Applications of Machine Learning in Mental Healthcare

Am Fachbereich Rechts- und Wirtschaftswissenschaften
der Technischen Universität Darmstadt

Dissertation

vorgelegt von

Elena Davcheva

geboren am 22.05.1990 in Kavadarci, Nordmazedonien

zur Erlangung des akademischen Grades
Doctor rerum politicarum (Dr. rer. pol.)

Erstgutachter: Prof. Dr. Alexander Benlian
Zweitgutachter: Prof. Dr. Peter Buxmann

Darmstadt 2021

Abstract

The global rise in individuals seeking treatment for mental health issues has presented a modern medical challenge for existing healthcare support structures. Currently, the number of practicing professionals in this field falls short of meeting rising demand; what is more, both healthcare systems as well as those seeking treatment face related challenges within various aspects such as high service costs, long waiting times for access to treatment, and long procedures to establish suitable treatments. With expectations of the demand trend to continue rising in upcoming years, it is clear that an innovative approach to this challenge is acutely needed.

On this backdrop, policymakers, practitioners, and affected individuals alike are turning to information technology as a possible aid in a global effort for better mental health. Digitalizing mental healthcare solutions holds a promise to amplify the reach of services to millions of affected people, to significantly cut costs (for users and service-providers alike), and to speed up access to treatments - primarily via the use of machine learning.

While there is great potential in information technology (IT) applications for mental health, and national and international initiatives to develop and test solutions are well underway, applications of machine learning within mental health present a nascent field within research as well as in practice. There is a pressing need to experiment with and evaluate the use of various possible digital approaches in terms of type and magnitude of effect on end-users, precision of data collection and data interpretation, and most of all – reliability of predictive data-driven solutions and treatments tailored to profiles of individual users. Furthermore, the digitalization of mental healthcare presents a cultural shift for medical professionals and machine learning practitioners, and synergies in such a collaboration have not yet been applied or explored, as pointed out in a number of studies. There are several milestones to reach in the adoption of psychological practices and their proper translation to machine learning algorithms, especially in the area of automated symptom detection, automating and personalizing digital treatments for users, as well as conclusive and clear understanding of how potential end-users would interact with and adopt such digital solutions.

This thesis addresses the outlined research gap by presenting three separate studies carried out in and connecting the domains of IT and mental healthcare by examining the interconnections of psychology, social media and machine learning. The included studies provide a comprehensive view on the current state of mental healthcare digitalization – from how, why,

and to what extent digital tools are used, to designing, testing, and evaluating models that solve specific tasks as part of treatment procedures.

Article 1 presents a longitudinal study of user posts on online mental health forums with the goal to map out the effects from forum participation based on user post content as well as the different role a user could assume. A regression and sentiment analysis on post content shows that a measurable effect of participation can be observed, and the magnitude and direction of this effect can furthermore be broken down per user role and per mental condition.

Article 2 attempts to map out successful and unsuccessful treatments based on user experiences shared in the form of online forum posts. The experiences analyzed are further segmented per user condition to arrive at clear recommendations for and against possible treatments depending on individual diagnoses.

Article 3 tests an implementation of a sequence-to-sequence recurrent bidirectional neural network model for symptom and conditions classification based on forum posts. The data used to train the neural network has been preprocessed and labelled according to input from mental health professionals. The entire procedure presents a novel classification approach where conditions are classified based on symptoms, allowing the model to capture possible comorbidities and generally explore a classification in much deeper detail than attempted before.

Thus, the thesis puts together a complete examination of digital mental health tools, providing a method to measure the effect of digital mental health tools on users, demonstrating how user-shared content can be used to harvest treatments based on individual conditions, and constructing models where this content can be used to automate certain processes such as symptom classification and tracking. This thesis contributes to the existing body of research in IS, precision psychiatry, as well as computational linguistics, by addressing gaps regarding digital tool applicability for mental healthcare, providing new avenues to harvest and utilize mental health information from text, and advancing the possibilities of automated diagnostics.

Zusammenfassung

Die zunehmende Nachfrage an psychotherapeutischen Behandlungsmöglichkeiten stellt das Gesundheitswesen vor eine Herausforderung: Derzeit ist die Kapazität an psychotherapeutischen Fachkräften ungenügend, um den Bedarf an Therapien zu decken. Darüber hinaus stehen sowohl das Gesundheitssystem als auch die Hilfesuchenden vor vielen Herausforderungen, beispielsweise durch hohe Kosten, lange Wartezeiten und viele Testverfahren, um die richtige Diagnose und dementsprechende Behandlung zu bestimmen. Der Bedarf an Therapien wird in den kommenden Jahren weiter steigen. Aus diesem Grund ist ein innovativer Ansatz erforderlich.

Politische Entscheidungsträger, Therapeuten, und Betroffene wenden sich an Informationstechnologien als möglicher Lösungsansatz zur Verbesserung der mentalen Gesundheit. Im Zuge der Digitalisierung kann die Reichweite von Therapiemöglichkeiten erhöht und Kosten erheblich gesenkt werden. Des Weiteren wird der Zugang zu Behandlungen erheblich beschleunigt. Hier spielt vor allem der Einsatz von maschinellem Lernen eine wichtige Rolle.

IT-Anwendungen haben ein großes Potenzial zur Behandlung von mentalen Problemen, und nationale sowie internationale Initiativen zur Entwicklung und Untersuchung von Lösungen werden bereits durchgeführt. Jedoch handelt es sich bei dieser Art von IT- Anwendungen um neues Anwendungsgebiet der IT-Forschung, dessen Forschung noch nicht sehr ausgereift ist. Daher ist es notwendig, mit verschiedenen digitalen Ansätzen zu experimentieren und ihre Rezeption und Wirkung auf Benutzer zu analysieren. Darüber hinaus muss die Zuverlässigkeit dieser Anwendungen untersucht werden, unter der Berücksichtigung, dass ihre Hilfeleistungen auf die Profile einzelner Benutzer zugeschnitten sind. Allgemein stellt die Digitalisierung von psychotherapeutischen Behandlungsmöglichkeiten einen kulturellen Wandel für Therapeuten und Nutzer des maschinellen Lernens dar. Die langfristigen Auswirkungen auf Therapeuten und Nutzer sind noch nicht abschätzbar und bedürfen weiterer Forschung und Studien. Um jedoch die Durchführbarkeit von digitalen Anwendungen im Bereich für mentale Gesundheit realisierbar zu machen, gibt es bereits einige Ansätze, um psychologische Praktiken in Algorithmen für maschinelles Lernen zu übersetzen, beispielsweise die automatisierte Symptomerkennung, die Automatisierung und Personalisierung digitaler Behandlungen sowie eine eindeutige und klare Einsicht, wie potenzielle Benutzer digitale Hilfeleistungen rezipieren würden.

Diese Arbeit setzt sich zum Ziel, die skizzierte Forschungslücke zu füllen, in dem sie drei separaten Studien analysiert, die in den Bereichen IT und geistige Gesundheit durchgeführt wurden, und die Zusammenhänge von Psychologie, sozialen Medien und maschinellem Lernen untersuchen. Die eingeschlossenen Studien bieten einen umfassenden Überblick über den aktuellen Stand der Digitalisierung in der Psychotherapie, vor dem Hintergrund auf welche Weise und in welchem Umfang digitale Tools verwendet werden, bis zur deren Entwicklung, Untersuchung und Bewertung von Modellen, die spezifische Aufgaben im Rahmen von bestimmten Behandlungsverfahren lösen.

Artikel 1 enthält eine Längsschnittstudie von Benutzerbeiträgen in Online-Foren zum Thema geistige Gesundheit, um deren Auswirkungen auf Benutzer zu ermitteln, basierend auf dem Inhalt von Beiträgen sowie unterschiedlichen Rollen, die Benutzer einnehmen können. Eine Regressions- und Stimmungsanalyse zeigt einen messbaren Effekt in der Teilnahme in Online-Foren; das Ausmaß und die Wirkung des Effekts können außerdem in Benutzerrollen und psychischen Zustand aufgeteilt werden.

Artikel 2 untersucht erfolgreiche und erfolglose Behandlungen auf Basis von Erfahrungen der Benutzer in den Foren. Die analysierten Erfahrungen werden weiter nach den Bedingungen der Benutzer segmentiert, um klare Empfehlungen für und gegen mögliche Behandlungen je nach individueller Diagnose zu erhalten.

Artikel 3 testet ein Neuronales Netzwerk, welches Symptome und Zustände basierend auf Forenbeiträgen klassifiziert. Die Trainingsdaten des neuronalen Netzwerks wurden von Psychologen vorverarbeitet. Das gesamte Verfahren stellt einen neuartigen Klassifizierungsansatz vor, bei dem Zustände anhand von Symptomen klassifiziert werden, damit das Modell mögliche Begleiterkrankungen erfassen kann und eine detailliertere Klassifizierung erstellen kann.

Insgesamt stellt die Arbeit eine vollständige Untersuchung digitaler Tools dar, die für die geistige Gesundheit von Relevanz sind. Die Studien beschreiben eine Methode zur Messung der Auswirkung digitaler Tools für geistige Gesundheit auf Benutzer. Es wird gezeigt, wie Benutzerinhalt verwendet werden kann, um Behandlungen basierend auf individuellen Bedingungen anzupassen und wie Modelle diese Inhalte verwenden, um bestimmte Prozesse zu automatisieren, beispielsweise die Klassifizierung und Analyse von Symptomen. Diese Arbeit trägt zur bestehenden Forschung im Bereich IS bei, indem sie Lücken in Bezug auf die Anwendbarkeit digitaler Tools für geistige Gesundheit anspricht, neue Wege zur

Datenerhebung und Nutzung von Informationen für geistige Gesundheit zeigt und die Möglichkeiten der automatisierten Diagnostik erweitert.

Table of Contents

Abstract ..... **I**

Zusammenfassung ..... **III**

Table of Contents ..... **VI**

List of Tables ..... **X**

List of Figures ..... **XI**

List of Abbreviations ..... **XII**

Chapter 1: Introduction **1**

1.1 Motivation and Research Question 1

1.1.1 Current State of Mental Healthcare Systems 1

1.1.2 Research in IS and Related Fields 3

1.2 Thesis Structure and Synopses 5

Chapter 2: Research Context **9**

2.1 Digitalization Efforts in Mental Healthcare 9

2.1.1 Government Initiatives 11

2.1.2 Fitting IT in the Mental Healthcare Ecosystem 11

2.2 Online Forums as a Digital Mental Health Platform 14

2.2.1 General Context of Using Social Media 14

2.2.2 Online Mental Health Forums 15

2.3 NLP Applications Within Mental Health Forums 16

2.3.1 Data Sources 17

2.3.2 Data Labeling 19

2.3.3 Applications 20

2.3.4 Data Analysis Techniques 21

2.4 Positioning of the Thesis 25

Chapter 3: Mapping User Roles and Dynamics in Mental Health Forums 27

3.1 Introduction 28

3.2 Conceptual Background 29

3.2.1 Online Mental Health Forums 29

3.2.2 Sentiment Analysis 30

3.3 Research Propositions 31

3.4 Methodology 33

3.5 Results 34

3.5.1 Thread Sentiment Progression of Posts by All Users 35

3.5.2 Sentiment Progression of Posts by Original Posters 36

3.5.3 Sentiment Progression of Forum-wide Posts by All Users 38

3.5.4 User Roles Within Forums 39

3.6 Discussion and Contributions 40

3.7 Limitations and Directions for Future Research 42

Chapter 4: Mining Experiences from Mental Health Forums 43

4.1 Introduction 43

4.2 Motivation and Design 45

4.3 Literature Review 45

4.4 Dataset and Methodology 47

4.5 Results 50

4.6 Method Evaluation 54

4.7 Conclusion 55

Chapter 5: Automating Symptom and Condition Classification with Neural Networks 57

5.1 Introduction 57

5.2 Background 60

5.2.1 Forums as a Mental Health Tool 60

5.2.2 Use of Machine Learning in Mental Health 61

5.2.3 Performance Comparison 63

5.2.4 Research Gap 64

5.3 Methodology 65

5.3.1 Data Preparation and Labelling 65

5.3.2 Symptom Classification Using a Neural Network 68

5.3.3 Condition Classification 70

5.4 Results 70

5.4.1 Symptom Classification 72

5.4.2 Condition Classification 76

5.5 Discussion 77

5.5.1 Contribution to IS in Healthcare 78

5.5.2 Practical Implications 79

5.5.3 Limitations and Future Work 79

5.6 Conclusion 80

Chapter 6: Thesis Conclusion and Contributions 81

6.1 Contributions to IS in Healthcare 81

6.2 Practical Contributions 83

6.3 Limitations and Directions for Future Research 84

References 85

Eidesstattliche Erklärung 102

List of Tables

Table 3-1 Regression results and p-values for analyzed mental conditions 35

Table 3-2 Average sentiment per user role across conditions 40

Table 4-1 Stanford CoreNLP dependency relations used in this study 49

Table 4-2 Average sentiment per concept across mental health forums 51

Table 4-3 Correlation Analysis 55

Table 5-1 Models performance comparison 64

Table 5-2 Symptoms derived from DSM-5 and LDA 67

Table 5-3 Annotation example 68

Table 5-4 Symptom classification model comparison 71

Table 5-5 Evaluation metrics for conditions and symptoms 72

Table 5-6 Symptom classification results for one user 76

List of Figures

Figure 2-1 Natural language processing workflow 21

Figure 2-2 High-level representation of a neural network with two hidden layers 23

Figure 2-3 Overview of the examined interactions in research articles 25

Figure 3-1 Research process and propositions 33

Figure 3-2 Average post sentiment score progression within threads 36

Figure 3-3 Average sentiment of OP posts within threads 37

Figure 3-4 Average post sentiment score per user post order within forums 39

Figure 3-5 Distribution of user roles across conditions 40

Figure 4-1 Analysis procedure 48

Figure 4-2 Sentiment scores 52

Figure 4-3 Precision and recall formulas 55

Figure 5-1 Data labelling process 65

Figure 5-2 Condition classification process 70

Figure 5-3 Depression evaluation metrics 73

Figure 5-4 Schizophrenia evaluation metrics 74

Figure 5-5 ADHD evaluation metrics 75

List of Abbreviations

ADAA	Anxiety and Depression Association of America
ADHD	Attention Deficit Hyperactivity Disorder
AMT	Amazon Mechanical Turk
ANN	Artificial Neural Networks
API	Application Programming Interface
BD	Bipolar Disorder
BPD	Borderline personality disorder
CBT	Cognitive Behavioral Therapy
IS	Information Systems
IT	Information Technology
LDA	Latent Dirichlet Allocation
ML	Machine Learning
NHS	National Health Service
NLP	Natural Language Programming
NLTK	Natural Language Toolkit
DSM-IV	Diagnostic and Statistical Manual of Mental Disorders
OCD	Obsessive-Compulsive Disorder
OP	Original Poster
OECD	Organization for Economic Development
PHQ-9	Personal Health Questionnaire

PTSD	Post-Traumatic Stress Disorder
SVM	Support Vector Machine
WHO	World Health Organization

Chapter 1: Introduction

1.1 Motivation and Research Question

1.1.1 Current State of Mental Healthcare Systems

Mental illness is defined as “the loss of mental health due to a mental disorder”, where mental disorders are clinical diagnoses based on formal psychiatric rules and procedures for classification, such as depression, anxiety, trauma, or substance abuse (OECD, 2018). Globally, mental illness affects one in three individuals at some point in their lives. Mental illnesses affect the thinking, perception, mood, and behavior of a person to the point that they affect proper day-to-day functioning. Affected persons may experience depression, reduced concentration, withdrawal from social life, fatigue, and substance abuse among many other symptoms. A prompt and correct diagnosis and timely recognition of symptoms is essential for a successful recovery process. A recovery is usually a long and complicated process, accompanied by potential relapses and episodes, for which constant monitoring is required. A mental health diagnosis can often be incorrect or delayed, thus complicating a possible recovery. Currently, relief for those who suffer from any kind of mental disorder can only be obtained from healthcare systems whose coverage is limited, where affected individuals rely on qualified professionals and one-on-one therapy.

In recent years, the world is facing a crisis in mental healthcare. On a global scale, demand for mental health services is rising rapidly, while the available number of support specialists is stagnating. Annually, one out of six citizens in the European Union (EU) suffers from a mental health disorder (OECD, 2018). Demand for mental healthcare services has seen a sharp increase in countries such as the US, UK, EU and beyond, and it is projected to steadily rise in the near future (Olfson, 2016). Suffering from a disorder can lead to affected individuals facing a declining quality of life, prolonged unemployment periods, and a severe decline in productivity. However, the number of formally trained professionals in the field is not nearly enough to cover these needs. If this trend continues, it is expected that by 2030 the global leading cause of disease burden will be depression (World Health Organization, 2011). This puts unmanageable burden onto government healthcare systems, which are not ready for the sharp increase in requests for treatment. There is a significant global gap between demand for mental health treatments and the ability to provision for treatments. For example, in

middle-income and low-income countries around the world, up to 85% of people with severe mental health issues have no access to treatment. For high-income countries this number sits at between 35% and 50% (World Health Organization, 2011).

Already, national costs for mental healthcare in many developed countries have ballooned. The OECD puts a 2018 estimate of the cost of mental health problems in Europe at 4.1% of annual GDP across the 28 EU member states, or €600 billion, and at 4.8% of annual GDP for Germany (OECD, 2018). Individuals seeking professional medical help face additional challenges in the form of financial healthcare costs, long waiting times to gain access to necessary resources and specialists (Olfson, 2016), as well as ever-present societal stigma and discrimination.

Faced with these mounting challenges, global institutions such as the World Economic Forum are turning to digital technologies and their potential to address the growing crisis (Doraiswamy et al. 2019). Technology can significantly cut the costs of providing mental healthcare services – both for government bodies as well as for individuals seeking help. The digitalization of mental health solutions also promises to considerably cut the time it takes for an individual to receive the right help for their individual case, while simultaneously preserving the anonymity of individuals.

Many government bodies within the EU, UK, and the US have started initiatives to explore the possible use and implementation of digital tools in the process to improve mental health support services. These initiatives often have the goal to act as a proof-of-concept for various novel technologies. IT has shown potential to create tools for automated condition diagnostics, remote chatbot support, and immediate interventions in emergency situations (Fortuna et al. 2019). Automating these tools has the potential to significantly decrease the cost of national mental healthcare, as well as the financial burden on individuals seeking help. These tools can be available around the clock regardless of location, thus maximizing reach to affected individuals.

The combination of availability of data and sophisticated machine learning models make it possible for this kind of automation in healthcare. In particular, text data in the form of testimonials and comments from affected individuals can be used to understand how individuals express themselves over particular symptoms, and to train models to automate many tasks that currently can be done only by trained medical professionals. For example,

condition diagnosis and symptom triaging are two areas of the treatment and therapy process that are already being integrated as automated processes in many innovative digital mental health services.

1.1.2 Research in IS and Related Fields

Online mental health forums have been the object of considerable information systems (IS) research as they contain personal and emotionally rich information which can be used for machine learning development, as well as to examine user behavior on digital mental health tools, and the impact of such tools on individual psychological well-being (Balani & De Choudhury, 2015). In the past, forum posts have been used to train machine learning models for triaging (scoring the severity of a person's condition), as well as mood and affect detection (determining the presence of specific emotions such as fear or anger) (Calvo et. al., 2017). IS literature has also looked into quantifying signals from social media with machine learning, such as sentiment scores or choice of vocabulary, as well as topics discussed (Coppersmith et. al., 2015a). In addition, first steps have been made into automating diagnostics, where recently neural networks have significantly boosted model performance and reliability (Guntuku et. al., 2017).

Despite the noticeable progress, there are open questions in the direction of integrating automated digital tools in mental healthcare. The first and most fundamental task is the identification of methods to capture the magnitude and type (positive or negative) of effects of digital tools for mental healthcare. Few studies evaluate the success of participating in digital mental health programs, mostly by using exit interviews with a small sample of participants, generally covering only depression and anxiety (Naslund et al., 2017). Questions persist on whether digital interventions have the capacity to provide treatment options that can be similarly useful to users as face-to-face therapy (Andersson et. al., 2018). Furthermore, there is sparse evidence that digital interventions would not cause unwanted side-effects, as IS researchers show that digital tools can often lead to negative consequences in work (Benlian, 2020) and home domains (Benlian et. al., 2019). Thus, digital interventions in mental health remain to be validated as favorable and sufficiently effective.

In a similar vein, literature is sparse on reports regarding digital interventions that are unsuccessful for certain groups of individuals. Studies rarely present comprehensive evaluations of the individual psychological outcome of such interventions (Lattie et al., 2019).

Therefore, IS literature lacks an integrated investigation into digital and automated tools to identify the type of and delivery for treatments suitable for particular user segments (Henson et. al., 2019). Given the complexity of mental health conditions and the individual nuances caused by specific symptoms and their varying severity, it is also necessary to examine the prospect of adjusting digital mental health services to individual needs.

In addition, past reviews of the field emphasize the need for algorithms with high generalizability and high specificity - in other words, algorithms that are widely applicable, and able to produce more accurate and simultaneously more detailed diagnoses (Dwyer et. al., 2018). This is in contrast to current efforts which mostly focus on single-class diagnoses. This would further enable the closing of a related research gap: enabling the selection and personalization of the correct treatment on an individual basis. This is needed because currently treatments must continually be adjusted for each individual based on a trial-and-error approach – a financially costly as well as time-consuming process (Wunderink et. al., 2009).

Research Question

This thesis is positioned at the interface of an interdisciplinary domain encompassing IS, psychology, and computational linguistics which in recent years has come to be known as precision psychiatry (Fernandes et al., 2017). The field has the goal to incorporate recent technological advances such as data collection and analytics and enable evidence-based psychological practice tailored to individual patients (Bzdok & Meyer-Lindenberg, 2018). This thesis contributes to the ongoing efforts in academia and practice of moving towards digitalizing mental healthcare – specifically in the areas of online peer support, affect detection, and automated diagnostics by conducting three consecutive studies to examine the usage pattern of mental health forums, the ability to automatically derive solutions from them, and to create automated diagnostics tools based on user posts content.

RQ: Can mental health forums be used to identify disorders and their potential solutions based on user input and interaction to different topics?

In this context, first an investigation is conducted on whether forum participation of different types (lurking, commenting, asking questions, etc.) at all affects individual forum participants; we further examine whether this affect is positive or negative in nature, in other words, does

forum participation improve or worsen user affect. Based on these conversations, we examine the possibility of extracting solutions per condition based on success and failure reports by users. Mining unstructured conversations of millions of users offers the benefit of learning from the trials and errors of many individuals worldwide over many conditions. This can generate many new features for a digital mental healthcare platform, such as automatic recommended courses of action on an individual user basis (Scholz et. al., 2017). Finally, the potential to automatically diagnose users with a mental health condition based on what they share about their experience is also tested via a state-of-the-art neural network. This completes the examination of the potential of a digital platform such as a forum to be an effective tool in addressing the global crisis in mental healthcare. By leveraging an automated classifier model, a forum can give clarity and guidance to those who, for whatever reason (financial, lack of resources), do not have access to formal medical care.

1.2 Thesis Structure and Synopses

This thesis contains six chapters. After the introduction, the second chapter presents the research context in which this thesis was conducted, providing an overview of machine learning techniques, online mental health forums, and the digitalization of mental health in general. Chapters three, four, and five present three research papers describing the three studies conducted as part of this thesis. Chapter six concludes the findings of the thesis and presents an outlook for future research in this domain.

Article 1

User Dynamics in Mental Health Forums – A Sentiment Analysis Perspective

Online forums are a widely used platform for individuals with mental health issues to connect anonymously. They offer further advantages such as location and time independence. Hundreds of thousands of active users give and seek advice, share experiences, making forums excellent data sources with a lot of text data rich in emotional expressions. Forums have previously been the subject of research within the IS field, in terms of the type of information users are willing to share publicly, or the factors that attract users to these platforms. Less attention has been paid to the effects of forum participation on users. The goal of this study is to explore whether forum participation has any effects on the expressed mood of participants via the application of machine learning, specifically sentiment analysis. What

is more, we explore the different roles users can adopt in a forum (original poster, commenter, lurker) and if and how the role influences the sentiment progression of individual user posts. This study confirms that forum participation does affect the expressed sentiment of users, mostly positively, however the polarity of sentiment as well as the magnitude in change are indeed affected by the combination of type of condition users discuss, as well as the role they assume. The study shows that depression and anxiety forum users derive the most benefits from forum participation, while attention deficit hyperactivity disorder (ADHD) users may derive no participation benefit. Specific users such as those in an autism sub-forum only mark improvement when participating in a specific role. The findings from this article point to the benefit that forum users stand to derive from increased customization supported by automated processing of their posts. Forum participation can have a more beneficial impact when users are encouraged to participate in a favorable way regarding their particular needs. In a more general regard, having the necessary tools and procedures in place to gauge the effect of digital treatments is an essential requirement that will drive the future direction of this research domain.

Article 2

Text Mining Mental Health Forums – Learning from User Experiences

With the amount of information shared directly by users, online mental health forums represent a great source to learn directly from affected individuals regarding effective and ineffective interventions for specific mental health conditions. So far research efforts within the IS and related fields have focused more on ways of forum use, as well as individual behavior. Forums and social media in general have not been harvested to analyze what users actually say in connection to their condition. Article 2 presents an aspect-based sentiment analysis on mental health forum posts with the goal of examining user sentiment towards various mental health treatments. We mine treatment outcomes from a large dataset spanning thousands of user experiences, thus demonstrating how social media can be harnessed and incorporated within mental health research, and embedded in the already existing body of research in IS regarding social media mining. The study introduces a novelty in the research domain by using a very large dataset representing many users across twelve different conditions, providing a rich opportunity for analysis and comparison. This study summarizes the different mechanisms affected individuals employ in order to tackle mental problems,

particularly users' reports on treatments tried in the past, and results achieved. Seven different coping mechanisms that are regularly mentioned on forums have been examined: family, pets, meditation, medications, sports, therapy, and medical professionals. These aspects have been evaluated with sentiment analysis for twelve different conditions. The analysis is conducted based on several linguistic aspects and semantic interrelations such as part-of-speech information as well as dependency relations. This comparison sheds light on the efficacy of each aspect per condition, showing which conditions most benefit from popular treatments, and which conditions need more innovative treatments, outside of the conventional choices available. The study demonstrates one way in which automated text processing of publicly available data can be applied within mental health research, as this field has not yet benefited from the digitalization efforts that other fields in medicine already have.

