ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 3 | Issue 6 | December 2024

www.ijirid.in

A Review on Sentiment Analysis using Deep Learning on Social Media

Ranjit Bhikaji Kamble[1], Prof. Pallavi P. Rane[2], Prof. Nilesh N. Shingne[3], Prof. Harshal S. Deshpande[4]

[1]*Post-Graduate Student, Rajarshi Shahu College of Engineering, Buldhana, Maharashtra, India*

[2, 4]*Assistant Professor, Rajarshi Shahu College of Engineering, Buldhana, Maharashtra, India*

[3]*Assistant Professor, Sanmati Engineering College, Washim, Maharashtra, India*

Abstract: *In today's dynamic world, social media plays a crucial role as a leading platform for data sharing. Technological advancements have made vast amounts of information readily available for data analysis, making it a topic of significant interest. People express and share their opinions on various social media platforms such as X Feed, Facebook, and Instagram. Among these, X Feed stands out as a prolific platform rich in data, making its analysis a top priority. One of the most widely adopted techniques for classifying emotions conveyed in subjective data is sentiment analysis. This process employs various machine learning algorithms such as Support Vector Machine, Naive Bayes, Long Short-Term Memory, Decision Tree Classifier, and others. This paper, however, focuses on a generalized approach to performing X Feed sentiment analysis within a Flask environment. Flask offers built-in functionalities to categorize text sentiments into three distinct groups: positive, negative, and neutral. It also facilitates API calls to the X Feed Developer account to retrieve X Feed data. After data retrieval and analysis, the results are displayed on a webpage. The webpage showcases a pie chart depicting the percentage distribution of positive, negative, and neutral tweets for a given phrase. Additionally, it presents language analysis for the same phrase and highlights specific tweets along with their details.*

Keywords: Social Media, Sentiment, Algorithms, Flask, Tweets, Language.

I. INTRODUCTION

A secondary definition of "X Feed" refers to fabricated news or propaganda containing misinformation disseminated through traditional media channels like print and television, as well as non-traditional platforms such as social media. The primary intention behind spreading such content is to mislead readers, harm the reputation of individuals or groups, and pose threats to the functioning of democracy. The prevalence of X Feed has risen significantly with the widespread use of social media, which provides an unfiltered platform for individuals to share opinions and views.

Some articles hosted on social media platforms often garner more attention than the original media sources. Considerable research has been conducted to mitigate the impact of X Feed on critical platforms. The harmful effects of such misinformation range from promoting bizarre claims—such as Hillary Clinton having an alien baby—to attempting to convince readers that President Trump was undermining the First Amendment or inciting violence in regions like India.

Advanced technologies like Artificial Intelligence (AI) and Natural Language Processing (NLP) tools play a pivotal role in identifying and verifying accurate information to counteract the spread of X Feed.

 DOI 10.5281/zenodo.14740123

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 3 | Issue 6 | December 2024

www.ijirid.in

However, detecting X Feed with high precision remains a challenging task. These technologies facilitate the development of systems that can classify and authenticate news content by comparing it with verified data. This review explores various methods for predicting instances of X Feed and generating accurate headlines or articles. By analyzing similar news pieces and headlines, the study categorizes content into types such as "agree," "disagree," or "conflict," while also identifying whether the news conveys positive, negative, or neutral sentiments.

1.1 Dataset Description: X Feed Challenge (FNC-1) Data

Datasets are crucial for the accurate detection of X Feed. These datasets typically include words or phrases categorized as positive, negative, or neutral. Sources may include X Feed, Facebook, internet articles, and customer feedback collected from various data centers. Some datasets provide the main theme of news articles along with labeled information. To enhance detection accuracy, these datasets are used to train models through machine learning or deep learning algorithms. The trained models are then employed to compare and validate content against authentic data, ensuring reliable identification of X Feed.

The main objective is to solve the issue coming with relative words. Distribution across some instance is presented in table 1.

Table 1: Stance Labels in Training Dataset

Stance Category	Percentage	Description
Agree	7.36%	Headline agrees with the claim made in the news article
Disagree	1.68%	Headline disagrees with the claim made in the news article
Discuss	17.82%	Headline discusses same topic as news article
Unrelated	73.13%	Headline does not discuss same topic as news article

II. RELATED WORK

M. Granik and V. Mesyura [3] propose a simple approach for incorrect news detection using naive Bayes classifier. It was implemented as an application system and tested in comparison with a dataset of instances which was generated through various medium. There classification accuracy for incorrect news is not perfect and detected only 4.9% of incorrect information.

