2025 Volume: 01 Issue: 07 www.ijrpa.com ISSN 2456-9995 Research Article

International Journal Research Publication Analysis

Page: 01-15

SMART SENTIMENT INTERPRETER USING ML

***Tirthi Jain, Dr.Vishal Shrivastava, Dr. Akhil Pandey**

Computer Science & Engineering, Arya College of Engineering & I.T. Jaipur, India.

Article Received: 03 November 2025

Article Revised: 23 November 2025

Published on: 13 December 2025

***Corresponding Author: Tirthi Jain**

Computer Science & Engineering, Arya College of Engineering & I.T. Jaipur, India. DOI: https://doi-doi.org/101555/ijrpa.1479

ABSTRACT

Social media platforms continuously produce vast amounts of unstructured textual data containing opinions and emotional expressions. Analyzing such information manually is impractical; therefore, automated sentiment analysis has become essential for understanding public attitudes and supporting data-driven decision-making. This paper introduces the SMART Sentiment Interpreter (SSI) — a machine learning–based system designed to accurately classify sentiments from social media comments as positive, negative, or neutral. The system employs a carefully designed preprocessing pipeline to minimize linguistic noise, extracts meaningful features using SentiWordNet, and performs classification through a Support Vector Machine (SVM) model. The proposed approach aims to convert unorganized social media text into actionable insights, enabling organizations to better understand user opinions and trends.

KEYWORDS: Sentiment Analysis, Machine Learning, SVM, Text Mining, Opinion Classification.

1. INTRODUCTION

1.1 Background and Problem Statement

Social media platforms bring people together and provide them with a way to share information, exhibit emotions, and state opinions on various subjects. Among the wide range of analytical applications emanating from such social media activities, sentiment analysis has become a key area in computer science. It is about detecting and classifying textual information into sentiments based on the polarity of emotions, which are generally positive, negative, or neutral.

Sentiment analysis has immense real-world applications related to recommender systems, market research, and business intelligence. For example, product reviews or tweets are studied by organizations to comprehend consumer preferences, enhance services, and predict market trends. Likewise, political analysts can gather an understanding of public opinion based on posts and comments made through social media.

Precisely, sentiment analysis is the interpretation of subjective data to identify the public's attitude towards a particular topical or entity issue. With the increased number of users online, especially on Twitter and Facebook, building an effective system that can analyze this ever-increasing text data represents both a challenge and a necessity.

1.2 The Computational Treatment of Opinion and Text Mining

Sentiment analysis basically deals with the use of NLP, machine learning, and text mining for understanding the opinion expressed in human language. It enables organizations to systematically monitor public perceptions of brands, products, or events. Text mining techniques analyze unstructured information to extract useful insights that can then be quantified and visualized.

In marketing and product development, sentiment analysis measures customer satisfaction, brand loyalty, and campaign effectiveness. The integration of computational tools in sentiment analysis changes it from qualitative feedback to measurable insights that support strategic decisions.

1.3 Aim and Objectives

The core goal of this research study is the design and development of a robust sentiment classification model that will classify user opinions with maximum accuracy from bulk volumes of social media data. Using NLP techniques, machine learning algorithms like Naïve Bayes and SVM, and lexical resources like SentiWordNet, this study attempts to:

1. Understand audience opinions about target topics or products.
2. Support organizations in decision-making based on consumer sentiment.
3. Monitoring and analyzing public attitudes on trending issues.
4. Develop an automated classifier of sentiments in real time.

The SSI model aims to achieve the above-mentioned objectives in order to transform unstructured textual data into meaningful sentiment-based intelligence.

2. LITERATURE SURVEY

2.1 Limitations of Traditional Sentiment Auditing

Traditional approaches to sentiment analysis usually rely heavily on lexical resources and manually predefined rules for opinion classification. Although these methods can provide some idea about the basic level of textual polarity, they often fail to capture subtlety and complexity in human language. User-generated content on social media tends to mix objective statements with subjective opinions, informal language, and even sarcasm. Consequently, rule-based or lexicon-dependent models frequently misclassify sentences with mixed sentiment or ambiguous expressions.

