Opinion mining on newspaper headlines regarding the US elections using NLP, SVM and Deep Learning

MSc Research Project
Data Analytics

Gandam Suresh Kailash
Student ID: X23177667

School of Computing
National College of Ireland

Supervisor: Dr Anu Sahni

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Gandam Suresh Kailash
Student ID:	X23177667
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Dr Anu Sahni
Submission Due Date:	12/12/2024
Project Title:	Opinion mining on newspaper headlines regarding the US elections using NLP, SVM and Deep Learning
Word Count:	XXX
Page Count:	17

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Gandam Suresh Kailash
Date:	29th January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Opinion mining on newspaper headlines regarding the US elections using NLP, SVM and Deep Learning

Gandam Suresh Kailash
X23177667

Abstract

This research investigates sentiment analysis on newspaper headlines concerning the 2024 U.S. Presidential Election using Natural Language Processing (NLP),Support Vector Machine(SVM). Multiple researches have been done in opinion mining for online blogs, Twitter, Facebook etc. using the public social media platforms but in this paper we are focused towards the headlines which first attracts the consumer to further read the content. The primary objective of the research is to predict the public sentiment and its potential influence on electoral outcomes by analyzing the important headlines from major news outlets. The study initially utilized Support Vector Machines (SVM) with TF-IDF vectorization with the further refinement was undertaken by incorporating Word2Vec embeddings with an improved accuracy.

To enhance performance and to understand the small nuances in the findings advanced transformers like BERT and RoBERTa were explored, leveraging their pretrained architectures for fine-grained sentiment classification. Despite the moderate gains with using the transformers, the results highlighted the inherent challenges of sentiment classification in nuanced, politically charged content. The project focused on the early stages such as feature engineering and preprocessing techniques, such as Named Entity Recognition (NER), to contextualize sentiment further.

1 Introduction

Opinion Mining or Sentiment Analysis is a way of analyzing the opinions or sentiments expressed in the textual data and is essential in various Natural Language Processing (NLP) applications. With the rapid growth of online social media, sentiment analysis has gained significant relevance, especially for analyzing reviews, feedback, and public opinions expressed on platforms like social media and blogs. In this modern society which has an shorter attention span news headlines are a important tool for influencing public opinion. Studies reveals that an individual often makes snap decisions by skimming news headlines instead of reading an whole article. Beacause of this, even small headlines can have a big social impact.

The sentiment analysis of newspaper headlines about the recent US elections in 2024 is the main focus of this paper. The main purpose of the paper is to examine how well the public perceptions of political parties and candidates are reflected and manipulated with that of the headlines that are published during the election season. A sentiment scale from 1 to 5, with 1 denoting extremely negative sentiment and 5 denoting extremely positive sentiment, is employed for this data with using the sentiment categorization was achieved using the vader sentiment and with some manual intervention through domain-specific analysis of each headline's tone and language.Data collection was conducted manually which spanned over 2 weeks and news sources which was published 2 months before the voting days has been collected starting from Sep5 to Nov 5. The headlines are preprocessed with the different stages such as lemmatization, removal of stopwords, and Named Entity Recognition (NER), the sentiment data was used to train different models.

The proposed method employs a combination of machine learning and deep learning techniques. Initially, Support Vector Machines (SVM) were applied using TF-IDF and Google Word2Vec embeddings for feature representation. The study further explores transformer-based models like RoBERTa to improve contextual understanding and accuracy. Models were evaluated using metrics such as accuracy, precision, recall, and F1-score, allowing a comparative analysis of their performance. This research aims to demonstrate how sentiment embedded in news headlines influences public opinion and provide insights into the effectiveness of various sentiment analysis techniques.

This paper is organized as follows: Section 2 reviews work on sentiment analysis and opinion mining. Section 3 explains the methodology and models used. Section 4 presents experimental results and compares the performance of SVM and RoBERTa. Finally, Section 5 concludes the study and provides future research directions.