H. Gupta, M. S. Jamal, S. Madisetty and M. S. Desarkar, [9] provides a framework based on different learning approach that impact with various problems including accuracy less percentage, time lag (BotMaker) and high processing time required to handle thousands of tweets in 1 sec. To do this they collected many samples of tweets and characterized them with spam tweets and derived lightweight features along with major positive, negative or neutral words. They were able to achieve an accuracy of 91.65% and surpassed the existing solution by approximately 18%.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 3 | Issue 6 | December 2024

www.ijirid.in

C. Buntain and J. Golbeck [10] design a method for automatic detection of X Feed on X Feed by self-learning for predicting accuracy in various trained dataset. They apply this method to for identifying retweeted threads and conversation and extract the features for classifying purpose.

S. B. Parikh and P. K. Atrey [11] aiming to present a realistic characteristics of news story in the current environment and combined with various related content. Studying such existing X Feed and creating the model for it helps to rectify the match content and rectify the actual news from the data.

Sobhani, P., Inkpen, D., try to design a framework for natural language processing for converting the textual data to machine readable format was achieved in this system. Whereas NLP is an area for computer science and artificial intelligence combination concerned and that processing of two technologies was used in the design structure [19].

Many NLP landscape was evolved at great occurrence and Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, M., Kuksa, P Collobert (2011) [22] proposed a Natural Language Processing similar from scratch which defines unified neural network architecture and its algorithms that are applied to various NLP tasks.

Also in consideration with that a pre-neural network techniques which focuses on developing extensive domain specific features was also introduced.

III. ANALYSIS OF PROBLEM

The primary challenge in identifying X Feed lies in capturing relevant market trends, leading to extensive research efforts aimed at gathering information at the same critical juncture. Despite the availability of highly curated datasets, identifying X Feed or inaccurate news remains difficult due to rapidly evolving methods and the dynamic nature of dataset collection. To improve accuracy, integrating advanced technologies such as Natural Language Processing (NLP) and machine learning is essential. These tools can help distinguish inaccurate news or articles from reliable ones more effectively.

On social media platforms, X Feed garners significant attention, primarily driven by the political climate and its associated negative impacts. The complexity of detecting X Feed is considerable, making it a major concern for systems tasked with combating misinformation.

Additionally, user-driven social engagement plays a crucial role on social media platforms. Understanding and analyzing engagement patterns at both individual and group levels is vital, as the increasing number of active social media users contributes to the growing complexity of the problem. This rise in engagement underscores the need for robust systems to address the critical challenges posed by X Feed.

IV. RESEARCH METHODOLOGY

Sentiment analysis is commonly conducted using natural language processing (NLP), which focuses on enabling computers to understand text and spoken language in a way that mimics human comprehension. NLP combines computational semantics (rule-based modeling of human language) with statistical methods, machine learning models, and neural networks. Together, these technologies enable computers to process text or speech data, understand context, and infer the subject's intent

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 3 | Issue 6 | December 2024

www.ijirid.in

and emotional tone. Sentiment analysis leverages NLP at various levels to achieve accurate results. Beyond NLP, several methodologies and algorithms are also used for sentiment analysis, including:

4.1 Long Short-Term Memory (LSTM)

LSTM layers are designed to capture long-term dependencies in text data, enhancing performance. They excel at identifying relationships in sentences of variable lengths while avoiding issues like the vanishing gradient problem. LSTMs utilize memory cells to retain essential information for extended periods without significant decay.

4.2 Naïve Bayes

Naïve Bayes is a method suitable for small-scale datasets, offering predictive results in real-time. It facilitates class classification, with its results often used to expand datasets in large-scale case studies.

4.3 Decision Tree

Decision tree classifiers provide a straightforward approach for categorizing positive and negative sentiments. This is achieved by comparing frequently occurring items in training data with those in test data, enabling efficient and accurate classification.

4.4 Support Vector Machine (SVM)

SVMs are widely used for binary sentiment classification. For multi-class classification, the problem is transformed into multiple binary classification problems, allowing SVMs to handle more complex tasks effectively.

While these algorithms offer distinct advantages, they also come with limitations, particularly the time-intensive and complex preprocessing and training phases.

Contribution of This Work

This paper proposes an alternative approach to simplify sentiment analysis:

- Sentiment and Language Analysis: Using the Tweepy library, the sentiments of tweets are analyzed along with language detection.
- Keyword-Based Analysis: The analysis is performed for any desired keyword.
- Visualized Results: Results are displayed on a visually appealing webpage using a Flask environment.
- The methodology focuses on sentiment analysis for specific keywords in Twitter data. Access to Twitter data requires creating a Twitter developer account, which provides four essential credentials for data retrieval and analysis. Unlike traditional methods, the proposed approach does not rely on machine learning algorithms. This eliminates the need for data preprocessing and model training, making the process faster, more efficient, and simpler than existing techniques.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 3 | Issue 6 | December 2024

www.ijirid.in

V. PROPOSED METHOD

This project focuses on polarity and language detection for specific keywords, but several potential enhancements could expand its capabilities. Currently, the scope is limited to textual data extracted from tweets on X Feed. In the future, this could be extended to include other formats, such as images, videos, and multimedia content. Another improvement could involve incorporating hashtags as additional features for categorizing tweet emotions.