Article 3

Classifying Mental Health Conditions Via Symptom Identification – A Novel Deep Learning Approach

Ever since machine learning and especially classification algorithms became widely used research tools in IS, many studies have been conducted to detect signs of depression or suicidal ideation on social media platforms such as Twitter. While these efforts have garnered attention both in research and practice, automated classification of mental health conditions is still an underexplored topic. The third article in this thesis presents a practice-oriented attempt at automated diagnosis of mental conditions on the basis of previously classifying underlying symptoms using a neural network. The study presents a three-step data labelling process that combines clinical psychiatric diagnostic rules with unsupervised machine learning and puts labeling results under review by a practicing psychologist. The diagnostic algorithm is a bidirectional recurrent neural network, which performs multiclass classification and introduces a novelty in that condition classification is based on a prior classification of symptoms. The condition classification derived afterwards is based on the neural network's symptom class probability output. The tested method improves upon the performance of methods reported in the past in the area of diagnostics. This approach also allows the diagnosis of a comorbid, or secondary, condition – a novelty in IS literature for our domain. It is also one of few studies to use unstructured text as feature input for model training. What is more, we show that demographic information on a user as well as external knowledge bases

are not necessary to build a robust classification model – two features that have been almost ubiquitous in previous academic writing on classification models for mental health. We further test and show that the data source as well as the jargon associated with that source are two features that greatly influence the classifier performance. The study makes a compelling case for the potential of IT in general and machine learning specifically to solve the lack of mental healthcare resources by automating services such as pre-screening or diagnostics.

Chapter 2: Research Context

In this chapter, the thesis is positioned in an academic and practical context within the interdisciplinary intersection of IS, precision psychiatry, and computational linguistics. Within IS, the use of social media for mental healthcare, especially for information sharing and seeking (De Choudhury et. al., 2014), and peer support (Ali et. al., 2015), has recently enabled the collection of large datasets on a multitude of topics that would become the foundation for machine learning modelling (Smailhodzic et. al., 2016). IS literature has framed social media technologies as "disruptive" when applied to mental health, due to the fundamental change in treatment paradigms that it enables. Psychology researchers have already provided initial evidence that internet-based interventions can have positive effect on users (Kuester et. al., 2016). The analysis of social media posts about personal mental health not only allow to deduct a user's mental health state, but also to gauge their information needs. The latest research in computational linguistics and precision psychology points out that there is much space to explore user input data with machine learning in terms of diagnosing and symptom identification (Chen et. al., 2018). A further look is needed into the effects of digital intervention (Naslund et al., 2019), as well as more detailed user profiling for improved diagnostics and treatment choices (Friedl et al., 2020). Therefore, academic work in this area will be an important driver forward for the further digitalization of mental healthcare.

From a practical perspective, machine learning holds the potential to enhance mental healthcare on an individual level and widen the accessibility of service. This is made possible through the availability of data, the possibility to empower consumers through self-paced and electable treatment components, and to provide measurable outcomes. This thesis reflects current efforts in rolling out digital innovations in partnerships with social institutions on a large scale.

2.1 Digitalization Efforts in Mental Healthcare

The pervasiveness of the internet and the widespread use of social media are likely to transform the way mental healthcare is researched, treated, and managed in the future. In the EU, 89% of the population has access to internet (Eurostat, 2017), and 43% make daily use of some of the most popular social media platforms (Media Use in the European Union, 2018). The intensive daily use of these technologies enables researchers and practitioners to amass

data from many individual users, which can then be utilized to train machine learning-based models to generate personalized solutions to the individual based on their condition. Large datasets allow algorithms to discover solutions tailored to various contexts, and eventually to build reliable digital tools for diagnostics, tracking, immediate help, among other issues. For example, the availability of data combined with sophisticated models allows the prediction and hence the prevention of suicidal ideations, treatment of depressive relapses, psychotic episodes, as well as panic and anxiety attacks. These tools could enable immediate intervention digitally as well as by reaching out to qualified clinicians.

Empowering users. The steady digitalization of mental healthcare is not just a question of effective use of new technologies - at its core it represents a **cultural shift** in how both medical professionals as well as patients approach the subject of therapy and symptom management. On the one hand, technology grants individuals greater transparency over the process as well as the possibility to choose elements of their treatment. Asynchronous communication and tools available 24/7 make this possible (Hollis et al. 2015). Thus, technology puts the individual in the center and enables them to play a more active role in their own treatment and recovery process. Especially social media platforms such as forums and chat groups allow individuals to build up a peer support network of other users who actually have been through or are undergoing similar challenges (Lisa & Gustafson, 2013). The drive to empower individuals is based on previous IS research efforts which show that digital tools can empower users psychologically to participate in the management of a condition more actively (Deng et. al., 2013).

Another **paradigm shift** that can be introduced by data in mental health is the possibility to track users' progress, record any fluctuations in symptom manifestation such as improvements or relapses, and correlate that to specific treatments attempted. This can result in more precise and measurable outcomes, which in the absence of technology are purely subjective and based on the emotional state of the individual as well as an assessment by a medical professional (Lisa & Gustafson, 2013). Previous research in psychology shows that even an identical diagnosis for two individuals would often require a different recovery approach (Friedl et al., 2020). This kind of data-driven progress tracking enables comprehensive insight into how every single person undergoing treatment reacts to different strategies, and adjustments can be more precise and tailored.

2.1.1 Government Initiatives

Many government bodies and national healthcare systems have begun to introduce digital tools within mental healthcare services, as reflected in the most recent IS literature (Binhadyan et. al., 2015). The Anxiety and Depression Association of America (ADAA) reviews and recommends mobile apps that can be used as tools in the process to overcome anxiety and depression disorders. For example, a chatbots app based on neural network models is able to mimic certain types of therapy (such as cognitive behavioral therapy (CBT)). By means of user input, the app is able to track symptom progressions, provide immediate relief (e.g. for anxiety attacks), remotely deliver CBT and meditation as tools to manage depression and anxiety. The app has more than one million downloads to date (Inkster et. al., 2018). In the UK, the National Health Service (NHS) partners with several providers of digital mental health solutions as a way to bridge the gap between capacity and demand. The goal of these digital communities is to provide a safe space monitored by moderators where users can express themselves and connect with similar individuals across the country, as well as find useful information and resources. The EU also has multiple initiatives meant to create, test, and explore the potential use of various digital solutions for mental health. They aim to create and disseminate multiple different digital applications in the areas of diagnostics, treatment, and prevention. Projects span several countries and applications covering depression, anxiety and post-traumatic stress disorder (PTSD). The different applications have been tested across several healthcare systems and settings in Europe.

The goal of all these initiatives is to explore how well users accept digital tools for mental healthcare, to gauge a potential adoption rate, to fine-tune elements of user interface and in general make these tools as easy to use as possible, and to identify success and failure factors for digital interventions. To a large extent, these are topics covered in this thesis, particularly identifying pitfalls and improvement opportunities in the technical implementation of such tools, but also in examining user dynamics on platforms and their own pros and cons.

2.1.2 Fitting IT in the Mental Healthcare Ecosystem

IS in healthcare promotes patient-centered solutions which are not exclusively based on the idea that IT can entirely replace the need for psychologists and other professionals as part of treatments. Digital healthcare tools are also seen as potentially significantly augmenting the

reach of formal healthcare services while simultaneously lowering the financial burden on healthcare systems and individuals, as well as improving the quality of these services (Robert et. al., 2011). From this perspective, three future paradigms of integration are currently emerging: blended treatment, passive/active user monitoring, and proactive vs. reactive interventions.

2.1.2.1 Blended Treatment

Blended treatment is a type of mental healthcare treatment where traditional resources such as face-to-face therapy are combined and complemented with automated technology in order to support a person on their way to recovery. This type of treatment enriches a user's road to recovery with the possibility to intensify the treatment if needed, as well as to immediately access resources in moments of crisis. Previous research at the intersection of IS and psychology shows that this approach is especially beneficial for more severe cases, where an individual is experiencing aggravated symptoms (Friedl et al., 2020). What is more, blended treatment offers the possibility of data collection which can feed objective information of a user's behavior, relapses, crises etc. to a therapist, thus greatly enriching the therapist's resources (Fairburn & Patel, 2017).

Blended treatment is already being considered and picked up by innovative care providers. In the UK, initiatives focus on the development, adoption and evaluation of new technologies for mental healthcare and dementia – looking into blended treatment to better understand the needs of patients, families, and professionals, as well as increase the adoption of digital tools by patients and professionals alike (Whelanabc, et al., 2015). Mobile apps offer assistance to users recovering from bipolar disorder, schizophrenia, and psychosis. Users are only able to see a care coordinator once a month, which may be insufficient since relapses in these conditions can take only a few days to develop. An app helps monitor users throughout their daily routine and alert their coordinators if relapses seem likely to occur.

Observations from using blended treatment show that patients overwhelmingly embrace the use of technology as an aid in recovery, however healthcare professionals have not been as receptive. Integrating professionals into the adoption of IT in mental healthcare is crucial for the success of the digitalization of the field in the global efforts to handle the mental health emergency (Fairburn & Patel, 2017).

2.1.2.2 Active vs. Passive Monitoring

Beyond official efforts, there are many informal digital means of support available, and individuals have flocked particularly to mobile apps. In their current form, using the apps requires active engagement from the user, in other words, the user must proactively track and manage the plans and activities set out in any app. This kind of app engagement is known as active monitoring (Wang et al., 2016). However, passive monitoring is also possible. This occurs when an app tracks user movements, messaging frequency or overall activity frequency, and other signals in the background in order to determine whether a user needs support (Wang et al., 2016). One example of using passive monitoring for depression in an app is a situation when a user is immobile and in bed for a long time, and does not engage with their mobile phone – this could be a signal for a depression relapse, and an app could proactively (e.g. via push messages) try to engage the user, or even alert a care provider (Canzian & Musolesi, 2015). Passive monitoring can be extremely helpful, as many times suffering individuals lack the motivation to ask for help (Canzian & Musolesi, 2015). Most literature on passive monitoring currently comes from IS-related computing fields, specifically focusing on pervasive computing or precision healthcare; within the IS field there is still a shallow understanding of the transformative role that IT-enabled continuous monitoring can take in mental healthcare (Wahle & Kowatsch, 2014). Passive monitoring can drive innovative digital interventions for immediate and situational support based on inconspicuous data gathering, which is why previous IS research has called for studies to quantify the individual impact as well as the intervention effects borne from such monitoring, and to examine the design of future technology-mediated interventions (Wahle et. al., 2017).

2.1.2.3 Proactive vs. Reactive Interventions

The use of data and IT opens the possibility to developing digital services that can enable early intervention (i.e. a person's critical state can be foreseen based on previously observed data) in critical moments such as depressive relapses, psychotic episodes, or suicidal attempts (Naslund et al., 2019). Knowing that there is a high probability of such an event happening around a certain date in the future would allow the scheduling of interventions, visits, or therapy sessions so that these events are avoided altogether.

The epidemic of opioid deaths is the number one cause of death for people under the age of 50 in the United States. Many pilot studies on addiction and substance abuse epidemics have

been completed mostly using Twitter data. They validate the possibility to use social media to monitor for and predict spikes in substance abuse. The issue in the US opioid crisis can be boiled down to lack of qualified medical staff. Many addicts who seek treatment are unable to get an appointment early enough, as the lack of local doctors who can prescribe the necessary medicine to battle addiction means that these individuals simply have no access to treatment. Adopting proactive monitoring can allow organizations to route resources to areas where spikes in use are expected (Saloner et al., 2018). This approach can be applied to any condition where practitioners and algorithms can learn from past data the conditions under which symptom worsening is observed, and the events that lead to it.

2.2 Online Forums as a Digital Mental Health Platform

2.2.1 General Context of Using Social Media

Social media have become popular channels to exchange and disseminate information and receive peer support in new formats and informal ways. These channels have also helped with the social isolation faced by many who suffer from mental conditions. In addition, health services can often be inaccessible for various reasons, in particular for young people; given that mental health issues disproportionately affect the younger population who are also the heaviest users of social media, information and support via this channel may help young people access information and resources (Hollis et al. 2015).

Social media users who suffer from some kind of mental disorder express interest in participating in treatments delivered via social media for the purposes of general well-being but also for coping with specific symptoms that they experience (Naslund et al. 2019). People have been shown to use social media especially in the beginning of receiving a diagnosis or experiencing symptoms, for several reasons. First of all, they seek out a social support group that can help them integrate the diagnosis as part of their self-identity, or help in accepting the diagnosis and coming to terms with its implications. Users also report that the engagement helps them in developing an understanding for the diagnosis and its implications by learning from third-party experiences (Shepherd et al. 2015).

Past studies from IS scholars into the usage of social media by people with mental conditions show that there are several benefits to be had from participation in these online communities, such as greater social integration, identifying as a member of a group, the possibility to share

the personal journey, as well as general support in daily coping and living with a mental health condition (Shepherd et al. 2015). Participation in online communities is rewarding as users can learn from other peers and gain understanding about important aspects of their diagnosis as well as general health care, thus this online environment in itself can promote mental healthcare seeking behavior in users.

It is possible for users to access valuable resources related to their mental well-being through social media. Social media is a suitable channel to deliver peer support, increase support and resource reach, and educate on available treatments. Inevitably, when using digital channels, users run the risk of accessing information that is wrong or incomplete, as well as receive misleading or abusive messages in online communities. However, studies to date point that the benefits of participation far outweigh potential risks (Naslund et al. 2016). Another different and potentially negative aspect of using social media as well as IT tools in mental healthcare comes from technology-driven stress (Benlian, 2020). This topic belongs to the IS research field, and findings and design principles from the field can be transposed and applied on mental health digital tools in order to mitigate potential user frustrations and feelings of technological alienation, e.g. communication with the tool over voice can help overcome a potential lack of anthropomorphic design features (Benlian et. al., 2019).

2.2.2 Online Mental Health Forums

Online forums dedicated to mental healthcare discussions allow users to start conversations about questions they might have regarding their own condition or that of a loved one, to answer or comment on the conversations started by others, or simply to derive benefit from the information already posted by passively consuming, or “lurking”. Forums have volunteer moderators who monitor message boards about possible abusive content, as well as users who may express suicidal thoughts and are in need of immediate help.

Forum boards exist in many different languages. The three largest boards in the English language have a combined user base of over 300,000 users and more than a million conversation threads. Forums offer several benefits to users. They are asynchronous i.e. time-independent, and users can start conversations or leave comments at any time. Forums are completely anonymous, which is important to affected individuals as they still face a stigma related to mental health disorders in society. They are also commonly free of charge.

Multiple IS studies have examined the benefits of active and passive participation in mental health forums (van Uden-Kraan et. al., 2009). Even just lurking or passively consuming information shared by others can lead to feeling less socially isolated, and more enabled to cope with a diagnosis (Preece et. al., 2004). Actively participating can help boost treatment-seeking behaviors and lead to a more positive outlook on life with a mental health diagnosis (van Uden-Kraan et. al., 2008). For all these benefits and its ease of use, mental health forums have in the past been actively supported and encouraged from national healthcare bodies (Naslund et. al., 2016).

The advantage of using data out of online forums is its source - it comes from the users, or the affected individuals, themselves. Forum users have usually undergone several different ways of therapy or treatments – prescribed by therapists, medications, and less traditional methods such as mindfulness, sports, etc. (Barak & Gluck-Ofri, 2007). Forums contain information on the type of intervention that helped a specific user profile – since symptoms even in one and the same condition can manifest differently from person to person, the level of detail that can be obtained, as well as the mass (hundreds of thousands of users) allow for a first-person experience and outcome check.

2.3 NLP Applications Within Mental Health Forums

Mental healthcare data contains many more variables over subjects, therefore predictive patterns can be discovered in large datasets by applying machine learning algorithms. The innovation in data collection in terms of volume, quantity, and sophistication, allow the enrichment of customary psychiatric epidemiology with data-driven analytical strategies. Traditional statistical methods of data modelling draw inferences from smaller sample data which is highly structured. On the other hand, machine learning methods can make use of structured as well as unstructured data (e.g. free text) with the goal to recognize patterns which can be used for prediction and classification. The application of these techniques in mental healthcare affords the augmentation of diagnostic decision-making with data-driven insights and offers clarity and transparency on potential reactions on specific treatment applications (Fortuna et al. 2019). This research approach of uniting psychology and data is the bedrock of precision psychiatry as an emerging interdisciplinary research field.

Rigorous and successful treatments that are thoroughly tested clinically exist for managing mental illnesses. However, diagnoses of two people for the same condition can be varying in symptoms, thus requiring personalization of treatments. The rich symptom heterogeneity of most mental disorders presents a significant barrier to developing such personalized treatments. With natural language programming (NLP) and machine learning, symptom heterogeneity can be captured and included in a treatment plan on an individual basis. Machine learning techniques are able to process high-dimensional data with many variables present, account for missing data, derive high-level abstractions, and make do without an a priori patient stratification. Stratification divides patients into groups or "stratas", for example based on age, ethnicity, or medical history. The application of machine learning in this way can enrich treatments to account for individual symptom manifestation and progression; it is also possible to create chatbots to mimic therapy conversations, as well as predict and anticipate social outcomes for users (Tan et al. 2019).

The NLP-driven research of mental health forums in the effort to digitalize mental healthcare applies text analytics techniques from computational linguistics within IS as a way to examine and explain the human and managerial aspects of technological applications within mental healthcare. Efforts in this interdisciplinary area both push forward the methods used to conduct the technical analyses (as we learn how to interpret text signals for our specific goal) (Chen et al. 2018) as well as drive forward our understanding of how users react to processes and user interfaces (Lecomte et al., 2020).

2.3.1 Data Sources

This sub-chapter presents the various data sources and collection methods utilized in studies where machine learning is applied to mental healthcare. Typical and commonly used sources are social media data scraping, standard or adjusted medical questionnaires used in clinical diagnostics, external linguistic resources that quantify specific features derived from text, as well as data blending – using any combination of primary and secondary data where the resulting dataset covers aspects present in multiple data sources.

Many machine learning (ML) studies use questionnaires in order to collect training and testing data, either standalone or in combination with another data source. Standard questionnaires are directly borrowed from psychological practice, such as the Patient Health Questionnaire, or PHQ-9 (Wang et al. 2017). The questionnaire consists of only nine

questions and is used to identify the presence and severity of depression in an individual. A diagnosis derived from the use of PHQ-9 is in accordance with DSM-IV, as it has been subject to reliability and validity tests involving several thousand test subjects (Braithwaite et al. 2016). The advantage of this approach is that it adheres to medical diagnostic standards, thus the resulting diagnosis is clinically valid; furthermore, the resulting data is highly structured and clean, therefore suitable for processing through an ML pipeline. Disadvantages in using questionnaire data is the difficulty and cost associated with data collection by means of a questionnaire. For example, in terms of active and passive monitoring (see above), filling in a questionnaire periodically requires active participation on the side of the user. Furthermore, questionnaires like the above-mentioned do not capture fine-grained information regarding a person's mental state in terms of symptom breakdown or comorbidity, as these instruments are designed with a specific condition in mind and specifically to provide a cutoff diagnosis (e.g. mild, severe, or no depression detected). However, data from medical questionnaires can be used to cross-check the clinical reliability of other datasets and results.

Large-scale surveys such as the National Survey of Children's Health (NSCH) have also been used as data sources for detection of specific disorders such as autism (van den Bekerom, 2017). These surveys are usually repeated on a yearly basis and are very extensive, covering hundreds of variables on a massive national or international scale. They provide a very rich source for extracting possible signals that can be used as predictors for the presence or absence of a certain disorder, however, they would be potentially very difficult and time-consuming to repeat for specific individuals, therefore they would be not well-suited for usage as-is in automated treatments on an individual basis.

The most popular data source in studies so far has been the scraping of social media posts available publicly. Data scraping is a technique that can fetch large datasets in a relatively short amount of time and inexpensively, spanning hundreds of thousands of data points, users, platforms, and countries. Twitter is a popular source because of the ease of collecting user posts (using the public API) as well as the fact that tweets are a convenient unit to work with, being limited to 280 characters per post. Past studies show that official US suicide statistics correlate with Twitter analyses on the use of negative emotions in tweets (Braithwaite et al. 2016). The challenges in using this approach for data collection lie in that the data are highly unstructured, noisy, and sometimes missing. Thus, data preprocessing becomes a necessary step in the analysis process. Furthermore, the resulting dataset will need to be additionally

labeled if it is to be used in supervised learning tasks. At the same time, web scraping provides deep granularity, in other words, the data captured contains a multitude of signals that can capture not only a person's potential condition diagnosis, but also multiple other factors that either influence a central concern or deliver more comprehensive information.

Although not a data source per se, the use of external resources such as the Linguistic Inquiry and Word Count (LIWC) or SenticNet is also a widespread practice in ML-driven psychological studies. These resources are used to extract signals from text data such as emotions, affect, and lexical information. They are usually used in combination with a text data source, providing the possibility to extract signals from text in a rule-based manner. For example, LIWC contains variables that cover behavioral descriptors such as attention, emotion, authenticity, or social functioning. Such variables can be used as features in training algorithms.

2.3.2 Data Labeling

In case of supervised ML, a training dataset will need to have appropriate labels, e.g. a category. For example, if the input data is sentence-based, then each sentence will need to be provided with a label, e.g. in a binary depression classification model, the label is normally 1 for depression, and 0 for no depression. Labeling, or annotation, is usually made either by domain experts or is crowdsourced on micro-tasking platforms such as AMT. Experts in annotation have previously been mostly machine learning researchers. There exists a gap in IS studies that apply ML in mental health, in that they are conducted without the participation of clinicians (Calvo et al., 2017). Data preparation is one step of the ML pipeline where they can contribute to data quality and validity.

On the other hand, micro-tasking platforms such as AMT are used to label data using many human workers who are not domain experts. Crowdsourcing data labeling is affordable and reliable in many cases, as it is used to create very large datasets for machine learning (Calvo et al., 2017). However, it may not be reliable in the case of mental health, as annotating sentences with symptoms or conditions would require expertise in psychology as well as machine learning (Chancellor & De Choudhury, 2020). For this task, the annotator must be able to identify medically applicable symptoms in user expressions.

2.3.3 Applications

Even though machine learning has been applied to mental health since very recently, the applications of the field have already undergone a visible and clear progress (Calvo et al. 2017). This section highlights the applications covered and aspects tackled in this thesis.

The proliferation of digital tools for mental health naturally warrants the examination of the effect of using such tools within IS literature. Specifically, it becomes necessary to establish whether digital intervention influences users, in what way (positive or negative), and to what extent (O'Raghallaigh & Frederic, 2017). Bringing an understanding of these aspects of digital mental health intervention will enable future improvements by distinguishing successful and failing methods and their contexts. Within a community context, these questions become more complex with the added layer of connections between users, assumed roles, usage patterns, and longitudinal aspects (Xi et. al., 2015). Online communities consist of intricate interdependencies – if understood correctly, each individual user can be guided as to how to more effectively make use of the community to their personal advantage.

Furthermore, with sentiment analysis showing high performance results even with earlier Bayesian models, the exploration of mood or affect has been one of the earliest fields where many studies have been conducted (Xu et. al., 2013). Being able to automatically discern a person's mood from text is very important as mood can often be a symptom. For example, in bipolar disorder, frequent mood swings are a symptom of the condition. Mood dips can also be a signal for a relapse, such as having a depressive episode or a panic attack.

Finally, current efforts are concentrated on automated diagnostics, where new models such as neural networks have improved performance significantly and open venues of exploring different approaches to this very complex and crucial task on the road to automating mental health. The technical modeling of diagnostics models is embedded within fields such as precision psychology and computational linguistics, whereas testing the functioning and acceptance of a finished tool falls within the IS domain. Studies have tackled this issue; however, current models still leave room for performance improvement. For example, many studies apply binary models to distinguish Twitter posts as showing signs of one particular condition or not, e.g. a post can be classified as depression or not depression, or suicidal or not suicidal, and no other condition would be considered. Furthermore, outside of depression and anxiety, other conditions are not well researched in terms of automated diagnostics

(Guntuku et. al., 2017). In both cases this thesis makes a contribution – by presenting a multiclass model as well as looking into a variety of conditions.

2.3.4 Data Analysis Techniques

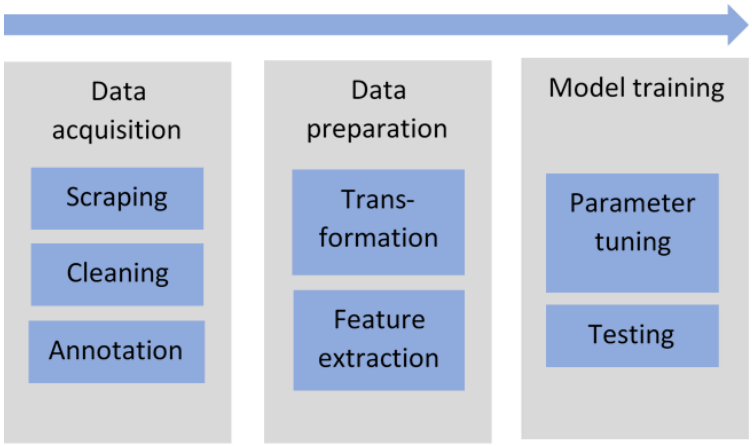

Figure 2-1 Natural language processing workflow

Studies included in this thesis apply semantic knowledge bases as well as machine learning algorithms within an NLP pipeline to extract information out of mental health forum posts. This section presents an overview of the techniques used throughout this thesis.

