News Articles Sentiment Analysis using Natural Language Processing

Aadya Mishra, Bhoomika Negi , Suhani Ahuja
School of Engineering and Technology, K.R. Mangalam University

Dr. Surabhi Shanker
Dept. of CSE, School of Engineering &
Technology, Centre of Excellence-Cybersecurity,
K.R. Mangalam University

Abstract
Digital news has revolutionized information consumption which had offered the instant access to global events through social media , websites and mobile apps . The rapid expansion of digital news media has led to an increased need for automated sentiment analysis to assess public opinions and detect biases. This research allow us to explore sentiment analysis in the news articles using natural language processing (NLPs) techniques. Various approaches, including lexicon-based models, are discussed, with a focus on their effectiveness in way to capture the sentiment nuances in journalistic text. The research focuses on analyzing the sentiment of news articles using nlp techniques to determine whether the content is positive, negative, or neutral. Preprocessing techniques such as tokenization, stopword removal, and lemmatization are applied to enhance the accuracy. This research also tries to explore for detecting the sentiment analysis of regional language news article whether the content is positive , negative or neutral . Results indicate that advanced nlp models achieve high accuracy, making sentiment analysis valuable for media monitoring as well as in decision-making.

Keywords: sentiment analysis, news articles, natural language processing, text mining

Introduction
News articles play a crucial role in shaping facts and public opinions and influencing the societal perspectives. With the proliferation of online news platforms, analyzing the sentiment conveyed in news content has become increasingly important. As news articles influence the decision-making in various sectors, including politics, finance, and public policy, understanding their sentiment can too provide us valuable insights.
Sentiment analysis, a subfield of natural language processing (nlp), is defined as the process of determining the sentiment polarity of textual data—whether it is positive, negative, or neutral. While sentiment analysis has been extensively applied to social media, product reviews, and customer feedback, its application to news articles presents unique challenges. News content is often written in a formal and objective manner, making it difficult to detect explicit sentiment. Additionally, implicit biases, contextual dependencies, and subtle linguistic nuances further complicate sentiment classification. Unlike social media posts, which are often emotionally charged and direct, news articles require more sophisticated nlp techniques to accurately determine their sentiment.

DOI: https://doi.org/10.5281/zenodo.15186992

This research paper aims to explore the effectiveness of nlp techniques in sentiment analysis of news articles. It examines existing literature to understand the strengths and limitations of various approaches. Traditional machine learning models, deep learning techniques, and lexicon-based approaches are
reviewed to determine their applicability to news sentiment analysis. The research paper tries to find out the sentiment analysis of the regional news article through various libraries of python.

By investigating various nlp models and evaluating their performance, this research aims to contribute to the advancement of sentiment analysis in journalism. The findings of this study can be beneficial for media monitoring, public sentiment tracking, and automated news evaluation systems, ultimately aiding informed decision-making.

Building on the reviews of machine learning techniques in order to generate the sentiments of the news content offers deeper insights into public opinion trends. Techniques such as naïve bayes and deep learning model effectively classifies the emotion from the text data enhancing decision-making in fields of research. Machine learning handles the complex patterns , large sets of data , ensuring higher accuracy.

Literature review:

Hossain, Arafat, et al. (2021) explored text mining and sentiment analysis of newspaper headlines, emphasizing the challenges associated with contextual interpretation. Previous studies had applied machine learning and deep learning techniques to classify sentiment ; however, difficulties arise because of implicit biases and linguistic complexities (Hossain et al., 2021). Lexicon-based approaches have also been utilized but often struggle with neutral tones in journalism, limiting their effectiveness (Hossain et al., 2021). Recent advancements in natural language processing (NLP) i.e. transformer- based models, have significantly improved the sentiment classification accuracy (Hossain et al., 2021). However, limitations still seen in detecting sarcasm, subjectivity, and context-specific meanings, highlighting the areas for further research and refinement (Hossain et al., 2021).