Another major challenge lies in handling the scale and diversity of online data. Social media posts, comments, and reviews appear in massive quantities, often written in various dialects and linguistic styles. Manual inspection of such data is infeasible, emphasizing the need for automated systems capable of capturing semantic nuances. Therefore, modern sentiment analysis research increasingly focuses on improving classification accuracy and adaptability through machine learning and deep learning–based approaches.

2.2 Foundational Research in Sentiment Analysis Techniques (Literature Review)

Several works have contributed a great deal to developing sentiment analysis frameworks and algorithms.

• Isah et al. (2014) proposed a framework for monitoring product safety through the sentiment analysis and text mining of social media. They utilized the Naïve Bayes classifier to compare the sentiment trends across many product categories and demonstrated how the automation of opinion tracking can help manufacturers and regulatory agencies in understanding brand perception.

• Agarwal et al. (2015) proposed a concept-level sentiment analysis method using dependency-based semantic parsing to extract meaningful concepts from text. Their system leverages the ConceptNet ontology to enhance semantic understanding, yielding better classification accuracy over word-based models.

• Haddi, Liu, and Shi (2017) indicated that text preprocessing is a crucial step in sentiment analysis. They demonstrated that filtering and data normalization significantly improve model performance by reducing noise and irrelevant features.

• Duwairi and Qarqaz (2014) investigated Arabic sentiment classification using three supervised algorithms: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and

Naïve Bayes. Their results illustrated the highest precision achieved by SVM, while KNN had the best recall, which showed certain trade-offs between different classifiers.

• Liu et al. (2013) proposed a scalable Naïve Bayes–based classifier integrated with the Hadoop framework for efficient handling of large-scale datasets. Their results reflected moderate accuracy, but as a proof-of-concept work, it demonstrated that distributed architectures can significantly enhance computational efficiency in big-data sentiment analysis.

• Wang and Manning (2012) studied N-gram–based supervised learning. They found that the combination of unigram and bigram features gives better performance. They also suggested that for short texts like tweets, Naïve Bayes often performs comparably to or better than SVM.

• Cambria et al. (2013) explored new trends in opinion mining by discussing the development of opinion sharing over the web, along with the limitations of the existing NLP tools. While machine algorithms efficiently detect binary sentiment-a positive/negative tone-they are still lagging with regard to recognizing complex human expressions.

• Gautam and Yadav (2014) worked on sentiment analysis for Twitter using semantic analysis in conjunction with machine learning. The preprocessing of text for extracting adjectives and the implementation of such algorithms like Naïve Bayes and SVM for classification indicate hybrid models' improvement in performance.

• Niu et al. (2016) proposed a multi-view sentiment analysis approach for the jointly textual-visual modalities that interpret posts containing text and images. This hence showed the rising importance of processing multimodal data in modern sentiment analysis.

• Mathieu, 2017, proposed a deep learning framework using Convolutional Neural Networks and Long Short-Term Memory networks for Twitter sentiment classification. Their results showed that neural networks trained over large, unlabeled datasets outperform classical machine learning methods in sentiment tasks on short texts.

Collectively, these studies underscore the transition from traditional lexicon-based methods to machine learning and deep learning approaches that can handle linguistic complexity, data volume, and contextual sentiment representation.

2.3 Summary of Foundational Research

Various deep learning architectures have been employed for sentiment analysis across the literature using a range of classifiers such as SVM, Naïve Bayes, KNN, Logistic Regression, and CNN-based architectures. Although traditional algorithms useful in structured data remain useful, state-of-the-art models lean towards deep learning models that have better performance in representing contextual dependencies and semantic nuances.

This trend signifies the necessity for flexible frameworks such as the proposed SMART Sentiment Interpreter, which integrates robust preprocessing, lexical enrichment, and supervised classification to provide accurate sentiment predictions on unstructured social media data.

