International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 9
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

Data Categorization and Review Identification on Twitter Using WordNet Implicit Aspect Sentiment Analysis

Kale Santosh Shivnath
Ph. D. Scholar
Department of Computer Science and Engineering
Dr. A. P. J. Abdul Kalam University, Indore MP.
kalesantosh0101@gmail.com

Dr. Sonawane Vijay Ramnath
Research Supervisor
Department of Computer Science and Engineering
Dr. A. P. J. Abdul Kalam University, Indore MP.
vijaysonawane11@gmail.com

***Abstract*—** Social media review analysis has developed into a fascinating profession that addresses important public safety issues that are respected globally. Sentiment analysis (SA) on Twitter is still a topic of ongoing attention in this profession. Tweet datasets for sentiment opposition bracket are subjected to aspect-grounded SA, a method that allows information to be extracted, dissected, and categorized in order to predict social media evaluations. The implicit aspect for social media review tweets that is implied by adjectives and verbs is the subject of this paper's aspect identification job. In order to improve training data for [1) Social media review Implicit Aspect Rulings Discovery (IASD) and Social media review Implicit Aspect Identification (IAI), a mongrel model is suggested. It is based on WordNet semantic relations and the Term-Weighting scheme. Three classifiers—Multinomial Naïve Bayes, Support Vector Machine, and Random Forest—are used to estimate the performance on three Twitter social media review datasets. The obtained results show the value of verbs in training data enhancement for social media review IASD and IAI, as well as the efficacy of WN reversal and description relations.

Keywords- Sentiment analysis based on implicit aspects, information retrieval, machine learning, supervised techniques, frequency model, WordNet, detection of hate speech on social media, and sentiment analysis on Twitter (HCTS).

I. INTRODUCTION

Because of the World Wide Web's massive proliferation, sentiment analysis, or SA, has become one of the most active motifs in information reclamation and textbook mining. The subject area known as SA examines how people automatically analyze their thoughts, feelings, views, assessments, positions, and feelings toward reality and the characteristics that are stated in written textbooks [1]. The reality may take the form of goods, services, relationships, personalities, occasions, or themes like to those seen in social media reviews. Three scenarios of granularity document, judgment, or aspect position have seen a significant amount of SA exploration. Position of the aspect The most granular type, known as SA, extracts viewpoints voiced against various facets of reality.

For the most part, it is insufficient to categorize opinion textbooks as positive or negative at the document or judgment positions. These groupings don't specify the subject matter of each opinion, or what each opinion is against. In fact, just because a document or decision assesses one reality, it doesn't follow that it covers all facets of that reality[2]. Aspects must be found before determining if the sentiment is favorable, negative, or neutral for each element for a more thorough examination. The application of aspect- grounded sentiment analysis, or ABSA, yields these forfeiture-grained results [3]. This final analysis takes into account the relationships between the document's elements and the object of the opinion. antagonism (either a favorable or negative sentiment conveyed in the viewpoint). An aspect is a notion that the author of the paper bases his or her opinions on. There are two categories of aspects: implicit and clear-cut aspects. Unambiguous elements match terms that are used specifically in the document. When there is a discrepancy, an implicit component is one that isn't stated clearly in the text. Adjectives, adverbs, verbs, and phrasal verbs can all be used to communicate implicit characteristics, which are incredibly significant because they can express opinions and improve the efficiency of SA systems.

In the near future, SA and IASA in particular are expected to introduce a viable strategy for social media review vaccination [4]. IASA is currently used for social media review forestallment systems that are meant to help social media review forestallment and fear reduction. These systems are comparable to neighborhood social media review standing systems and safety of academy platforms. Relating the set of married social media reviews based on their types, locations, and individualities is the most difficult assignment in the

***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

social media review validation space. This is especially difficult when the information is implied rather than stated clearly in the data. This script uses Implicit Aspect Grounded Sentiment Analysis (IASA) to break apart the social media review patterns..

