IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125264

SENTIMENTAL ANALYSIS OF NEWS ARTICLES USING NAÏVE BAYES

Mrudul Khairkar[1], Dhwani Waghela[2], Manas Bhilare[3], Aman Shriyan[4], Chitralekha Dwivedi [5]

Student, Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pune, India[1,2,3,4]

Assistant Professor, Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pune, India[5]

Abstract: Sentiment analysis is one of the recent technologies under NLP (an application of Artificial intelligence and Machine Learning). It is used in many applications for recommendation and feedback analysis. In this paper, from defining sentiment analysis, to algorithms for sentiment analysis are discussed with practical results. The results declared in this paper are from the implantation of sentiment analysis on the news articles dataset using Naïve Bayes classifier. Additionally, the paper explores the various techniques employed in sentiment analysis, and delves into the challenges faced in accurately determining sentiment polarity. The experimental results demonstrate the effectiveness of the Naïve Bayes classifier in sentiment analysis, shedding light on its potential for enhancing decision-making processes in industries such as marketing, customer service, and public opinion analysis.

Keywords: Sentimental Analysis, News Articles, Target level Sentiment, Opinion Mining

I. INTRODUCTION

News analysis and calculations of news sentiment are now routinely used by the people in all business categories including buy-side and sell-side of the stock market compliance. In our paper, we have focused on news articles. The main errands identified for news opinion mining consists of extracting sentences from online published news articles that mention corporation news, and identifying positive and negative sentiments in that article and further summarizing the article polarity. The utilization of news sentiment analysis has become an integral part of decision-making processes across various business sectors, ranging from financial institutions to stock market compliance entities. Our research specifically targets news articles as the primary source of analysis. The key objectives of news opinion mining revolve around the extraction of relevant sentences from online published news articles pertaining to corporate news, discerning positive and negative sentiments expressed within those articles, and ultimately generating a comprehensive summary of the article's overall polarity. By employing advanced NLP techniques, including text mining, sentiment classification, and sentiment summarization, our study aims to provide a comprehensive understanding of the sentiment landscape within news articles. This research contributes to the existing body of knowledge by offering insights into the effectiveness of sentiment analysis algorithms in discerning nuanced sentiments expressed in the news, ultimately empowering businesses with valuable information for strategic decision-making and risk assessment. Overall, this paper serves as a significant step towards harnessing the power of news sentiment analysis for informed business decisions.

II. LITERATURE SURVEY

Sentiment analysis on news articles has gained significant recognition in recent years due to the immense data generated by news agencies and the need to analyze them efficiently. In this literature review, we will discuss some of the studies that have been conducted in the field of sentiment analysis on news articles.

"The emergence of social media data and sentiment analysis in election prediction " by Chauhan, P., Sharma, N. & Sikka, G. [1]: In this study, the authors assess the power of various volumetric, sentiment, and social network approaches to predict pivotal decisions from social media platforms. The paper highlighted some state-of-the-art studies related to sentiment analysis using deep learning and word embedding methods. This paper also gave some future directions in the related field. "Multimodal Sentiment Analysis: A Survey of Methods, Trends and Challenges " by Ringki Das and Thoudam Doren Singh [2]: In this study, the authors proposed a Multimodal sentiment analysis for sentiment analysis is shifting from unimodality to multimodality. "Deep Learning Approach for Sentiment Analysis of COVID-19 Reviews " by Singh, C.; Imam, T.; Wibowo, S.; Grandhi, S. [3]: In this review paper, the authors provided a deep learning approach for sentiment analysis of Twitter data related to COVID-19 reviews. The algorithm was based on an LSTM-RNN-based network and enhanced featured weighting by attention layers."Sentiment Analysis for Fake News Detection" by Alonso,

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125264

M.A.; Vilares, D.; Gómez-Rodríguez, C.; Vilares, J [4]: In this paper, the authors studied the different uses of sentiment analysis in the detection of fake news, alongwith relevant elements and shortcomings. "ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis" by Basiri, Mohammad Ehsan [5]: In this study, the authors proposed that by utilizing two independent bidirectional LSTM and GRU layers, ABCDM will extract both past and future contexts by considering temporal information flow in both directions. “A comprehensive survey on sentiment analysis: Approaches, challenges and trends, Knowledge-Based Systems”, by Marouane Birjali (et al) [6]: The paper provides a comprehensive study on sentiment analysis, exploring its applications, generic process, and various approaches used in the field. It also discusses the challenges associated with sentiment analysis and offers insights for future research directions, aiming to provide researchers with a global survey of sentiment analysis and related fields. “Sentiment Analysis of Online Course Evaluation Based on a New Ensemble Deep Learning Mode: Evidence from Chinese” by Xiaomin Pu (et al) [7]: The paper presents a new ensemble model for sentiment analysis in online course comments, utilizing advanced techniques such as Word2Vec, Glove, bidirectional long- and short-time networks, and convolutional neural networks, achieving high accuracy and stability in sentiment recognition compared to other models.