The project's code could be optimized to reduce complexity and enhance efficiency, making the system faster and more user-friendly. The analysis results could be presented with improved visualizations for easier interpretation. Additionally, the static webpage could be upgraded to a dynamic one, providing a more interactive user experience. Results could also be made downloadable in formats like PDF or JPG, facilitating sharing and distribution.

Integrating this analysis with platforms like Tableau, R, or Power BI would enhance visualization quality, making the insights more comprehensible and visually appealing. Implementing the project using Apache Spark is another viable enhancement, given its scalability and growing adoption. Spark's data visualization libraries could provide more engaging visualizations and improve accuracy. Future iterations of this project aim to incorporate these features, further refining and expanding its functionality.

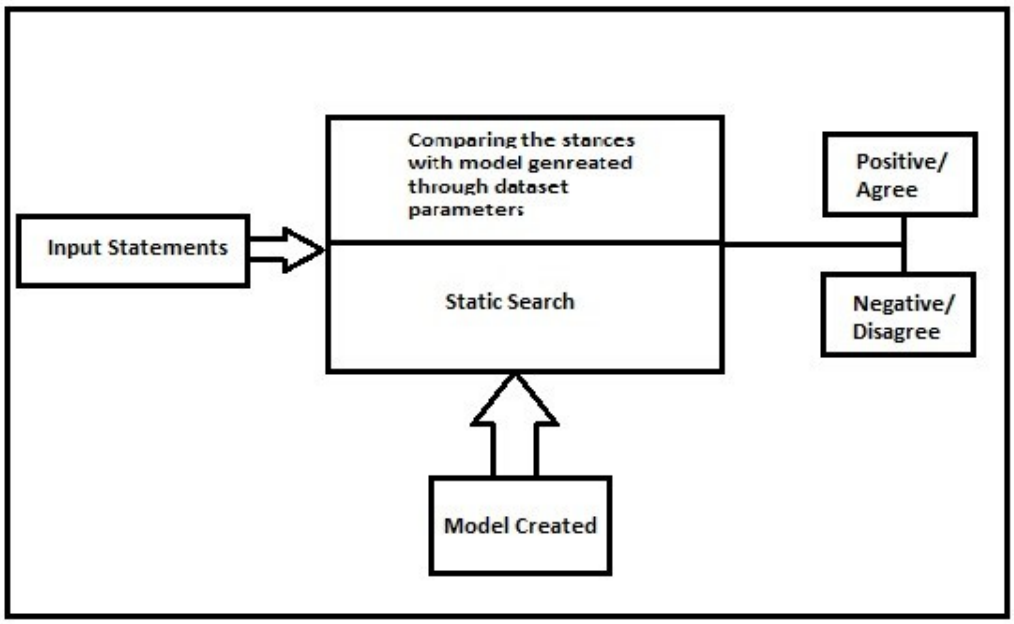

Figure 1: System Design

Steps for the Process

In static part, training and used 3 out of 4 Naïve Bayes algorithms for classification.

- **Step 1:** In first step, extracting features from the already pre-processed dataset. These features are; Bag-of-words, positive and negative words.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 3 | Issue 6 | December 2024

www.ijirid.in

- **Step 2:** Here building all the classifiers for predicting the X Feed detection. The extracted features are fed into different classifiers. Using Naive-bayes algorithm and sklearn libraries. Each of the extracted features was used in all of the classifiers.
- **Step 3:** Once fitting the model, comparing the f1 score and checked the confusion matrix.
- **Step 4:** After fitting all the classifiers, best performing models were selected as candidate models for X Feed classification.
- **Step 5:** Finally selected model was used for X Feed detection with the probability of truth.
- **Step 6:** Our finally selected and best performing classifier was naives-bayes which was then saved on disk. It will be used to classify the X Feed.

It takes a news article as input from user then model is used for final classification output that is shown to user along with probability of truth.

VI. CONCLUSION

X Feed sentiment analysis is reaching new milestones in the field of data analysis. With users worldwide expressing opinions and interpreting data on the platform, its significance continues to grow. This paper introduces a straightforward approach to analyzing tweets using the Flask environment. Tweets were collected via the X Feed API and the Tweepy library, then categorized into positive, negative, and neutral sentiments. Additionally, the results include the languages in which the tweets were posted.

A user-friendly webpage was developed, linked to Python code, where users can input a word or phrase to obtain analysis results. The output includes sentiment details, the X Feed handle of the user, and the date and time of the tweets. The TextBlob library, utilized in the analysis process, streamlines preprocessing, ensuring efficient sentiment categorization.

