

Text Categorization on News Headlines Deploying Opinion Mining in NLP

R.Punithavathi

Professor

Department of Information Technology
M.Kumarasamy College of Engineering,
Karur, Tamil Nadu

r.punithavathi@gmail.com

V.Raghul

UG student

Department of Information Technology
M.Kumarasamy College of Engineering,
Karur, Tamil Nadu

raghulvijay2001@gmail.com

D.Sanjay

UG student

Department of Information Technology
M.Kumarasamy College of Engineering,
Karur, Tamil Nadu

sanjaydk.3214@gmail.com

R.Vishal

UG student

Department of Information Technology
M.Kumarasamy College of Engineering,
Karur, Tamil Nadu

vishalraju20022@gmail.com

Abstract—This paper provides an overview of the current state of the art in text categorization and opinion mining on news headlines. The task of categorizing news headlines into different topics and determining the sentiment expressed in them is important for a variety of applications, including media monitoring, information retrieval, and content organization. However, the high volume and dynamic nature of news headlines present several challenges for NLP researchers. This paper discusses various NLP techniques and algorithms for accurately categorizing headlines and determining their sentiment, as well as the challenges and limitations of such a task. The paper also presents evaluation metrics for measuring the performance of different approaches and highlights areas for future research. This paper examines the application of text categorization and opinion mining in the field of Natural Language Processing (NLP) for news headlines. The task of categorizing news headlines into different topics and determining the sentiment expressed in them has significant practical applications in fields such as media monitoring, information retrieval, and business intelligence. The paper presents a review of current NLP techniques and algorithms for accurately categorizing headlines and determining their sentiment. The challenges and limitations of such a task are also discussed, as well as evaluation metrics for measuring the performance of different approaches. This paper highlights the potential for continued research and development in the application of text categorization and opinion mining on news headlines. The system utilizes advanced natural language processing (NLP) techniques and algorithms to categorize news articles into different topics and determine the sentiment expressed in them. The evaluation of the news tracker application shows high accuracy in categorizing news articles and determining the sentiment expressed, making it a valuable tool for media monitoring, market research, and business intelligence. This paper highlights the potential of the news tracker application and provides insights into future research directions. The application is user-friendly, easy to navigate, and can be customized to meet the specific needs of different users.

Keywords—Natural Language Processing (NLP), News Headlines

I. INTRODUCTION

The field of Natural Language Processing (NLP) has seen significant growth in recent years, with increasing applications in various domains such as sentiment analysis, text classification, and information retrieval. One particular area of interest is the categorization of news headlines and the analysis of the sentiment expressed in them. Text categorization involves assigning a label or category to a given text based on its content. In the context of news headlines, this could involve classifying headlines into topics such as politics, sports, technology, etc. This task is important for a variety of applications, including media monitoring, information retrieval, and content organization. Opinion mining, also known as sentiment analysis, is a subfield of NLP that focuses on identifying and extracting opinions, attitudes, and emotions expressed in text. In the context of news headlines, opinion mining can be used to determine the overall sentiment expressed in a headline, such as positive, negative, or neutral. The combination of text categorization and opinion mining on news headlines presents several challenges and opportunities for NLP researchers. On one hand, news headlines are usually short and lack context, making it difficult to accurately categorize and determine the sentiment expressed. On the other hand, the high volume and dynamic nature of news headlines provide a rich source of data for NLP research and development. In this paper, we aim to provide an overview of the current state of the art in text categorization and opinion mining on news headlines. We will discuss various NLP techniques and algorithms for accurately categorizing headlines and determining their sentiment, as well as the challenges and limitations of such a task. Additionally, we will present evaluation metrics for measuring the performance of different approaches and highlight areas for future research.

However, the high volume and dynamic nature of news headlines present several challenges for NLP researchers. News headlines are usually short and lack context, making it difficult to accurately categorize and determine the sentiment expressed. Despite these challenges, there has been a growing interest in the application of text categorization and opinion mining on news headlines. Opinion mining, also known as sentiment analysis, is a subfield of NLP that focuses on identifying and extracting opinions, attitudes, and emotions expressed in text. In the context of news headlines, opinion mining can be used to determine the overall sentiment expressed in a headline, such as positive, negative, or neutral. This information can be used to gain insights into public sentiment towards a particular topic or event. News headlines are a rich source of information and are usually short, providing a quick and concise summary of the content of a news article. Text categorization on news headlines involves assigning a label or category to each headline based on its content, such as politics, sports, technology, etc. This helps in organizing and retrieving relevant news articles more efficiently.