To summarize, sentiment analysis on news articles is a predominant area of research that has received engrossment and awareness in recent years. Various techniques such as machine learning, deep learning, rule-based, lexicon-based, hybrid, and ensemble approaches have been proposed and studied in the literature. Machine learning-based approaches are generally found to be the most effective for sentiment analysis on news articles, and ensemble and hybrid approaches can further improve accuracy.

III. AIM & OBJECTIVES

The motivation behind this project is:

1. To find optimal polarity of the news articles
2. Predict in what way it affects the taxpayer and companies alike.
3. Perform a comparative analysis of the algorithms
4. Evaluation of results of comparison
5. Identify research gaps from literature survey

We aim to implement the sentiment analysis for recommendation / identification of news article into positive, negative, neutral, from the user’s perspective. Lastly, we would like to determine the exact outlook of the mass towards the subject or the news.

IV. PROPOSED SYSTEM

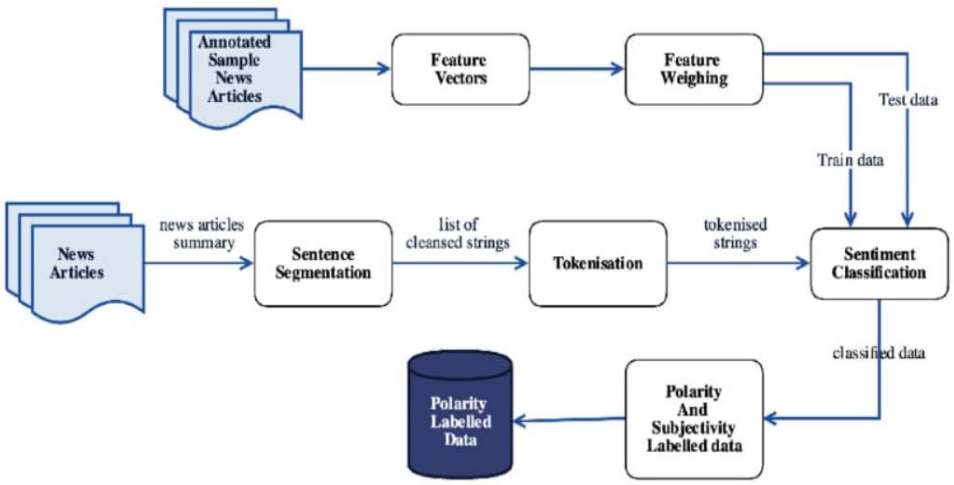

Fig. 2 System Architecture (SpringerLink, Sentiment Analysis of Twitter Data)

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023
DOI: 10.17148/IJARCCE.2023.125264

The system architecture for sentimental analysis of news articles typically involves several components, each with a specific role in the analysis process. These components can be divided into three main categories: data collection and preprocessing, sentiment analysis, and visualization and reporting.

A. Data Collection and Preprocessing

The first step in analyzing news articles for sentiment is to collect and preprocess the data. This involves gathering articles from various sources such as news websites, social media platforms, and blogs. The collected data is then processed to extract relevant information such as the article title, author, publication date, and text content. Text preprocessing techniques are then applied to clean the data and prepare it for sentiment analysis. This can include removing stop words (common words that do not convey much meaning), stemming (reducing words to their root form), and removing special characters and punctuation.

B. Sentiment Analysis:

Once the data is accumulated and preprocessed, next, we perform sentiment analysis on the text content of the articles. This is done using Naïve Bayes algorithm.

C. Visualization and Reporting:

The final step in the sentiment analysis of news articles is to visualize and report the results. This can involve creating charts and graphs to display the distribution of sentiment across different articles or sources. It can also involve generating reports that summarize the sentiment of the news articles and provide insights into trends and patterns.

V. METHODOLOGY

A. Dependencies

We first import the necessary modules from NLTK. NLTK is a widely used open-source Python library for NLP. It provides a wide range of tools and resources for tasks such as tokenization, stemming, lemmatization, POS tagging, and sentiment analysis. We also import the requests module to retrieve the HTML content of the news article.

B. Analyzer Used

The SentimentIntensityAnalyzer is a tool used for sentiment analysis in natural language processing. It is a part of the Natural Language Toolkit (NLTK) library in Python. The analyzer is designed to determine the sentiment polarity of a given text, indicating whether the text expresses a positive, negative, or neutral sentiment. The SentimentIntensityAnalyzer uses a pre-trained model to assign sentiment scores to individual words and phrases within the text. These scores are then combined to calculate an overall sentiment score for the entire text. The sentiment scores range from -1 (negative sentiment) to +1 (positive sentiment), with 0 indicating a neutral sentiment.