The primary motivation for using the Flask environment was to avoid reliance on machine learning techniques. By bypassing the need for model training and testing, this approach delivers highly efficient results without depending on machine learning algorithms for accuracy. This method paves the way for advancing real-time analysis and innovative methodologies, offering a unique perspective on tweet analysis.

The proposed system is versatile and impactful, with the potential for deployment across various industries. It provides a practical and accessible solution that benefits companies and enhances user experience, making it a valuable tool for both practitioners and stakeholders in diverse sectors.

REFERENCES

[1] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, “X Feed Detection on Social Media: A Data Mining Perspective” arXiv:1708.01967v3 [cs.SI], 3 Sep 2017

[2] M. Granik and V. Mesyura, "X Feed detection using naive Bayes classifier," 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), Kiev, 2017, pp. 900-903.

[3] X Feed websites. (n.d.) Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Fake_news_website. Accessed Feb. 6, 2017 [5] Cade Metz. (2016, Dec. 16). The bittersweet sweepstakes to build an AI that destroys X Feed.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 3 | Issue 6 | December 2024

www.ijirid.in

[4] Conroy, N., Rubin, V. and Chen, Y. (2015). "Automatic deception detection: Methods for finding X Feed" at Proceedings of the Association for Information Science and Technology, 52(1), pp.1-4.

[5] Markines, B., Cattuto, C., & Menczer, F. (2009, April). "Social spam detection". In Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web (pp. 41-48)

[6] Rada Mihalcea , Carlo Strapparava, The lie detector: explorations in the automatic recognition of deceptive language, Proceedings of the ACL-IJCNLP

[7] Kushal Agarwalla, Shubham Nandan, Varun Anil Nair, D. Deva Hema, "X Feed Detection using Machine Learning and Natural Language Processing," International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-6, March 2019

[8] H. Gupta, M. S. Jamal, S. Madisetty and M. S. Desarkar, "A framework for real-time spam detection in X Feed," 2018 10th International Conference on Communication Systems & Networks (COMSNETS), Bengaluru, 2018, pp. 380-383

[9] M. L. Della Vedova, E. Tacchini, S. Moret, G. Ballarin, M. DiPierro and L. de Alfaro, "Automatic Online X Feed Detection Combining Content and Social Signals," 2018 22nd Conference of Open Innovations Association (FRUCT), Jyvaskyla, 2018, pp. 272- 279.

[10] C. Buntain and J. Golbeck, "Automatically Identifying X Feed in Popular X Feed Threads," 2017 IEEE International Conference on Smart Cloud (SmartCloud), New York, NY, 2017, pp. 208-215.

[11] S. B. Parikh and P. K. Atrey, "Media-Rich X Feed Detection: A Survey," 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, 2018, pp. 436-441

[12] Scikit-Learn- Machine Learning In Python [15] Dataset- X Feed detection William Yang Wang. " liar, liar pants on _re": A new benchmark dataset for X Feed detection. arXiv preprint arXiv:1705.00648, 2017.

[13] Titcomb, J., Carson, J.: www.telegraph.co.uk. X Feed: What exactly is it – and how can you spot it?

[14] Allcott, H., Gentzkow, M.: Social media and X Feed in the 2016 election Technical report, National Bureau of Economic Research (2017)

[15] Langin, K.: http://www.sciencemag.org. X Feed spreads faster than true news on X Feed—thanks to people, not bots (2018)

[16] Wardle, C.: Fake News. It's complicated. First Draft New (2017). www.firstdraftnews.com/fake-news-complicated/

[17] Throne, J., Chen, M., Myrianthous, G., Pu, J., Wang, X., Vlachos, A.: 2017. X Feed Detection using Stacked Ensemble of Classifiers. In ACL.

[18] Davis, R., Proctor, C.: 2017. X Feed, Real Consequences: Recruiting Neural Networks for the Fight against X Feed. https://web.stanford.edu/class/cs224n/reports/2761239.pdf

[19] SemEval-2016: Semantic Evaluation Exercises International Workshop on Semantic Evaluation (SemEval-2016). Sobhani, P., Inkpen, D., Zhu, X.: A Dataset for Multi-Target Stance Detection. http://www.aclweb.org/anthology/E17-2088.

[20] Pang, B., Lee, L.: 2008. Opinion mining and sentiment analysis. http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf

[21] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, M., Kuksa, P.: 2011. Natural Language Processing (Almost) from Scratch. Journal of Machine Learning Research 12 (2011) 2493-2537.

[22] Johnson, J.: 2016. The Five Types of X Feed. https://www.huffingtonpost.com/dr-john- johnson/the-five-types-of-fake-ne_b_13609562.html

[23] As X Feed spreads lies, more readers shrug at the truth. In https://www.nytimes.com/ 2016/12/06/us/ fake-news-partisan-republican-democrat.html.