IASA functions in three ways when it comes to social media review vaticination: implicit aspect rulings discovery (IASD), implicit aspect identification (IAI), and sentiment bracket.

Twitter is a credible and logical source of data for social media review datasets that is often employed in pattern detection and forestallment techniques. The biggest problem with extracting implicit aspect rules from this widely used social media platform is the vast amount of tweets that have incorrect spelling, hashtags, URLs, and/or incorrect alphabetization. Thus, in order to categorize applicable and inapplicable rulings, preparatory treatment and information reclamation techniques are needed during the compilation of implicit aspect social media review datasets. "Implicit aspect tweets or rulings discovery" is the term for this procedure.
The next step is to do Implicit Aspect Identification (IAI) after creating social media review datasets. Implicit aspect term (IAT) aggregation and birth are included in IAI. IAT birth tries to root adjectives and verbs inferring aspects for each implicit aspect evaluation. Following this, concepts that have been uprooted and suggest the same aspect are combined into a single implicit aspect in the IAT aggregate. Sentiment brackets can be used to categorize opinions toward each implicit aspect into positive and negative classifications after the aspect has been identified.

The focus of this study is on implicit aspect identification and implicit aspect judgment discovery. In order to accommodate both IASD and IAI, a cold-blooded model that couples WordNet Synonym and Definition semantic relations with a term weighting method is presented for training data enhancement. Three classifiers—Multinomial Naïve Bayes (MNB), Support Vector Machine (SVM), and Random Forest (RF)—are empirically used to estimate the suggested mongrel model on three Twitter social media review datasets. The analysis demonstrates how our method aids in the three classifiers' successful completion of IASD and IAI tasks.

II. RELATED WORKS

Considerable research has been published in the field of aspect-based sentiment analysis [10, 11], and several have attempted to solve implicit aspect identification. The two main styles used for this work are supervised learning approaches and verbal predicated styles. Semantic exposure styles are employed in verbal predicated techniques to accommodate double type [12]. One of the most common verbal approaches in this discipline is dictionary-based. In [13], authors test a novel verbal system based on part-speech tracking, SentiWordNet, WordNet, and a weighted model provided by sentiment-type natural language processing (NLP) weight assignment tools. Their outcomes surpass the initial use of WEKA Naïve Bayes Classifier and validate the efficacy and contribution of the While several have attempted to solve the implicit aspect identification, a sizable number of works have been published in aspect-predicated sentiment analysis[10],[11]. The two primary techniques used for this assignment are verbal predicated and supervised learning approaches. The semantic exposure styles are employed to promote double type among the verbal predicated approaches[12]. In this sector, dictionary-based methods are among the most widely utilized linguistic approaches. The authors of [13] attempt a novel verbal system technique based on part-speech tracking, SentiWordNet, WordNet, and a weighted model provided by natural language processing NLP (weight assignment programs) in sentiment type. They demonstrate the efficacy and generosity of the verbal
A significant number of recent research on SA focus specifically on opinions and trends on Twitter. Scholars and researchers from all around the world have conducted a great deal of discourse in this area [18]. Tweet sentiment analysis is done in three main ways: by linking opinion target, by classifying the sentiment opposition of tweets, and by identifying clear or implicit aspects. The following procedures were used in the end of the discussion to perform sentiment type on Twitter: data collection, information recovery, and sentiment type.Term frequency-Inverse Document frequency (TF-IDF) is one of the most often utilized methods for tweet selection and text classification in information recovery.. This weighting technique is simple to calculate, use, and comprehend. However, its flaw is widely acknowledged. The TF-IDF needs to be improved for unbalanced datasets in order to achieve better results [22]. Sentiment analysis is quickly becoming an indispensable tool for data discovery and social media review containment. In order to predict the timing and location of a particular social media review, writers in[4] lay out a sentiment analysis approach based on wordbook styles and combine it with kernel viscosity estimation based on actual social media review incidents. Their methodology yields a noteworthy accomplishment when juxtaposed with the conventional model. Using a cold-blooded model, others in[8] tackled aspect- based sentiment analysis for social media review tweets. Based on SentiWordNet and Natural Language Processing techniques, the mongrel model predicts the opposition to hateful social media reviews by tweets and identifies the subjectivity of social media reviews..

