

SOCIAL MEDIA OPINION ANALYSIS

Using NLP

G. Shiva prasad¹, Ch. Navya², G. Sahithi³, S. Shivamani⁴, K. Rakesh⁵, Ramdas Vankdothu⁶,

P.Nagaraju⁷

^{2,3,4,5}BTech Student, Department of CSE, Balaji Institute of Technology and Science, Laknepally,
Warangal, India

^{1,7} Assistant Professor, Department of CSE, Balaji Institute of Technology & Science, Laknepally,
Warangal, India

ABSTRACT

Sentiment analysis on social media contributes towards public opinion perception, guiding market strategies, political commentary, as well as buyer sentiment. Attitudes and brand loyalty are crafted by social networking sites like Facebook and "X" (Twitter), and having sentiment analysis on these is instrumental in tracking directions. The present study analyzes over 1000 newscast-related posts for the purpose of identifying sentiment direction. Social networking not only facilitates communication but affects user behavior too. Sentiment analysis is vital for organizations, policymakers, and businesses to take decisions based on facts, improve customer engagement, and react to public matters effectively. With user-generated content expanding exponentially, sentiment analysis has become a must to make effective use of the digital world.

The proposed system enhances existing methodologies by incorporating advanced word embedding and transformer-based models like BERT to improve accuracy and context understanding.

1. INTRODUCTION

In the context of contemporary socio-technical systems, social media has transformed profoundly in the means of communication and expression of feeling and sharing information. The social media platforms like Facebook, X (formerly Twitter), Instagram, and YouTube with their millions of active users produce massive amounts of content that captures emotions, attitudes, and opinions. These sentiments are crucial in shaping public opinion, brand image, and buying behavior.

The volume of data produced every day per user, however, has accelerated to such an extent, that it poses a challenge to individuals or businesses attempting to gain insights from it. Sentiment Analysis, which is a subfield of Artificial Intelligence (AI) and Natural Language Processing (NLP), solves this problem through automated extraction and sorting of feelings into positive, negative, or neutral. It is now possible for researchers, business people, and policymakers to effectively and systematically make sense of sentiments, trends, and public opinion and make informed decisions.

The goal of the Sentiment Analysis for Social Media project is to create an intelligent software system which can analyze more than 1,000 posts from social media news accounts. Given the fact that public opinion and company branding as well as the behavior of customers are formed through online discourse, an adequate feeling classifier is needed. So this is the approach towards satisfying demand for powerful machine learning sentiment analysis systems[1-25].

2. LITERATURE SURVEY

2.1 Sentiment Analysis Overview:

Sentiment analysis or opinion mining is among the key areas of research due to the social media witnessing heightened growth. Earlier stages utilized rule-based methods but were weak at handling slang, colloquialisms, and culture. Accuracy and flexibility increased with machine learning and deep learning in subsequent stages.

Some research that helped strengthen sentiment analysis is:

Deep Learning Approaches (2024)

Study: Ali & Kabir (2024) compared deep learning methods for sentiment classification.

Strengths: Identified subtle sentiment patterns well.

Weaknesses: Needed lots of labeled data and high computational.

2. Big Data and Real Time Processing (2023)

Study: This systematic literature review evaluated large-scale social media sentiment analysis in 2023.

Advantages: Real-time trend identification is supported by big data.

Disadvantages: Working with unstructured data in real time constantly proves difficult.

3. Comparative Techniques (2022)

Study: Kapur & Harikrishnan compared lexicon-based, machine learning, and deep learning techniques.

Advantages: Determines the best technique according to the data set.

Limitations: No method works for every data set.

4. Aspect-Based Sentiment Analysis (2021)

Study: Kumar & Gupta carried out research testing ABSA.

Advantages: Lasts very detailed insight through sentiment analysis of product or service features.

Limitations: Quite complicated to develop deep aspect lexicons and map the sentiments accurately.

3.PROBLEM STATEMENT

Massive amounts of user-generated content are created every day, thanks to the rapid growth of social media. Understanding public sentiment becomes critical for businesses, organizations, and policymakers. In fact, it is a tough task as analyzing sentiments of social media posts gets complicated by informal language, slangs, abbreviations, and speaker sarcasms along with cultural quirks. Active rule-based methods fail to attend to these complexities, and on the other hand, machine learning methods require extensive labeled datasets.