C. Algorithm

Naïve Bayes Algorithm:

The Naïve Bayes algorithm is a popular machine learning algorithm used for classification tasks. It is based on the Bayes' theorem, which calculates the probability of a certain event happening given prior knowledge. In Naïve Bayes, the " naïve " assumption is made that all features in the data are independent of each other, meaning that the presence or absence of one feature does not affect the presence or absence of another.

The algorithm works by estimating the probability of a data point belonging to each class based on the observed feature values. It calculates the conditional probability of each class given the feature values and then assigns the data point to the class with the highest probability.

This is done by multiplying the individual probabilities of each feature given the class. Naïve Bayes is efficient and scalable, making it particularly useful for large-scale data sets. It is commonly used in text classification tasks, such as spam detection or sentiment analysis, where each word or term is treated as a feature.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 5, May 2023

DOI: 10.17148/IJARCCE.2023.125264

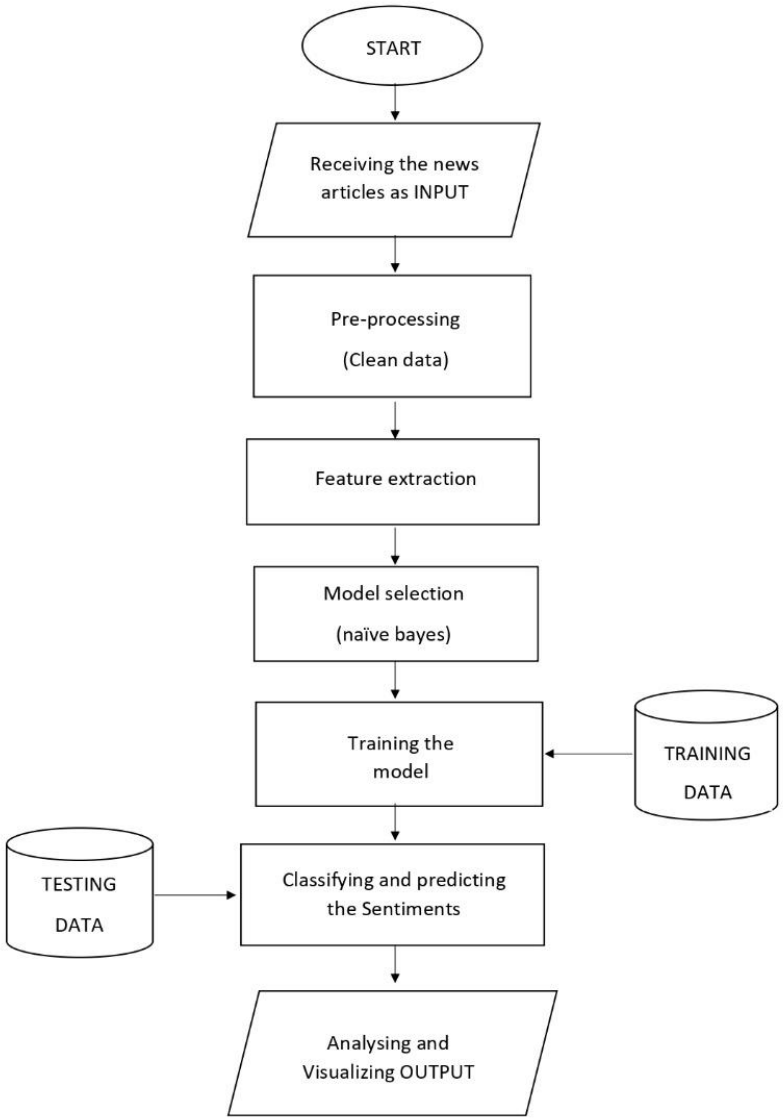

Fig. 1 Flowchart of Sentimental Analysis

D. Data Collection

Input of news articles taken from user. Test data was run on news articles database. Many news organizations provide APIs that allow developers to access their news articles. By using these APIs, you can retrieve news articles along with their metadata, such as headlines, publication date, and content.

E. Output

Polarity is the output that lies between [-1,1], where -1 refers to negative sentiment and +1 refers to positive sentiment. Subjectivity is the output that lies within [0,1] and refers to an individual's opinions and judgements. Another common approach is to assign categorical labels to represent the sentiment polarity.