III. PROPOSED FRAMEWORK

In order to reflect implicit aspects and improve training data, the proposed study draws inspiration from the first phase of WordNet retrieved terms for adjectives and verbs according to Synonym and Definition subsets, as well as a new Term Weighting model. The motivation behind this comes from two interests: [1] how to best represent implicit characteristics by combining these WN extracted items with their corpus verbs and adjectives, and [2] how to make this combination as informative for both goals as possible: implicit facet Sentence detection and implicit aspect identification in social media reviews

***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

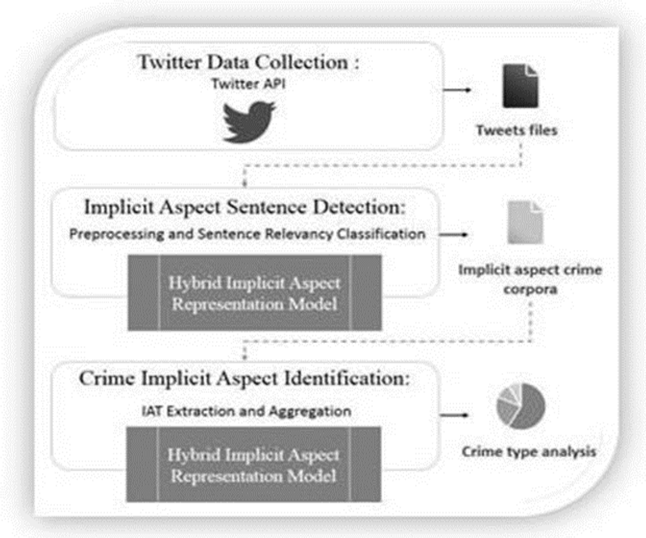

Twitter Data Collection :
Twitter API
Tweets files
Implicit Aspect Sentence Detection:
Preprocessing and Sentence Relevancy Classification
Hybrid Implicit Aspect Representation Model
Implicit aspect crime corpora
Crime Implicit Aspect Identification:
IAT Extraction and Aggregation
Hybrid Implicit Aspect Representation Model
Crime type analysis