Machine learning algorithms can be descriptive (unsupervised learning) or predictive (supervised learning). Descriptive techniques construct clusters based on similarity between data points and without answering a pre-defined question, rather exploring relationships that underlie the data (Góngora et al., 2018). Supervised learning tasks require a dataset that has both the learning inputs as well as the desired outputs (labels), whereas unsupervised methods do not require labeled data. Supervised learning is broadly subdivided into classification and regression techniques. Classification methods categorize inputs, i.e. they are discrete, while regression methods predict a continuous output. On the other hand, unsupervised algorithms are used to map out structures in a dataset, such as clusters of data points. This thesis makes use of both supervised (classification with neural networks) and unsupervised algorithms (topic modelling with LDA), explained below. In addition to algorithms, **two semantic knowledge bases have been used** to conduct sentiment analysis on data, namely the Python module Vader (Valence Aware Dictionary and sEntiment Reasoner) and SenticNet.

In the case of Vader, it is a lexicon and rule-based sentiment analysis tool which can be used to calculate sentiment on a sentence level and has been specifically designed for social media text processing. It builds on and improves a bag-of-words model approach to sentiment analysis by using heuristics, i.e. rule-based processing in order to account for word order in a sentence. Degree modifiers (words which increase the intensity of a word e.g. 'angry' vs. 'very angry') are also accounted for in the lexicon. Vader was built by using human raters to rate over 9000 words on a scale from -4 to +4, where a negative score is associated with a negative sentiment and vice versa. Vader is a unique resource in that it also has the ability to parse the sentiment of emoticons, acronyms such as "LOL", as well as commonly used slang. Vader has been empirically tested with a wisdom-of-the-crowd approach.

Sentic Net is built on the bag-of-concepts model, which differs from a bag-of-words model in that it goes beyond counting word co-occurrence frequencies - the bag-of-concepts model is generated on top of the word2vec document representation model, which uses a simple neural network to embed words into continuous vector space (Kim et. al., 2017). A bag-of-concepts model makes use of linguistic patterns such as part-of-speech order, to allow sentiments to flow from concept to concept by inspecting dependency relations between clauses, in other words – to create a network of concepts where the sentiment of a concept is related to that concept's position in the general concept network and its relation to other concepts and their sentiments (Poria et al. 2014).

Resources such as Vader and SenticNet make certain types of semantic analysis, e.g. aspect-based sentiment analysis (with SenticNet) possible without the need to collect and label data for algorithm training. This accelerates the ML process. These resources are also rigorously tested and checked by the broader scientific community in various research tasks, adding to their reliability.

Recent studies have made use of **artificial neural networks** (ANN) – models loosely based on networks in the central nervous system. Neural networks process input data with a series of transformations into computational units called neurons, which are organized in several layers (Góngora et al. 2018).

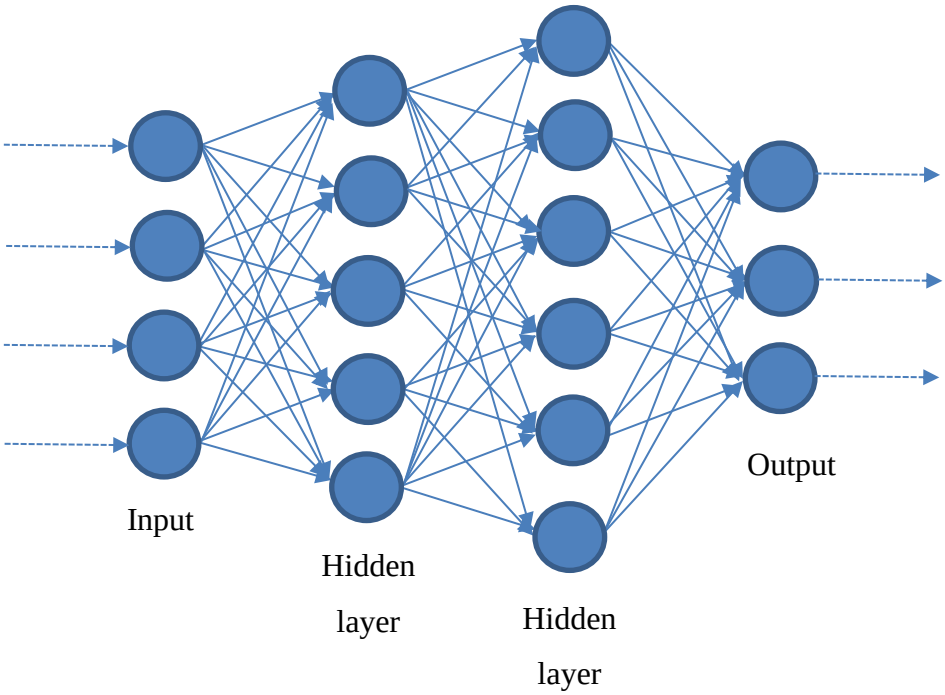

Figure 2-2 High-level representation of a neural network with two hidden layers

Neural networks are flexible models where the number of hidden layers and neurons, as well as the types of activation functions and optimizers, can vary, thus neural networks can be customized for optimal performance on different types of data (text, image, time series, etc.). Neural networks are replacing other models for diagnostics modeling because of their improved performance.

Data input for training neural networks must always be numeric, hence when working with text data, words are first converted into a numeric representation that can be as simple as assigning numbers to words (tokenization) or as complicated as assigning arrays containing several thousand numbers per word and capture intricate lexical and semantic relationships among words (word embedding). This thesis uses word embeddings to convert text input to numbers.

As the data input to train the neural network passes through its several computational layers, each neuron focuses on a particular feature present in the data, and how this feature impacts or

is impacted by other data features – this is where an activation function is used. An activation function usually outputs a number between 0 and 1 where 0 would mean that a certain feature can be ignored and 1 would mean that this feature should be further processed in the next computational layer. Thus, the neural network is able to focus on those input data characteristics that make a difference in the end result of a task, e.g. classification. In this thesis, neural networks are used to classify text data into classes, which falls under supervised learning.

One unsupervised technique used in this thesis is topic modelling by way of Latent Dirichlet allocation (LDA). LDA is a Bayesian probabilistic model which represents a topic distribution in text as a word mixture modeled by a Dirichlet distribution. The requires an estimate of topics expected to be encountered within text inputs, and it output clusters of words that are associated with a certain topic. It is important to note that LDA does not explicitly name the topics discovered, rather it outputs the associated words and it is up to a human expert to name the topics. In this thesis, LDA is used to extract names of or references to symptoms which can be used to label data for supervised learning.

2.4 Positioning of the Thesis

Figure 1-1 positions each study within the research context in terms of machine learning techniques used and which research aspect each technique was used to explore.

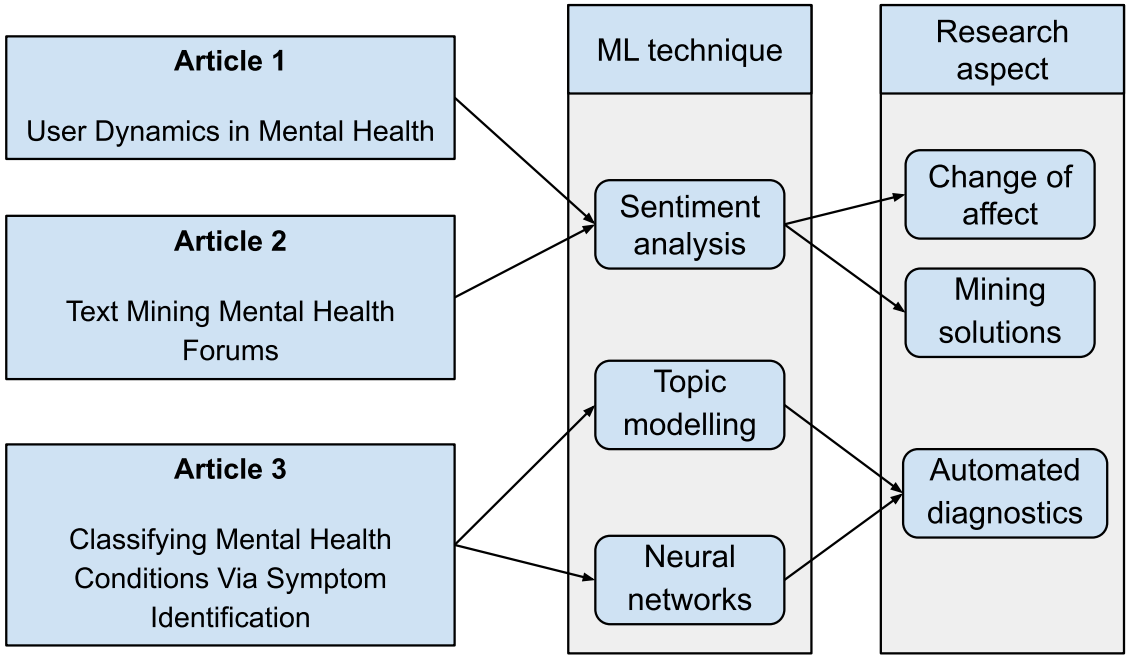

Figure 2-3 Overview of the examined interactions in research articles

On the backdrop of rising numbers of individuals who report experiencing mental health disorders, the demand for mental healthcare services has risen beyond current capabilities, and this rising trend is expected to persist in the near future. IT, and in particular the combination of social media, big data, and machine learning, has been researched in academia and is being implemented in practice as a possible resolution to mitigate the imbalance in supply and demand. Online digital services for mental health offer several advantages to deal with the current lack of resources such as wider reach and lower service costs. They can also enhance current recovery procedures by offering users additional on-demand support in times of crisis, as well as empowering users to have a more personalized treatment by self-selecting digital support tools. Machine learning algorithms when applied to psychological tasks such as diagnosis or mood classification have not yet reached optimal performance, and there is the need to provide more scientific evidence that digitally administered interventions would be as effective as one-on-one therapy. This thesis shows that participation in digital platforms has a significant influence on participants' mood and highlights the need to continue experimenting

with state-of-the-art machine learning models to integrate psychological practices and practitioners alike in the development of these models as a way to showcase consistency and reach high performance. We further contribute to the growing body of literature in this interdisciplinary domain by improving on previously reported performance metrics for diagnostics, as well as captured level of detail and complexity.

Chapter 3: Mapping User Roles and Dynamics in Mental Health Forums

Title: User Dynamics in Mental Health Forums – A Sentiment Analysis Perspective (2019)

Authors: Elena Davcheva, Technische Universität Darmstadt, Germany
Martin Adam, Technische Universität Darmstadt, Germany
Alexander Benlian, Technische Universität Darmstadt, Germany

Published in: 14th International Conference on Wirtschaftsinformatik, (WI 2019) Siegen, Germany

Abstract

Individuals around the world in need of mental healthcare do not find adequate treatment because of lacking resources. Since the necessary support can often not be provided directly, many turn to the Internet for assistance, whereby mental health forums have evolved into an important medium for millions of users to share experiences. Information Systems research lacks empirical evidence to analyze how health forums influence users' moods. This paper addresses the research gap by conducting sentiment analysis on a large dataset of user posts from three leading English-language forums. The goal of this study is to shed light on the mood effects of mental health forum participation, as well as to better understand user roles. The results of our exploratory study show that sentiment scores develop either positively or negatively depending on the condition. We additionally investigate and report on user forum roles.

Keywords: Mental Health, Sentiment Analysis, Big Data, Forums, Natural Language Processing

3.1 Introduction

Mental disorders are defined as “a combination of abnormal thoughts, perceptions, emotions, behavior and relationships with others” (World Health Organization, 2018a). The term comprises depression, bipolar disorder, schizophrenia, dementia, and developmental disorders such as autism. An estimated 300 million people are affected by depression alone (worldwide) and 15% of people aged 60 and over suffer from one or more disorders (World Health Organization, 2018b). What is more, the world population is aging rapidly. Between 2015 and 2050, the proportion over 60 years will nearly double, from 12% to 22% (World Health Organization, 2018b). Health systems have not yet adequately answered the growing burden. Between 35% and 85% of affected individuals receive no official treatment (World Health Organization, 2018c). In many countries, less than one psychiatrist per 100,000 people is available (Mathers et. al., 2004). Moreover, the fear of stigma discourages people from seeking assistance (Crabtree et. al., 2010). Even if willing, those affected often cannot afford the medical treatment (Saleem et al., 2012), since professional help involves expensive clinical procedures (Weathers et. al., 2001). The medical cost of mental health exceeds $200 billion in the U.S., making it the costliest medical condition in the country (Roehrig, 2016).

The digitalization of health information has created opportunities for individuals to seek self-help and connect directly to other affected individuals (Lluch, 2011). Online information is free, anonymous, and time- and location-independent (Winzelberg, 1997). 80% of the U.S. population with Internet access gather information from mental health discussions, and 34% of those read others’ personal stories (Malmasi et. al., 2016). Mental health forums are particularly appealing to those individuals who are afraid that coming out of the cloak of anonymity may expose them to stigma (Kummervold et al., 2002). Online information is by no means limited only to the younger generation, as the usage of social media by adults aged 60 and over nearly doubles on a yearly basis (Madden, 2010). Furthermore, studies show that users are more honest and more likely to share personal stories online than in-person (Barak & Gluck-Ofri, 2007). The chance to share experiences, connect with others with similar conditions as well as gain insights from their stories, creates a rewarding experience for the users of these forums (Malmasi et. al., 2016).

This paper looks into online tools for mental health by longitudinally analyzing sentiment development of user posts in online mental health forums. We investigate (1) the sentiment

progression in forums over time, and (2) the user role dynamics, as well as (3) the relation between user role and sentiment. We apply sentiment analysis on a dataset of 500,754 individual posts across 8 mental health conditions collected from 3 leading English-language forums. We show that the longitudinal development of user sentiment differs across conditions and types of engagement. We compare user roles and their correlation to sentiment.

This study contributes to existing research by exploring how engagement in virtual healthcare communities affects users, and the potential to empower patients to self-management. We especially address the application of behavioral analytics for mental health, which has been more widely adopted for commercial purposes (Monteith et. al., 2015), by illustrating how text mining can help practitioners and policymakers in understanding the value and risks of using user-led online tools (Smailhodzic et. al., 2016).

3.2 Conceptual Background

3.2.1 Online Mental Health Forums

Previous IS research has focused on trust formation in online health communities, triaging of symptoms, user roles in forums (Huh & Pratt, 2014), but not on forum influence on well-being. We address this research gap by looking into conversation dynamics and their influence on user sentiment. Prior research shows that users recognize risks of posting personal medical information online, they nevertheless share as potential rewards outweigh risks (Kordzadeh & Warren, 2017). User roles have been researched in terms of super-users (frequent posters), however not in terms of sentiment and its correlation to user roles.

Prior research has found online communities to be not only helpful for mitigating various mental conditions but also dangerous, as users can be influenced to commit potentially life-threatening actions, as in pro-suicide or pro-anorexia groups (Bell, 2007). Many of these studies have applied manual analysis, such as user surveys or discourse analysis based on a small sample of posts. Johnsen, Rosenvinge and Gammon (2002) used human readers to classify interactions in mental health forums as helpful or unhelpful based on only 102 posts. By manual assessment, Spijkerman, Pots and Bohlmeijer (2016) investigated advantages of online mindfulness and meditation practice for depression and anxiety. In a comparable paper, Mitchell et. al. (Mitchell et. Al., 2016) used human coders to analyze 401 posts from 55 forum

threads and found that 25% of users with ADD reported positive effects of self-medicated cannabis on their illness. Although human readers provide reliable analysis, the amount of processed posts is very limited. In this study, we analyze 500,754 forum user posts. Such a vast dataset, to the best of our knowledge, has not been analyzed yet in this context, thus allowing us to explore the shared stories and experiences of tens of thousands of individuals.

3.2.2 Sentiment Analysis

Recently, natural language processing (NLP) techniques such as sentiment analysis (SA), emotion classification, and stigma measurements have been increasingly applied in research on various data sources such as Twitter, Facebook, and online forums (Calvo et. al., 2017). SA can determine sentiment polarity in written text (Tao, 2014), using a classifier such as Naive Bayes to train with a pre-annotated dataset of sentences, and then apply to new data. Sentiment in psychological literature is referred to as mood or affect. There is a long-standing position in psychological research that a well-adjusted mood is crucial for good mental health (Bradburn, 1969) (Kahneman et. al., 1999). A person in good mental health would express about “three times more positive than negative affect" (Diehl et. al., 2011). Diehl et. al. formally showed that the absence or presence of positivity or negativity can be used to distinguish an individual's mental health status (waning vs. healthy). Therefore, while sentiment is not a tell-all signal of mental well-being, it is a fundamental indicator of the progression of a mental state.

Only few papers use SA in the context of online mental health forums. Nguyen et. al. (2014) conducted a fundamental effort by comparing the sentiment expressed on depression forums with sentiment in non-depression forums, demonstrating that individuals without depression express themselves more positively. The study directly backs the application of SA as an appropriate method to analyze data from mental health forums. SA has also been applied to pharmacovigilance in social media (Twitter) – the identification of adverse drug effects – by detecting posts with negative sentiment towards specific medications (Korkontzelos et al., 2016), allowing researchers to explore drug effects across a large and diverse population. The study presents an important example in using text mining to identify potentially negative effects of treatments that might have been otherwise considered safe. Twitter alone has been used in a dozen studies where healthcare issues in general have been explored using sentiment analysis (Gohil et. al., 2018). Furthermore, Cobb, Mays and Graham (Cobb et. al., 2013)

demonstrated by using SA that talking positively about quitting smoking influences social media users to quit in real life. Via SA, the study shows that user-to-user communication can affect life choices. Aspect-based sentiment analysis has also been applied in order to examine how forum users express themselves on specific concepts such as family or therapy (Davcheva, 2018). Finally, Wang et. al. (Wang et al., 2013) applied SA to classify a user's condition. These ways of using SA have had great success and will be even more useful in the future, considering the development of automated online diagnostic tools (Saleem et al., 2012). What has not been addressed is the effect of participating in mental health forum conversations, and if and how this depends on the way a user engages with the forum and the role they assume.

3.3 Research Propositions

Many studies have provided evidence that exchanging support and talking to peers can improve well-being (van Uden-Kraan et al., 2008). In fact, even if people seek treatment from professionals, they often still participate in online discussions (Kummervold et al., 2002). Participants also enhance their psychological well-being by providing support (Riessman, 1997). As a result, they feel "better informed, confident with the physician, treatment and social environment, improved acceptance of the illness, increased optimism and control, enhanced self-esteem" (Kummervold et al., 2002). Disadvantageous posts may also occur due to the lack of control of quality of information, potentially destructive content that reinforces negative emotions (Finn, 1999) (Johnsen et. al., 2002). Some evidence of the possible benefits from participation exists specifically for mental health forums, although findings are mixed mainly due inconclusive studies and a lack of quantitative research (Eysenbach et. al., 2004).

Owing to the exploratory nature of the study, we use propositions to frame our central premises. In our study we look explicitly at forum users who post, therefore in this context users (or participants) are those who post on the forums, whether to ask a question, to share an experience, to provide answers or comment. By applying sentiment analysis, we test if through participation in forum conversations the affect of a participant is improved. Since threads in mental help forums are started with the purpose to solve or at least mitigate the problem of the original poster (Eysenbach et. al., 2004) (van Uden-Kraan et al., 2008), the sentiment is expected to improve throughout the thread if the conversation is to be proven as helpful.

P1a: The sentiment score of posts within a thread in mental health forums increases throughout the thread with each subsequent post.

Moreover, we expect that after receiving advice and support, the original poster (OP) will improve their mood, e.g. with regards to confidence, optimism, or ability to cope with their mental condition (Eysenbach et. al., 2004) (van Uden-Kraan et al., 2008) (van Uden-Kraan et. al., 2009).

P1b: The sentiment score of posts by the original poster within a thread in mental health forums increases throughout the thread with each subsequent post.

Second, we investigate whether the proposed positive sentiment progression is true outside of single threads, i.e. throughout a user's general forum participation. Studies (van Uden-Kraan et. al., 2009) found that "helping others" was one of the empowering processes that occurred the least frequently in social support groups, since users have to actively post to assist others. High-frequency posters were found to provide more advice than asking for help (van Uden-Kraan et. al., 2009). Therefore, if a user is commenting on threads by other users, they will improve sentiment and keep the solution-oriented perspective.

P2: The sentiment score of user posts in mental health forums increases throughout the general participation in the forum with each subsequent post.

Third, we look into the roles that users assume in mental health forums, and the relation between roles and sentiment. Initially, users participate by passively reading (Preece et. al., 2004). In urgent cases, they will post their problem fast and receive encouragement or advice. However, users also increase their psychological well-being by commenting on and engaging with issues posted by others (Riessman, 1997), and providing help to others will indirectly assist in coping with their initial problem. Therefore, we propose that users of mental health forums normally fall into one of two roles, either original posters or commenters. The interaction between these two user groups is what drives and defines mental health forum dynamics (Becker, 2014). Proposition P3 is sentiment-independent and helps establish the dynamics of a user's forum lifetime. The analysis procedure for P3 is to look at the timeline and type of user comments - specifically original poster versus commenter.

P3: Users start their forum participation as original posters, then go on to participate and reply to threads opened by other users.

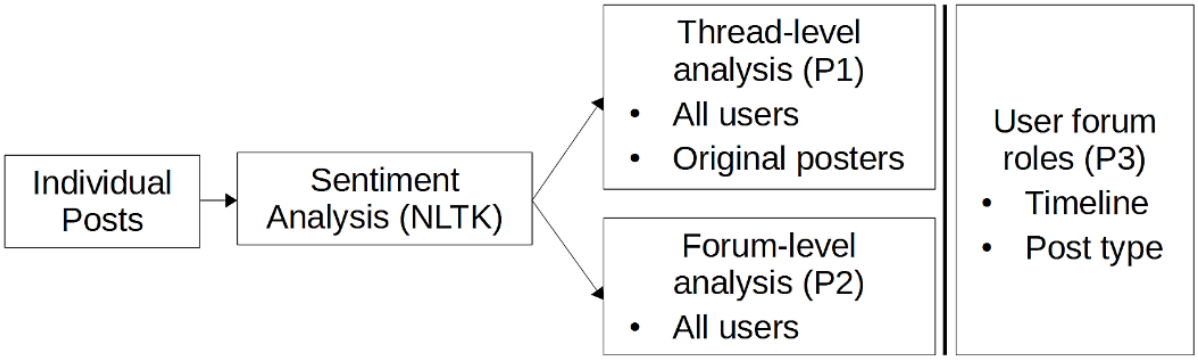

Figure 3-1 Research process and propositions

3.4 Methodology

To test our propositions, we apply sentiment analysis on peer-to-peer mental health forum posts. Three of the leading English-language mental health forums were scraped to create a combined set of 49,113 threads containing 500,754 individual posts from about 75,000 users across 8 conditions (i.e., depression, bipolar disorder (BD), anxiety and panic attacks, attention deficit hyperactivity disorder (ADHD), Borderline personality disorder (BPD), obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD)). After removing outlying users who have posted more than 30 times, each user posted on average 6,7 posts throughout a forum. The data were extracted in August 2017 and encompass all publicly available posts on the respective websites. For each post, the dataset contains information on forum username, date posted, thread name, and post content. The forums have moderators whose task is to make sure conversations do not go off-topic; thus, we can be sure that in our research we are considering discussions relevant to each condition. Additionally, moderators remove offensive or damaging material (e.g. posts that encourage self-abusive behavior). However, they do not provide advice, as the forum is a place of discussion among the users, i.e. a forum is not meant or seen as a tool to replace established practices such as therapy.

The sentiment score of a post is scored between -1 (negative) and +1 (positive). We use the Python Natural Language Toolkit (NLTK) sentiment analysis implementation with a lexicon-based classifier. Each forum post is given a score. In order to trace the sentiment progression

of user posts throughout forums and threads, we map the average sentiment score of individual users' posts based on the post order within threads (P1) and forums (P2); for proposition P1b we look exclusively at posts made by the original poster within the threads the OP started.

The sentiment is calculated using the Valence-Aware Dictionary for Sentiment Reasoning (Vader), which offers several unique advantages to other models. The sentiment of a post equals the sentiment valence (or score) of each word recognized by the lexicon. If a word is part of a negation structure (e.g. neither…nor), it's valence will be reversed. If a word is used in combination with a booster word (e.g. "amazingly", "awfully"), its valence will be intensified. The Vader lexicon was built on social media data, namely Twitter tweets, New York Time articles, and online movie and product reviews. The sources capture a variety of aspects of social media writing, as well as more analytical texts, thus generalizing well to mental health forum data as it captures features from informal online discussions, e.g. conventional use of punctuation ("good" vs. "good!!!"), capitalization, emoticons, degree modifiers ("good" vs. "very good"), as well as common slang and abbreviations ("this sux", "lol") to signal sentiment intensification. Since its creation in 2014, Vader has been iteratively empirically validated by human judges.

In order to determine the extent of improvement resulting from a prolonged forum posting, we test sentiment score trends by applying regression analysis. For P3, we test the association between user roles and conditions by applying chi-square analysis.

3.5 Results

Using linear regression, we analyze the sentiment score of a sequence of posts made over a period of time, either within threads or on a per-user basis. Equation (1) models the forum-wide sentiment of depression.

$$y = 0.0011 * x + 0.09 \quad (1)$$

Table 3-1 presents the regression results and significance for degree of improvement for sentiment trends from three points of view: (1) thread sentiment progression for OP posts only, (2) thread sentiment progression for all thread participants, and (3) sentiment progression per user throughout a forum. Table 3-1 shows a mild improvement in affect as

users keep posting for some conditions, and deterioration in other conditions. The posts of original posters show the highest improvement. The pace of change in sentiment on forums (whether positive or negative) is slightly slower compared to patient improvement observed in therapy. For comparison, a moderate therapy treatment would last 3-4 months and would include 15-20 sessions for 50% of patients to report significant symptom improvement. According to our analysis, anxiety, bipolar, and depression forum users must post 100 posts in order to improve their expressed sentiment by 15%. For example, if a depression forum user posts as an OP in threads, the average sentiment improvement per post is 0,003, thus adding +0,3 to positive sentiment after 100 posts. The low R2 values point to high variability in individual user post sentiments. This is consistent with observations from previous studies with a psychological or sociological aspect, as individuals are relatively meandering in their responses (Bedeian & Mossholder, 1994). Testing for weekday effects on sentiment change did not prove to be significant.