3. Proposed Methodology

3.1 Introduction

This proposed SMART Sentiment Interpreter classifies social media comments into sentiment polarity. The text generated by users usually consists of noisy, non-standard language with many abbreviations and emojis. Hence, effective data preprocessing is needed. The SSI framework has the following three major stages:

1. Pre-processing of raw text
2. Feature extraction
3. Sentiment classification

These stages work sequentially to clean the input data, extract relevant linguistic features, and finally predict the sentiment polarity by utilizing an SVM-based classifier.

3.2 Block Diagram and System Description

Social media posts or tweets serve as inputs to the system architecture. During preprocessing, text cleaning eliminates irrelevant symbols, URLs, and stop words. Expanding abbreviations and slang expressions is done using a predefined dictionary. The cleaned text is tokenized and normalized to ensure uniformity. During the feature extraction phase, identification of entities such as nouns, adjectives, and verbs will be determined through using the SentiWordNet lexical database. These words have sentiment scores associated with them, reflecting positive or negative tendencies. The scores are then aggregated and passed on to the classification module. In the classification stage, a Support Vector Machine algorithm classifies the processed text into one of three classes: positive, negative, or neutral. The margin-based optimization gives SVM robustness and accuracy even on noisy textual data.

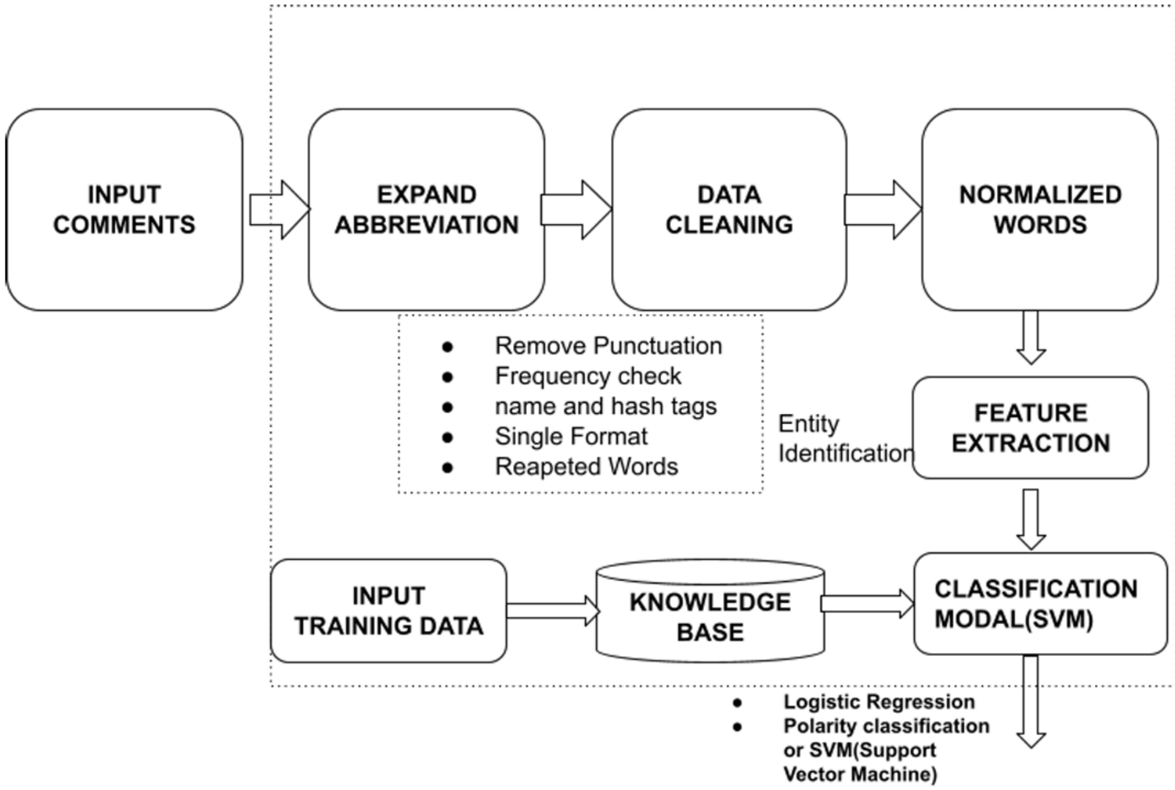

Fig 3.1: Block diagram of the system.