DOI: 10.17148/IJARCCE.2023.125264

VI. CONCLUSION

In summary, the main aim of this study is to explore and complete classification methods in sentiment analysis, considering their advantages and disadvantages. The research begins by discussing various levels of sentiment analysis and providing an overview of necessary procedures like data collection and feature selection. The study then classifies and compares different methods used in sentiment categorization systems, highlighting their strengths and weaknesses. Supervised machine learning methods, particularly the NB and SVM algorithms, are widely utilized due to their simplicity and high accuracy, often serving as benchmarks for new approaches. The study also examines common application areas and explores the significance and consequences of sentiment analysis challenges in sentiment evaluation. By comparing the structure of sentiment reviews with the difficulties faced in sentiment analysis, the research reveals the importance of domain dependence in identifying sentiment issues. Moving forward, future work will involve expanding the comparison scope with additional findings. The challenges encountered in this research demonstrate that sentiment analysis remains a relatively unexplored field of study.

ACKNOWLEDGMENT

We would like to show our heartfelt indebtedness to our guide **Mrs. Chitralekha Dwivedi** and also to our project coordinator Mrs. Sunita Patil, whose valuable guidance has helped us to make this project a reality. And we are grateful to **Dr. Vinod Kimbahune**, HOD of Computer Department and **Dr. Lalit Kumar Wadhwa**, Principal of **Dr. D. Y. Patil** Institute of Technology, Pimpri, whose constant encouragement and motivation inspired us to do our best. Last but not the least, we sincerely, thank our colleagues, the staff and all others who directly or indirectly helped us and made numerous suggestions which have surely improved the quality of our work.

REFERENCES

[1] Chauhan, P., Sharma, N. & Sikka, G. The emergence of social media data and sentiment analysis in election prediction. J Ambient Intel Human Computer,2021.

[2] Ringki Das and Thoudam Doren Singh. 2023. Multimodal Sentiment Analysis: A Survey of Methods, Trends and Challenges. ACM Computer Survey, 2023

[3] Singh, C.; Imam, T.; Wibowo, S.; Grandhi, S. A Deep Learning Approach for Sentiment Analysis of COVID-19 Reviews. Appl. Sci. 2022

[4] Alonso, M.A.; Vilares, D.; Gómez-Rodríguez, C.; Vilares, J. Sentiment Analysis for Fake News Detection. Electronics, 2021

[5] Basiri, Mohammad Ehsan, et al. "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis." Future Generation Computer Systems, 2021

[6] Marouane Birjali, Mohammed Kasri, Abderrahim Beni-Hssane, “A comprehensive survey on sentiment analysis: Approaches, challenges and trends, Knowledge-Based Systems”, 2021

[7] Pu, X.; Yan, G.; Yu, C.; Mi, X.; Yu, C. Sentiment Analysis of Online Course Evaluation Based on a New Ensemble Deep Learning Mode: Evidence from Chinese. Appl. Sci. 2021

[8] S. Poria, D. Hazarika, N. Majumder and R. Mihalcea, "Beneath the Tip of the Iceberg: Current Challenges and New Directions in Sentiment Analysis Research," in IEEE Transactions on Affective Computing, vol. 14, no. 1, pp. 108-132, 1 Jan.-March 2023

[9] P. Karthika, R. Murugeswari and R. Manoranjithem, "Sentiment Analysis of Social Media Network Using Random Forest Algorithm," IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS), 2019

[10] Dang, N.C.; Moreno-García, M.N.; De la Prieta, F. Sentiment Analysis Based on Deep Learning: A Comparative Study. Electronics 2020

DOI: 10.17148/IJARCCE.2023.125264

[11] Sohangir, S., Wang, D., Pomeranets, A. et al. Big Data: Deep Learning for financial sentiment analysis. J Big Data, 2018

[12] Seyed Mahdi Rezaeinia, Rouhollah Rahmani, Ali Ghodsi, Hadi Veisi, "Sentiment analysis based on improved pre-trained word embeddings", Expert Systems with Applications, Volume 117,2019

[13] Zhou Y, Zhang Q, Wang D, Gu X., "Text Sentiment Analysis Based on a New Hybrid Network Model" Computer Intel Neuroscience. 2022 Dec 28;2022:6774320. doi: 10.1155/2022/6774320.

[14] Basiri, Mohammad Ehsan, et al. "ABCDM: An attention-based bidirectional CNN-RNN deep model for sentiment analysis." Future Generation Computer Systems, 2021

[15] Drus, Zulfadzli, and Haliyana Khalid. "Sentiment analysis in social media and its application: Systematic literature review." Procedia Computer Science, 2019

[16] Zad, Samira, et al. "A survey on concept-level sentiment analysis techniques of textual data." 2021 IEEE World AI IoT Congress (AIIoT). IEEE, 2021.