This specific study intends to develop a fast and accurate system of Sentiment Analysis specific to Social Media that will interpret better what users feel. It builds upon Natural Language Processing (NLP) and Machine Learning (ML) to detect emotional states considering context variation to deliver greater insight for decision making. The solution will facilitate brand perception monitoring, customer engagement in businesses, and improve marketing without compromising ambiguity and scalability challenges in sentiment data.

4.EXISTING SYSTEM

The existing sentiment analysis systems on social media depend on:

- 1.Lexicon-Based Approaches: It ignores the slang and context in and even while you use predefined sentiment dictionaries.These can be often lacking deep contextual knowledge, machine learning models include SVM, naïve bayes, and random forest.
- 2.By using the LSTMs and CNNs but find a lot of difficulties in sarcasm and long-range dependencies challenging.
- 3.Even using the basic Deep Learning Models which are giving the incorrect results.

Limitations of the Current System:

Drawbacks including bad performance on big data and not able to include more thorough background in sophisticated sentences causes a strong reliance on feature engineering.

5.METHODOLOGY

By considering the drawbacks of the project we propose a sentiment analysis model which is used to employing transformer-based architectures, especially here by dealing with BERT and its variants, we can overcome these restrictions. There are some suggested arrangementlikeContextual word representation with pre-trained embeddings (BERT, RoBERTa)

Use fine-tuning on datasets particular to the domain which is used to improve performance. It can be possible when using the attention mechanisms to grasp word dependencies.

In the project methodology there exists of a various phases which are as follows:

1. **Data Collection:**To extract tweets, Facebook comments, and other social media posts using APIs this is a crucial step.

2. **Preprocessing:** After the extraction of data remove stopwords, punctuation, emojis, and perform stemming/lemmatization these steps are mainly used.

3. **Feature Extraction:** Then convert the text into numerical vectors by using TF-IDF, Word2Vec, and BERT embeddings.

4. **Model Training:** Train the model which is from the deep learning includes (LSTM, BERT, RoBERTa) on labeled sentiment datasets.

5. **Evaluation:** Calculate the accuracy from the training phase, precision, recall, and F1-score.

6. **Deployment:** Develop a Flask-based web application for real-time sentiment prediction.