	OP Thread Posts		**All Thread Posts**		**All Forum Posts**	
	Coeff.	R2	Coeff.	R2	Coeff.	R2
ADHD	0.007	0.2	0.0001	0.2	0	0.2
Anxiety	0.003*	0.3	0.001*	0.4	0.001*	0.3
Autism	0.002*	0.3	-0.006**	0.3	0.002	0.3
Bipolar	0.003**	0.4	0.001	0.2	0.001**	0.2
Borderline	0.001*	0.4	0.0001	0.2	0.0001	0.2
Depression	0.003**	0.8	0.001*	0.4	0.001**	0.4
OCD	0.0001*	0.2	-0.001*	0.2	-0.001	0.1
PTSD	0.0001	0.4	0.003**	0.5	0.001**	0.4
*Note: N = 500,754; * p < 0.05; ** p < 0.01; *** p < 0.001; Coeff. = Coefficient*						

Table 3-1 Regression results and p-values for analyzed mental conditions

3.5.1 Thread Sentiment Progression of Posts by All Users

Overall, 5 out of 8 tested conditions (i.e., depression, anxiety, autism, PTSD, OCD) show statistically significant sentiment trend. For depression, anxiety, and PTSD, as the conversation thread progresses, post sentiment improves marginally for every subsequent

post. PTSD threads have the steepest improvement per post. On the other hand, autism and OCD mark a downward trend, with autism sentiment deteriorating the fastest per post.

Also, OCD and autism threads show significantly more fluctuation within thread posts, whereas conversations in depression, autism, and PTSD have relatively stable trend progressions.

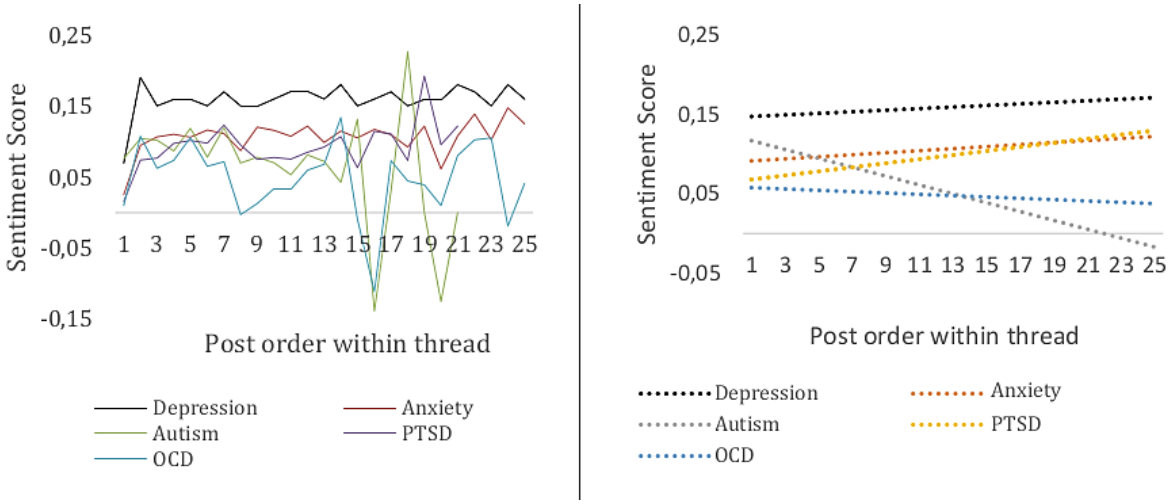

Figure 3-2 Average post sentiment score progression within threads

Based on these findings, Proposition P1a can be confirmed only for depression, anxiety, and PTSD. Thus, mental health forums for these three conditions increase the expressed sentiment in threads that originally started with a lower sentiment score.

3.5.2 Sentiment Progression of Posts by Original Posters

Across all conditions, OP posts within threads have a positive sentiment score development for every subsequent OP post made (see Fig. 3). However, only 5 out of 8 conditions show significant trends (depression, anxiety, borderline, autism, bipolar). Autism, anxiety, and depression sentiment improve the fastest, whereas depression and borderline forum sentiment improves still, however at a slower pace.

For the most part, the depression posts by OPs have the highest average score per post order. The anxiety forum exhibits the highest sentiment fluctuation per OP post; interestingly, if an OP opens a thread about anxiety and then immediately follows their own opening post by a second post (usually to provide more detail to their situation), the second post normally has a much lower sentiment score. Thus, those anxiety forum users may be feeling particularly

negative when opening a thread, making the subsequent sentiment improvement within their own thread posts that much more substantial.

Based on these results, proposition P1b is supported for 5 out of the 8 conditions tested, showing that, as OPs or advice seekers become more engaged with the thread conversations they start, their expressed sentiment increases.

In terms of improvement per OP post, in all statistically significant conditions, autism forums show the fastest positive sentiment development. This is starkly opposed to general autism posts in a thread, where autism had the worst performance. It is important to note that OCD forums exhibit the same pattern, but without as drastic a change as in the case of autism.

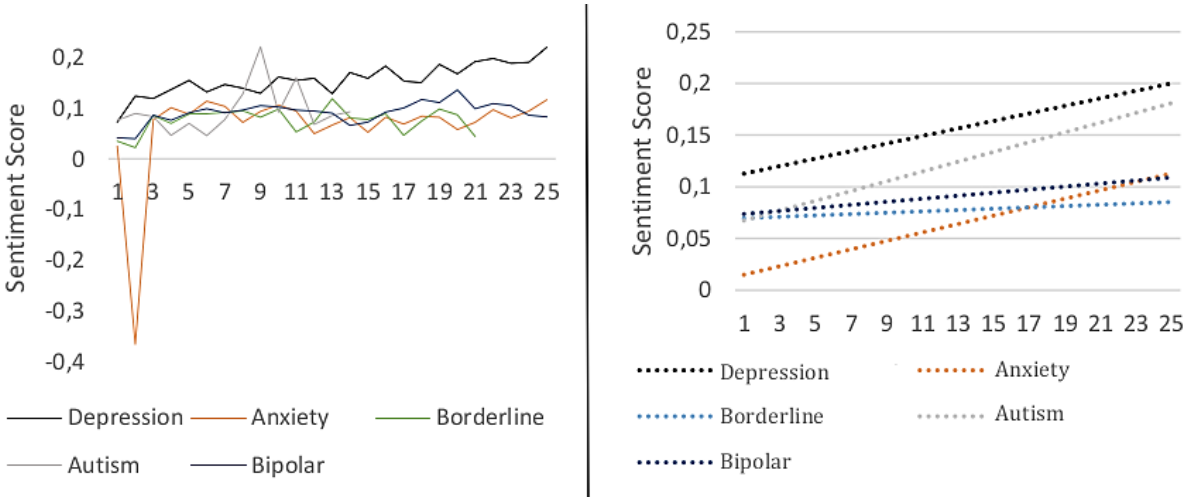

Figure 3-3 Average sentiment of OP posts within threads

Compared to OP-only thread posts, overall thread posts still do improve sentiment score with each subsequent post, however at a slower pace. The coefficients suggest that a forum participant benefits much more from OP posts. Therefore, the forum platforms could encourage users to open their own posts as a way to accelerate and strengthen the increased positivity that results from this interaction.

Additionally, many OPs (50% across all conditions) post only once within the threads they open. However, analysis results indicate that a prolonged participation in the thread which an OP starts actually improves the sentiment within OP posts. Therefore, assuming that the personality of forum members does not play a confounding role, our results suggest that users

should be incentivized or encouraged to remain active posters in the threads they create, to draw maximum benefit from their participation.

3.5.3 Sentiment Progression of Forum-wide Posts by All Users

The 4 out of 8 tested conditions with statistically significant trends are depression, anxiety, PTSD and bipolar disorder, where all show a positive change in sentiment score as a user writes more posts. PTSD marks the highest improvement rates per comment. Of all conditions, only OCD forum users show a negative sentiment score trend with subsequent posts, however the trend is statistically insignificant.

With regards to depression forums, those users with 10 or more posts express a consistently positive sentiment score. Furthermore, the sentiment improves continuously as users post more frequently (regardless of whether as OP or commenter). In the anxiety forum, those users with 42 or more posts express a consistently positive sentiment score. The results point that depression users on average achieve consistently more positive sentiment much faster than anxiety users. PTSD and OCD forum posts show much more variability and fluctuation between positive/negative sentiment. When discussing OP posts vs. thread posts vs. Forum posts (OP and commenter) – the forum-wide posts for the 4 conditions above, even though in general showing positive trend development, mark much more fluctuation in sentiment than the other two post types.

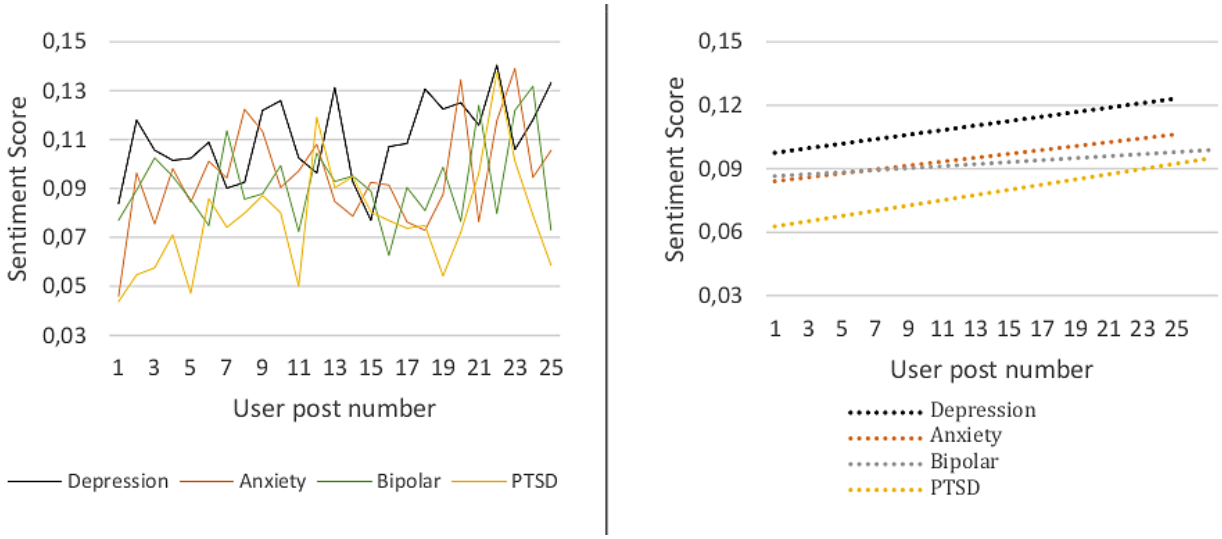

Figure 3-4 Average post sentiment score per user post order within forums

Based on the results presented, proposition P2 is supported for depression, anxiety, PTSD and bipolar forums, where exposure to mental health forums is shown to be beneficial in terms of improving a user's post sentiment per each subsequent post. For propositions P1 and P2, we can rule out familiarity as a reason for increased sentiment, as the sentiment does not appear to increase or decrease significantly faster depending on how long a user has used a forum.

3.5.4 User Roles Within Forums

Figure 5 shows the partition of users according to roles for each of the eight conditions examined (chi-square test significance < 0.01). We recognize three user roles: users who exclusively use forums as original posters (OP), users who exclusively act as commenters, and users who take on both roles. A very low percentage of users (15%) take on both roles, which suggests that a transition from OP to commenter or vice versa does not happen for most forum users. PTSD forums have the highest percentage of users who only post as OP (39%), as opposed to bipolar forums, where only 20% of all users are only OP.

Overall, around half of participants across forums use the forums only in a capacity of commenters, with bipolar forums having the highest percentage (64%) of users in that category. The results are concurrent with the reader-to-leader framework (Preece & Shneiderman, 2009) in that only few individuals make the transfer from one group to another. Looking into the connection between user roles and sentiment scores, the average OP post

sentiment across conditions is always lower than the average sentiment for commenters (see Table 3-2).

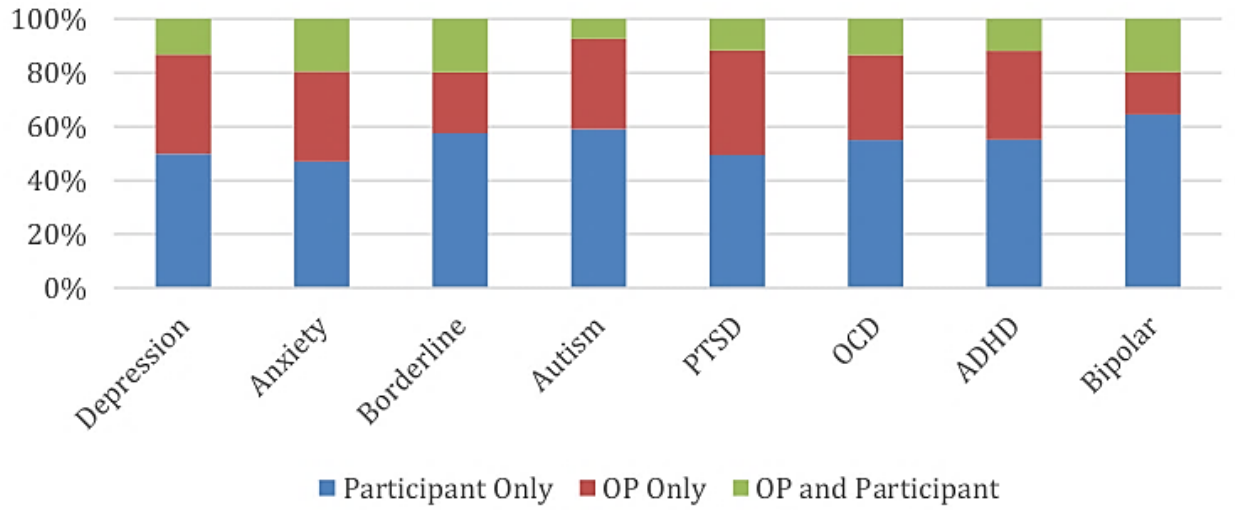

Figure 3-5 Distribution of user roles across conditions

According to the analysis results, there is a correlation between user role and sentiment score. Psychological literature supports the fact that commenters gain benefits by helping other users (van Uden-Kraan et at., 2009), which is also supported by our analysis. Therefore, forums may be able to further help users in best utilizing forum resources by encouraging participation into the questions and concerns posed by other users.

	Depression	**Anxiety**	**Border-line**	**Autism**	**PTSD**	**OCD**	**ADHD**	**Bipolar**
OP	0.13	0.07	0.07	0.08	0.04	0.03	0.06	0.08
Commenter	0.16	0.11	0.08	0.10	0.10	0.08	0.09	0.10

Table 3-2 Average sentiment per user role across conditions

3.6 Discussion and Contributions

Based on our results, in their current form, mental health forums are most beneficial for depression and anxiety sufferers for all roles examined: original poster, commenter, and general forum participant. Some, yet limited benefit was observed for bipolar, borderline, PTSD, OCD, and autism forums, and no benefit for ADHD forums. Borderline forum users benefit only by posting as OP, as opposed to OCD users who only significantly benefit from general thread participation. OCD and autism forum users are the only ones to mark a

worsening when participating in threads, specifically in the role of posters (not OP). Research into the content and discourse of the threads and a comparison with contents of other forums can uncover potential conversational and topical differences which might be the reason for opposing sentiment trends. Autism forum users only mark sentiment increase from OP threads. Consequently, autism forum users may benefit from having more options to share their personal stories. These differences in results are important to understand how current forum formats and specific features affect help seekers, and what new features may be beneficial to introduce in order to facilitate forum usage and maximize benefit for all users. The observations made in this study point to a need to provide more customized options for forum participation based on a user's condition – not every type of participation is equally beneficial for all types of conditions and such customizations could optimize a user's forum experience.

In this exploratory study, we apply automated text mining techniques (sentiment analysis) to provide evidence of the benefit of mental health forums. Specifically, we expand the practical application of text mining to mental health behaviors online, thus showing the potential of this technique to not only describe the behavior of thousands of users online, but also to shed light on the environment in which specific users benefit from forums the most. We show that not all users benefit the same, i.e. that mental health conditions as well as user roles are factors related to expressed sentiment. So far, mental health forums have not in any way been optimized or open to customization and personalization, and this paper shows that the individual users stand to gain from such a thing.

We contribute to research by exploring how engagement in virtual user-driven healthcare communities affects users who suffer from mental health conditions, and the potential to empower patients to self-management. Forum creators and administrators can learn from forum content with behavioral analytics in order to adjust forum mechanisms so that they register positive effects across all conditions, and not just a few. At the example of three of the leading English-language mental health forums, we show which forums are beneficial for specific conditions, so that forum creators can guide the future development of these platforms accordingly. We contribute to healthcare analytics research by demonstrating that machine learning and text analytics can uncover new information on user behavior to be used by practitioners and policy-makers in order to advance forum design. We specifically show that forums can in fact have a negative effect on user sentiment for conditions such as autism

or OCD, with a further important observation that a change in user roles (e.g. from commenter to original poster) can play a significant role in sentiment development over time.

3.7 Limitations and Directions for Future Research

The conducted study presents an initial investigation and, thus, needs to be understood with respect to limitations. These limitations simultaneously represent opportunities for future research.

Our study does not control for all factors internal and external to mental health forums, e.g. moderation and feature effects. Also, we did not control whether users received therapy during participation in the forums, which may have affected forum sentiment. Furthermore, using different measures can help form a more detailed picture of forum effects across conditions. Given the difference in positive vs. negative sentiment development in different conditions and for different user roles, future studies can look into the specific forum mood drivers and mechanisms on a per-condition basis. Each of the investigated illnesses are independently unique, with nuances that cannot be fully captured by sentiment analysis alone; the application of various measures and an inquiry specifically regarding forum design and moderation may help to provide more detailed answers to this question. Adopting a user perspective, future studies can also address sentiment as a function of user forum life. Finally, this study presents an application of sentiment analysis on a large dataset with sensitive personal data. This raises privacy concerns which require a detailed separate study.

The digitization of health information has created opportunities for individuals to take control of their health, highlighting the evolving socio-technical change that occurs within healthcare (Lluch, 2011). Although our findings highlight the potential applicability of machine learning within mental healthcare practice and research, our analyses are still an initial endeavor in form of an exploratory approach. Further investigation is needed to understand how and why sentiment develop the way they do, and might be a helpful undertaking to comprehend why health forums are visited so often for advice.

Chapter 4: Mining Experiences from Mental Health Forums

Title: Text Mining Mental Health Forums – Learning from User Experiences (2018)

Authors: Elena Davcheva, Technische Universität Darmstadt, Germany

Published in: Twenty-Sixth European Conference on Information Systems (ECIS2018), Portsmouth, UK

Abstract

Mental healthcare today represents a serious global challenge with not enough resources to allow for adequate patient support. As a result, many turn to the Internet for help, where mental health forums have become a rich resource of experiences shared by millions of users. This study applies aspect-based sentiment analysis on mental health forum posts in order to examine user sentiment regarding different mental health treatments. We shed light into the practices used by affected individuals to cope with mental issues and generate possible treatment recommendations.

Keywords: text mining, mental health, sentiment analysis, big data

4.1 Introduction

Mental health today has become a global issue. The World Health Organization reports that around a quarter of all years lived in disability is owed to mental disorders such as depression, bipolar disorder, or substance abuse, to name a few. Each year, around 800,000 persons in the world commit suicide, making it the second leading cause of death among 15-29-year-old individuals globally (World Health Organization, 2018a). Regardless of these numbers, with 60 countries reporting the availability of less than one psychiatrist per 100,000 population (Mathers et. al., 2004), the number of mental healthcare professionals is not nearly enough in order to adequately address the needs of all those seeking help (Kauer et. al., 2014). Of those affected, many are unable to afford professional mental healthcare, as it also happens to be the

costliest medical condition to treat (Saleem et. al, 2012). Professional psychological help usually requires patients to undergo expensive and time-consuming clinical tests and in-person interviews (Weathers et. al, 2001). In the US alone, providing mental healthcare costs $200 billion a year (Roehrig, 2016), thus causing also a burden on healthcare planners and providers. Finally, regardless of their personal financial situation, many affected individuals would not even consider professional mental healthcare because of fear of stigma and peer rejection (Crabtree et. al., 2010).

As a result of the state of the global mental healthcare system, more and more people turn to the Internet in a bid to get the help they need rapidly, affordably, and anonymously. Online forums devoted to mental health discussions are growing. In the US alone, 80% of those with Internet access use it to get health-related information, and 34% of those look up specifically personal stories from other users (Malmasi et. al., 2016). Forums are free of charge and open for anyone to join discussions, share experiences, provide answers, or ask their own questions. Users flock to forums thanks to the anonymity of the Internet, the ease of use, as well as the convenience of time and location independence (Kummervold et. al, 2002). Users feel that forum participation is a rewarding experience because it makes them feel as part of a group; they get the chance to share the burden of living with a certain condition, connect to others with similar experiences, and learn from others' successes and mistakes (Eysenbach et. al, 2004). It is a place that demands low investment, but participation can prove very rewarding.

Health forums prove to be an obligatory source of information for mental health related questions, as even when people do visit a medical professional, they still participate in forum discussions nonetheless (Kummervold et. al, 2002). Even more importantly, the age distribution of forum participants suggests that as digital natives grow up, online forums will become even more relevant in the future (Kummervold et. al, 2002). Online tools such as the forums have the potential to become a tool to address the shortcomings of the mental healthcare system and the ever-growing number of people looking for help.

Past studies show that users are more honest and more prone to sharing personal stories online (Barak & Gluck-Ofri, 2007). The data within the forum discussions is therefore a valuable source of information to researchers wishing to understand more about people suffering from certain conditions.

4.2 Motivation and Design

This study makes use of user posts on online forums about mental health and tries to makes sense of what users talk about – namely, how individuals suffering from different conditions express themselves about various topics that can have a positive influence on their condition, e.g. therapy, doctors, meditation, or sports. The goal of the study is to produce a cross-condition comparison of the sentiments expressed for these concepts. For that goal we employ aspect-based sentiment analysis based on linguistic modelling techniques for natural language.

Natural language data are both very valuable and difficult to process because of their inherent lack of structure and formality. Text mining and processing large volumes of forum posts requires constructing a sophisticated data processing pipeline which can identify complex grammar structures and word interactions within a specific language, as well as having the capacity to discern different emotional nuances in words and phrases (Saleem et. al, 2012). Such analyses are time-consuming and require much computing time and power. This paper combines the application of state-of-the-art NLP techniques onto a large dataset with a novel research question in order to advance our understanding of mental health experiences and provide recommendations to enhance future treatment approaches.

The rest of the paper is ordered as follows: Section 2 presents a detailed review of research into online mental health interactions, focusing specifically on the use of sentiment analysis and NLP techniques. Section 3 presents the methodology of the study and the data used. Section 4 presents the analysis results. Section 5 summarizes findings and limitations, as well as suggests next steps.

4.3 Literature Review

Research into the use of online mental health spaces represents a recent effort which is gaining more and more traction as the Internet becomes an important space for mental health information and communication. Even though participation in forums is not a replacement for therapy, the goal of therapy is to induce a positive change in behavior, and there is already evidence that usage of online mental health aids such as forums, social media, or chatbots, leads to changes in behavior. A few studies show that both online aids with and without medical professionals' participation can lead to comparatively effective results, particularly in

the cases of alcoholism (Riper et. al., 2014), smoking (Aveyard et. al, 2012), anxiety (Cuijpers et. al., 2009), and PTSD (Kuester et. al., 2016).

In terms of research methods, researchers have frequently employed manual techniques as tools to analyze forum conversations, such as forum user surveys or discourse analysis on a small sample of user posts. Recently, automated text mining techniques such as sentiment analysis and NLP have also been applied. Thematically, research subjects vary from classifying post helpfulness to measuring and comparing sentiment, to identifying the presence of specific content in single posts. Previous studies find online communities generally helpful for various mental conditions. Johnsen et. al. (2002) use human readers to classify mental-health forum interactions as either helpful or unhelpful. While using human readers is regarded as a reliable analytical tool, human processing speed limits the amount of posts that can be processed in the analysis – in this case to only 102, which is why automated analysis is recently introduced as a way to take advantage of available big data sets. Specifically, regarding sentiment analysis, even though it has been established as a reliable technique, there are still very few papers using the method in the context of online mental health forums. A fundamental effort by Nguyen et. al. (2014) compares the sentiment expressed on depression forums with sentiment on non-depression forums, showing that individuals who are not depressed generally express themselves more positively. Thus, the study directly supports the use of sentiment analysis (SA) as a viable tool in analyzing mental health forums. SA has also been applied in pharmacovigilance, or the identification of adverse drug effects, by automatically identifying medications with negative opinions on social media (Korkontzelos et. al. 2016) – thus granting researchers the possibility to explore drug effects over a large and diverse population. The study is also an important step in using text mining to detect possible negative effects of treatments that might generally be considered safe. Further uses of SA in mental health forums include monitoring the influence of social media messages on potential behavioral changes. Namely, through the application of sentiment analysis, Cobb et. al. (2013) showed that positively discussing certain actions which lead to quit smoking will lead to social media users actually implementing these actions in real life, a testament to the influence that online user-to-user communication has on people's life choices. Wang et. al. (2013) used SA as a classification tool to identify if a user has a certain condition – such studies may lead to future development of automated online diagnostic tools (Saleem et. al., 2012).