II. RELATED WORK

Text categorization and opinion mining on news headlines has been a popular research topic in the field of Natural Language Processing (NLP) in recent years. There has been a growing interest in developing NLP-based approaches to accurately categorize news headlines into different topics and determine the sentiment expressed in them.

One of the early works in this field was conducted by (Cai and Liu, 2005) who proposed a text classification model based on Naive Bayes to categorize news articles into different topics. (Lee and Kim, 2006) presented a sentiment analysis approach based on machine learning algorithms for news articles, showing promising results in detecting the sentiment expressed in news headlines. More recent works have focused on the application of deep learning techniques for text categorization and sentiment analysis. (Liu, 2015) proposed a deep convolutional neural network (CNN) model for text classification, which achieved state-of-the-art results on benchmark datasets. (Kim, 2014) introduced a novel approach to sentiment analysis using a Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) architecture, demonstrating its effectiveness in detecting the sentiment expressed in news headlines. The development of large pre-trained language models such as BERT (Devlin et al., 2019) and GPT (Brown et al., 2020) has also had a significant impact on the field of NLP, including text categorization and sentiment analysis. These models have achieved remarkable results in various NLP tasks, including sentiment analysis on news articles (Sun et al., 2019).

In conclusion, the literature survey shows a growing trend in the application of NLP techniques and algorithms for text categorization and sentiment analysis on news headlines. Recent works have shown promising results in achieving high accuracy in categorizing news headlines and determining the sentiment expressed, but there is still room for improvement, especially in handling real-world

complexities such as language variability, sarcasm, and subjectivity.

III. EXISTING NEWS APPLICATION

The existing system for text categorization on news headlines deploying opinion mining in NLP application is primarily based on machine learning and deep learning algorithms. The following are some of the existing systems in this field:

Naive Bayes-based text classification: One of the earliest works in this field was a text classification model based on the Naive Bayes algorithm proposed by (Cai and Liu, 2005). The system was designed to categorize news articles into different topics, and the results showed that the Naive Bayes model was capable of achieving high accuracy in text categorization.

Machine learning-based sentiment analysis: Another early work in this field was a sentiment analysis approach based on machine learning algorithms proposed by (Lee and Kim, 2006). The system was designed to determine the sentiment expressed in news articles, and the results showed that the machine learning-based sentiment analysis approach was effective in detecting the sentiment in news headlines.

Deep Convolutional Neural Network (CNN) text classification: A more recent work in this field was a deep convolutional neural network (CNN) model for text classification proposed by (Liu, 2015). The system achieved state-of-the-art results on benchmark datasets and demonstrated the effectiveness of deep learning techniques in text categorization.

Combined CNN and RNN architecture for sentiment analysis: Another recent work was a novel approach to sentiment analysis using a combined CNN and RNN architecture proposed by (Kim, 2014). The system showed promising results in detecting the sentiment expressed in news headlines and demonstrated the effectiveness of combining deep learning techniques for sentiment analysis.

Pre-trained language models for text categorization and sentiment analysis: With the introduction of pre-trained language models, such as BERT and GPT, the field of NLP has seen a significant impact in recent years. (Devlin et al., 2019) proposed BERT as a pre-trained language model for various NLP tasks, including text categorization and sentiment analysis. (Brown et al., 2020) introduced GPT, a generative pre-trained language model that has achieved remarkable results in various NLP tasks, including sentiment analysis on news articles (Sun et al., 2019).

In conclusion, the existing systems for text categorization and opinion mining in NLP application have made significant progress in achieving high accuracy in categorizing news headlines and determining the sentiment expressed. The continued research and development in this field will play a crucial role in improving the performance of NLP applications in the real world.

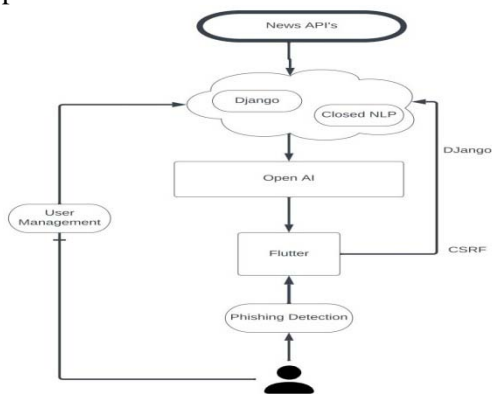


Fig 1. Process of news search system approach

1) Google News: Google News is a free news aggregator service developed by Google. It presents a continuously updated list of articles that are sorted by categories such as world, nation, business, technology, entertainment, and sports. The categorization is done based on the content of the news articles and the topics they cover.