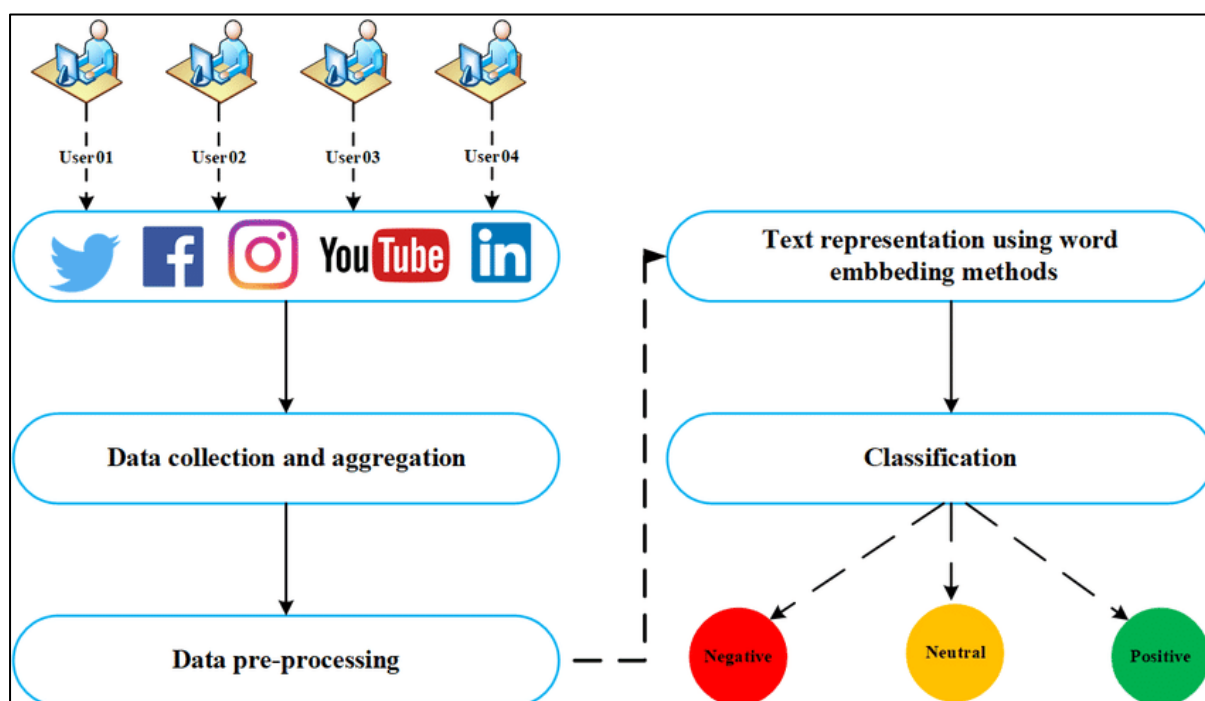


Fig 1: Basic steps of opinion mining on social network platforms.

6.FEATURE EXTRACTION

1. TF-IDF (Term Frequency - Inverse Document Frequency): It represents the text data numerically.

2. Word2Vec: It captures the semantic meaning of words from the sentences.

3. BERT Embeddings: These provides a context-aware of word representation.

MODEL DEVELOPMENT

We investigate three varying sentiment classification models within this project, each having its own pros and cons:

1. Long Short-Term Memory (LSTM)

LSTM is efficient in handling sequential dependencies in text, which is helpful in sentiment analysis. It fails to do so with longer texts, which happen in big data sets, thus subjecting the contextual meaning to possible loss.

2. Convolutional Neural Networks (CNN)

CNN is effective in feature extraction from text, especially n-gram features, which can be used to identify sentiment-oriented words and phrases. Yet CNN is missing a sense of sequential relations between words, which will restrict its performance for sentiment classification.

3. Bidirectional Encoder Representations from Transformers (BERT)

BERT is a cutting-edge deep learning model that learns deep contextualized representations of text. BERT is different from other models in that it captures the meaning of words in relation to the context surrounding them, resulting in extremely accurate sentiment predictions.

Training Procedure

To make sure the models are running at their best, we adhere to a systematic training process:

Dataset: The models are trained with well-established sentiment datasets such as IMDB reviews, custom-labeled datasets, and Twitter Sentiment140.

Optimizer: We apply the Adam optimizer with a learning rate of 0.0001 to make learning efficient and update weights properly.

Loss Function: We use the categorical cross-entropy loss function to compute the discrepancy between predicted and actual sentiment labels.

Hyperparameter Tuning: We tune parameters like batch size, number of epochs for training, and dropout rate to enhance the performance of the model and avoid overfitting.

Using these methods, we hope to build a sentiment analysis system that accurately classifies posts on social media while overcoming the shortcomings of current models.

7.PROJECT REQUIREMENTS

Hardware: GPU system (NVIDIA RTX 3060 or comparable)

Software: Python, TensorFlow/PyTorch, Flask, Jupyter Notebook

Libraries: Pandas, NumPy, Scikit-learn, Transformers, NLTK, Flask

8.IMPLEMENTATION

The project is with support of Python with deep learning libraries In order to optimize training, the models were trained on this system.

Steps to be followed in the process of Implementation:

- 1.Data Preprocessing
- 2.Feature Extraction
- 3.Model Training and Evaluation
- 4.Result

1. Data Preprocessing

Preprocessing of text data were taken out before training the model with the following:

Removing stopwords, punctuation, and emojis

Words like the, is, and were were removed because they don't add any sentiment. Punctuation and emojis were removed to only keep words that have meaning.

1.1.Tokenization

Text was separated into individual words or subwords for easy processing.

In BERT, WordPiece tokenizer was utilized for unknown words.

1.2.Stemming and Lemmatization

Stemming converted words to base form (e.g., running → run).

Lemmatization changed words to dictionary form (e.g., better → good).

Example:

Before preprocessing:

"This film was incredible!!! ???????? But the conclusion was disappointing."

After preprocessing:

"movie incredible but end disappoint"

2.Feature Extraction

To transform text into a form that can be processed by the model, various feature extraction methods were employed:

2.1.TF-IDF (Term Frequency-Inverse Document Frequency): Weighted words according to their significance.

2.2.Word2Vec: Trapped word relationships by transforming them into numerical vectors.