Coppersmith et. al. (2015a) show that the presence of positive emotion is not an indicator for the presence of a mental health condition. On the other hand, expressing a variety of negative emotions can be an indicator for all of the conditions examined in this study. Furthermore, through an analysis of social media messages, De Choudhury et. al. (2013) also show that the positive or negative affect of written language is one of the best language features to predict depression.

More recent studies have begun focusing on determining beneficial practices for specific conditions by applying content and topic analysis. Spijkerman et. al. (2016) demonstrated the benefits of meditation and mindfulness practices delivered online for people suffering from depression and anxiety by utilizing manual assessment. A similar study investigates the perception that cannabis can successfully treat ADHD – by using human coders to analyse natural text, it shows that 25% of ADHD users who have self-medicated with cannabis reported positive experiences (Mitchell et. al., 2016). As it can be seen, many of these efforts rely on manual assessment, which presupposes working with limited-size datasets. Automated NLP has also been applied in mental health forum analysis to research diverse topics such as the presence and effects of stigmatizing individuals with mental health conditions, efforts to automatically identify suicide ideations (with the goal of timely intervention and eventual prevention), topic modelling, and identifying specific emotions within user posts (Calvo et. al., 2017). Many of these efforts have focused on Twitter, where posts are limited to 140 characters. In terms of data size, the research field has yet to make proper use of the large amount of data available online. Johnsen et. al. (2002) used only a small sample of a one-month dataset to be processed and analyzed by human readers. Mitchell et. al. (2016) uses only 55 threads and 401 posts altogether.

To the best of our knowledge, there exist no previous studies which have attempted to create a cross-condition comparison of potentially helpful treatments for mental health via aspect-based sentiment analysis. Previous studies also tend to focus on single conditions or specific features, while the goal of our study is to compare several features across several conditions.

4.4 Dataset and Methodology

We use aspect-based sentiment analysis (ABSA) to determine the sentiment forum users express on various aspects or concepts such as family or doctors, which can potentially help

them on their path to healing. SA can be subject-dependent and subject-independent (Wang et. al., 2013). A subject-independent SA measures the sentiment of chunks of text, e.g. a sentence or a post. On the other hand, subject-dependent SA measures the sentiment expressed regarding a particular subject. The latter is also known as ABSA and is based on parsing natural text through linguistic NLP dependency parsers that match subject words with other words that directly describe or relay a quality of the subject words. For example, after parsing a simple sentence such as "These apples are green", an NLP dependency parser would return the pair (apples, green), where "apples" is recognized as the subject, and "green" as the descriptive word.

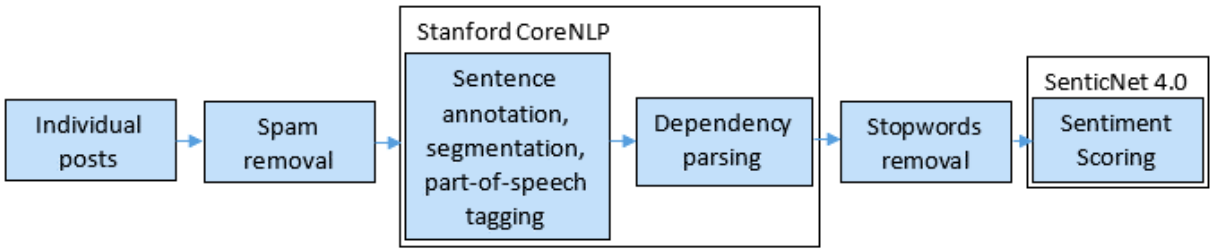

Figure 4-1 Analysis procedure

In this study, our subject words are all words connected to a concept whose sentiment we want to measure, as seen in (Wang et. al., 2013) and other studies. The concepts we measure SA for are family, medications, therapy, pets, sports, and meditation. Each of the concepts is a sum of sentiments expressed for all words related to itself, e.g. the sentiment for family is the sum of sentiments for all family-related words, e.g. family, parents, siblings, mother, sister, son, etc. For the specific calculation of the sentiment score, we use the SenticNet 4.0 dictionary by following the approach of Taboada et. al. (2011). That is, apart from using word scores, we also account for intensification (e.g. "bad" vs. "very bad") and negation (e.g. "good" vs. "not good"). Figure 1 describes the analysis process used in this paper.

The first step in the data processing is removing spam, or in our case – posts that have been quoted in later replies. Then the posts are tokenized (words are replaced with numeric representations), segmented into sentences, and tagged with part-of-speech information, e.g. noun, verb, adjective. We use the Stanford CoreNLP models for the tasks of tokenization, part-of-speech tagging, sentence segmentation, and dependency parsing. Namely, CoreNLP provides an English-language syntactic dependency parser based on a recurrent neural networks model. The dependency parser is a crucial part of this analysis, as it models the

grammatical structure of a sentence and provides information as to relationships between different words (Neural Networks Dependency Parser).

Relation	CoreNLP Definition	Example Sentence
Adjectival modifier	any adjectival phrase that serves to modify the meaning of a noun	Sam eats red meat (meat, **red**)
Nominal modifier	nominal dependents of another noun or noun phrase that functionally correspond to an attribute or genitive complement	The Chair‘s office (chair, **office**)
Nominal subject	a nominal which is the syntactic subject and the proto-agent of a clause	The baby is cute (cute, **baby**)
Open clausal complement	predicative or clausal complement	He says that you like to swim (**like**, swim)

Table 4-1 Stanford CoreNLP dependency relations used in this study

The neural network models and rules are described in-depth in the paper by Chen and Manning (2014). In this way, for each concept and its related terms we can put together a sub-selection of relevant descriptive words. The sentiment of each of the 7 concepts examined in this paper is calculated as the average sentiment (from SenticNet) of all descriptive words for that concept as identified by the dependency parser; i.e., for each concept SC_i the sentiment is found by the formula $SC_i = \frac{\sum SC_j}{n}$, where n is the number of related terms to that concept, and SC_j is the sentiment score of an SC_i related term as provided by SenticNet. The syntactic relations we chose to focus on are presented in Table 4-1 with their respective CoreNLP documentation definition. These relations output word pairs where a subject word is described or modified by another word – these relations contain the most relevant word pairings for the needs of sentiment scoring.

We further focus our choice of words by taking into consideration only nouns, adjectives, and verbs within the above relations. This measure reduces noise by removing words in the above relations which do not bear any descriptive qualities. As a result, sentiment accuracy will be improved. The SenticNet dictionary version 4.0 is what enables the aspect, or word-based sentiment scoring. In a nutshell, SenticNet is a list of 50,000 English-language words and their appropriate sentiment scores. We use SenticNet following the example of Wang et. al. (2013). Scores are adjusted if words are preceded by a negation. Three of the leading English-

language mental health forums were scraped to create a combined set of 132,072 threads containing 1,155,403 individual posts across 12 conditions (depression, bipolar disorder (BD), anxiety and panic attacks, schizophrenia, attention deficit hyperactivity disorder (ADHD), Asperger's Syndrome, Borderline personality disorder (BPD), obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), self-harming, substance abuse). The data were extracted in August 2017 and encompass all publicly available posts on the respective websites.

The forums have administrators and moderators whose task is to make sure conversations do not go off-topic; thus, we can be sure that in our research we are considering discussions relevant to each condition. Additionally, moderators remove offensive or damaging material (e.g. posts that encourage self-abusive behavior). However, the role of moderators is not to provide advice, as the goal of a mental health forum is to be a place of discussion among the users, and not between a user and a medical professional, i.e. a forum is not meant or seen as a tool to replace established medical practices such as therapy.

4.5 Results

In terms of conditions, bipolar disorder forum users have expressed the highest sentiment across conditions (average 0.16), whereas autism forum users have the lowest average score of -0.04. The autism forum posts are the only ones to score an average negative sentiment, while simultaneously expressing the best sentiment for family and pets, while scoring meditation and spirituality the lowest. Clearly distinguishing family and pets as a positive presence suggests that emotional support by loved ones is very important for people with autism. This finding is backed by psychological research (Solomon, 2010) as well as recommendations of leading organizations as to the benefits of pets for autistic individuals (Autism Speaks, 2014).

Concept / Forum	Family	Sports / Exercise	Meditation / Spiritu-ality	Pets	Therapy	Medication	Medical Professionals
Anxiety	0,086	0,077	0,184	0,098	0,225	0,075	0,123
ADHD	0,109	0,215	-0,016	0,117	0,255	0,047	0,144
Depression	0,074	0,090	0,179	-0,067	0,186	0,116	0,109
Asperger's	0,082	0,102	0,150	-0,021	0,162	0,055	0,176
OCD	0,037	0,097	0,051	0,048	0,233	0,150	0,145
BD	0,088	0,154	0,250	0,109	0,263	0,071	0,163
BPD	0,098	0,137	0,242	0,060	0,180	0,071	0,148
Self-harm	0,078	0,186	0,006	0,092	0,263	0,117	0,149
PTSD	0,062	0,030	0,063	-0,086	0,199	0,108	0,050
Schizo-phrenia	0,047	0,120	0,146	0,069	0,188	0,128	0,145
Subs. Abuse	0,064	0,320	-0,063	0,090	0,287	0,083	-0,023
Autism	0,176	0,018	-0,355	0,149	0,0007	-0,161	-0,089

Table 4-2 Average sentiment per concept across mental health forums

Users in the substance abuse as well as self-harm forums have distinguished sports and therapy as the two most positive concepts. Substance abuse posters negatively score meditation, as well as medical professionals other than therapists, which again draws similarities with self-harm, where meditation is the worst-performing concept.

Although all concepts score positively in the schizophrenia forums, therapy is most positively regarded, while family – the least. This may indicate that dealing with loved ones with this condition is not undesirable, but it remains challenging. Literature on the matter considers family relationships a necessary part of schizophrenia treatment, but a difficult and complex one (Motlova, 2007). Forum users express an extreme range of emotions regarding family – from a need for understanding and support to disinterest and rejection, however detailed research into the conversations is needed to extract signals as to ways in which a schizophrenic can better communicate with their closest ones while undergoing treatments. Regarding PTSD, therapy stands out as positive, while pets are the single negative concept. This finding contradicts a long-established practice to treat PTSD with pets such as dogs or horses (Altschuler, 1999). Looking more closely, the low sentiment score in our dataset is

owed by the fact that many of the PTSD forum users report the loss or death of pets as one of several reasons that have triggered PTSD.

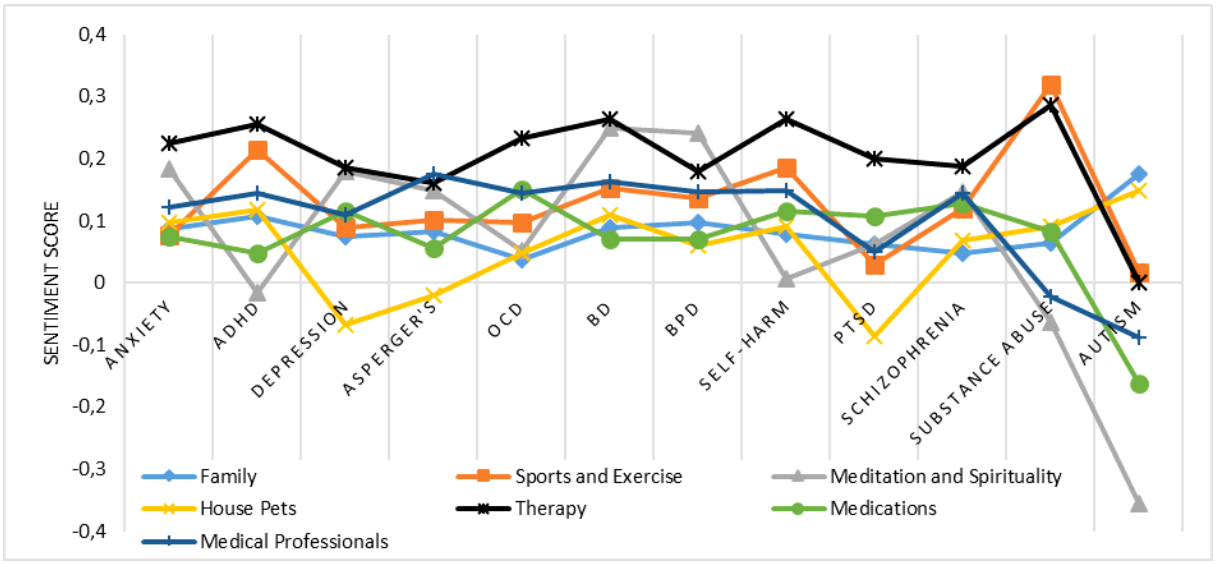

Figure 4-2 Sentiment scores

BPD forum posters talk prominently most positively regarding meditation and spirituality. Meditative and spiritual healing practices vary widely between conditions, with BD and BPD individuals finding it the most helpful, the lowest scores coming from ADHD and substance abuse forums users. Despite some research efforts on the subject (Perich et. al., 2013), meditation is not yet established as a positive treatment for BD, and has not been considered for BPD. BD forum users highly recommend meditation to those seeking help and advice, characterizing it as “particularly helpful”, and report that meditation helps them in controlling the mood swings associated with the condition. BPD forums users report meditation to be a tool for avoiding psychotic states and dissociation. Most users report picking up meditation online, by using YouTube videos or mobile apps.

In terms of concepts, pets have the lowest sentiment score across all conditions, with an average score of 0,05 and a negative score in 3 out of the 12 conditions. PTSD and depression forum users often mention the loss of pets as a condition trigger. On the other hand, many in the Asperger’s forum share a genuine dislike of pets. Therapy is by a wide margin the most positively talked-about concept of all with an average sentiment of 0,2, and a highest sentiment score in 8 out of 12 conditions. This is a testament that professional one-on-one conversations are still the best treatment for most conditions, however this is not a rule of

thumb, as 1/3 of forum users in all conditions surveyed feel more positive about another treatment. Therapy does score considerably lower within autism, having a neutral score of almost 0.0, despite many different types of therapy for autistic individuals, including speech, sensory, occupational, and cognitive-based therapy, to name a few (Blumberg et. al., 2016). This is backed by the fact that about 13% of autistic individuals ever lose their diagnosis later in life, after going through the rigorous therapy plan (Blumberg et. al., 2016). Therefore, implementing therapy alternatives such as family support, house pets, and even physical activity, may be more beneficial for people with autism, as suggested also by our results. Some on the autism forum report being "hurt" by certain types of therapy, and many say that therapy was helpful but "not too much".

Looking into sports and exercise, people with substance abuse, ADHD, or self-harm, stand to gain the most out of being physically active. For those with substance abuse this is the number one positive concept, and for ADHD individuals this is especially sensible, as being physically active plays into their hyperactivity (Lufi & Parish-Plass, 2011). There is not much in the literature as yet connecting substance abuse and self-harm with sports as a possible remedy for these behaviors.

Sports and meditation are concepts that score significantly high scores in certain conditions, which indicates that it is worthwhile to further investigate the possibility of incorporating these practices in the appropriate condition therapies also from a formal and professional point of view. On the other hand, getting pets appears undesirable across the 12 conditions, especially for Asperger's forum users. Autism users are the exception when it comes to pets, where besides family, pets are the leading positive concept. It is also important to note that medications, albeit not negatively scored, are also not regarded with significantly high sentiment, consistently scoring significantly lower than therapy. Medications have the lowest sentiment among concepts within anxiety and BD forums, suggesting that these conditions gain the least advantage from using medication. Anxiety forum users report having mild anxiety even when using medications. An exception in this case is only the OCD forum, where medications are the second most positive concept.

Furthermore, in order to reveal the similarities between scores of different conditions, we conducted a two-tailed Pearson correlation analysis. Anxiety is strongly correlated with BD (Pearson 0.923, $p<0.01$) and BPD (Pearson 0.771, $p<0.05$). These findings are supported by

psychological literature, as anxiety has high comorbidity with both conditions (Keller, 2006) (Zanarini et. al., 1998). Furthermore, ADHD is correlated to self-harming behavior (Pearson 0.922, $p<0.01$) as well as substance abuse (Pearson 0.826, $p<0.05$). The link between these two conditions and ADHD has also been well-documented (Wilens, 2004). This suggests that these conditions may gain similarly positive or negative results from the same treatments.

It is interesting to note that self-harm and BPD – two conditions usually associated with each other (Chapman et al, 2005), are strongly uncorrelated in our analysis. The difference is owed to the sentiment expressed on meditation and spirituality. In this case this may indicate that even though self-harm and BPD have a high comorbidity, different concepts or treatments may still elicit different responses.

4.6 Method Evaluation

Evaluating the accuracy of dictionary-based sentiment analysis requires evaluating the sub-tasks of aspect term extraction as well as aspect term sentiment evaluation. Both task results were evaluated using the standard metrics precision and recall (Salas-Zárate et. al., 2017). The system-generated aspect pairs and scores are checked against a subset of 2500 human-annotated pairs from the forum data. In the first sub-task, the goal is to make sure that aspect terms are related to and relevant for the subject term. In the second task, even though the SenticNet dictionary comes with word sentiment scores, it is necessary to check whether the aspect term scores make sense in the context of the subject term evaluated as well as the forums; for example, the pair (therapy, continued) was scored with a -0,04, while in context it had a more positive meaning.

The precision measure for aspect term extraction was 71.83%, and recall was 70.48%, whereas for sentiment scoring it was 78.12% for precision and 75.49% for recall. The lower measures regarding aspect term extraction as a sub-task signal that rules for extracting descriptive aspect terms specifically meant for a SA use must be narrowed down more precisely than those presented in our study. Nevertheless, the measures are encouraging when compared to other ABSA accuracy scores and render reliability to the study results (Da Silva et. al., 2014).

	Anxiety	ADHD	Depress.	Asper-ger's	OCD	BD	BPD	Self-harm	PTSD	Schiz.	Subs. Abuse	Autism
Anxiety	1	,157	,625	,618	,473	,923**	,771*	,240	,568	,716	,027	-,380
ADHD	,157	1	-,002	,194	,559	,211	-,072	,922**	,256	,247	,826*	,568
Depres-sion	,625	-,002	1	,850*	,533	,653	,766*	,210	,866*	,769*	,032	-,661
Asperger's	,618	,194	,850*	1	,479	,714	,808*	,284	,655	,723	-,048	-,495
OCD	,473	,559	,533	,479	1	,382	,160	,823*	,756*	,799*	,449	-,145
BD	,923**	,211	,653	,714	,382	1	,907**	,243	,458	,758*	,116	-,463
BPD	,771*	-,072	,766*	,808*	,160	,907**	1	-,043	,430	,657	-,124	-,658
Self-harm	,240	,922**	,210	,284	,823*	,243	-,043	1	,499	,498	,802*	,330
PTSD	,568	,256	,866*	,655	,756*	,458	,430	,499	1	,687	,278	-,312
Schizo-phrenia	,716	,247	,769*	,723	,799*	,758*	,657	,498	,687	1	,197	-,622
Subs. Abuse	,027	,826*	,032	-,048	,449	,116	-,124	,802*	,278	,197	1	,436
Autism	-,380	,568	-,661	-,495	-,145	-,463	-,658	,330	-,312	-,622	,436	1

Table 4-3 Correlation Analysis (** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed))

$$Precision = \frac{True\ Positives}{True\ Positives - False\ Positives} \qquad Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Figure 4-3 Precision and recall formulas

4.7 Conclusion

Mental health forums are a vast data source of information that formal psychology has not yet tapped into. This paper shows how natural text processing of large datasets can automate and accelerate data collection, processing, and insight generation from millions of posts by tens of

thousands of individuals. Being able to back up many of the findings of this study with previous findings in psychological literature and in a few cases also practice, demonstrates how and why text mining of large public datasets is a viable and dependable tool that can bring about a veritable change in how we research mental health, and how we approach the question of improving and administering therapy and treatments.

The study singles out meditation and spirituality practices as well as sports and exercise as helpful practices in a variety of conditions such as BD and BPD. Although they are found to be positive by forum users, these practices have yet to become an established part of formal treatments. The results point to shortcomings of therapy in autism. As next steps the authors will take a deeper look into how exactly these concepts are being discussed within the forum data, in order to provide precise and detailed answers as to why sentiment scores appear to be what they are in this study. We will furthermore examine user-related variables such as location, mental health expertise and similar, in order to better understand the progression of sentiment. Finally, since autism is the most complex condition we look into in this study, there is much space to conduct a separate study focused only on this condition.

Although users express themselves positively for some concepts tested in this study, the current study does not measure exactly how effective users found these concepts to be in regards to their specific condition. More detailed linguistic analysis needs to follow in order to assess degree of effectiveness, as well as exact symptoms alleviated by each concept. Finally, this study raises the need to investigate the transfer of forum advice into the real world, specifically in terms of trust issues that may arise in an impersonal online environment (Benlian & Hess, 2011), as well as in terms of appropriate interface tools to aid users in finding the information that is pertinent to them (Benlian, 2015).

Chapter 5: Automating Symptom and Condition Classification with Neural Networks

Title: Classifying Mental Health Conditions Via Symptom Identification: A Novel Deep Learning Approach (2019)

Authors: Elena Davcheva, Technische Universität Darmstadt, Germany

Published in: Fortieth International Conference on Information Systems (ICIS 2019), Munich, Germany

Abstract

An increasing number of individuals around the world suffer from mental health issues but are unable to access professional help as resources are lacking on a global level. Information technology and especially machine learning has shown great potential as a basis for automated digital services that can act as a support tool for affected individuals. This study presents a state-of-the-art deep learning model for multiclass classification of mental health conditions based on a novel approach of prior classification of symptoms. We contribute to existing research on IS applications in healthcare by improving upon the performance of similar previously reported models, as well as showing that unstructured text can be used to reliably extract not only a primary but also a secondary condition classification. We show that classifying symptoms of individual conditions first and based on that result extracting conditions leads to better model performance.

Keywords: mental health, machine learning, neural network, social media

5.1 Introduction

We live in a world where more and more people struggle with mental health problems. Currently, 700 million people around the globe are affected by mental health issues. In China only, there was a 25% increase in the number of individuals seeking help for mental health problems between 2014 and 2016 – a total of 173 million. Only 20 million of those have been

treated professionally (World Economic Forum, 2019). At least one in six people within the European Union are affected by mental health issues (OECD, 2018). Not only does this adversely affect individual well-being, but also it is very costly for national healthcare systems - mental healthcare within the Union is estimated to cost at least 600 billion euros, which is 4% of GDP. What is more, the deterioration in individual mental health adversely affects society and economy as well (World Economic Forum, 2019). This global mental healthcare crisis in recent years has spurred an international effort in academia and practice to create automated tools for mental healthcare, capable of reaching an unlimited number of users around the world. These efforts mark an increased interest in automated triaging (classifying the severity of a mental condition), automated mental health condition classification, as well as a proliferating number of apps offering psychological diagnostics and chatbots that attempt to mimic the experience of therapy. The need for more and better digital mental health services and solutions is fueled by the fact that mental health worldwide is costly for individuals to address, it is still associated with social stigmas, and it is underserved with resources and trained professionals (Saxena et. al., 2007).

On this backdrop, information technologies have shown potential to fill in an acute need with services where deep learning and specifically natural language understanding (NLU) techniques have shown great potential (Wahle et. al. 2017) for automated diagnostics and treatment. However, automatically diagnosing mental health conditions as well as their symptoms still presents an open challenge for researchers and practitioners. Prior research has shown that automatic mental health condition classification is possible, however conditions themselves are highly complex – some have a severity spectrum, such as autism, others have different forms, such as ADHD (attention deficit and hyperactivity) vs. ADD (attention deficit, but no hyperactivity). Many times the same symptom (e.g. depressive feelings) can be present in several conditions.

Texts written by individuals in an attempt to explain how they feel contain rich information, including identifiers for specific symptoms. Thus, an automated classification can leverage this information to capture not only the overall condition, but also the specific symptoms. This study presents a novel approach to classifying conditions by classifying underlying symptoms of a mental health condition first, and classifying the condition based on them. Based on the symptoms, a primary and a secondary condition classification also become possible. To the best of the authors' knowledge, this is the first attempt to classify mental

health conditions based on symptom modelling first, as well as one of the first studies to classify a secondary condition.

There are many benefits of being able to automatically deduce both the symptoms and the condition expressed by a forum user. For example, by looking into symptoms, an IT service is able to let a user know which forum is most suitable for their question or message (as many forum users are not diagnosed by a professional). Furthermore, knowing that someone has a condition (e.g. depression) does not also show if this individual is experiencing a specific symptom such as suicidal ideations – being able to capture these details will allow an automated service to offer urgent and concrete help. Additionally, being able to classify both the condition and the symptom from user text input opens new avenues for mental health chat assistants – while not a formal clinical diagnosis, a chatbot with the ability to recognize not only possible conditions, but even symptoms, can be much more helpful in pointing the user to emergency help, online resources, as well as specialists in their local area specific to that user's input. A combined condition and symptom modelling can also be greatly beneficial for screening, i.e. if a user feels discomposed but is unsure if he or she is experiencing signs of a mental illness.

In this exploratory study, we apply a state-of-the art neural network model to classify mental health symptoms based on textual input from a user, and based on the symptoms to then derive a condition. A mental health condition can be "depression", and indicative symptoms could be "fatigue", or "disinterest". In order to build a model, we collect data from online mental health forums. These forums are one of the most popular digital tools for mental health, where affected individuals discuss their mental health issues. Users visit the forums in order to share personal experiences, ask for advice, and offer support and solutions for others affected. The anonymity provides forum users with protection from social stigma; in addition, studies suggest that people tend to be more open and honest when discussing personal mental issues online (Barak & Gluck-Ofri, 2007). Given the quantity and type of data available, all this makes forums a good place to collect data.