Fig. 1. Abstract Process of the Proposed Framework

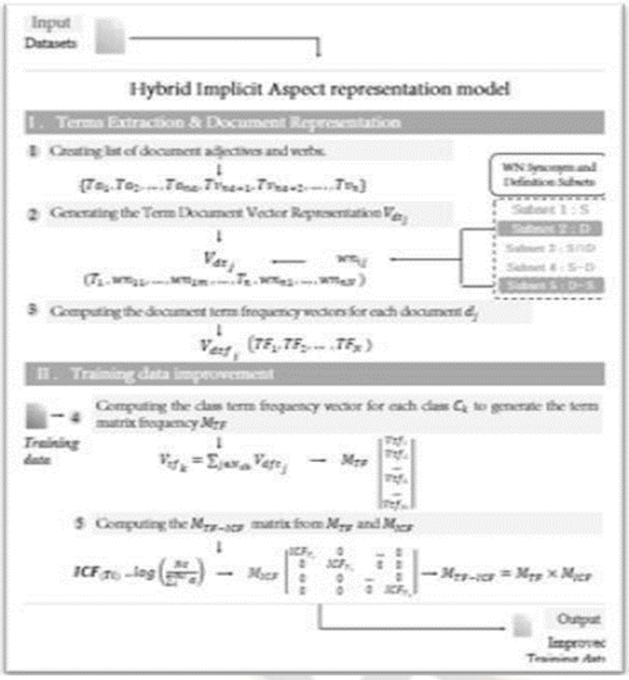

Input
Datasets
Hybrid Implicit Aspect representation model
I. Terms Extraction & Document Representation
1 Creating list of document adjectives and verbs.
2 Generating the Term Document Vector Representation
WN Synonym and Definition Subsets
Subset 1 : S
Subset 2 : D
Subset 3 : S\D
Subset 4 : S−D
Subset 5 : D−S
3 Computing the document term frequency vectors for each document
II. Training data improvement
4 Computing the class term frequency vector for each class to generate the term matrix frequency
Training data
5 Computing the matrix from and
Output
Improved Training data

Fig. 2. Summary of the Proposed Hybrid Implicit Aspect Representation Model.

Aspect Tweets are reviewed on social media. Additionally, some tweets contain hashtags, URLs, grammar and spelling errors, and sources of data. The preprocessing stage of the IASD phase removes these obstacles in order to guarantee improved implicit aspect detection for social media reviews.

Phase 1: Implicit Aspect Sentence Detection

IASD phase, as shown in figure 3, consists of preprocessing and sentence relevancy classification process:

Preprocessing

Eliminating noisy data is the first stage in the preparation process. The first step in the process is to remove URL[17], @usernames, and #hashtags. Subsequently, tweets are parsed using the Part of Speech tagger (POS) to extract verbs and adjectives that may indicate implicit aspect phrases that allude to social media evaluations. We utilized the popular compression words approach for tweets to extract elongated terms that had more than three consecutive occurrences of the same letter [17]. It is employed to retrieve the correct spelling of a word that the WordNet dictionary accepts. Finally, the stop words are eliminated from.

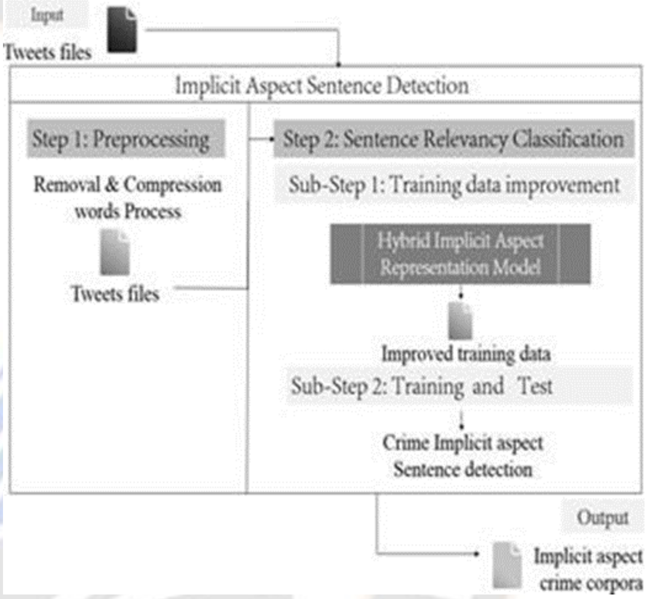

Input
Tweets files
Implicit Aspect Sentence Detection
Step 1: Preprocessing
Removal & Compression words Process
Tweets files
Step 2: Sentence Relevancy Classification
Sub-Step 1: Training data improvement
Hybrid Implicit Aspect Representation Model
Improved training data
Sub-Step 2: Training and Test
Crime Implicit aspect Sentence detection
Output
Implicit aspect crime corpora

Fig. 3. Social media review Implicit Aspect Sentences Detection using Hybrid Model

3.1 Sentence relevancy classification

The goal of Sentence Relevancy Classification, which consists of two sub-steps, is to construct an implicit aspect social media review corpus from each dataset by classifying tweets as relevant or irrelevant. The suggested hybrid model is used in the first sub-step to improve training data by preprocessing tweet datasets. In the second sub-step, social media review implicit aspect corpora are created by using the enhanced training data to construct a classification model for social media review implicit aspect phrases.

i i $\sum Nc\ \alpha$

Where α takes 0 if term Ti does not appear in class Ck , and 1 in otherwise. The new ICF boosts the importance of terms appearing only at one class and penalizes irrelevantterms.