1.1 Research Questions

RQ1: How do sentiment trends in newspaper headlines leading up to the 2024 US election influence public opinion and voting behavior?
RQ2: Can advanced NLP techniques, such as BERT and RoBERTa, significantly improve sentiment analysis accuracy compared to traditional methods (e.g., TF-IDF and SVM) in the context of political news?
RQ3: What is the impact of named entity recognition (NER) on the performance of sentiment analysis models for political news?

s

2 Related Work

Agarwal et al. [4] proposed a method which is a combination of the two algorithms—one for data preprocessing and another for detecting the polarity value of words, in their approach they made use of the NLTK, a Python-based library, for tokenization, POS tagging, lemmatization, and stemming, along with SentiWordNet, which assigns numerical scores to words based on their sentiment polarity. Their whole added sentiment score of headlines determines their classification as positive or negative. This categorization is effective for lexicon-based methods, the approach struggles with context-dependent sentiment expressions. categorized the data in to 5 categories ranging from 1 to 5.

Yang et al. [5] created an hybrid model which is tailored to do sentiment analysis within a single domain. They used a variety of classification techniques, which increased the effectiveness of dividing the data into neutral, negative, and positive groups. However, its scalability was limited as the complexity increased when applied across domains.

Rana and Singh [7] conducted a comparative analysis of Linear SVM and Naive Bayes for sentiment classification on movie reviews. Using Porter Stemmer for preprocessing and RapidMiner for model building, their study demonstrated that SVM outperformed Naive Bayes in terms of precision and recall. In this paper we know about the effectiveness of the SVM for text classification tasks, which aligns with its adoption for the study of election related sentiment analysis over the headlines.

Bakshi et al. [8] focused on analysing Twitter data for Samsung Electronics Ltd. by segregating tweets into positive, negative, and neutral categories. Although limited to a single company, this study displayed the effective use of the sentiment analysis in the targeted domains, this showed the importance of preprocessing and categorization techniques.

Aroju et al. [9] applied SVM and Naive Bayes for opinion mining across news headlines from The Hindu, The Times of India, and Deccan Chronicle. They showed the different levels of differences in the positivity across the newspapers, which showcased the influence of the source bias but also this had only the small dataset of the 105 headlines limits the study's generalizability. [10]In this project of using the naive bayes in order to perform the opinion mining. on amazon reviews in both English and Bangla. Their results displayed the importance of clearing the noise in translated datasets, underscoring the importance of preprocessing for cross-language sentiment analysis. In election headlines sentiment analysis Chaudhary et al. analysed the use of the Linear SVM with the n-gram and CoreNLP for the newspaper headlines in using the appropriate feature selection. The authors of this research paper points out that these kinds of NLP project require long process and time consuming processing procedure. Similar to this, my research uses SVM as the main classifier and Google News Word2Vec embeddings which has like 300 dimensional vectors to improve feature representation in order to know about the complete essence of the sentiment in news headlines prior to the 2024 U.S. elections. Lastly, research by Susanti et al. [13] and Arora et al. [12]demonstrated novel sentiment classification methods like the Multinomial Naive Bayes Tree for textual frequency computations and Cross BOMEST for cross-domain analysis. Even though these are useful methods which can be used to integrate in our research but the domain specificity limited their wider applicability, despite their notable accuracy. My research focuses to overcome these constraints by improving the contextual comprehension and generalizability in headline sentiment analysis through the integration of Word2Vec embeddings with the tf-idf in order to achieve the better understanding of the essence of the headlines.

2.1 BERT AND RoBERTa based model:

BERT AND RoBERTa were used by Nemkul [20] for Nepali news classfication, They demonstrated with their work that how well these tranformer based models can handle the liguistic difficulties and provided insightful information how to fit the previously trained models with the dataset of their own, demonstrated the work of versatility in low resource and multilingual environments. Our effort, which drew inspiration from this study, also used transformer-based models to tackle the difficulties presented by the complex and context-rich nature of political news headlines.

Jiao and zhao[21] experimentation of the collecting the real world news event extraction of using BERT's was highlighted beacause of the model's capacity to handle and recognize the contextual linkages and effectively extract important information. With the idea of this integration it helped us to integrate the (NER) with the contextual comprehension from their work. Our work shows similarities to this method in extracting sentiment-related insights from news articles by using tools such as spaCy for NER and incorporating embeddings from transformer-based models.