3.3 Core Processing Stages: Pre-Processing, Feature Extraction, and Classification

A. Pre-Processing:

This stage reduces noise and standardizes textual data. The steps include:

- Expanding abbreviations and acronyms (e.g., “btw” → “by the way”).
- Removing punctuation, URLs, and non-text elements.
- Converting all characters to lowercase for consistency.
- Correcting repeated letters and misspellings (e.g., “goooood” → “good”).
- Eliminating stop words while retaining essential negations such as “not.”

B. Feature Extraction:

Key linguistic features are extracted using **SentiWordNet**, which provides semantic and sentiment orientation for words. Adjectives, verbs, and nouns are identified, and their corresponding sentiment scores are calculated. This approach helps capture contextual polarity and enhances the model’s precision.

C. Classification:

The extracted features are then classified using the SVM classifier. SVM separates data points into sentiment categories using optimal hyperplanes, ensuring maximum margin between classes. This results in improved generalization performance when compared with traditional models.

The classifier output is a sentiment label—positive, negative, or neutral—along with a confidence score. The results can be visualized using graphical representations such as bar or pie charts for intuitive interpretation.

4. Experimentation and Technical Specification

4.1 Introduction to Experimentation

Sentiment analysis is the process of using natural language processing, text analysis, and statistics to analyse customer sentiment. The best businesses understand the sentiment of their customers-what people are saying, how they are saying it, and what they mean. Customer sentiment can be found in tweets, comments, reviews, or other places to mention the brand. Sentiment Analysis is the domain of understanding these emotions with software, and it is a must-understand for developers and business leaders in a modern workplace.

4.2 Technical Specifications and Experimental Flow

The implementation of the system utilized standard software and libraries within a controlled environment.

Software Requirements:

Operating System	Windows 10
Technology	Machine Learning and Deep Learning
Operating system	32bit
Coding Language	Python
IDE	Pycharm, Syder
Database	Sqlite

Core Technical Components:

- **Pandas:** Used for data manipulation and analysis, offering data structures and operations for numerical tables and time series.

- **NumPy:** A Python package for numerical processing, consisting of multidimensional array objects (ndarrays) and a collection of routines for array processing. It is essential for handling multi-dimensional arrays, vectors, and matrices.
- **Matplotlib:** A plotting library used for data visualization.
- **Scikit-learn:** A machine learning library providing various classification and regression algorithms, including Support Vector Machines and Logistic Regression. It also provides the **Count Vectorizer** utility, which transforms text into numerical frequency vectors.
- **SVM (Support Vector Machine):** A powerful tool for pattern recognition using a discriminative approach. SVMs use linear and nonlinear separating hyper-planes for data classification. It is a generalized linear classifier with maximum-margin fitting functions that provides regularization to help the classifier generalize better. SVM controls model complexity by regulating the VC dimension.
- **Logistic Regression:** An appropriate predictive analysis to conduct when the dependent variable (sentiment polarity) is dichotomous (binary).
- **Decision Tree:** A potent tool for classification and prediction, structured as a flowchart-like tree.
- **Random Forest:** A supervised learning technique based on ensemble learning, combining multiple decision trees to improve predictive accuracy.