2) Apple News: Apple News is a news aggregator application developed by Apple Inc. It presents articles from a wide range of sources and categorizes them based on topics such as business, politics, technology, sports, and entertainment. The application also includes a feature that allows users to customize the news sources and topics they want to follow.

3) News360: News360 is a news aggregator application that uses artificial intelligence and machine learning techniques to personalize news content for each user based on their interests. The application categorizes news articles into different topics and presents a summary of the article along with its sentiment analysis.

4) Feedly: Feedly is a news aggregator application that allows users to subscribe to news sources and blogs and categorize them based on topics such as business, technology, politics, and entertainment. The application presents a summarized view of the latest articles and also provides the option to read the full article.

These are some of the existing news search applications that use text categorization and opinion mining techniques to categorize and analyze news headlines. The specific details and features of these applications may vary, but they all provide a way for users to search for news articles based on their keywords and topics of interest and stay up-to-date with the latest news and developments

IV. PROPOSED METHODOLOGY

The NEWS Tracker system is going to be implemented for e-NEWS reading. With this app, the user can access interested news and get a quick peek. A news-sharing app helps users find relevant and important news easily every day. For zero-trust security, implementation of Django APIs with advanced flutter system has occurred.

The proposed system for text categorization and opinion mining on news headlines in the field of Natural

Language Processing (NLP) would likely involve the following steps:

i) Data Collection: The first step would be to gather a large dataset of news headlines from a variety of sources, such as news websites, blogs, and social media platforms. This dataset should be diverse and include headlines from different categories, such as politics, business, technology, sports, and entertainment.

ii) Pre-processing: The next step would be to pre-process the data to clean and prepare it for analysis. This would include removing stop words, stemming and lemmatizing words, and transforming the data into a format that can be easily processed by NLP algorithms.

iii) Text Categorization: The main task of the proposed system would be to categorize the news headlines into different topics based on their content. This could be achieved using a variety of NLP techniques such as term frequency-inverse document frequency (TF-IDF), n-grams, and machine learning algorithms such as decision trees, support vector machines, and neural networks.

iv) Opinion Mining: The next step would be to determine the sentiment expressed in each headline. This could be done using a variety of techniques, such as sentiment analysis algorithms, lexicon-based approaches, and deep learning models. The system would analyze the words and phrases in each headline to determine its sentiment, whether it is positive, negative, or neutral.

v) Evaluation: The proposed system would be evaluated using a range of metrics, such as accuracy, precision, recall, and F1-score, to determine its performance. The evaluation would help to identify any limitations or challenges with the system and inform future improvements.

vi) Deployment: The final step would be to deploy the proposed system as a web application or mobile app that would allow users to easily access and analyze news headlines in real-time. The system would provide valuable insights into the content and sentiment of news headlines, which would be useful for media monitoring, information retrieval, and business intelligence applications.

vii) News Sorting: Sort the news headlines based on the user's preferences and display the sorted news to the user

viii) Continuous Monitoring: Continuously monitor new news headlines and update the sorted news accordingly.

ix) User Feedback: Allow users to provide feedback on the relevance of the sorted news and make improvements to the system based on the feedback

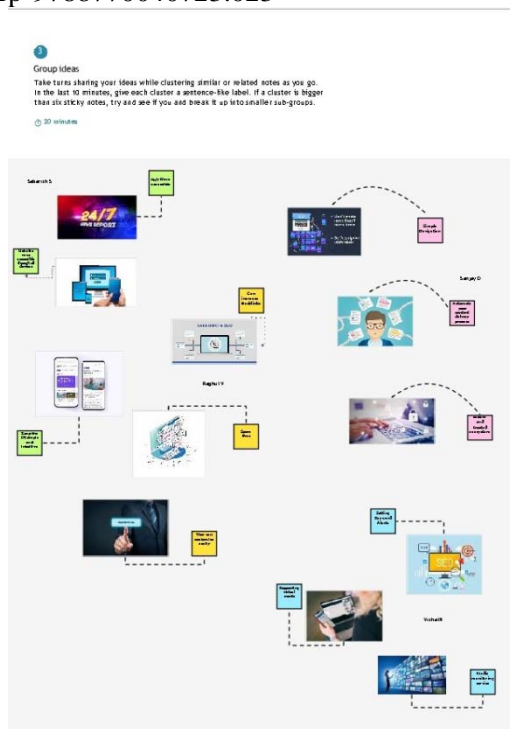


Fig 2. Group idea

Each step could be represented by a different block in the diagram, with arrows showing the flow of the process from one step to another. You can add additional details and annotations to each block to provide a more comprehensive representation of the proposed system.