The final $MTF{-}ICF$ (NC, N) matrix is obtained by
$MTF{-}ICF = MTF \times MICF$

Where the $MICF$ (N, N) is the diagonal matrix of ICF.

As mentioned earlier, our approach proceeds in three phases (shown in Fig.1) as follows:

Phase 1: Twitter Data Collection

Using the official Twitter Search API v1.1, the data is collected from Twitter. Real-time access to and extraction of tweets in response to a particular query is made possible by the Twitter API. We generate three distinct social media review datasets based on over fifty requests [32]. The first two datasets examine the four main categories of social media reviews: assault, robbery, rape, and homicide.

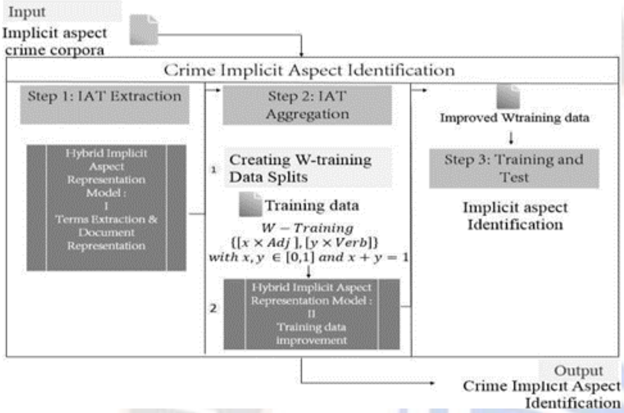

Input
Implicit aspect crime corpora
Crime Implicit Aspect Identification
Step 1: IAT Extraction
Hybrid Implicit Aspect Representation Model : I Terms Extraction & Document Representation
Step 2: IAT Aggregation
1
Creating W-training Data Splits
Training data
W − Training {[x × Adj], [y × Verb]} with x, y ∈ [0,1] and x + y = 1
2
Hybrid Implicit Aspect Representation Model : II Training data improvement
Improved Wtraining data
Step 3: Training and Test
Implicit aspect Identification
Output
Crime Implicit Aspect Identification

Fig. 4. Social media review Implicit Aspect Identification using Hybrid Model.

Phase 2: Social Media Review Implicit Aspect Identification

The challenge, as depicted in Fig. 4, attempts to extract implicit characteristics of social media reviews from phase 2 corpora. The suggested hybrid approach addresses IAT extraction and aggregation through two phases.

The hybrid model's Terms Extraction & Document Representation phases are used in IAT Extraction to extract potential implicit aspects that are suggested by verbs and adjectives. The hybridj model then returns the document term frequency vectors $Vdtf$ for each dataset document, representing the contribution of verbs and adjectives along with associated WN extracted terms for that particular document[26].

Several W-Training data splits are used in IAT aggregation to implement the hybrid model's training data enhancement phases. Using a weighing system that allocates distinct weights to verbs and adjectives, these splits are produced. The effect of employing various ratios of adjectives and verbs on the enhancement of training data for social media review datasets is assessed using this weighting system. Equation 9 calculates each W-Training data split as follows:$W - Training\ split\ = *, \times Adj$ -, $,y \times Verb$-+
$where\ x, y \in$,0,1- $and\ x + y = 1$

Table I. Size Of Datasets

SIZE OF DATASETS	Social media review dataset1	Social media review dataset2	Hate social media review dataset
Number of sentences	2k	3k	6k
Number of implicit aspect sentences	357	641	648
Number of irrelevant sentences	1643	2359	5352
Number of Training data for implicit aspect	180	350	300
Number of Training data for implicit aspect	670	1500	3500