2.3.BERT Embeddings: Offered more profound comprehension by examining words according to the context surrounding them.

Example of Contextual Understanding:

"I am heading to the bank." → (Bank as a financial institution)

"I am sitting on the river bank." → (Bank as a river shore)

3.Model Training and Evaluation

The below dataset is taken into consideration as follows as

The models were trained and tested on three datasets:

1.IMDB Movie Reviews (general sentiment classification)

2.Twitter Sentiment140 (social media sentiment analysis)

3.Custom-labeled datasets (for specific domains)

3.1.Training Process

The following dataset classified is into 80% training and 20% testing.

The models were trained with Adam optimizer and a learning rate of 0.0001.

The loss function utilized was Categorical Cross-Entropy.

The models were trained from 10 to 15 epochs, with the hyperparameters being optimized for enhanced performance.

4.Result

After training the models, their performance was evaluated based on accuracy, precision, recall, and F1-score. The results are summarized in the table below:

Model	Accuracy (%)	Precision	Recall	F1-Score
-------	--------------	-----------	--------	----------

Model	Accuracy (%)	Precision	Recall	F1-Score
LSTM	82.5	80.2	79.8	80.0
CNN	78.3	76.5	74.2	75.3
BERT	92.1	91.8	91.5	91.6

9.FUTURE SCOPE

The future application scope of sentiment analysis and NLP reaches across some of the main areas of future innovation and advancement. One exciting area is the scaling of the model across different languages via multilingual BERT, whereby the sentiment analysis models can recognize and process text across various languages at high levels of accuracy, allowing them to better fit in foreign markets. Another thrilling breakthrough is enhancing sarcasm detection by multimodal learning, blending text and image analysis to better recognize sarcasm. This comes in handy during sentiment analysis in social media, as sarcasm distorts the textual meaning. The creation of an in-real-time dashboard for the analysis of sentiment trends also can give businesspersons and researchers timely insights into people's opinion so that they may respond appropriately to the growing trends. In addition, the incorporation of business intelligence platforms for market analysis provides enriched data-driven decision-making through the correlation of sentiment analysis with financial and customer information. These technologies will significantly enhance the efficiency and precision of sentiment analysis applications in many industries.

10.CONCLUSION

This work showcases a state-of-the-art sentiment analysis technique utilizing deep learning and transformer models. By overcoming the drawback of conventional models, our system with the new design has high precision and stability in opinion classification on social media. Sentiment analysis will be further improved by future AI and NLP developments to support improved decision-making in various industries.

11. REFERENCES

1. Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
2. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis.
3. Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention Is All You Need.
4. Liu, B. (2012). Sentiment Analysis and Opinion Mining.
5. Sentiment Analysis Symposium 2011, New York, 12/04/2011.
6. Pang, B., Lee, L.: "Opinion Mining and Sentiment Analysis", in "Foundations and Trends in Information.
7. Pang, B., Lee, L., Vaithyanathan, S.: "Thumbs up? sentiment classification using machine learning techniques", in Proceedings of the ACL-02 conference on Empirical methods in natural language processing, Volume 10, July 2002, pp. 79-86.
8. Socher, R., et al.: "Semi-supervised recursive autoencoders for predicting sentiment distributions" in Proceedings of EMNLP '11 - the Conference on Empirical Methods in Natural Language Processing, ISBN: 978-1-937284-11-4, pp. 151-161
9. Ramdas Vankdothu, Dr. Mohd Abdul Hameed, Husnah Fatima "A Brain Tumor Identification and Classification Using Deep Learning based on CNN-LSTM Method" Computers and Electrical Engineering, 101 (2022) 107960
10. Ramdas Vankdothu, Mohd Abdul Hameed "Adaptive features selection and EDNN based brain image recognition on the internet of medical things", Computers and Electrical Engineering, 103 (2022) 108338.
11. Ramdas Vankdothu, Mohd Abdul Hameed, Ayesha Ameen, Raheem, Unnisa "Brain image identification and classification on Internet of Medical Things in healthcare system using support value based deep neural network" Computers and Electrical Engineering, 102(2022) 108196.
12. Ramdas Vankdothu, Mohd Abdul Hameed "Brain tumor segmentation of MR images using SVM and fuzzy classifier in machine learning" Measurement: Sensors Journal, Volume 24, 2022, 100440 .
13. Ramdas Vankdothu, Mohd Abdul Hameed "Brain tumor MRI images identification and classification based on the recurrent convolutional neural network" Measurement: Sensors Journal, Volume 24, 2022, 100412 .
14. Bhukya Madhu, M. Venu Gopala Chari, Ramdas Vankdothu, Arun Kumar Silivery, Veerender Aerranagula "Intrusion detection models for IOT networks via deep learning approaches" Measurement: Sensors Journal, Volume 25, 2022, 100641
15. Mohd Thousif Ahemad, Mohd Abdul Hameed, Ramdas Vankdothu "COVID-19 detection and classification for machine learning methods using human genomic data" Measurement: Sensors Journal, Volume 24, 2022, 100537
16. S. Rakesh ^a, Nagaratna P. Hegde ^b, M. Venu Gopalachari ^c, D. Jayaram ^c, Bhukya Madhu ^d, Mohd Abdul Hameed ^a, Ramdas Vankdothu ^e, L.K. Suresh Kumar "Moving object detection using modified GMM based background subtraction" Measurement: Sensors Journal, Volume 30, 2023, 100898