The proposed approach uses a neural network to model and classify symptoms, and conditions are classified based on the presence and probability of symptoms which are part of a specific condition. The approach is tested on a dataset with three different mental health conditions and achieve high accuracy on both condition and symptom classification, thus

succeeding to deeply enrich the information we get from the model. The model performance is achieved without adding complexity, in other words, model features are learned independently of external tools like the LIWC (Linguistic Inquiry and Word Count) toolkit, which is a popular resource for studies modeling emotions in text (Goldbeck, 2016). Thus we manage to extract much more information from our model while simultaneously simplifying the data preparation process.

The rest of the paper is set up as follows: section 2 presents previous studies in the area of automatic mental condition classification; section 3 presents the methodology of the study, section 4 presents experiment results; section 5 provides a discussion of findings, and section 6 provides a conclusion.

5.2 Background

5.2.1 Forums as a Mental Health Tool

Mental health forums have long been used as a viable source of data for research forays into digital mental health (Ali et. al., 2015). Research into the effects of forum participation shows evidence of positive effects on the mental state of its users. For example, users of forums show a high degree of disinhibition, or a willingness to share personal information, encouraged by the anonymity that these platforms provide. This kind of anonymity was surprisingly found to elicit more emotional and engaging comments and reactions. (De Choudhury et. al., 2014). Furthermore, sentiment analysis of forum posts shows gradual improvement in the sentiment that forum users express when using the forums for a longer period of time (Davcheva et. al., 2019). Thus, past research shows that mental health forums can have a positive influence on users and can truly be an emotional support tool.

At the same time, the tools that forums currently provide to users outside of a typical conversation are very limited (e.g. links to general information online), when in fact the plethora of comments hide potential for a much wider variety of digital support mechanisms. By automatically examining conversations, texts, etc., forum platforms can connect similar users into support groups, point a user to conversations that may be useful to them, suggest resources (online as well as in their local area) pertinent to the specific symptoms they have expressed, and more. Deep learning can help unlock the potential of forums as mental health support platforms.

5.2.2 Use of Machine Learning in Mental Health

Machine learning as well as deep learning (neural network models) in particular is increasingly applied in the area of mental health. Recent efforts have focused on exploring lexical features of mental health texts on social media, triaging (categorizing the severity of a person's mental state), and condition classification. To the best of the authors' knowledge, so far there has been no study using mental health text datasets to classify a primary and secondary condition based on symptoms extracted using a neural network.

Past studies have already begun to show the potential of applying machine learning within mental health. Cohan et. al. triage the severity of online forum posts into four different categories based on ideations for self-harm expressed in a user's forum post (Cohan et. al., 2017). Being able to perform automatic triaging can allow forum moderators to quickly identify users in need and perform an urgent intervention. For example, if the triaging of a suicidal user shows that this user is in a severe mental condition, the moderator can point them to a crisis center. The authors train a tree boosting model with data from online mental health forums and achieve an F1 measure of 75%. The model is built entirely on heavily engineered features such as the emotions expressed in a post, the level of subjectivity in a post, or post topics mined by a separate topic modelling algorithm. To create these features, the study depends on external data sources that may not necessarily be specific to psychological data, as well as forum metadata such as total number of user posts etc. To collect such features, substantial feature engineering is required. This approach presents a situation where in order to triage content, one must first build a very elaborate data pipeline in order to first create or extract all these necessary features. With neural networks, the need to engineer features separately from the data to be modelled is eliminated, as the network automatically learns to model a specific feature, in other words, a deep learning model does not specifically need pre-derived features from text, rather it expects only labelled data.

In another study on triaging, researchers used a mix of supervised learning and rule-based classification to triage mental health forum posts (Almeida et. al., 2016). Their approach also relied on pre-engineered features such as part-of-speech tags or n-grams, as well as a sentiment dictionary. However, rule-based classification has the drawback that it is rigid and generates classifications according to rules, thus the capacity of such a system to learn from

new data is very limited. Furthermore, rule creation requires time-consuming manual work as well as deep domain expertise.

In terms of classification, only in the past 1-2 years has automatic classification of mental health conditions been studied more rigorously. Many studies describe classifiers that can distinguish between text exhibiting no mental health distress versus texts showing indications for only one specific condition (e.g. ADHD vs. no-ADHD), or one specific symptom such as suicidal ideations (suicidal vs. not suicidal) (Guntuku et. al. 2017). Many of these efforts still rely on structured data from formal medical sources as modelling input. For example, Walsh et. al. (2017) used random forests and logistic regression models to predict suicide risk based on electronic health records data. Oh et. al. (2017) used neural networks to identify and classify suicidal ideations by asking survey participants to complete an elaborate self-report consisting of 31 psychiatric scales as well as additional questions regarding participants' history of suicide attempts. Such rich, detailed, and structured input (coming from a survey) is usually not available for users of online mental health services, including forum users. A survey also allows to directly query information needed for a diagnosis or to pinpoint symptoms – another advantage normally not available on online services that allow users to express themselves freely and without constraints. These types of services need models that can use unstructured data and are able to extract specific knowledge from it.

Duda et. al. (2016), show that machine learning can be applied to automatically differentiate between autism and ADHD with high performance and based on a small number of symptomatic behaviors. These types of models can perform "preliminary risk evaluation and pre-clinical screening and triage that could help to speed the diagnosis of these disorders". However, they also rely on structured input data in the form of a questionnaire.

A few recent studies have shown that deep learning models can learn to model a mental health condition automatically from unstructured data. One such study is presented by Tran and Kavuluru, who used recurrent neural networks to predict psychiatric conditions for a patient based on their history of present illness (Tran & Kavuluru, 2017). The history was presented in a short note as well as questions indicating whether an illness was present in the past or not. The history of present illness - a short text with 300 words on average – proved to be a good predictor for some conditions such as depression, anxiety, or ADHD. The applied recurrent neural network outperformed previously studied methods. This study confirms that it is

possible to train a neural network to differentiate between mental health conditions based on unstructured short texts. In another example, Orabi et. al. (2018) used Twitter data to classify depressed vs. not depressed users. Using a convolutional neural network model, they achieved an F1 measure of 87%. This is one of the best-performing models so far reported in literature on the subject of mental health condition classification.

Gkotsis et. al. (2017) have employed a two-step approach for their multiclass model in that first a comment is classified as either containing signs of a mental disorder or not, and only then proceeding to classify within one of 11 considered disorders. Both this model as well as the one by Almutairi et. al. (2018) deal with a binary classification use case (e.g. depressed vs. not depressed). Delahunty et. al. (2019) propose a neural network approach that uses a screening process for four different symptoms indicative of depression and anxiety (by using a clinical questionnaire), and only then proceeding to using a neural network classifier. Ive et. al. (2018) on the other hand test novel neural network mechanisms in order to show that their utilization improves the performance of a baseline RNN model.

5.2.3 Performance Comparison

Comparing different machine learning approaches for mental health condition classification can be challenging as many studies only present binary models which are not directly comparable to a multiclass approach such as the one presented in this study. Additionally, studies usually measure performance with accuracy or an F1 score, while not breaking down for precision, recall, or other metrics, making it challenging to compare model performance. This section presents a broad comparison of the latest models from literature which address mental healthcare condition classification by means of modelling text data.

Study	Classification type	Model	Metric	Performance
Orabi et. al. (2018)	Binary (depression)	Convolutional neural network	F1	87%
Almutairi et. al. (2018)	Binary (schizophrenia)	Support vector machine	Accuracy	90%
Delahunty et. al. (2019)	Multiclass	Neural network	Accuracy	43%
Gkotsis et. al. (2017)	Multiclass	Convolutional neural network	Accuracy	71%
Ive et. al. (2018)	Multiclass	Recurrent neural network	F1	76%

Table 5-1 Models performance comparison

5.2.4 Research Gap

Machine learning applications within mental health are generally observed in four areas: diagnostics, treatment, public healthcare, and clinical research (Shatte et. al., 2019). This study falls into the area of diagnostics. The focus within this area in terms of previous research conducted so far has specifically fallen on depression detection, suicidal ideations, as well as cognitive decline (Wongkoblap et. al., 2017).

There is a recognized research gap in the literature when it comes to diagnosing disorders other than depression, which we directly address with this study by evaluating a model that can differentiate between several conditions. We furthermore contribute to closing the existing research gap in collaboration between mental health professionals and machine learning engineers, as prior studies note that a greater cohesion is required between professionals of both fields in order to arrive at optimal digital mental health solutions (Shatte et. al., 2019). We do this by consulting with a psychologist in terms of the quality and reliability of the data that we use to train the proposed model. More importantly, we follow clinical practice in that we first look at the symptoms, or the smaller components of a mental health condition classification, and then proceed to build a big picture by extrapolating a condition label.

This paper further contributes to the existing literature of IS applications in healthcare by showing that unstructured text can successfully be used within automated applications meant for psychological diagnosis. Many previous studies still used structured data input. These

approaches are unfit for use in a digital service such as chatbots. We further contribute to existing literature by providing a state-of-the-art multiclass model in the field of automated mental health diagnostics – based on Shatte et. al. (2019), most machine learning applications within mental disorder detection have been binary in nature – i.e. a classification would show whether a person has a certain condition such as depression or not. This study expands upon these research efforts by presenting a multiclass model, i.e. our model differentiates between multiple possible conditions. While there are many models addressing binary classification of mental health conditions, classifiers capable of differentiating among multiple conditions are as yet few.

Finally, the study takes an original approach in classifying conditions by first classifying the underlying symptoms for each condition. This allows for a comorbidity classification – i.e. a classification of both a primary and a secondary condition, which to the best of the author's knowledge has not been reported before.

5.3 Methodology

5.3.1 Data Preparation and Labelling

As a first step, a dataset of approximately 50,000 forum posts is scraped from several English-language mental health forums using the Python library Scrapy. The dataset contains posts, username and date posted, as well as post order within a thread. Next, a subset of 3,551 posts is manually selected from 554 users who self-disclose a diagnosis made by a medical professional. These posts are later used to train the model since they are more reliable

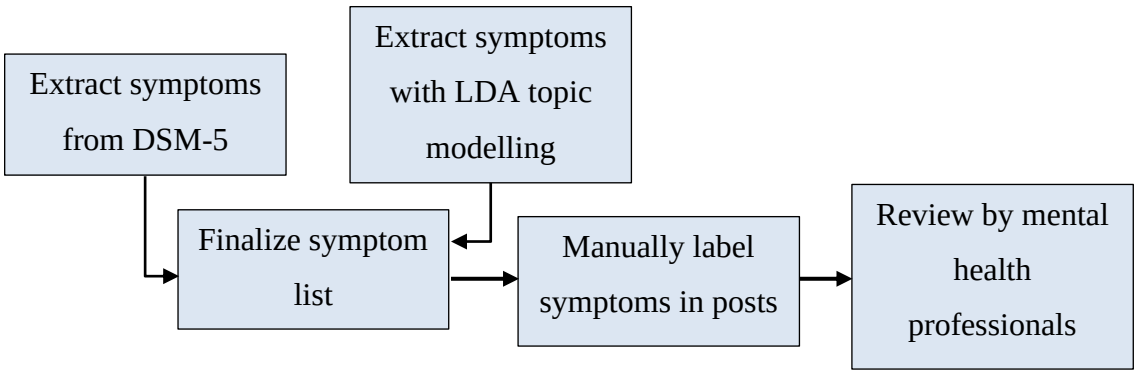

Figure 5-1 Data labelling process

representatives of a certain condition than any forum post in general, as many forum users do not have a formal diagnosis (Coppersmith et. al., 2015b). The data are labelled only for symptoms and not for conditions. In order to carry out the labelling of symptoms, we conduct the process depicted in Figure 5-1.

The symptom-labelling process starts by extracting symptoms for each of the three considered conditions from the Diagnostic and Statistical Manual of Mental Disorders, fifth edition (DSM-5) (American Psychiatric Association). The DSM is a manual compiled and published by the American Psychiatric Association used by mental health professionals around the world to diagnose patients, therefore it is an excellent source of symptom information, as it is very recent, rigorously updated, and succinct (Cavazos-Rehg et. al., 1996). However, the language of the DSM-5 is highly formal and medical, while forum communication is colloquial and casual. For example, for schizophrenia, one symptom from DSM-5 is "anhedonia" (feeling of apathy), however forum users would express this sentiment with more colloquial phrases, such as "nothing gives me pleasure". This difference in the verbiage creates a gap in the way DSM-5 would word a symptom, and the way this symptom would be discussed by forum users. To bridge the gap, we additionally conduct unsupervised topic modelling with an LDA (latent Dirichlet allocation) algorithm, in order to find alternative expressions that users use to designate their symptoms. LDA is an unsupervised bag-of-words generative probabilistic algorithm which represents posts as mixtures of topics, and topics are represented by a distribution over words (Blei et. al. 2003).

LDA processes each post p as follows:

1. Decide on a number N of topics to be extracted per post
2. For each post, assign each word randomly to one of the N topics, thus building topic representations as well as word distributions
3. Improve on the representations and distributions by going through each word and:
    a. for each topic t, calculate:
        i. p (topic t | post d), or prevalence of words in post d assigned to topic t
        ii. p (word w | topic t), or prevalence of assignments to topic t from all posts based on the current word w. Select a new topic t for word w, where topic t has probability p (topic t | post d) * p (word w | topic t)

In step three, the algorithm operates with the assumption that all topic assignments are correct, with the exception of the current word, then updates the word's assigned topic via the LDA topic representations and word distributions. Step three is iterated until satisfactory performance is achieved.

We apply LDA with the Python package lda, using a collapsed Gibbs sampling as a parameter fitting technique. The result of LDA is not a direct word for the topics – rather, words in a post are grouped in three clusters, and then it is up to the user to name each topic cluster. For example, a frequent cluster in schizophrenia is made up of the words "voices", "hearing", etc. Thus, we find that "voices" is a very frequently used way for schizophrenia forum users to talk about experiencing auditory hallucinations. In other words, our symptom labeling is derived both from professional diagnostic tools, as well as from LDA in order to account for colloquial verbiage from forum users, thus covering a wide array of possible symptom expressions.

ADHD	**Schizophrenia**	**Depression**
Hyperactivity	Delusions	Depressive mood
Lack of focus	Visual hallucinations	Anhedonia / apathy
Attention problems	Auditory hallucinations	Weight and appetite fluctuation
Forgetfulness	Disorganized speech	Fatigue
Disorganization	Anhedonia / apathy	Feelings of worthlessness
Impulsivity	Depersonalization	Suicidal ideations
Problems finishing things	Derealization	

Table 5-2 Symptoms derived from DSM-5 and LDA

The derived symptoms underwent a revision by a medical professional in the mental health field, namely a practitioner in a public practice in the city of Mainz, Germany. Table 5-2 summarizes the results after the revision. Thus, we make sure that the end list of derived symptoms, both from DSM-5 and LDA, is medically reliable.

Based on the final list of symptoms per condition, 1800 sentences in total were labelled - 600 per condition. The sentences were labelled manually by four machine learning researchers. Annotation by way of crowdsourcing, namely Amazon Mechanical Turk, was attempted,

however upon inspection the labelling results were unsatisfactory. Manually labelling these texts using crowdsourcing platforms is challenging as crowdsourcing workers do not possess expert knowledge in the domain, which can lead to many false positives or false negatives in the training data. Therefore, the sentences were re-labelled by experts under direct supervision of the study authors. More specifically, the experts were instructed on the possible symptom classes, as well as provided with labelled sentence examples from the forum data. One forum sentence can contain expressions of more than one symptom. As a neural network is capable of modelling several labels per input sequence, the annotators were instructed to label at most two symptoms per sentence, thus maximizing the knowledge that can be learned by the input data. Table 5-3 presents an example sentence and its annotation.

Example sentence	Label
I feel tired constantly, and I have no desire for anything	Fatigue, apathy

Table 5-3 Annotation example

5.3.2 Symptom Classification Using a Neural Network

Neural network models are undergoing rapid development and have widespread usage in text classification problems due to their excellent capability of self-learning and self-adapting (Watson 2019). A neural network is a computational model made of a combination of computational units called neurons and their connections, called weights. During training, the network is trained using labeled data in order to learn how to map input words to a desired output category. Weights have a numeric value which is adjusted during training based on input labeled data – if the current weights lead to a bad result in an iteration, they will be adjusted slightly (according to a learning rate parameter) so that the network performs better in the following iterations. Equation one shows the update rule for a single weight w_i after a training iteration, where a stands for learning rate, and gradient is the rate of change of the total error with respect to w_i. Weights are the primary product of neural network training, which will be used for classification of new texts.

$$w_i = w_i - a * gradient(w1) \quad (1)$$

In order to predict the mental health symptoms from forum posts, we construct and train a state-of-the-art deep learning model. We use a sequence-to-sequence bidirectional recurrent neural network with a long short term memory (LSTM) cell. This model is currently widely

used for working with text sequences (Lipton et. al. 2015). A recurrent neural network in combination with an LSTM cell has the ability to remember and process longer sequences. It is necessary for our model to be able to process longer sequences since the average length of the extracted forum sentences is 16 words.

This model setup and parameters value has been reached after experimentation with different options for activation, optimization, learning rate, and dropout. In order to avoid model overfitting (able to classify only what has been observed in training and cannot generalize) to the training data, we use neuron dropout of 0.8 - a technique which allows the network to "forget" a few neurons on each training pass, thus enhancing generalization (Srivastava et. al., 2014). We use a decaying learning rate, starting with 0.001 and reducing it further (no more than 75% reduction) with each iteration. Using a decaying learning rate allows the model to start learning at a higher learning rate, and then to gradually reduce the rate, which increases the chance for model convergence (An et. al., 2017).

The model has two hidden layers with 96 neurons in each layer. Hyperbolic tangent (equation 2) is used as the activation function – namely, the activation takes the value x of a neuron and maps it to a space between -1 and 1. A value above 1 activates a neuron, essentially signaling that the feature the neuron has learned has been recognized in the current input the neuron has processed.

$$tanh = f(x) = \frac{e^{2x}-1}{e^{2x}+1} \quad (2)$$

The model is fitted with conditional random fields (CRF) as a last layer, as they have been empirically shown to improve the modeling of text sequences, which is the goal of this study.

The model has been trained with a batch size of 10 sentences. Before passing a batch to the network, the sentences are converted into a numeric representation. An efficient way to do this is to use word embeddings, or numerical vector representations of words. For this we use the Embedding Language Model (ELMo) through its Tensorflow implementation (Peters et. al., 2018). ELMo provides state-of-theart embeddings for the English language and it allows converting each word in a sentence into a 1024-dimensional numeric representation of that word and its meaning. One advantage of using ELMo is that it is context-dependent. Thus, the word "apple" can have a different numeric word embedding, depending on whether it is used in context as the fruit, or as the technological company, for example.

We construct the neural network in Python using the deep learning framework Tensorflow version 1.10, as well as the Python modules numpy and scikit-learn.

5.3.3 Condition Classification

Conditions will be classified based on all the symptoms identified within all available posts by one user. We look at two things: the ratio of symptoms classified that fit into one of the three conditions (ADHD, depression, schizophrenia), as well as the probability output from

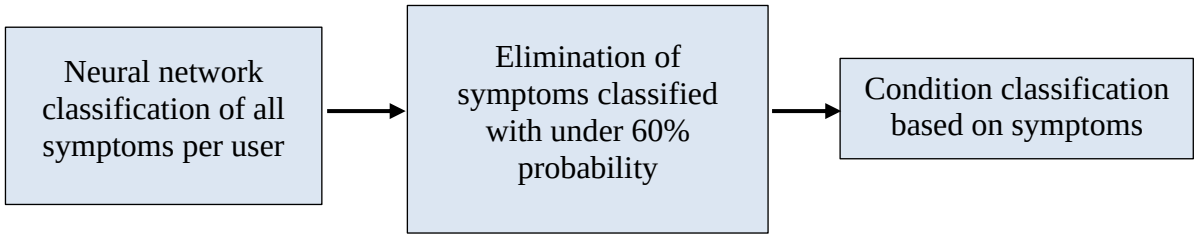

Figure 5-2 Condition classification process

the neural network model for each of these symptoms. First, symptoms from a user's posts are classified by the neural network. Then, we want to only account for those symptoms that are classified with relative certainty. We arrive at a threshold probability of 0.6. The threshold value is established empirically after repeated experimentation – it is applied as a heuristic measure with the goal to improve performance by lowering false positives when the final label prediction depends on class probabilities lower than 0.6, or in other words, when the certainty that the predicted label is correct is low. Any symptoms classified with a probability less than 0.6 are discarded. Then, a primary and potentially a secondary condition classification is arrived at based on the number of symptoms present from each of the three conditions reviewed.

5.4 Results

In this section results are presented for the classification of symptoms, as well as the computation of a condition based on the prior symptom classification. The model performance is tested with a ten-fold cross-validation with a random dataset split of 90% for training and 10% for testing. Cross validation is used to evaluate the predictive performance of the models and to estimate how the model performs when classifying a dataset not seen in training. The training was conducted in batches of 10 sentences chosen randomly from the training set, over 10 training epochs. Additionally, the testing samples include sentences

where users do not describe any symptoms, so that the model metrics account for symptom predictions which are false positives.

Model performance is measured using precision, recall, accuracy, and F1. Recall measures the fraction of symptoms in the test dataset that were identified correctly to be symptoms, while precision measures the fraction of recalled symptoms that was classified into the correct class. Accuracy reflects the correct predictions over all possible predictions, while F1 represents the harmonic mean of precision and recall. The equations for the metrics are presented below.

$$precision = \frac{true\ positives}{true\ positives + false\ positives} \tag{3}$$

$$recall = \frac{true\ positives}{true\ positives + false\ negatives} \tag{4}$$

$$accuracy = \frac{number\ of\ correct\ predictions}{total\ number\ of\ predictions} \tag{5}$$

$$F1 = 2 * \frac{precision * recall}{precision + recall} \tag{6}$$

The following table compares the performance of a baseline model with models that incrementally add more methods. The baseline model does not make use of CRF, ELMo, dropout, or decaying learning rate. Methods such as CRF have been added and tested incrementally. The table shows that when all methods are used in combination, the model performs best (83% and 84% for precision and recall, respectively).

	Precision	**Recall**	**Accuracy**	**F1**
Baseline model	60%	64%	59%	62%
Adding ELMo	67%	69%	65%	68%
Adding CRF layer	72%	74%	74%	73%
Adding dropout	81%	82%	83%	81%
Adding decaying learning rate (final model)	83%	84%	84%	83%

Table 5-4 Symptom classification model comparison

The overall performance metrics for all conditions and symptoms are given in Table 5-5. On average, for all conditions the precision and recall stand at 87% and 85% respectively, and for symptoms at 83% and 84% respectively. This sets our model at comparatively high performance with other similar models reported in literature. Compared to previous models for mental health condition classification, the multiclass model presented in this study does not outperform binary class models, however it does improve upon previous multiclass models' performance (as seen in Table 5-1). In terms of model architecture, the current best-performing multiclass models make use of a recurrent neural network architecture, which is used in our model as well as the model presented by Ive et. al. (2018).

	Precision	**Recall**	**Accuracy**	**F1**
All Conditions	87%	85%	84%	86%
ADHD	85%	87%	86%	86%
Depression	84%	80%	86%	82%
Schizophrenia	92%	89%	91%	90%
All symptoms	83%	84%	84%	83%
ADHD	84%	89%	84%	86%
Depression	82%	81%	82%	81%
Schizophrenia	81%	82%	82%	81%

Table 5-5 Evaluation metrics for conditions and symptoms

5.4.1 Symptom Classification

Depression

The histogram for depression symptoms shows unsurprisingly that depression forum users mostly express depressive moods and feelings of sadness. Within depression sentences, the depressive mood symptom appears frequently in sentences with other symptoms – most frequently with fatigue, confusion, and feeling lost.

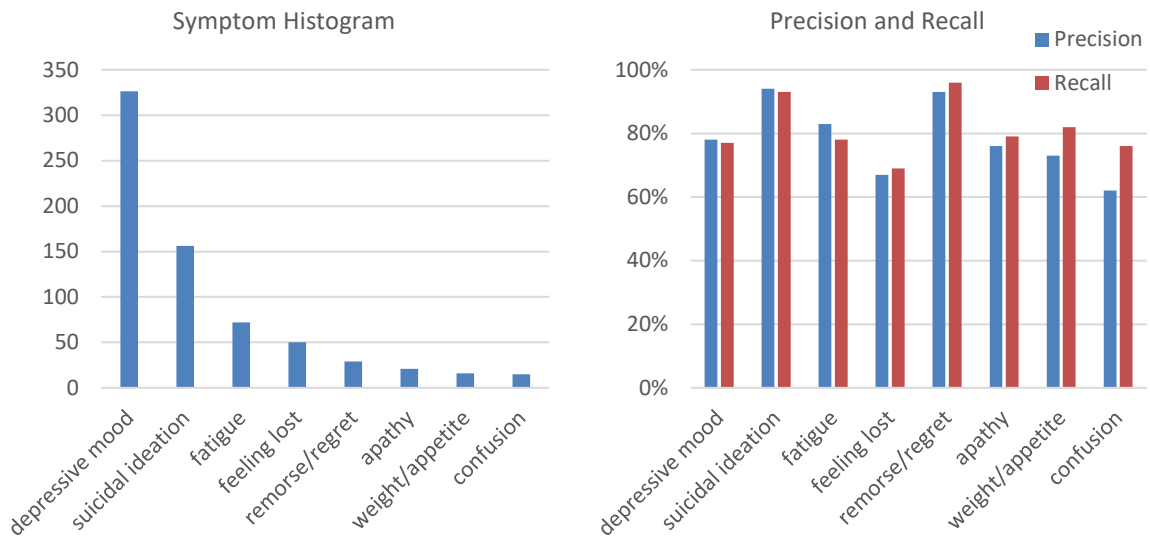

Figure 5-3 Depression evaluation metrics

What is more surprising is that the second leading symptom expressed in the forums is suicidal ideations (capturing phrases with the words “suicide” and/or “death”, e.g. "I want to die" or "I have been thinking of ways to kill myself"). The third most frequent symptom is feelings of fatigue, depletion, and lack of energy. It is interesting to note that even though feelings of worthlessness are emphasized in the DSM-5, depression forum users do not discuss this subject so much. Feeling lost or disoriented as well as discussing appetite and weight were challenging to model. Often, the vocabulary annotated for the “lost” symptom class would also be used in a literal sense (“I lost my keys”).