Nasukawa and Yi (2003) proposed an NLP-based method for sentiment extraction, highlighting the importance of linguistic structures in determining sentiment polarity. Early approaches depends on rule- based and lexicon-based techniques; but, these methods were found struggling with contextual ambiguity (Nasukawa & Yi, 2003). The introduction of machine learning models enhanced sentiment analysis by incorporating statistical methods, thereby improving accuracy for the analysis (Nasukawa & Yi, 2003). More recent advancements, such as deep learning and transformer-based architectures, had further refined sentiment classification by studying complex linguistic patterns (Nasukawa & Yi, 2003). Despite these improvements, challenges such as sarcasm detection and implicit sentiment understanding continue to require further research (Nasukawa & Yi, 2003).

Balahur et al. (2013) examined sentiment analysis in the news domain, and highlighting challenges in identifying implicit sentiments. Traditional methods approaches often struggle with the neutral tone of the articles, which makes the sentiment classification difficult and complicated (Balahur et al. 2013). On the other hand, Deep learning models have improved the performance just by capturing contextual nuances, yet they still face difficulties in detecting subtle biases and indirect expressions of sentiment (Balahur et al. 2013). Recent advancements in NLP, offer promising solutions, but further refinement is needed to enhance the accuracy in handling complex linguistic structures in the content of news article (Balahur et al. 2013).

Yi et al. (2003) developed a sentiment analyzer using NLP techniques to extract sentiments from topic- related texts, emphasizing the importance of linguistic structures. Early sentiment analysis methods were relied on rule-based and lexicon-based approaches but it struggled with contextual interpretation (Yi et al., 2003). With the rise of deep learning, models have significantly improved accuracy (Yi et al. 2003), However,

DOI: https://doi.org/10.5281/zenodo.15186992

challenges remain in detecting implicit sentiment, and sarcasm, necessitating further advancements in NLP-based sentiment analysis Yi et al. (2003)

Shanker and Bhushan (2020) analyses social media data to support policymaking by reviewing various machine learning techniques. It routes over methods such as natural language processing (NLP), sentiment analysis and clustering algorithms which emphasizes the effectiveness in identifying public opinion, detecting the trends and helping the decision-making processes (Shanker & Bhushan,2020). The authors also highlights the significance of leveraging social media insights for informed policy decisions in dynamic socio-political environments (Shanker & Bhushan,2020). The authors also highlights the significance of leveraging social media insights for informed policy decisions in dynamic socio-political environments (Shanker & Bhushan,2020). Khoo et al. (2012) applied theory to online news article for its sentiment analysis, providing insights into the author's perspectives and evaluative language. By analyzing the judgmental, and appreciational elements in news articles, they demonstrated how sentiments are not only expressed through direct sentiment words but also via subtle evaluative tones (Khoo et al. 2012). This approach deepens the understanding of how sentiment is conveyed beyond the simple polarity, and through also offering a more nuanced interpretation of the news contents (Khoo et al. 2012). Despite advancements, challenges also remain in the detection and exploration of these complex linguistic features(Khoo et al. 2012). Žitnik et al. (2022) surveyed the target level sentiment analysis to study and enhance the sentiment percision by associating sentiments with specific entities mentioned in the articles. This approach may allow for more accurate results of sentiment classification by just focusing on the specified or particular targets like individuals , organizations , present within the text. (Žitnik et al., 2022).

Unlike traditional sentiment analysis, which evaluates the overall sentiment, target-level analysis captures nuanced opinions toward distinct entities and domains (Žitnik et al., 2022).

Although this method helps to improve accuracy, challenges remain in identifying and disambiguating targets, as well as handling the complex sentence structures in news content (Žitnik et al., 2022)

Singh and Jain (2021) examined transformer models for sentiment analysis of news headlines, which achieve high accuracy in sentiment classification. By leveraging the power of transformer architectures, they enhance the ability of the model to study contextual relationships, often ambiguous news headlines (Singh & Jain, 2021). Their approach exhibit significant improvements over traditional machine learning models and lexicon-based methods (Singh & Jain, 2021). However, challenges remain in handling diverse linguistic features, and varying news styles, requiring further fine-tuning of transformer models for optimal performance (Singh & Jain, 2021).

Srusti (2024) prospects NLP techniques for sentiment analysis on financial news, which focuses their role in prediction of the market trend. The study explores the traditional methods alongside with advanced models like BERT (Srusti,2024). It highlights the significance of domain-specific language models in order to improve accuracy. It layers the impact of data

.