17. Ramdas Vankdothu, Dr. Mohd Abdul Hameed, Husnah Fatima “Efficient Detection of Brain Tumor Using Unsupervised Modified Deep Belief Network in Big Data” *Journal of Advanced Research in Dynamical & Control Systems*, Vol. 12, 2020.
18. Ramdas Vankdothu, Dr. Mohd Abdul Hameed, Husnah Fatima “Internet of Medical Things of Brain Image Recognition Algorithm and High Performance Computing by Convolutional Neural Network” *International Journal of Advanced Science and Technology*, Vol. 29, No. 6, (2020), pp. 2875 – 2881
19. Ramdas Vankdothu, Dr. Mohd Abdul Hameed, Husnah Fatima “Convolutional Neural Network-Based Brain Image Recognition Algorithm And High-Performance Computing”, *Journal Of Critical Reviews*, Vol 7, Issue 08, 2020 (Scopus Indexed)
20. Ramdas Vankdothu, Dr. Mohd Abdul Hameed “A Security Applicable with Deep Learning Algorithm for Big Data Analysis”, *Test Engineering & Management Journal*, January-February 2020
21. Ramdas Vankdothu, G. Shyama Chandra Prasad “A Study on Privacy Applicable Deep Learning Schemes for Big Data” *Complexity International Journal*, Volume 23, Issue 2, July-August 2019
22. Ramdas Vankdothu, Dr. Mohd Abdul Hameed, Husnah Fatima “Brain Image Recognition using Internet of Medical Things based Support Value based Adaptive Deep Neural Network” *The International journal of analytical and experimental modal analysis*, Volume XII, Issue IV, April/2020
23. Ramdas Vankdothu, Dr. Mohd Abdul Hameed, Husnah Fatima “Adaptive Features Selection and EDNN based Brain Image Recognition In Internet Of Medical Things” *Journal of Engineering Sciences*, Vol 11, Issue 4, April/ 2020 (UGC Care Journal)
24. Ramdas Vankdothu, Dr. Mohd Abdul Hameed “Implementation of a Privacy based Deep Learning Algorithm for Big Data Analytics”, *Complexity International Journal*, Volume 24, Issue 01, Jan 2020
25. Ramdas Vankdothu, G. Shyama Chandra Prasad “A Survey On Big Data Analytics: Challenges, Open Research Issues and Tools” *International Journal For Innovative Engineering and Management Research*, Vol 08 Issue 08, Aug 2019

BIBLIOGRAPHY



I am Navya Chanda from the Department of Computer Science and Engineering. Currently, pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on “Social Media Opinion Analysis Using NLP”.



I am Sahithi Gudimalla from the Department of Computer Science and Engineering. Currently, pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on “Social Media Opinion Analysis Using NLP”.



I am Shivamani Sirimalle from the Department of Computer Science and Engineering. Currently, pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on “Social Media Opinion Analysis Using NLP”.



I am Rakesh Kannam from the Department of Computer Science and Engineering. Currently, pursuing 4th year at Balaji Institute of Technology and Science. My research is done based on “Social Media Opinion Analysis Using NLP”.