The model performs best when classifying suicide and remorse/regret symptoms. These symptoms are expressed by a very distinct jargon, making it more likely for the model to capture them (high recall) and to classify them correctly (high precision).

Schizophrenia

Overall, schizophrenic symptoms were classified with precision and recall of 81% and 82% respectively. Auditory hallucinations, or "voices", as the forum users refer to them, are by far the most common symptom discussed in the schizophrenia forums. Since hearing voices is so often discussed, there is an established way among forum users how to talk about it, e.g. by calling them simply voices, or using the word "hearing" in some context. This vocabulary is almost never used in a different context. As a result, the two most discussed symptoms are

modelled successfully. The performance worsens when considering less frequently discussed symptoms such as feelings of depression or anxiety.

Similarly, the symptom of "lacking insight" into one's own schizophrenia does not have many examples, but they are all very uniformly expressed on the forums, e.g. "misdiagnosed" or "I am not schizophrenic". Interestingly, schizophrenia sentences have the least symptom overlap, in other words, most sentences have only one symptom label.

The symptom of disorganized speech is difficult to annotate, as there is no established way in expressing this symptom. For example, words such as "mumble" are too general and used in a wide array of occasions; words like "incoherent" are often used to describe auditory hallucinations rather than speech. The lack of specificity combined with fewer training examples leads to a lower performance in classifying these symptoms in particular.

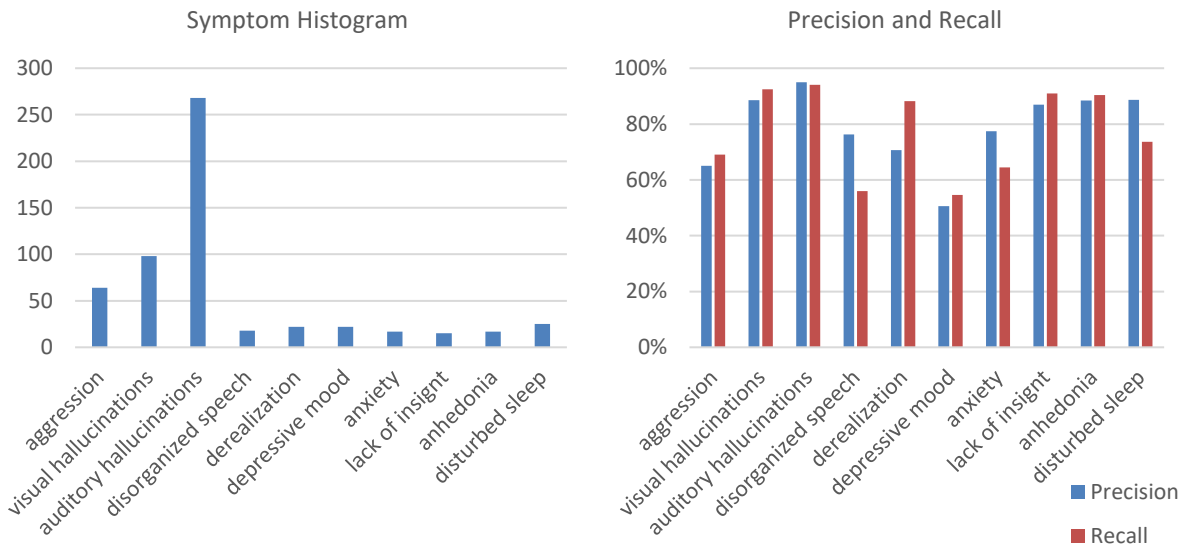

Figure 5-4 Schizophrenia evaluation metrics

What makes schizophrenic texts difficult to work with for deep learning is that users more often recount the story of their delusion, hallucination, or other symptomatic behavior rather than discussing the presence of the symptom directly. For example, is not possible to annotate users' recounts of their delusions as such, there must be a clear expression of the symptom of delusions, as in "I am delusional" instead of a recount of a delusion as in "They are in my head". This makes the annotation as well as subsequent classification of schizophrenic

symptoms particularly challenging. This is why this condition has the least annotated symptoms of the three conditions reviewed.

ADHD

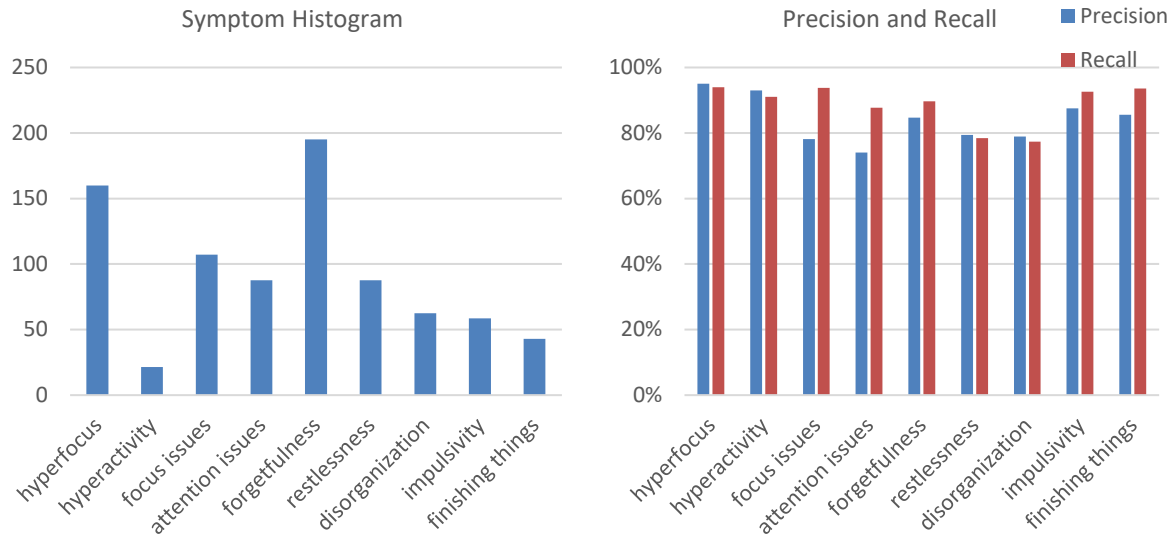

Figure 5-5 ADHD evaluation metrics

ADHD sentences were annotated with at most two symptoms per sentence, as users in these forums tend to jump from one subject to another in the same sentence. They also write the longest sentences on average. ADHD forum conversations show the highest balance among the number of times a class was discussed. The better balance in representation of the symptoms in the training data contributes to a fairly consistent classification performance for all symptoms. The classification performance of the symptoms hyperfocus and hyperactivity stand out as they are almost always expressed exactly with those words.

Forgetfulness and hyperfocus were the most discussed symptoms. The forgetfulness symptom has a high recall because the jargon is restricted to a handful of words, but those words often can be used in a different context in the other conditions (e.g. “to forget”), therefore this class has high recall and lower precision. It is interesting to note that even though hyperactivity is one of the most distinguishing traits of ADHD (according to the DSM-5), it is one of the least discussed symptoms on the forum. This could signal that the users are less worried about being hyperactive than they are about other symptomatic behavior, therefore do not seek assistance with this trait in particular.

5.4.2 Condition Classification

To classify a condition, the classified symptoms from all posts for one user are considered in order to arrive at the most detailed and precise classification possible. We look at both the amount of symptoms classified per condition, as well as the probability assigned to each symptom by the neural network at the point of inference. The more sentences expressing symptoms a user has written, the more accurate the classified condition will be. We observe that at least 7 such sentences are needed to arrive at a condition classification precision above 50%. However, since only every 2,3 out of 10 sentences would show a symptom, this requires at least 44 sentences per user from the forum to arrive at a precision above 50% for a condition classification.

Furthermore, using a threshold for a minimum probability for the classified symptoms is a key mechanism in the classification of conditions. Using all classified symptoms without doing a cut-out below a specific probability leads to many false positives. In other words, many sentences that do not describe symptoms get classified as such. Therefore, we experiment with a cutout for different probabilities, and we arrive at appropriate performance with a cutout of 60% or above probability. After the cutout, approximately 26% of the false positives are eliminated. For example, after removing those classified symptoms with a probability under the threshold of 60%, one user had the following symptom classifications:

Classified Symptom	Probability	Symptom appears in conditions
Auditory hallucinations	71%	Schizophrenia
Visual hallucinations	76%	Schizophrenia
Fatigue	79%	Depression
Anhedonia / apathy	63%	Depression, schizophrenia
Anger	67%	Schizophrenia

Table 5-6 Symptom classification results for one user

Since four out of the five symptoms identified by the neural network are part of schizophrenia, this user's primary condition was classified as schizophrenia. The results suggest that this user also possibly suffers from depression as a secondary condition. This

result is in fact unsurprising, as the comorbidity of depression and schizophrenia is found to exist in 50% of those diagnosed as schizophrenic (Buckley et. al. 1998).

According to the results in Table 5-5, schizophrenia was the most accurately classified of the three examined conditions. Of the three conditions, depression proves to be the most challenging to predict by means of the described approach. This comes as a result of depressive mood or feelings of sadness often being expressed in the other two conditions, as well as some of those expressions being labelled with different symptoms. As a result, the recall for depression is the lowest of the three conditions.

In cases where classes have low recall, a problem can occur due to high variance. High variance occurs in models when there is much noise in the training data. This does not have to mean that there is much irrelevant data in the datasets, but rather that the symptoms with low recall have been expressed with a very diverse vocabulary. In this case, the model has seen too many variations of a symptom, but not enough repetitions of each of those variations. Consequently, adding more examples of this class in the training data would help stabilize the variance and increase the recall.

In other cases, a combination of high recall but low precision for some classes is observed, for example the “lost” class for depression. This occurs with a class such as “lost” because the variations of the class in the text are not so large, however they are too general as to be only symptom-specific. There are many cases where vocabulary has been tagged in some context as a symptom, and in another context not as a symbol.

The model successfully picks up all true positives, but also some false positives, leading to the high recall / low precision performance.

5.5 Discussion

In this paper, we present a novel approach to classifying multiple mental health conditions and their underlying symptoms using a state-of-the-art neural network model. Based on text input by an online forum user, the model classifies not the condition, but the symptoms. The derived symptoms are then discarded if they were classified by the neural network with a probability lower than 60%; based on the remaining symptoms, we can classify not only a primary condition, but also a secondary one.

By first identifying which symptoms affected individuals are specifically looking for help with, our model captures much more information from the underlying data regarding the needs and issues a particular forum user is experiencing. This is achieved with a single model and without the need for separate feature engineering or external information sources such as the popular LIWC lexical resource. It is also important to note that our model does not rely on demographic data such as age or gender, thus this approach allows for more privacy preservation, which is of high importance to affected individuals who also struggle with stigma and potential peer rejection. Thus, our approach simplifies the training of mental health condition classification models, while simultaneously significantly enriching the information derived from the model.

5.5.1 Contribution to IS in Healthcare

With this study we contribute to further academic understanding of possible support tools in digital mental health. We contribute to the existing studies exploring the application of deep learning in mental health by showing that conditions can be successfully classified based on symptom classification; we also successfully model comorbidities, or co-occurrences of two conditions. We also show that it is not necessary to rely on demographic data or external lexical resources in order to build a classification model.

In our exploratory approach, we show that the data source and the respective jargon play an important role in the performance of deep learning models for mental health condition and symptom classification. Given that this model has been trained on forum data in particularly, this model would most probably deteriorate if transferred to classify data from a different source, for example chatbots text input; models must be re-trained with data for the appropriate source in order to accurately capture the many diverse forms of expressing one and the same symptom or emotional state.

To the best of the authors' knowledge, this study describes one of the first models that are able to account for condition comorbidity, which is ubiquitous in diagnoses made by mental health professionals. Furthermore, this is also one of few studies that model a condition's symptoms. Since a mental health condition can be very complicated and manifest in different forms, it is essential for a digital mental health service to be able to differentiate details about a user's mental state beyond just the condition itself. Knowing the specific symptoms a user is

experiencing can allow for an automated response tailored to this symptom. Thus forums can be significantly enhanced beyond only conversation to active automated support for its users.

With this paper we contribute to academic literature by first of all constructing and testing a model for other conditions and not only depression. We further close a research gap of collaboration between psychologists and machine learning practitioners.

5.5.2 Practical Implications

Processing personal mental health texts with similar deep learning models leads to much greater detail in what has been learned about an individual user of an IT mental health service such as a forum, a chatbot, a journaling app and similar. This allows the IT service to then automatically address the identified issues with a variety of tools. For example, by learning that a user struggles with suicidal thoughts (regardless of an identified condition), the IT service can automatically suggest resources such as hotlines or crisis centers for this individual; or, by identifying that a user with ADHD struggles specifically with forgetfulness and lack of organization, the IT service can automatically set that user in contact with therapists who specialize in the treatment of those exact symptoms and condition (Mädche et al., 2019). Essentially, these models can significantly alleviate users in the search for resources, qualified therapists, etc. While our model is not a formal means of diagnosis, this exploratory study shows the enormous potential of deep learning applications within psychology and mental healthcare, thus a much more concentrated research effort is required in order to one day have these models as aides to medical professionals as well. This might help in allowing healthcare systems to reach much more affected individuals, offer cheaper resources, offer anonymity. This all will become even more important as we head into a future where projections are that more and more individuals will suffer from mental health conditions.

5.5.3 Limitations and Future Work

Our model classifies only three conditions. In reality, a model will have to distinguish between dozens of conditions. Therefore, a study that considers more conditions is necessary in order to arrive at an assessment for this approach that is closer to a practical deployment. The model performance can be improved by training a model with more data for all

symptoms. Furthermore, we have only tested this approach on forum data, it will have to be tested on data from other platforms in order to confirm its potential generalizability.

During the study, we identified bottlenecks in the neural network modeling that are particularly challenging to solve. First, it is difficult to label as well as model a certain symptom which is not being expressed or discussed with uniform language by the forum users. For example, the ADHD symptom “focus issues” can be expressed in a multitude of ways: “I just couldn’t focus”, “I was so distracted”, “I wasn’t able to concentrate”, etc. In these cases, more training example will be of benefit, so that the model is presented with enough examples of a large number of expressions for the same symptom.

The model performance also suffers from difficulty sorting symptoms into conditions when symptoms (such as feelings of depression) are shared across several conditions or are labelled as non-symptoms in the other conditions. This leads to many false negatives and low recall. A joint modelling of symptoms and conditions, i.e. adding a label for condition when annotating the training and testing data may help the model distinguish between cases when a symptom should be labelled, and cases when it should be ignored.

5.6 Conclusion

As deep learning is increasingly applied in the field of psychology and mental health, new opportunities appear in automating diagnostics as well as treatments of mental health conditions by means of IT and digital services. In this study we build and test a novel classification approach based on a state-of-the-art neural network model for symptom and primary and secondary condition classification, trained on online mental health forum data. This is one of the first studies to model symptoms as well as secondary conditions by using only one model. This study contributes to academic research on deep learning applications within the mental health field. We contribute to practice by showing how deep learning models can enable potential automated applications that can serve as digital support tools for those affected by mental health issues.

Chapter 6: Thesis Conclusion and Contributions

Individual mental health is poised to become a more prevalent issue on a global scale. Already, conventional approaches to treatment and therapy fall short in providing timely and effective support for all affected individuals. This thesis sheds light to the possibility of IT-based interventions to fill in the healthcare gap and improve services in terms of amplified reach, improved efficiency, lower costs. To this end, three different studies have been conducted at the intersection of IS, psychology, and machine learning. Specifically, machine learning automation in the areas of diagnostics, mood tracking, and peer support has been examined. The studies show that NLP techniques such as sentiment analysis, topic modelling, and neural networks can achieve high accuracy in automating important practices in the mental illness recovering process. Compared to previous studies, the articles included in this thesis advance classification in terms of a more detailed diagnosis and model precision, providing high generalizability and specificity; we add to ongoing efforts to better select treatments for individuals, as well as to enable patient self-management, through large-scale data analysis and continuous monitoring; we design and test a procedure to evaluate the effects on a user of continuously using a digital tool for mental health, which is a bedrock of further embedding IT in mental healthcare; Nevertheless, certain adoption barriers still persist in terms of integrating these techniques in healthcare services, namely in the area of ethics, as well as adoption by medical professionals. This thesis documents different aspects of machine learning automation and the ways it can support individual mental health in ways that are compliant with formal treatment and therapy, but also more generally in accessing information and communities, receiving a diagnosis, and starting the right treatment with no time delay.

6.1 Contributions to IS in Healthcare

We investigate the current and potential impact of digitalization in mental healthcare by exploring three different aspects of the process: the effects of using digital mental health tools on users, treatment selection and personalization based on user input, and automatic symptom and condition diagnosis based on user input.

The first study contributes to ongoing research focused on the dynamics of user-driven online support platforms. It presents several implications for the design and usability of digital

mental health tools. First and foremost, we confirm that digital solutions can be employed as support tools in mental healthcare, as we observe significant and mostly positive changes in affect in users. This is a crucial finding as the conversation around the implications of the digitalization in mental health continues. We add to the growing body of research that confirms that IT-based interventions can lead to treatment results that are comparable or complementary to established medical treatments such as face-to-face therapy (Ebert et al., 2018).

We show that digital platforms are largely beneficial for the end-user, however not equally for all users, and some even experience a negative effect of participation. With this in mind, our findings paint a complex picture of potentially effective digital tools in this area, and the standards such tools would need to uphold to be deployed to end-users. Standardized services would be inapplicable for a digital mental health tool – even individuals who suffer from one and the same condition differ significantly in their needs and ways of interacting with a digital support tool. In line with the above findings, the study examines and confirms that the role (type of user engagement), duration of engagement, as well as the condition a user suffers from, largely shape the current experience with digital tools by help-seekers. The interplay between these factors is a stepping stone to design better and more effective digital tools. Further researching the factors that contribute to these differences can help optimize the usability and effectiveness of such tools in the future. This conclusion is consistent with the broader literature on personalization in IS (Benlian, 2015) - service personalization plays an important role in ensuring a positive therapeutic effect for all users (Rüegger et. al., 2017).

We further contribute to research in healthcare information systems by corroborating the use of machine learning analysis of text user input as a valid research tool to capture knowledge from thousands of individuals as well as analyze the complex environment and interactions that happen on platforms such as forums. Analysis of big data which centers around mental health is not prevalent in IS or other research fields touched upon in this thesis, however our studies show that much can be learned from this new research paradigm for the field. We make an interdisciplinary contribution to both IS and precision psychology by showcasing an effective employment of machine learning of large datasets. Automated analysis of large datasets presents a shift in the research paradigm of the field, which still depends on more traditional research methods that allow the inclusion of small groups of individuals (Dwyer et. al., 2018).

Furthermore, through employing machine learning research techniques, we add to the body of knowledge in IS in healthcare and related fields in that we integrate formal diagnostic procedure from medical practitioners into our data models. This led to a model which first examines possible distress symptoms as a first step in a screening process. Then, based on the identified symptoms, it is possible to derive a diagnosis that is now much more thorough and comprehensive, giving the opportunity to prioritize treatment for those symptoms which occur most frequently or are most acutely present. This type of diagnosis is medically sound (following DSM-IV diagnostic guidelines). With this approach, we address a stark gap in this highly interdisciplinary field – the integration of medical procedure into the evaluated digital tools in healthcare and in the computational approach of building those tools (Calvo et. al., 2017).

Likewise, **accounting for comorbidity** is another step in the direction of machine learning techniques catching up with practitioners' diagnoses. In this context we already show that it is possible to identify and discern between primary and secondary mental health concerns of a user. Thus, we show that processing user input with machine learning techniques can capture a full psychological profile, providing a multifaceted perspective of a person's mental state, and capturing details and nuances that remained unaccounted for by previous models.

A further academic contribution to precision psychology as well as computational linguistics is the **testing of a multiclass classification model**, i.e. a model with capabilities to differentiate between multiple conditions. This represents an important milestone to reach in this domain, as binary models (models that differentiate whether a user does or does not exhibit symptoms of one particular condition, e.g. depression) do not provide a diagnostic gateway that can be implemented in practice in the initial stages of treatment such as pre-screening. Adding more classes will make future models even more applicable in a real-world scenario, as in reality a user's input can be distributed among dozens of conditions and hundreds of symptoms.

6.2 Practical Contributions

There is space for improvement in terms of how online peer support platforms facilitate user engagement. Using natural language processing, peer support platforms can maximize participation benefits for users by constructing personalized spaces within the wider

community in terms of clustering users with similar symptoms for support and experience exchange; usually in forums users must seek out the content that is relevant to them, while machine learning allows automated content discovery and recommendation, where, based on user input such as profile settings, likes, or post content, algorithms can deliver and recommend potentially useful and relevant content, thus simplifying and accelerating content and peer discovery. The type of content could be for example suggesting resources that are appropriate for a user within their local area, thus localizing the platform experience and augmenting it outside the digital environment.

Machine learning and automation open the door for digital solutions that offer pro-active rather than reactive monitoring. Reactive monitoring would only trigger actions once a user has self-reported difficulties, relapses, and similar. Proactive monitoring on the other hand could track user progress and anticipate future events, thereby raising actions ahead of time to mitigate potential challenges, e.g. by monitoring for emergencies and routing users to the appropriate specialists or locations to receive help in real time. This is also an opportunity to explore the potential of blended treatment in mental health, where a digital service acts as a support tool in the recovery process lead by a medical professional. Blended treatment offers all the benefits of digitalization and automated services and data collection as an enhancement to traditional treatment.

The studies included in this thesis show that automated diagnostics can provide significantly more details about a person's mental health than past experiments have shown. These details can be used to create a treatment plan that is much more personalized and precise – arguably within a shorter timeframe than standard procedures that are unsupported by digital aids.

6.3 Limitations and Directions for Future Research

Research into treatments making use of machine learning has been very promising. Current efforts are concentrated in maximizing the information gain as well as the accuracy from ML techniques and different types of data serving as input. The domain is already at the junction where future studies can assess the degree of effectiveness either of fully digital treatments or blended treatments, as well as pinpoint the exact symptoms alleviated by each concept. In order for the field to continue improving, the next step in research is to learn to what extent

and how rapidly these symptoms are alleviated, and juxtapose these findings to standard practice.

One big concern – shared by governments, medical professionals, and potential end-users alike – is the ethical aspect of users having to share a lot of personal information in order to be able to derive benefits from digital services. Mental health data are very sensitive personal information in an area that is still very much associated with stigma and discrimination (both personal and professional), therefore maintaining privacy and security is of utmost importance for the credibility of digital efforts in mental health. Future efforts must showcase and demonstrate that privacy is by no means at stake in this context. Fortunately, deep learning models which rely on nothing more than user text input (not requiring demographic data that can potentially lead to identification) allow for a sweeping anonymization of user data, without incurring any information loss for model development. Data storage location can be employed as a further safeguard against data breaches – personal data can be stored on users' devices instead of central storage, thus removing the possibility of a mass-scale breach (Manji & Saxena, 2019).

Furthermore, many gains can be made in bridging the gap between healthcare professionals and ML specialists, where collaboration has been very rare. Making inter-field collaboration not only a standard practice, but a necessity, will accelerate the improvements we stand to gain in technology performance, as well as raise the credibility of applying these techniques in practice (Marsch & Gustafson, 2013). What is more, there is still resistance from the medical community in adopting novel technologies. This is where opening ML-based studies to input from psychologists can be immensely beneficial in arriving at an integrated interdisciplinary approach (Marsch & Gustafson, 2013).

References

Adjerid, I., & Kelley, K. (2018). Big data in psychology: A framework for research advancement. *American Psychologist, 73*(7).

Ali, K., Farrer, L., Gulliver, A., & Griffiths, K. M. (2015). Online peer-to-peer support for young people with mental health problems: a systematic review. *JMIR mental health, 2*(2).

Almeida, H., Queudot, M., & Meurs, M. J. (2016). Automatic triage of mental health online forum posts: CLPsych 2016 system description. *Proceedings of the Third Workshop on Computational Lingusitics and Clinical Psychology,* (pp. 183-187).

Almutairi, M. M., Alhamad, N., Alyami, A., Alshobbar, Z., Alfayez, H., Al-Akkas, N., & Olatunji, S. O. (2019). Preemptive Diagnosis of Schizophrenia Disease Using Computational Intelligence Techniques. *2nd International Conference on Computer Applications & Information Security.*

Altschuler, E. L. (1999). Pet-facilitated therapy for posttraumatic stress disorder. *Annals of Clinical Psychiatry, 11*(1), 29-30.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5).* American Psychiatric Pub.

An, W., Wang, H., Zhang, Y., & Dai, Q. (2017). Exponential decay sine wave learning rate for fast deep neural network training. *IEEE Visual Communications and Image Processing (VCIP)*, 1-4.

Andersson, E., Holmes, E., & Kavanagh, D. (2018). Innovations in digital interventions for psychological trauma: harnessing advances in cognitive science. *mHealth, 4*, 47.

Autism Speaks. (2014, April 15). *Autism and Pets: More Evidence of Social Benefits*. Retrieved from https://www.autismspeaks.org/science/science-news/autism-and-pets-more-evidence-social-benefits

Aveyard, P., Madan, J., Chen, Y. F., Wang, D., Yahaya, I., Munafo, M., & Welton, N. (2012). Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. *Health technology assessment, 16*(38).

Balani, S., & De Choudhury, M. (2015). Detecting and characterizing mental health related self-disclosure in social media. *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, (pp. 1373-1378).

Barak, A., & Gluck-Ofri, O. (2007). Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior, 10*(3), 407-417.

Becker, L. C. (2014). *Reciprocity (Routledge Revivals).* Routledge.