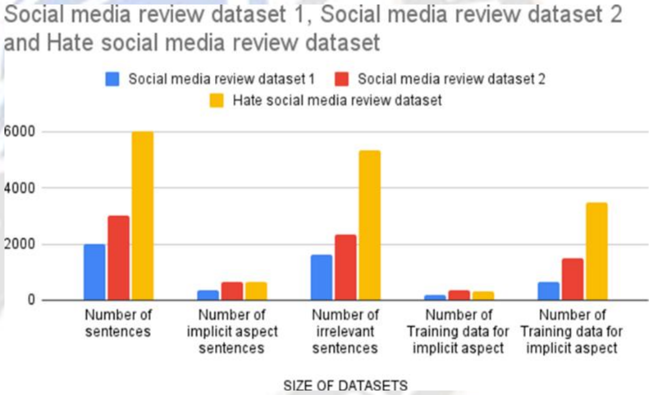


Fig. 5. Social media review dataset 1

Table II MNB, SVM And RF For Relevant / Irrelevant Classification

MNB, SVM AndRF	*Social media review dataset 1*			*Social media review dataset 2*			*Hate social media review dataset*		
	MNB	***SVM***	***RF***	***MNB***	***SVM***	***RF***	***MNB***	***SVM***	***RF***
(1)	0.51	0.63	0.63	0.52	0.65	0.63	0.63	0.69	0.68
(2)	0.57	0.65	0.65	0.59	0.68	0.68	0.74	0.77	0.75
(1)/(2)	11.6%	3.1%	3.1%	13.4%	4.6%	7.9%	17.4%	11.5%	10.2%
(3)	0.71	0.68	0.68	0.69	0.71	0.69	0.76	0.78	0.78
(4)	0.74	0.71	0.71	0.74	0.74	0.72	0.74	0.78	0.78
(5)	0.78	0.71	0.71	0.74	0.73	0.73	0.80	0.81	0.80
(6)	**0.83**	**0.88**	**0.87**	**0.79**	**0.80**	**0.79**	**0.82**	**0.89**	**0.87**
(3)/(5)	9.8%	4.4%	4.4%	7.2%	2.8%	5.7%	5.2%	3.8%	2.5%
(4)/(6)	12.1%	23.9%	22.5%	6.7%	8.1%	9.7%	10.8%	14.1%	11.5%

***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 September 2023*

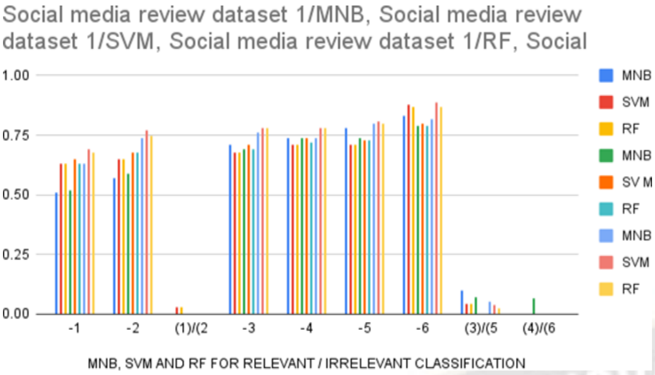
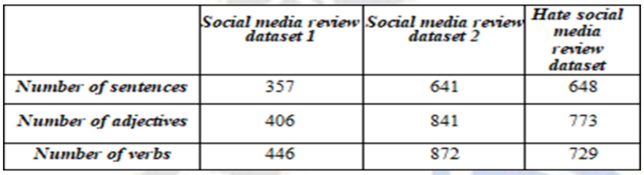


Fig. 6. Social media review dataset 1/MNB

Table III. Number Of Adjectives And Verbs Implying Implicit Aspect For Each Social MediaReview Dataset



	Social media review dataset 1	*Social media review dataset 2*	*Hate social media review dataset*
Number of sentences	357	641	648
Number of adjectives	406	841	773
Number of verbs	446	872	729