DOI: https://doi.org/10.5281/zenodo.15186992

processing and feature engineering in order to enhance the model performance, also offering overview of advancements of sentiment analysis in financial context (Srusti,2024).

Albladi et al. (2025) provide us a comprehensive review of the NLP models for sentiment analysis of data on Twitter, emphasizing the strengths and limitations about various approaches. The study focuses on traditional machine learning models, deep learning models, and transformer-based models such as (Albladi ,2025). The authors also states that transformer models helps to excel its accuracy upto their contextual understanding capabilities. The review underscores the importance of data preprocessing and model fine-tuning for the optimal results (Albladi ,2025).

S.No.	Research Paper Title And Year	Journal/Conference	Methodology	Result
1.	Text Mining And Sentiment Analysis Of Newspaper Headlines." Information 12.10 (2021): 414.	Journal: Information	Applied Methods Like Text Mining And Sentiment Analysis On Newspaper Headlines Using Lexicon-Based And Machine Learning Models To Classify Sentiments Of The Headline.	Achieved High Accuracy In Detecting Sentiment Trends, With Ml-Based Approaches Performing Better Than Lexicon-Based Methods. It Also DemonstratedPractical Applications In Media Monitoring.
2.	"Sentiment Analysis: Capturing Favorability Using Natural Language Processing." Proceedings Of The 2nd International Conference On Knowledge Capture. 2003.	Conference: 2nd International Conference On KnowledgeCapture	Proposed An Nlp-Based Sentiment Analysis Model WhichIdentifies Favourability In Text Using Semantic Analysis.	Showed That Deep Linguistic Processing ImprovesAccuracy Of Sentiment Detection, Particularly For Subjective Expressions, Outperforming Keyword-Based Approaches.
3.	"Sentiment Analysis In The News." Arxiv Preprint Arxiv:1309.6202 (2013).	Journal: Arxiv Preprint	Evaluated Sentiment Analysis TechniquesFor News Articles, Compares Lexicon-Based And Machine Learning Methods, Alongside Analyzing Linguistic Challenges In News Reporting.	Highlighted Biases AndComplexities In News Sentiment, Showing That Ml Models May Struggle With Neutral Sentiment Detection But Perform Well For Polarity Classification.

DOI: https://doi.org/10.5281/zenodo.15186992

4.	**"Sentiment Analyzer: Extracting Sentiments About A Given Topic Using Natural Language Processing Techniques." Third Ieee International Conference On Data Mining. Ieee, 2003.**	**Conference: Third IEEE International Conference On Data Mining**	**Developed A Sentiment Extraction System Which Detects The Emotional Expressions Using NLP BasedSyntactics And Semantic Analysis.**	**The System Was Able To Effectively Extract Sentiment At A Topic Level, Along With Improved Context Understanding Compared To That Of Traditional Keyword-Based Sentiment Methods.**
5.	**"A Review On Machine Learning Techniques On Social Media Data For Policy Making." International Research Journal Of Engineering And Technology (IRJET), 7(6), 4573–4578.**	**Journal: International Research Journal Of Engineering And Technology(IRJET).**	**Author Reviewed On Machine Learning Techniques To Social Media Data By Analyzing Algorithms Like Decision Trees , NLP For Sentiment Analysis.**	**Machine Learning Significantly Improve Policymakers To Identify Trends, Analyse Social Media Data And Understand Public Sentiment.**
6.	**"Sentiment Analysis Of Online News Text: A Case Study Of Appraisal Theory." Online Information Review 36.6 (2012): 858-878.**	**Journal: Online InformationReview**	**Applied Appraisal Theory For Sentiment Analysis Of News, Also Analyzing Textual Tone, Modality, And Stance Of The Article Rather Than Just Polarity.**	**Demonstrated That News Articles Contains Complex Sentiment Structures, Which Require Context-Aware Nlp Approaches For Accurate Analysis.**
7.	**"Target-Level Sentiment Analysis For News Articles." Knowledge-Based Systems 249 (2022): 108939.**	**Journal:Knowledge-Based Systems**	**Conducted Target-Level Sentiment Analysis Reviewing Deep Learning**	**The Result Showed That Context-Aware Nlp Models Can Significantly Enhance Sentiment**