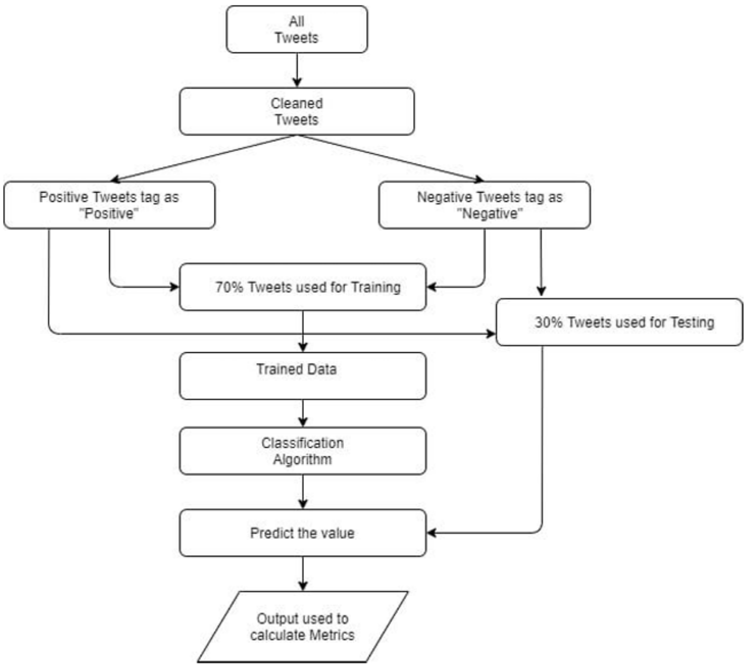

Fig 4.1: Flowchart of sentiment analysis

4.3 Test Cases

The model utilizes specific test cases for validation, ensuring robustness in handling user input and system conditions, including validation for the login screen, username/password, and registration details such as mobile number and email ID.

Test case 1: Login Screen-Sign up

The objective is to check required/mandatory fields when signing up. The expected result is that required mandatory fields should display with the symbol "*" and an instruction line "* field(s) are mandatory".1 The validation for Password and Confirm Password should ensure they are the same, displaying "Password and confirm password should be same" if not.1

Test Case 2: Login test case.

Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Enter The Wrong username or password click on submit button	Username or password	Error comes	Error Should come	P
002	Enter the correct username and password click on submit button	Username and password	Accept	Accept	P

Test Case 3: Registration test case.

Test Case ID	Test Case	Test Case I/P	Actual Result	Expected Result	Test case criteria(P/F)
001	Enter the number in username, middle name, last name field	Number	Error Comes	Error Should Comes	P

001	Enter the character in username, middle name, last name field	Character	Accept	Accept	P
002	Enter the invalid email id format in email id field	Kkgmail.com	Error comes	Error Should Comes	P
002	Enter the valid email id format in email id field	kk@gmail.com	Accept	Accept	P
003	Enter the invalid digit no in phone no field	99999	Error comes	Error Should Comes	P
003	Enter the 10 digit no in phone no field	9999999999	Accept	Accept	P

These conditions are applied to the model, and upon successful satisfaction of all conditions, the main page is accessed for sentiment analysis and comparison with graphs.

5. RESULTS AND DISCUSSION

5.1 Quantitative Results and Polarity Mapping

The system generates a numerical sentiment score for each phrase, usually on a scale of 0 (neutral), 1 (positive), and 2 (negative). This numerical score facilitates the calculation of performance metrics and the visualization of analysis outcomes.

Table of Polarity Scoring

Score	Sentiment Polarity	Example
0	Neutral	"Food is tasty"
1	Positive	"ram is good"

2	Negative	"Ram is bad"

The results are displayed with the statistical information from the dataset, generating a CSV file with categorized details, including the tweeting person's username or comment, type of the tweets and comments, the tweet itself or comment, and finally, sentiment. This output allows for visualization, such as graphs illustrating the ratio of positive, negative, and neutral comments in the dataset (Figure 8.7).

