X., Tao, X., Rahman, M. M., & Zhang, J. (2017). Coupling topic modelling in opinion mining for social media analysis. In *Proceedings of the International Conference on Web Intelligence* (pp. 533-540). Retrieved from <https://doi.org/10.1145/3106426.3106459>

Zhu, Y., Moh, M., & Moh, T. S. (2016). Multi-layer text classification with voting for consumer reviews. In *IEEE International Conference on Big Data* (pp. 1991-1999). US/Canada: IEEE.

Zivanovic, S., Martinez, J., & Verplanke, J. (2020). Capturing and mapping quality of life using Twitter data. *GeoJournal*, 85(1), 237-255.

Zucco, C., Calabrese, B., Agapito, G., Guzzi, P. H., & Cannataro, M. (2020). Sentiment analysis for mining texts and social networks data: methods and tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(1), e1333. Retrieved from <https://doi.org/10.1002/widm.1333>

APPENDICES

Appendix A: Ethical Clearance letter



02 September 2021

Mr Japheth Kiplang'at Mursi (216068813)
School of Management, IT & Governance
Pietermaritzburg Campus

Dear Mr Mursi,

Protocol reference number: HSS/2067/018D

Project title: Designing a conceptual model for predicting narcissistic traits among South Africans through Social Media

Amended title: Fuzzy-based machine learning for predicting narcissistic traits among Twitter users

Approval Notification – Amendment Application

This letter serves to notify you that your application and request for an amendment received on 26 August 2021 has now been approved as follows:

- Change in title

Any alterations to the approved research protocol i.e. Questionnaire/Interview Schedule, Informed Consent Form; Title of the Project, Location of the Study must be reviewed and approved through an amendment /modification prior to its implementation. In case you have further queries, please quote the above reference number.

PLEASE NOTE: Research data should be securely stored in the discipline/department for a period of 5 years.

All research conducted during the COVID-19 period must adhere to the national and UKZN guidelines.

Best wishes for the successful completion of your research protocol.

Yours faithfully

.....
Professor Dipane Hlalele (Chair)

/dd

Cc Supervisor: Dr Prabhakar Rontala Subramaniam and Professor Irene Govender
cc Academic Leader Research: Professor Isabel Martins
cc School Administrator: Ms Debbie Cunynghame

Humanities & Social Sciences Research Ethics Committee
UKZN Research Ethics Office Westville Campus, Govan Mbeki Building
Postal Address: Private Bag X54001, Durban 4000
Tel: +27 31 260 8350 / 4557 / 3587

14 November 2018

Mr Japheth Kiplang'at Mursi (216068813)
School of Management, IT & Governance
Pietermaritzburg Campus

Dear Mr Mursi,

Protocol reference number: HSS/2067/018D

Project title: Designing a conceptual model for predicting narcissistic traits among South Africans through Social Media

Approval Notification – Expedited Application

In response to your application received on 09 November 2018, the Humanities & Social Sciences Research Ethics Committee has considered the abovementioned application and the protocol has been granted **FULL APPROVAL**.

Any alteration/s to the approved research protocol i.e. Questionnaire/Interview Schedule, Informed Consent Form, Title of the Project, Location of the Study, Research Approach and Methods must be reviewed and approved through the amendment/modification prior to its implementation. In case you have further queries, please quote the above reference number. PLEASE NOTE: Research data should be securely stored in the discipline/department for a period of 5 years.

The ethical clearance certificate is only valid for a period of 3 years from the date of issue. Thereafter Recertification must be applied for on an annual basis.

I take this opportunity of wishing you everything of the best with your study.

Yours faithfully



Professor Shenuka Singh (Chair)

/ms

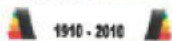
Cc Supervisor: Dr Prabhakar Rontala Subramaniam and Professor Irene Govender
cc Academic Leader Research: Professor Isabel Martins
cc School Administrator: Ms Debbie Cunynghame

Humanities & Social Sciences Research Ethics Committee
Professor Shenuka Singh (Chair) / Dr Shamila Naidoo (Deputy Chair)
Westville Campus, Govan Mbeki Building






Postal Address: Private Bag X54001, Durban 4000

Telephone: +27 (0) 31 260 3587/8350/4557 Facsimile: +27 (0) 31 260 4609 Email: simbap@ukzn.ac.za / snymartini@ukzn.ac.za / mohunod@ukzn.ac.za

Website: www.ukzn.ac.za



100 YEARS OF ACADEMIC EXCELLENCE

Founding Campuses:  Edgewood  Howard College  Medical School  Pietermaritzburg  Westville

Appendix B: Twitter Gatekeeper's letter (developer account approval)

Twitter developer account application [ref:00DA0000000K0A8.5004A00001Svv33:ref] Inbox x



developer-accounts <developer-accounts@twitter.com>

to me ▾

9:21 PM (10 hours ago) ☆



Your Twitter developer account application has been approved!

Thanks for applying for access. We've completed our review of your application, and are excited to share that your request has been approved.

Sign in to your [developer account](#) to get started.

Thanks for building on Twitter!