Bedeian, A., & Mossholder, K. (1994). Simple question, not so simple answer: Interpreting interaction terms in moderated multiple regression. *Journal of Management*, *20*(1), 159-165.

Bell, V. (2007). Online information, extreme communities and internet therapy: Is the internet good for our mental health? *Journal of mental health*, *16*, 445-457.

Benlian, A. (2015). Web Personalization Cues and Their Differential Effects on User Assessments of Website Value. *Journal of Management Information Systems, 32*(1), 225-260.

Benlian, A. (2020). A daily field investigation of technology-driven stress spillovers from work to home. *MIS Quarterly, 44*(3), 1259-1300.

Benlian, A., & Hess, T. (2011). The signaling role of IT features in influencing trust and participation in online communities. *International Journal of Electronic Commerce, 15*(4), 7-56.

Benlian, A., Hinz, O., & Klumpe, J. (2019). Mitigating the intrusive effects of smart home assistants by using anthropomorphic design features: A multimethod investigation. *Information Systems Journal*.

Binhadyan, B., Peszynki, K., & Wickramasinghe, N. (2015). The Effect of e-Mental Health Services on Saudi General Mental Health. *BLED*, (p. 34).

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 993-1022.

Blumberg, S. J., Zablotsky, B., Avila, R. M., Colpe, L. J., Pringle, B. A., & Kogan, M. D. (2016). Di-agnosis Lost: Differences between Children who Had and who Currently Have an Autism Spec-trum Disorder Diagnosis. *Autism : The International Journal of Research and Practice, 20*(7), 783-795.

Bradburn, N. M. (1969). *The structure of psychological well-being.*

Braithwaite, S. R., Giraud-Carrier, C., West, J., Barnes, M. D., & Hanson, C. L. (2016). Validating machine learning algorithms for Twitter data against established measures of suicidality. *JMIR mental health, 3*(2), e21.

Buckley, P., Miller, B. J., Lehrer, D. S., & Castle, D. J. (2008). Psychiatric comorbidities and Schizophrenia. *Schizophrenia bulletin, 35*(2), 383-402.

Bzdok, D., & Meyer-Lindenberg, A. (2018). Machine learning for precision psychiatry: opportunities and challenges. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 3*(3), 223-230.

Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering, 23*(5), 649-685.

Canzian, L., & Musolesi, M. (2015). Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, (pp. 1293-1304).

Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., & Bierut, L. J. (2016). A content analysis of depression-related tweets. *Computers in human behavior*, 351-357.

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine, 3*(1), 1-11.

Chapman, A. L., Specht, M., & Cellucci, T. (2005). Borderline personality disorder and deliberate self-harm: does experiential avoidance play a role. *Suicide and Life-Threatening Behavior, 35*(4), 388-399.

Chen, D., & Manning, C. (2014). A Fast and Accurate Dependency Parser Using Neural Networks^. *Proceedings of EMNLP*.

Chen, X., Sykora, M., Jackson, T., Elayan, S., & Munir, F. (2018). Tweeting Your Mental Health: an Exploration of Different Classifiers and Features with Emotional Signals in

Identifying Mental Health Conditions. *Proceedings of the 51st Hawaii International Conference on System Sciences.*

Cobb, N. K., Mays, D., & Graham, A. L. (2013). Sentiment analysis to determine the impact of online messages on smokers' choices to use varenicline. *Journal of the National Cancer Institute Monographs*, 224-230.

Cohan, A., Young, S., Yates, A., & Goharian, N. (2017). Triaging content severity in online mental health forums. *Journal of the Association for Information Science and Technology*, *68*(11), 2675-2689.

Coppersmith, G., De Choudhury, M., Dredze, M., & Mrinal, K. (2015a). Detecting changes in suicide content manifested in social media following celebrity suicides. *Proceedings of the 26th ACM conference on Hypertext & Social Media*, (pp. 85-94).

Coppersmith, G., Dredze, M., Harman, C., & Hollingshead, K. (2015b). From ADHD to SAD: Analyzing the language of mental health on Twitter through self-reported diagnoses. *Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, (pp. 1-10).

Crabtree, J. W., Haslam, S. A., & Postmes, T. (2010). Mental health support groups, stigma, and self-esteem: positive and negative implications of group identification. *Journal of Social Issues, 66*, 553-569.

Cuijpers, P., Marks, I. M., van Straten, A., Cavanagh, K., Gega, L., & Andersson, G. (2009). Computer-Aided Psychotherapy for Anxiety Dis-orders: A Meta-Analytic Review. *Cognitive Behavior Therapy, 38*, 66 – 82.

Da Silva, N. F., Hruschka, E. R., & Hruschka, E. R. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems, 66*, 170-179.

Davcheva, E. (2018). Text Mining Mental Health Forums - Learning from User Experiences. *European Conference on Information Systems (ECIS).*

Davcheva, E., Adam, M., & Benlian, A. (2019). User Dynamics in Mental Health Forums - A Sentiment Analysis Perspective. *14. Internationale Tagung Wirtschaftsinformatik (WI 2019).*

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. *Seventh international AAAI conference on weblogs and social media*.

De Choudhury, M., Morris, M. R., & White, R. W. (2014). Seeking and sharing health information online: comparing search engines and social media. *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, (pp. 1365-1376).

Delahunty, F., Johansson, R., & Arcan, M. (n.d.). Passive Diagnosis incorporating the PHQ-4 for Depression and Anxiety. *Proceedings of the Fourth Social Media Mining for Health Applications (# SMM4H) Workshop & Shared Task*, (p. 2019).

Deng, X., Khuntia , J., & Ghosh, K. (2013). Psychological Empowerment of Patients with Chronic Diseases: The Role of Digital Integration. *Thirty Fourth International Conference on Information Systems*, (pp. 1-20). Milano.

Diehl, M., Hay, E. L., & Berg, K. M. (2011). The ratio between positive and negative affect and flourishing mental health across adulthood. *Aging & Mental Health, 15*, 882-893.

Doraiswamy, P. M., London, E., Varnum, P., Harvey, B., Saxena, S., Tottman, S., & Campbell, P. (2019). Empowering 8 Billion MindsEnabling Better Mental Health for All via the Ethical Adoption of Technologies. Geneva: World Economic Forum.

Duda, M., Ma, R., Haber, N., & Wall, D. P. (2016). Use of machine learning for behavioral distinction of autism and ADHD. *Translational psychiatry*, *6*(2).

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual review of clinical psychology*, *14*, 91-118.

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, *14*, 91-118.

Ebert, D., Van Daele, T., Nordgreen, T., Karekla, M., Compare, A., Zarbo, C., & Brugnera, A. (2018). Internet-and mobile-based psychological interventions: applications, efficacy, and potential for improving mental health. *European Psychologist*.

Eurostat. (2017). Digital economy and society statistics-households and individuals.

Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., & Stern, A. (2004). Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *BMJ, 328*, 1166.

Fairburn, C. G., & Patel, V. (2017). The impact of digital technology on psychological treatments and their dissemination. *Behaviour research and therapy, 88*, 19-25.

Fernandes, B. S., Williams, L. M., Steiner, J., Leboyer, M., Carvalho, A. F., & Berk, M. (2017). The new field of 'precision psychiatry. *BMC medicine, 15*(1), 80. doi:10.1186/s12916-017-0849-x

Finn, J. (1999). An exploration of helping processes in an online self-help group focusing on issues of disability. *Health & Social Work, 24*, 220-231.

Fortuna, K. L., Torous, J., Depp, C. A., Jimenez, D. E., Areán, P. A., Walker, R., & Ajilore, O. (2019). A Future Research Agenda for Digital Geriatric Mental Health Care. *The American Journal of Geriatric Psychiatry, 27*(11).

Friedl, N., Krieger, T., Chevreul, K., Hazo, J. B., Holtzmann, J., Hoogendoorn, M., & Kleiboer, A. (2020). Using the Personalized Advantage Index for Individual Treatment Allocation to Blended Treatment or Treatment as Usual for Depression in Secondary Care. *Journal of Clinical Medicine, 9*(2), 490.

Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T. J., Dobson, R. J., & Dutta, R. (2017). Characterisation of mental health conditions in social media using Informed Deep Learning. *Scientific Reports*, 45141.

Gohil, S., Vuik, S., & Darzi, A. (2018). Sentiment Analysis of Health Care Tweets: Review of the Methods Used. *JMIR Public Health and Surveillance, 4*(2).

Goldbeck, J. (2016). Predicting personality from social media text. *AIS Transactions on Replication Research, 2*(1).

Góngora, A. S., de la Torre-Díez, I., Hamrioui, S., López-Coronado, M., Barreno, D. C., Nozaleda, L. M., & Franco, M. (2018). Data mining algorithms and techniques in mental health: A systematic review. *Journal of medical systems, 42*(9), 161.

Guntuku, S. C., Yaden, D. B., Kern, M. L., Ungar, L. H., & Eichstaedt, J. C. (2017). Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences, 18*, 43-49.

Henson, P., Wisniewski, H., Hollis, C., Keshavan, M., & Torous, J. (2019). Digital mental health apps and the therapeutic alliance: initial review. *BJPsych open, 5*(1).

Hollis, C., Morriss, R., Martin, J., Amani, S., & Cotton, R. (2015). Technological innovations in mental healthcare: harnessing the digital revolution. *The British Journal of Psychiatry, 206*(4), 263-265.

Huh, J., & Pratt, W. (2014). Weaving clinical expertise in online health communities. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1355-1364). ACM.

Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth, 6*(11), e12106.

Ive, J., Gkotsis, G., Dutta, R., Stewart, R., & Velupillai, S. (2018). Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health. *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, (pp. 69-77).

Johnsen, J. A., Rosenvinge, J. H., & Gammon, D. (2002). Online group interaction and mental health: An analysis of three online discussion forums. *Scandinavian Journal of Psychology, 43*, 445-449.

Kahneman, D., Diener, E., & Schwarz, N. (1999). *Well-being: Foundations of hedonic psychology*. Russell Sage Foundation.

Kauer, S. D., Mangan, C., & Sanci, L. (2014). Do online mental health services improve help-seeking for young people? A systematic review. *Journal of Medical Internet Research, 16*(3).

Keller, M. B. (2006). Prevalence and impact of comorbid anxiety and bipolar disorder. *The Journal of clinical psychiatry, 67*, 5-7.

Kim, H., Kim, H., & Cho, S. (2017). Bag-of-concepts: Comprehending document representation through clustering words in distributed representation. *Neurocomputing, 266*, 336-352.

Kordzadeh, N., & Warren, J. (2017). Communicating Personal Health Information in Virtual Health Communities: An Integration of Privacy Calculus Model and Affective Commitment. *Journal of the Association for Information Systems, 18*.

Korkontzelos, I., Nikfarjam, A., Shardlow, M., Sarker, A., Ananiadou, S., & Gonzalez, G. H. (2016). Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. *Journal of biomedical informatics, 62*, 148-158.

Kuester, A., Niemeyer, H., & Knaevelsrud, C. (2016). Internet-based interventions for posttraumatic stress: A meta-analysis of randomized controlled trials. *Clinical Psychology Review, 43*, 1-16.

Kummervold, P. E., Gammon, D., Bergvik, S., Johnsen, J., Hasvold, T., & Rosenvinge, J. H. (2002). Social support in a wired world: use of online mental health forums in Norway. *Nordic journal of psychiatry, 56*, 59-65.

Lattie, E. G., Adkins, E. C., Winquist, N., Stiles-Shields, C., Wafford, E. Q., & Graham, A. K. (2019). Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review. *Journal of medical Internet research, 21*(7).

Lecomte, T., Potvin, S., Corbière, M., Guay, S., Samson, C., Cloutier, B., . . . Khazaal, Y. (2020). Mobile Apps for Mental Health Issues: Meta-Review of Meta-Analyses. *JMIR Mhealth Uhealth, 8*(5).

Lisa, M. A., & Gustafson, D. H. (2013). The role of technology in health care innovation: a commentary. *Journal of dual diagnosis, 9*(1), 101-103.

Lluch, M. (2011). Healthcare professionals' organisational barriers to health information technologies - A literature review. *International journal of medical informatics, 80*, 849-862.

Mädche, A., Legner, C., Benlian , A., Gimpel, H., Hess, T., Hinz, O., . . . Söllner, M. (2019). AI-Based Digital Assistants. *Business & Information Systems Engineering, 61*(4), 535-544.

Madden, M. (2010). *Older adults and social media: Social networking use among those ages 50 and older nearly doubled over the past year.* Pew Internet & American Life Project.

Malmasi, S., Zampieri, M., & Dras, M. (2016). Predicting post severity in mental health forums. *Proceedings of the third workshop on computational lingusitics and clinical psychology*, (pp. 133-137).

Manji, H., & Saxena, S. (2019, January). *The Power of Digital Tools to Transform Mental Healthcare*. Retrieved from World Economic Forum: www.weforum.org/agenda/2019/01/power-digital-tools-transform-mental-health-care-depression-anxiety/

Marsch, L. A., & Gustafson, D. H. (2013). The Role of Technology in Health Care Innovation: A Commentary. *Journal of Dual Diagnosis, 9*(1), 101-103.

Mathers, C., Boerma, T., & Fat, D. M. (2004). *The global burden of disease: 2004 update.* World Health Organization.

Media Use in the European Union. (2018, May 15). Standard Eurobarometer 88.

Medina-Moreira, J., Lagos-Ortiz, K., Luna-Aveiga, H., Rodríguez-García, M., Valencia-García, R., & Salas-Zárate, M. (2017). Sentiment Analysis on Tweets about Diabetes: An Aspect-Level Approach. *Computational and Mathematical Methods in Medicine*.

Mitchell, J. T., Sweitzer, M. M., Tunno, A. M., Kollins, S. H., & McClernon, F. J. (2016). "I use weed for my ADHD": a qualitative analysis of online forum discussions on cannabis use and ADHD. *PLoS One, 11*.

Monteith, S., Glenn, T., Geddes, J., & Bauer, M. (2015). Big data are coming to psychiatry: a general introduction. *International journal of bipolar disorders, 3*(21).

Motlova, L. (2007). Schizophrenia and family. *Neuro endocrinology letters, 28*(1), 147-159.

Naslund, J. A., Aschbrenner, K. A., Marsch, L. A., & Bartels, S. J. (2016). The future of mental health care: peer-to-peer support and social media. *Epidemiology and psychiatric sciences, 25*(2), 113-122.

Naslund, J. A., Aschbrenner, K. A., McHugo, G. J., Unützer, J., Marsch, L. A., & Bartels, S. J. (2019). Exploring opportunities to support mental health care using social media: A survey of social media users with mental illness. *Early intervention in psychiatry, 13*(3), 405-413.

Naslund, J. A., Aschenbrenner, K. A., Araya, R., Marsch, L. A., Unützer, J., Patel, V., & Bartels, S. J. (2017). Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. *The Lancet Psychiatry, 4*(6), 486–500.

Naslund, J. A., Gonsalves, P. P., Gruebner, O., Pendse, S. R., Smith, S. L., Sharma, A., & Raviola, G. (2019). Digital innovations for global mental health: opportunities for data science, task sharing, and early intervention. *Current Treatment Options in Psychiatry*, 1-15.

Nguyen, T., Phung, D., Dao, B., & Venkatesh, S. (2014). Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing, 5*, 217-226.

O'Raghallaigh, P., & Frederic, A. (2017). A Framework for Designing Digital Health Interventions. *Journal of the Midwest Association for Information Systems (JMWAIS), 2017*(2).

OECD. (2018). Promoting mental health in Europe: Why and how? In *Health at a Glance: Europe 2018: State of Health in the EU Cycle.* Paris: OECD Publishing.

Oh, J., Yun, K., Hwang, J. H., & Chae, J. H. (2017). Classification of suicide attempts through a machine learning algorithm based on multiple systemic psychiatric scales. *Frontiers in psychiatry, 8*, 192.

Olfson, M. (2016). Building the mental health workforce capacity needed to treat adults with serious mental illnesses. *Health Affairs, 35*(6), 983-990.

Orabi, A. H., Buddhitha, P., Orabi, M. H., & Inkpen, D. (n.d.). Deep learning for depression detection of Twitter users. *Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic*, (pp. 88-97).

Parish-Plass, J., & Lufi, D. (2011). Sport-based roup therapy program for boys with ADHD or with other behavioral disorders. *Child & Family Behavior Therapy, 33*(3), 217-230.

Perich, T., Manicavasagar, V., Mitchell, P. B., & Ball, J. R. (2013). The association between medita-tion practice and treatment outcome in Mindfulness-based Cognitive Therapy for bipolar disorder. *Behaviour research and therapy,, 51*(7), 338-343.

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep Contextualized Word Representations. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1*, pp. 2227-2237.

Poria, S., Cambria, E., Winterstein, G., & Huang, G.-B. (2014). Sentic patterns: Dependency-based rules for concept-level sentiment analysis. *Knowledge-Based Systems, 69*, 45-63.

Preece, J., & Shneiderman, B. (2009). The reader-to-leader framework: Motivating technology-mediated social participation. *AIS transactions on human-computer interaction, 1*(5).

Preece, J., Nonnecke, B., & Andrews, B. (2004). The top five reasons for lurking: improving community experiences for everyone. *Computers in human behavior*, 201-223.

Riessman, F. (1997). Ten self-help principles. *Social Policy, 27,* 6-12.

Riper, H., Blankers, M., Hadiwijaya, H., Cunningham, J., Clarke, S., Wiers, R., & Cuijpers, P. (2014). Effectiveness of guided and unguided low-intensity internet interventions for adult alcohol misuse: a meta-analysis. *PLos One, 9*(6).

Robert, G. F., Kohli, R., & Krishnan, R. (2011). Editorial overview—the role of information systems in healthcare: current research and future trends. *Information Systems Research*, 419-428.

Roehrig, C. (2016). Mental disorders top the list of the most costly conditions in the United States: $201 billion. *Health Affairs, 35*(6), 1130-1135.

Rüegger, D., Stieger, M., Flückiger, C., Allemand, M., & Kowatsch, T. (2017). Leveraging The Potential Of Personality Traits For Digital Health Interventions : A Literature Review On Digital Markers For Conscientiousness And Neurotism. *MCIS 2017 Proceedings.*

Saleem, S., Pacula, M., Chasin, R., Kumar, R., Prasad, R., Crystal, M., . . . Speroff, T. (2012). Automatic detection of psychological distress indicators in online forum posts. *Signal & Information Processing Association Annual Summit and Conference* (pp. 1-4). IEEE.

Saloner, B., McGinty, E. E., Beletsky, L., Bluthenthal, R., Beyrer, C., Botticelli, M., & Sherman, S. G. (2018). A public health strategy for the opioid crisis. *Public Health Reports, 133*(1), 24S-34S.

Saxena, S., Thornicroft, G., Knapp, M., & Whiteford, H. (2007). Resources for mental health: scarcity, inequity, and inefficiency. *The Lancet, 370*(9590), 878-889.

Scholz, M., Dorner, V., Schryen, G., & Benlian, A. (2017). A configuration-based recommender system for supporting e-commerce decisions. *European Journal of Operational Research, 259*(1), 205-215.

Shatte, A., Hutchinson, D., & Teague, S. (2019). Machine learning in mental health: A scoping review of methods and applications. *Psychological Medicine, 49*(9), 1426-1448.

Shepherd, A., Sanders, C., Doyle, M., & Shaw, J. (2015). Using social media for support and feedback by mental health service users: thematic analysis of a twitter conversation. *BMC psychiatry, 15*(1), 29.

Smailhodzic, E., Boonstra, A., & Langley, D. (2016). Social media disruptive change in healthcare: Responses of healthcare providers? *European Conference on Information Systems (ECIS).*

Solomon, O. (2010). What a dog can do: Children with autism and therapy dogs in social interaction. *Ethos, 38*(1), 143-166.

Spijkerman, M., Pots, W. T., & Bohlmeijer, E. T. (2016). Effectiveness of online mindfulness-based interventions in improving mental health: A review and metaanalysis of randomised controlled trials. *Clinical psychology review, 45*, 102-114.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929-1958.

Standard Eurobarometer 88. (2018, May 15). Media Use in the European Union. Standard Eurobarometer 88.

Stanford. (2017, October 15). *Neural Networks Dependency Parser*. Retrieved from https://nlp.stanford.edu/software/nndep.html

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sen-timent analysis. *Computational linguistics, 37*(2), 267-307.

Tan, J., Rogers, C. R., Israel, S., & Benrimoh, D. (2019). Primed for Psychiatry: The role of artificial intelligence and machine learning in the optimization of depression treatment. *University of Toronto Medical Journal, 96*(1).

Tao, Y. (2014). *Sentiment Analysis Method Review in Information Systems Research.*

Tran, T., & Kavuluru, R. (2017). Predicting mental conditions based on history of present illness in psychiatric notes with deep neural networks. *Journal of biomedical informatics, 75*, S138-S148.

van den Bekerom, B. (2017). Using machine learning for detection of autism spectrum disorder. *Proc. 20th Student Conf. IT*, (pp. 1-7).

van Uden-Kraan, C. F., Drossaert, C. H., Taal, E., Seydel, E. R., & van de Laar, M. A. (2009). Participation in online patient support groups endorses patients' empowerment. *Patient education and counseling, 74,* 61-69.

van Uden-Kraan, C. F., Drossaert,, C. H., Taal, E., Shaw, B. R., Seydel, E. R., & van de Laar, M. A. (2008). Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. *Qualitative health research, 18,* 405-417.

Villamil, M. B., & Garcia, E. (2017). Virtual Articulator – Aid Simulator at Diagnosis, Pre-Surgical Planning and Monitoring of Bucomaxilofacial Treatment. *50th Hawaii International Conference on System Sciences.*

Wahle, F., & Kowatsch, T. (2014). Towards the Design of Evidence-based Mental Health Information Systems: A Preliminary Literature Review. *Thirty Fifth International Conference on Information Systems.* Auckland.

Wahle, F., Bollhalder, L., Kowatsch, T., & Elgar, F. (2017). Toward the design of evidence-based mental health information systems for people with depression: a systematic literature review and meta-analysis. *Journal of medical internet research, 19*(5).

Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. *Clinical Psychological Science, 5*(3), 457-469.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., . . . Campbell, A. T. (2017). StudentLife: Using smartphones to assess mental health and academic performance of college students. *Mobile Health*, 7-33.

Wang, R., Min, A. S., Abdullah, S., Brian, R., Campbell, A. T., Choudhury, T., & Hauser, M. (2016). CrossCheck: toward passive sensing and detection of mental health changes in people with schizophrenia. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, (pp. 886-897).

Wang, X., Zhang, C., Ji, Y., Sun, L., Wu, L., & Bao, Z. (2013). A depression detection model based on sentiment analysis in micro-blog social network. *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 201-213). Springer.

Watson, H. J. (2019). Update Tutorial: Big Data Analytics: Concepts, Technology, and Applications. *Communications of the Association for Information Systems, 44*(1), 21.

Weathers, F. W., Keane, T. M., & Davidson, J. R. (2001). Clinician-Administered PTSD Scale: A review of the first ten years of research. *Depression and anxiety, 13*, 132-156.

Weathers, F. W., Keane, T. M., & Davidson, J. R. (2001). Clinician-Administered PTSD Scale: A re-view of the first ten years of research. *Depression and anxiety, 13*(3), 132-156.

Whelanabc, P., Machinabc, M., Lewisabc, S., Buchanabc, I., Sandersab, C., Applegated, E., & Stocktona , C. (2015). Mobile early detection and connected intervention to coproduce better care in severe mental illness. *MEDINFO 2015: EHealth-enabled Health: Proceedings of the 15th World Congress on Health and Biomedical Informatics.* IOS Press.

Wilens, T. E. (2004). Impact of ADHD and its treatment on substance abuse in adults. *Journal of Clinical Psychiatry, 65*, 38-45.

Winzelberg, A. (1997). The analysis of an electronic support group for individuals with eating disorders. *Computers in Human Behavior, 13*, 393-407.

Wongkoblap, A., Vadillo, M. A., & Curcin, V. (2017). Researching mental health disorders in the era of social media: systematic review. *Journal of medical Internet research, 19*(6), 228.

World Economic Forum. (2019). *Global Risks Report 2019.* Retrieved from http://www3.weforum.org/docs/WEF_Global_Risks_Report_2019.pdf

World Health Organization. (2011). Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level. In R. b. Sekretariat. Geneva: World Health Organization.

World Health Organization. (2018a). *Mental Disorders.* Retrieved 6 11, 2018, from https://www.who.int/news-room/fact-sheets/detail/mental-disorders

World Health Organization. (2018b). *Mental Health of Older Adults.* Retrieved 6 11, 2018, from https://www.who.int/news-room/fact-sheets/detail/mental-health-of-older-adults

World Health Organization. (2018c). *Suicide Data.* Retrieved 6 11, 2018, from http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/

Wunderink, L., Systema, S., Nienhuis, F. J., & Wiersma, D. (2009). Clinical recovery in first-episode psychosis. *Schizophrenia Bulletin*, 362-369.

Xi, W., Zhiya, Z., & Kang, Z. (2015). The Evolution of User Roles in Online Health Communities – A Social Support Perspective. *PACIS 2015 Proceedings*, (p. 121).

Xu, H., Phan, T. Q., & Tan, B. (2013). How does online social network change my mood? An empirical study of depression contagion on social network sites using text-mining. *International Conference on Information Systems.*

Zanarini, M. C., Frankenburg, F. R., Dubo, E. D., Sickel, A. E., Trikha, A., Levin, A., & Reynolds, V. (1998). Axis I comorbidity of borderline personality disorder. *American Journal of Psychiatry, 155*(12), 1733-1739.

Zhang, C., & Kamal, M. (2013). A LENS INTO INVESTIGATING PATIENT ENGAGEMENT USING HEALTH INFORMATION TECHNOLOGY. *SAIS 2013 Proceedings, 41*.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit selbstständig angefertigt habe. Sämtliche aus fremden Quellen direkt und indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher nicht zu Prüfungszwecken verwendet und noch nicht veröffentlicht.

_____

Elena Davcheva