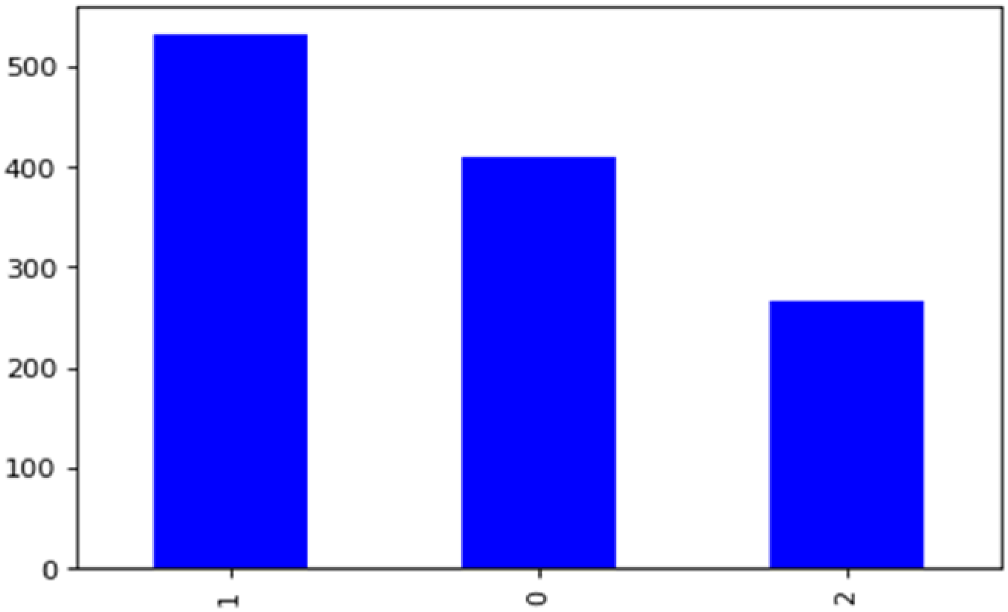

Figure 5.1: Ratio of positive, negative and neutral comments.

5.2 Practical Implications and Advantages

The SSI model provides a valuable tool for strategic intelligence derived from social media data.

Advantages :

- Easy to Operate.
- Understand Customer reviews and emotions.
- Allows us to understand the broader public opinion behind specific topics.
- Can create better services and products with the help of analysis in Business.

Applications :

- Social media Monitoring
- Customer feedback
- Product analysis

- Brand monitoring and reputation management

5.3 DISCUSSION OF LIMITATIONS

Despite its utility, the system has several limitations inherent to computational language processing:

- It can not handle different dialects and informal or slang words.
- It does not contain acronyms and shorthand.
- The proposed system is currently incapable of interpreting sarcasm.
- It is not a hundred per cent efficient.
- Transferring to another language suffers from translational problems and cultural differences.

6. CONCLUSION

6.1 Summary of Contributions

This project represents an approach for text mining and sentiment analysis using the SVM (Support Vector Machine) classifier algorithm and two more classifiers throughout this work. The collected tweets and comments from different resources merged the individual data and created a complete dataset. Sentiment analysis is one of the essential or practical methods for industries and businesses, and many more. These data are initially pre-processed, followed by extracting features from the pre-processed data. Moreover, after the complete process, this will show the opinion on every comment, tweet and many more.The primary constraints of the current SSI architecture—specifically its documented inability to interpret complex linguistic structures such as sarcasm, evolving slang, and various dialects —establish the critical objectives for future research efforts. The system's performance boundary, set by its dependence on explicit lexical matching, necessitates a strategic shift towards computational models capable of capturing semantic context.

The proposed roadmap for future work involves leveraging advanced deep learning techniques, which recent literature indicates are setting new standards for accuracy in SA . Studies comparing traditional SVM/Naive Bayes models against modern Transformer models (like BERT) show that the latter significantly outperforms classical methods, especially in nuanced tasks like sarcasm and complex sentiment classification.

1. **Sarcasm Detection via Contextual Embeddings:** To resolve the current failure in interpreting sarcasm, future research must integrate deep learning models, such as Transformer-based architectures. These models learn complex contextual and sequential

relationships, enabling them to capture the subtle semantic dissonance when positive language is used with negative intent—a task that exceeds the capacity of current, localized feature weighting.

2. **Adaptive Slang and Dialect Integration:** To mitigate the critical out-of-vocabulary challenge posed by rapidly evolving slang and community-specific dialects, the system requires an adaptive learning mechanism. This involves moving beyond static dictionaries to incorporate self-supervised learning methods that can dynamically update the lexicon and word embeddings based on constantly emerging social media usage patterns. This evolution would transform feature extraction from a fixed, rule-based approach into a fluid, data-driven one.