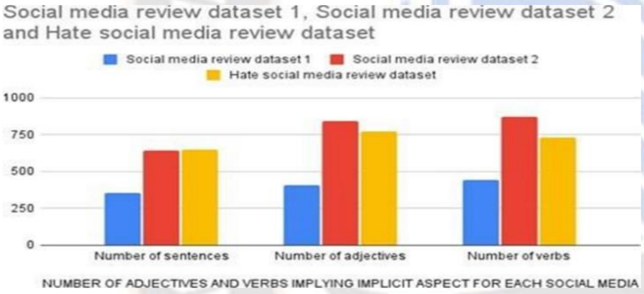


Fig. 6. Social media review dataset 1 and 2

IV. CONCLUSION

We talked about a hybrid approach that uses MNB, SVM, and RF classifiers with superior training data to perform aspect-based sentiment analysis for social media review datasets. We do an empirical and analytical study at the level of:

The IASD phase of the social media assessment involves conducting experiments based on three criteria: using WordNet Synonym relations of adjectives and verbs in the document representation model; using TF-IDF rather than TF-ICF for document representation; and incorporating the best WN subsets of adjectives and verbs. The two main areas of experimentation in the social media implicit aspect identification (IAI) phase are the use of adjectives and verbs to improve training data and the absence of WN words for adjectives and verbs in document representation.

References

1. M. Hu and B. Liu, “Mining and summarizing customer reviews,” in Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, 2004, pp. 168–177.
2. Doaa Mohey El-Din, “Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis” International Journal of Advanced Computer Science and Applications(IJACSA), 7(1), 2016.
3. B. Liu, Sentiment analysis: mining opinions, sentiments, and emotions. New York, NY: Cambridge University Press, 2015.
4. X. Chen, Y. Cho, and S. Y. Jang, “Social media review prediction using Twitter sentiment and weather,” 2015, pp. 63–68.
5. Jermy Prichard, Paul Watters, Tony KRONE, Caroline Spiranovic, and Helen Cockburn, “Social Media Sentiment Analysis: A New Empirical Tool for Assessing Public Opinion on Crime?,” pp. 217–236, 2015.
6. Nisal Waduge, “Machine Learning Approaches For Detect Crime Patterns - Data Gathering and Analysing Techniques,” 2017.
7. P. Burnap et al., “Detecting tension in online communities with computational Twitter analysis,” Technol. Forecast. Soc. Change, vol. 95, pp. 96–108, Jun. 2015.
8. N. Zainuddin, A. Selamat, and R. Ibrahim, “Improving Twitter Aspect-Based Sentiment Analysis Using Hybrid Approach,” in Intelligent Information and Database Systems, vol. 9621, N. T. Nguyen, B. Trawiński, H. Fujita, and T.-P. Hong, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 151–160.
9. Hissah AL-Saif and Hmood Al-Dossari, “Detecting and Classifying Social medias from Arabic Twitter Posts using Text Mining Techniques” International Journal of Advanced Computer Science and Applications(IJACSA), 9(10), 2018.
10. B. Keith, E. Fuentes, and C. Meneses, “A Hybrid Approach for Sentiment Analysis Applied to Paper Reviews,” N. S., p. 10, 2017.
11. K. Schouten, P. O. Box, and D. Rotterdam, “Supervised and Unsupervised Aspect Category Detection for Sentiment Analysis with Co-occurrence Data,” IEEE Trans. Cybern., p. 13, 2017.
12. V. S. Jagtap and K. Pawar, “Sentence-Level Analysis of Sentiment Classification,” Natl. Conf. Emerg. Trends Eng. Technol. Archit., p. 6, 2013.
13. K. Gull, S. Padhye, and D. S. Jain, “A Comparative Analysis of Lexical/NLP Method with WEKA‟s Bayes Classifier,” Int. J. Recent Innov. Trends Comput. Commun., vol. 5, no. 2, p. 7, 2017
14. B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: sentiment classification using machine learning techniques,” in Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, 2002, pp. 79–86.
15. A. Alghunaim, M. Mohtarami, S. Cyphers, and J. Glass, “A Vector Space Approach for Aspect Based Sentiment Analysis,” 2015, pp. 116– 122.
16. G. Abdulsattar A. Jabbar Alkubaisi, S. Sakira Kamaruddin, and H. Husni, “Conceptual Framework for Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naïve Bayes Classifiers,” Int. J. Eng. Technol., vol. 7, no. 2.14, p. 57, Apr. 2018.
17. Devidas S Thosar, Rajashree R Shinde, User Controlling System Using LAN , Asian Journal For Convergence In Technology (AJCT) ISSN -2350-1146: Vol 2 (2016): Issue I
18. V. N. Patodkar and S. I.R, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” IJARCCE, vol. 5, no. 12, pp. 320–322, Dec. 2016.
19. S. Rosenthal, N. Farra, and P. Nakov, “SemEval-2017 Task 4: Sentiment Analysis in Twitter,” p. 17, 2017.
20. B. Gokulakrishnan, P. Priyanthan, T. Ragavan, N. Prasath, and As. Perera, “Opinion mining and sentiment analysis on a Twitter data stream,” 2012, pp. 182–188.
21. M. Ishtiaq, “Sentiment Analysis of Twitter Data Using Sentiment Influencers,” vol. 6, no. 1, p. 9, 2015.
22. H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, “A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle,” p. 6, 2012.
23. Y. Liu, H. T. Loh, and A. Sun, “Imbalanced text classification: A term weighting approach,” Expert Syst. Appl., vol. 36, no. 1, pp. 690–701, Jan. 2009.
24. El Hannach, H. and Benkhalifa , M., “Using Synonym and Definition WordNet Semantic relations for implicit aspect identification in Sentiment Analysis,” Pap. Present. 1st Int. Conf. Netw. Inf. Syst. Secur. NISS 2018 Conf. Tangier Morocco., p. 8, 2018.
25. Junseok Song, Kyung Tae Kim, Byungjun Lee, Sangyoung Kim, and Hee Yong Youn, “A novel classification approach based on Naïve Bayes for Twitter sentiment analysis,” KSII Trans. Internet Inf. Syst., vol. 11, no. 6, Jun. 2017
26. R. Sergienko, M. Shan, and A. Schmitt, “A Comparative Study of Text Preprocessing Techniques for Natural Language Call Routing,” in