DOI: https://doi.org/10.5281/zenodo.15186992

			Models, Focusing On Specific Entities In News Articles.	**Accuracy, Outperforming Traditional Approaches In TargetedOpinion Extraction.**
8.	**"Sentiment Analysis Of News Headlines Using Simple Transformers." 2021 Asian Conference On Innovation In Technology (Asiancon). Ieee, 2021.**	**Conference: Asian Conference On Innovation In Technology (Asiancon)**	**Utilized Simple Transformers For Sentiment Analysis On The News Headlines.**	**Achieved High Sentiment Classification Accuracy, Proving The Effectiveness Of Transformer-Based Nlp Models And Success In Analyzing Short News Texts.**
9.	**“Srusti, R. "NLP-Based Sentiment Analysis Of Financial News." Journal Of Financial Technologies And NLP (2024).”**	**Journal: International Journal Of Engineering Research & Technology (IJERT)**	**The Study Employs Natural Language Processing Techniques And A Custom K-Nearest Neighbours Classifier To Categorize Financial News Into Sentiments I.E. Positive , Negative And Neutral.**	**The Classifier's Performance Is Illustrated Through A Confusion Matrix, Revealing Relationships Between Sentiment Classes. It Provides Insights Into The Correlation Between Sentiment And The Vocabulary Usage In Financial News.**
10.	**Albladi, Aish, Minarul Islam, And Cheryl Seals. "Sentiment Analysis Of Twitter Data Using NLP Models: A Comprehensive Review." IEEE Access (2025).**	**Conference: IEEE**	**NLP Models Applied To Twitter Sentiment Analysis Which Includes Machine Learning, Deep Learning, And Hybrid Approaches.**	**Study Highlights That Transformer Models Like BERT Outperform The Traditional Methods In Accuracy And The Efficiency Of Analysis.**

DOI: https://doi.org/10.5281/zenodo.15186992

Research methodology:

This study examines natural language processing (NLP) to perform sentiment analysis on news articles, categorizing them as positive, negative, or neutral and providing an associated sentiment score along with the analysis. Firstly , the regional news articles are first translated into english language using the googletrans python library to standardize the text. Then, the data undergoes pre-processing, including tokenization, stopword removal, and lemmatization, to enhance the accuracy. Sentiment scores are then computed using lexicon-based techniques such as vader or textblob. The analyzed results help the user in understanding the overall sentiment trend seen in news reporting, ensuring reliable insights into public perception and media bias.

Result and analysis

This section presents the results of sentiment analysis on English and regional language (hindi, maithili, and haryanvi) news articles. The analysis evaluates sentiments int the article as positive, negative, or neutral, using polarity scores computed via library such as textblob and vader. The system accepts the news article in the form of text file . Then the text file undergo the analysis . For regional languages, googletrans, a python library, is being used for translation before sentiment analysis to make it easy and understandable for the user.

Result for English news article –

The news article as follows: “the economy is thriving, with industries expanding and creating new job opportunities. Unemployment rates have dropped significantly, and wages are rising. Investors are optimistic, and businesses are flourishing. Government initiatives are further strengthening economic stability, fostering innovation, and enhancing the quality of life for millions. ”

The text file:

```
The economy is thriving, with industries expanding and creating new job
opportunities. Unemployment rates have dropped significantly, and wages
are rising. Investors are optimistic, and businesses are flourishing.
Government initiatives are further strengthening economic stability,
fostering innovation, and enhancing the quality of life for millions.
```

```
Requirement already satisfied: nltk in /usr/local/lib/python3.11/dist-packages (3.9.1)
Requirement already satisfied: textblob in /usr/local/lib/python3.11/dist-packages (0.19.0)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk) (8.1.8)
Requirement already satisfied: joblib in /usr/local/lib/python3.11/dist-packages (from nltk) (1.4.2)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk) (2024.11.6)
Requirement already satisfied: tqdm in /usr/local/lib/python3.11/dist-packages (from nltk) (4.67.1)
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Package punkt is already up-to-date!
[nltk_data] Downloading package averaged_perceptron_tagger to
[nltk_data]     /root/nltk_data...
[nltk_data]   Package averaged_perceptron_tagger is already up-to-
[nltk_data]       date!
'Sentiment: Positive (Polarity: 0.7333333333333333)'
```


DOI: https://doi.org/10.5281/zenodo.15186992

The information about expansion, investments, stability has positive sentiment polarity.