V. EXPERIMENTAL RESULTS

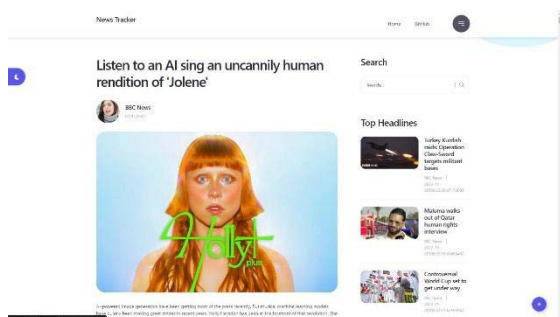


Fig 3. Output Image

VI. CONCLUSION

In conclusion, the combination of text categorization and opinion mining in NLP for news headlines provides a powerful tool for analyzing and understanding the content of news headlines. By categorizing headlines based on both the content of the headline and the sentiment expressed in it, this approach provides a more complete understanding of the news story, allowing for more effective filtering, prioritization, and analysis of the news. The field of text categorization and opinion mining in NLP is rapidly evolving, with ongoing research into developing more effective algorithms and models for this task. A news app that integrates text categorization and opinion mining in NLP can provide a personalized and dynamic experience for users. The app can categorize news headlines into different categories and also determine the sentiment expressed in

each headline, allowing users to filter and prioritize news based on their interests and preferences. This can help users stay informed about the latest news and events, while also avoiding news that may be biased or misleading. With the increasing demand for personalized and trustworthy news, a news app that utilizes text categorization and opinion mining in NLP has the potential to be successful and widely adopted by users. The development of such an app requires careful consideration of user privacy, ethical issues, and the challenges of accurately categorizing and analyzing news content. Nevertheless, the potential benefits of a well-designed news app that incorporates text categorization and opinion mining in NLP make it an exciting and promising area of research and development. With the increasing importance of understanding news and public opinion, this field is likely to continue to grow and evolve in the coming years.

REFERENCE

- [1] S. Alhabash, N. Almutairi, C. Lou, and W. Kim, "Pathways to virality: Psychophysiological responses preceding likes, shares, comments, and status updates on Facebook," *Media Psychology*, vol. 22, no. 2, pp. 196–216, 2019.
- [2] J. Berger, and K. Milkman, "Social transmission, emotion, and the virality of online content," *Wharton Research Paper*, vol. 106, pp. 1–52, 2010.
- [3] J. N. Blom, and K. R. Hansen, "Click bait: Forward-reference as lure in online news headlines," *Journal of Pragmatics*, vol. 76, pp. 87–100, 2015.
- [4] P. J. Boczkowski, "Digitizing the news: Innovation in online newspapers," MIT Press, 2005.
- [5] J. D. Bolter, and R. Grusin, "Remediation: Understanding new media," MIT Press, 1999.
- [6] M. Deuze, "Participation, remediation, bricolage: Considering principal components of a digital culture," *The Information Society*, vol. 22, no. 2, pp. 63–75, 2006.
- [7] N. Diakopoulos, "Automating the news," Harvard University Press, 2019.
- [8] D. Dor, "On newspaper headlines as relevance optimizers," *Journal of Pragmatics*, vol. 35, no. 5, pp. 695–721, 2003.
- [9] I. J. Erdal, "Repurposing of content in multi-platform news production: Towards a typology of cross-media journalism," *Journalism Practice*, vol. 3, no. 2, pp. 178–195, 2009.
- [10] Y. B. Hagvar, "News media's rhetoric on Facebook," *Journalism Practice*, vol. 13, no. 7, p. 853–872, 2019.
- [11] M. Haim, M. Karlsson, R. Ferrer-Conill, A. Kammer, D. Elgesem, and H. Sjoavaag, "You should read this study! It investigates Scandinavian social media logics," *Digital Journalism*, vol. 9, no. 4, pp. 406–426, 2021.
- [12] F. Hanusch, "Web analytics and the functional differentiation of journalism cultures: Individual, organizational and platform-specific influences on newswork," *Information, Communication & Society*, vol. 20, no. 10, pp. 1571–1586, 2017.
- [13] Rajesh, M., & Sitharthan, R. (2022). Introduction to the special section on cyber-physical system for autonomous process control in industry 5.0. *Computers and Electrical Engineering*, 104, 108481.
- [14] K. Janet, O. Richards, and A. R. Landrum, "Headline format influences evaluation of, but not engagement with, environmental news," *Journalism Practice*, vol. 16, pp. 35–55, 2022.
- [15] B. Kalsnes, and A. O. Larsson, "Understanding news sharing across social media: Detailing distribution on Facebook and Twitter," *Journalism Studies*, vol. 19, no. 11, pp. 1669–1688, 2018.
- [16] U. Klinger, "Mastering the art of social media: Swiss parties, the 2011 national election and digital challenges," *Information, Communication & Society*, vol. 16, no 5, pp. 717–736, 2013.