Dialogues with Social Robots, vol. 427, K. Jokinen and G. Wilcock, Eds. Singapore: Springer Singapore, 2017, pp. 23–37.

27. Devidas S Thosar, Rajashree R Shinde, Prashant J Gadakh, Pratibha V Kashid, Secure kNN Query Processing in Entrusted Cloud Environments , Asian Journal For Convergence In Technology (AJCT) ISSN -2350-1146: Vol 2 (2016): Issue I.
28. G. Loosli, S. Canu, and L. Bottou, “Training Invariant Support Vector Machines using Selective Sampling,” p. 26, 2005.
29. S. S. Keerthi, S. K. Shevade, C. Bhattacharyya, and K. R. K. Murthy, “Improvements to Platt‟s SMO Algorithm for SVM Classifier Design,” Neural Comput., vol. 13, no. 3, pp. 637–649, Mar. 2001.
30. B. Xu, X. Guo, Y. Ye, and J. Cheng, “An Improved Random Forest Classifier for Text Categorization,” J. Comput., vol. 7, no. 12, Dec. 2012.
31. LEO BREIMAN, “Random Forests,” Machine Learning, The Netherlands, pp. 45, 5–32, 2001.
32. C. J. van RIJSBERGEN, “Information Retrieval,” 2nd ed. Butterworth-Heinemann, 1979.
33. M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, “The WEKA data mining software: an update,” ACM SIGKDD Explor. Newsl., vol. 11, no. 1, p. 10, Nov.2009