Result for regional news article –

the news article as follows: :

“ 3

the text file:

Sentiment analysis of the maithili news article:

```
Device set to use cpu
'Sentiment: NEGATIVE, Score: 0.9967567324638367'
```

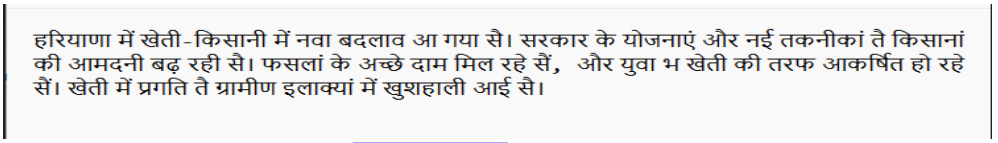
Economic issues like unemployment ,

Result for regional news article –
Say haryanvi,

the news article as follows: : “

T

the text file:



DOI: https://doi.org/10.5281/zenodo.15186992

Sentiment analysis of haryanvi news article –

```
Device set to use cpu
'Sentiment: POSITIVE, Score: 0.9978960752487183'
```


DOI: https://doi.org/10.5281/zenodo.15186992

News about development , increment had positive score .

Conclusion

This research highlights the significance of sentiment analysis in news articles using NLP techniques, particularly in the era of digital media expansion. By leveraging preprocessing techniques such as tokenization, stopword removal, and lemmatization, sentiment analysis became more effective in distinguishing between the positive and the negative sentiments. The study also tied to extends its scope to regional language news articles by utilizing translation tools, ensuring a broader application of sentiment analysis. Results demonstrate that advanced nlp models achieve high accuracy in detecting sentiment, making them valuable and effective for media monitoring and decision-making. Machine learning and its model have significantly enhanced sentiment analysis in news articles by improvement in accuracy, scalability, and context understanding. Techniques like deep learning and transformer models outperform traditional NLP methods by capturing complex language patterns, sentiment nuances, and contextual dependencies. These advancements enable more precise sentiment classification, benefiting financial forecasting and market analysis.

However, a key limitation of the system is its inability to effectively analyze neutral sentiment. Many news articles contain balanced or factual reporting without strong emotional tones, making it difficult for the model to classify them accurately. This limitation affects the overall sentiment distribution and may lead to biased interpretations. Future research may focus on improving the detection of neutral sentiments to enhance the system's reliability and precision.

References

1. hossain, arafat, et al. "text mining and sentiment analysis of newspaper headlines." information 12.10 (2021): 414.
2. nasukawa, tetsuya, and jeonghee yi. "sentiment analysis: capturing favorability using natural language processing." proceedings of the 2nd international conference on knowledge capture. 2003.
3. Balahur, alexandra, et al. "sentiment analysis in the news." arxiv preprint arxiv:1309.6202 (2013).
4. Yi, jeonghee, et al. "sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques." third ieee international conference on data mining. Ieee, 2003.
5. Bhushan, S., & Shanker, S. (2020). A review on machine learning techniques on social media data for policy making. International Research Journal of Engineering and Technology (IRJET), 7(6), 4573– 4578.
6. Soo-guan khoo, christopher, armineh nourbakhsh, and jin-cheon na. "sentiment analysis of online news text: a case study of appraisal theory." online information review 36.6 (2012): 858-878.
7. Žitnik, slavko, neli blagus, and marko bajec. "target-level sentiment analysis for news articles." knowledge-based systems 249 (2022): 108939.
8. Singh, anurag, and goonjan jain. "sentiment analysis of news headlines using simple transformers." 2021 asian conference on innovation in technology (asiancon). Ieee, 2021.

DOI: https://doi.org/10.5281/zenodo.15186992

9. Srusti, R. "NLP-Based Sentiment Analysis of Financial News." Journal of Financial Technologies and NLP (2024).
10. Albladi, Aish, Minarul Islam, and Cheryl Seals. "Sentiment Analysis of Twitter data using NLP Models: A Comprehensive Review." IEEE Access (2025).

DOI: https://doi.org/10.5281/zenodo.15186992