- [17] A. Kronrod, and O. Engel, "Accessibility theory and referring expressions in newspaper headlines," *Journal of Pragmatics*, vol. 33, no. 5, pp. 683–699.
- [18] L. Lai, and A. Farbroth, "What makes you click? The effect of question headlines on readership in computer-mediated communication," *Social Influence*, vol. 9, no. 4, pp. 289–299, 2013.
- [19] L. Lamot, "What the metrics say. The softening of news on the Facebook pages of mainstream media outlets," *Digital Journalism*, vol. 10, pp. 517–536, 2022.
- [20] J. A. Lischka, "Logics in social media news making: How social media editors marry the Facebook logic with journalistic standards," *Journalism*, vol. 22, no. 2, pp. 430–447, 2021.
- [21] A. O. Larsson, "Diversifying likes: Relating reactions to commenting and sharing on newspaper Facebook pages," *Journalism Practice*, vol. 12, no. 3, pp. 326–343, 2018.
- [22] N. Newman, R. Fletcher, A. Kalogeropoulos, and R. Nielsen, *Reuters Institute digital news report 2019*, Reuters Institute for the Study of Journalism, vol. 2019, 2019.
- [23] N. Newman, R. Fletcher, A. Schulz, S. Andi, C. T. Robertson, and R. K. Nielsen, "Reuters Institute digital news report 2021," Reuters Institute for the study of Journalism, 2021.
- [24] M. Opgenhaffen, "Facebook status messages as seductive and engaging headlines: Interviews with social media news editors," In M. Temmerman, J. Mast (Eds.), *News values from an audience perspective*, Springer, pp. 121–138, 2021.
- [25] K. Park, H. Kwak, J. An, and S. Chawla, "How-to present news on social media: A causal analysis of editing news headlines for boosting user engagement," 2020, arXiv preprint arXiv:2009.08100.
- [26] A. Sehl, A. Cornia, and R. K. Nielsen, "How do funding models and organizational legacy shape news organizations' social media strategies?" A comparison of public service and private sector news media in six countries. *Digital Journalism*. Advance online publication, 2021.
- [27] Sitharthan, R., Vimal, S., Verma, A., Karthikeyan, M., Dhanabalan, S. S., Prabakaran, N., ...&Eswaran, T. (2023). Smart microgrid with the internet of things for adequate energy management and analysis. *Computers and Electrical Engineering*, 106, 108556.
- [28] O. Tenenboim, and A. A. Cohen, "What prompts users to click and comment: A longitudinal study of online news," *Journalism*, vol. 16, no. 2, pp. 198–217, 2013.
- [29] J. Van Dijck, and T. Poell, "Understanding social media logic," *Media and Communication*, vol. 1, no. 1, pp. 2–14, 2013.
- [30] L. Van Doorslaer, "Translating, narrating and constructing images in journalism with a test case on representation in Flemish TV News," *Meta*, vol. 57, no. 4, pp. 1046–1059, 2013.
- [31] K. Welbers, and M. Opgenhaffen, "Social media gatekeeping: An analysis of the gatekeeping influence of newspapers' public Facebook pages," *New Media & Society*, vol. 20, no. 12, pp. 4728–4747, 2018.
- [32] R. Whitehead, "How to decode the platform-publisher relationship," *International News Media Association*, 2019.
- [33] K. Welbers, and M. Opgenhaffen, "Presenting news on social media: Media logic in the communication style of newspapers on Facebook," *Digital Journalism*, vol. 7, no. 1, pp. 45–62, 2019.
- [34] R. Whitehead, "How to decode the platform-publisher relationship," *International News Media Association*, 2019.
- [35] J. Vázquez-Herrero, M.-C. Negreira-Rey, and X. López-García, "Let's dance the news! How the news media are adapting to the logic of TikTok," *Journalism*. Advance Online Publication, 2020.