7. REFERENCES

1. Derbala, Y. A., Slim, S. O., & Aboutabl, A. E. (2024). A Survey of Sentiment Analysis and Sarcasm Detection: Challenges, Techniques, and Trends. International Journal of Electrical and Computer Engineering Systems, 15(1), 2953. [https://doi.org/10.32985/ijeces.15.1.7]
2. Bandyopadhyay, I., & Karmakar, R. (2024). Deep Learning-based Sentiment Analysis of Olympics Tweets. arXiv preprint arXiv:2407.12376. [https://doi.org/10.48550/arXiv.2407.12376]
3. D'Souza, S., & Balamurugan, V. (2024). Comparative Analysis of SVM and BERT for Sentiment and Sarcasm Detection in the Boycott of Israeli Products on Platform X. Journal of Intelligent Systems: Theory and Application, 7(1).
4. Rahman, S. M. A. et al. (2023). Sentiment Analysis: A Survey. arXiv preprint arXiv:2311.11250. [https://doi.org/10.48550/arXiv.2311.11250]
5. Liao, Y., et al. (2021). Comparison of Naive Bayes and SVM Algorithm based on sentiment analysis using review dataset. Proceedings of the 2019 8th International Conference System Modeling and Advancement in Research Trends (SMART).
6. Haddi, Emma, Xiaohui Liu and Yong Shi, "The Role of Text Preprocessing in Sentiment Analysis", Procedia Computer Science. Elsevier, Vol. 17, pp. 26-32, 2017.
7. Cliche, Mathieu. "BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs." arXiv preprint arXiv:1704.06125 (2017). [https://arxiv.org/abs/1704.06125]
8. Niu, Teng, et al. "Sentiment analysis on multi-view social data." International Conference on Multimedia Modelling. Springer, Cham, 2016.

9. Zhu, Maoran, et al. "Identification of Opinion Leaders in Social Networks Based on Sentiment Analysis: Evidence from an Automotive Forum." (2016).
10. Agarwal, Basant, Soujanya Poria, Namita Mittal, Alexander Gelbukh and Amir Hussain, "Concept-level Sentiment Analysis with Dependency-based Semantic Parsing: A Novel Approach", Cognitive Computation. Springer, Vol. 7, No. 4, pp. 487-499, 2015. [https://doi.org/10.1007/s12559-014-9316-6]
11. Isah, Haruna, Paul Trundle and Daniel Neagu, "Social Media Analysis for Product Safety using Text Mining and Sentiment Analysis", In Computational Intelligence (UKCI) 14th UK Workshop onBayes Workshop IEEE, pp. 1-7, 2014. [https://doi.org/10.1109/UKCI.2014.6930158]
12. Duwairi, Rehab M and Islam Qarqaz, "Arabic Sentiment Analysis using Supervised Classification", International Conference on Future Internet of Things and Cloud (FiCloud). IEEE, pp. 579-583, 2014.
13. Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of Twitter data using machine learning approaches and semantic analysis." Contemporary Computing (IC3), 2014 seventh international conference on IEEE, 2014.
14. Liu, Bingwei, Erik Blasch, Yu Chen, Dan Shen and Genshe Chen, "Scalable Sentiment Classification for Big Data Analysis using Naive Bayes Classifier", International Conference Big Data. IEEE, pp. 99-104, 2013.
15. Cambria, Erik, et al. "New avenues in opinion mining and sentiment analysis." IEEE Intelligent Systems 28.2 (2013): 15-21.
16. Wang, Sida, and Christopher D. Manning. "Baselines and bigrams: simple, good sentiment and topic classification." Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Short Papers-Volume 2. Association for Computational Linguistics, 2012.
17. Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." (2007): 324.
18. Liu, Bing, Minqing Hu, and Junsheng Cheng. "Opinion observer: analysing and comparing opinions on the web." Proceedings of the 14th international conference on the World Wide Web. ACM, 2005.
19. Kim, Soo-Min, and Eduard Hovy. "Determining the sentiment of opinions." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

20. Pang, B., L. Lee, and S. Vaithyanathan, Thumbs up?: "sentiment classification using machine learning techniques", Acl-02 Conference on Empirical Methods in Natural Language Processing, 2002, p. 79--86.

21. McCallum, Andrew, and Kamal Nigam. "A comparison of event models for Naive Bayes text classification." AAAI-98 Workshop on learning for text categorisation. Vol. 752. 1998.