Mengi, Ghorpade, and Kakade [22] focused on fine-tuning RoBERTa for sentiment analysis and text summarization, showcasing its superior performance in capturing intricate sentiments and summarizing textual content. Their work reinforced the importance of RoBERTa's optimized pretraining for tasks requiring a high degree of contextual understanding. Inspired by this, our research implemented RoBERTa to explore its potential in enhancing the accuracy of sentiment classification, even with complex and sentiment-heavy political headlines.

3 Research Methodology

The structure of this research employs an methodology which is divided into three key processes such as data collection and preprocessing, model building and feature engineering, and model evaluation and optimization in which all the section of the methodology has been rigorously tried working with other possible methodologies and chosen the best of the methods at each stage to ensure robust implementation.

3.1 Data Pre-processing and Model Building

Since all the drama, election-related activities, and public speaking take place during this time, the first stage is to compile news items on the U.S. elections from credible sources during a two-month period beginning on September 5 and ending on November 5. In order to create an comprehensive news channel and article published are notably collected with most relevant to the elections, the headlines are gathered every day from credible sources such as CNN, Fox News, BBC, Al Jazeera, Reuters, and The Guardian. While further pre-processing of the dataset removal of components like stopwords, removal of special characters, the HTML tags while scrapping or any symbols that comes with headlines which doesn't play any major role in deciding the sentiment factors are removed.

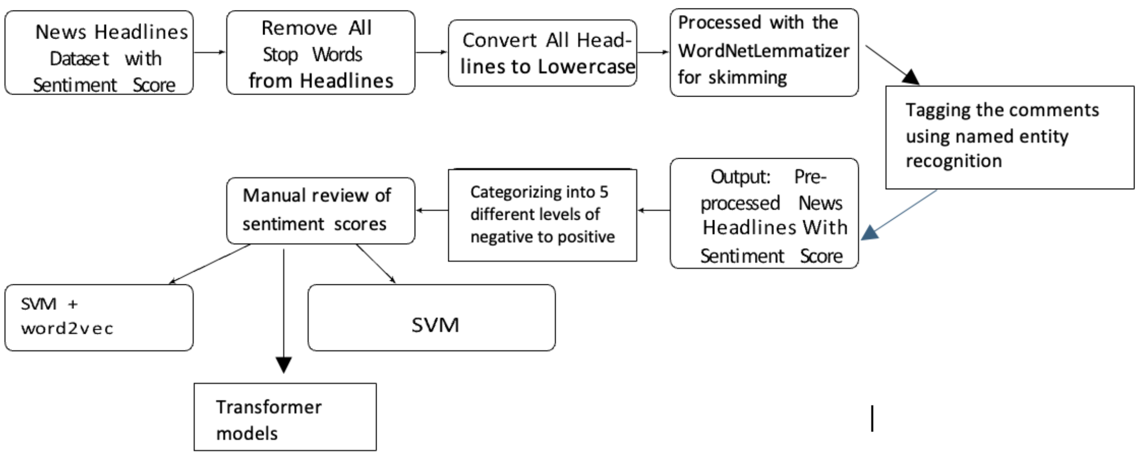

Figure 1: Flow diagram of data pre-processing and model building

3.1.1 Data Pre-processing

To remove the unwanted components which was mentioned before that might impede the effective analysis, since the stopwords like"and,""the,"and "or" dont add any value in predicting the sentiment of the headlines neither doesn't produce any emotion over there. In In order to prevent inconsistent word representation (such as treating "Trump" and "trump" as separate words), the dataset was normalized by converting all text to lowercase. By reducing words to their most basic forms (such as "running" to "run"), lemmatization helped reduce dimensionality without sacrificing semantic meaning.

Normalization: This step involves converting all text to lowercase to avoid treating the same word in different cases (e.g., "Trump" and "trump") as distinct entities.

Named Entity Recognition (NER): To identify and tag persons mentioned in the headlines, we use Named Entity Recognition (NER). This is an another crucial step involved in the process as this helps us to understand the collected headlines refers about who(its either the trump or harris or any other contestants competing in the election).This step is crucial for understanding the entities (such as Trump and Kamala Harris) associated with each headline. For achieving this entity tagging a customer entity ruler known as spaCy has been very effective and also it ensured accuracy in labelling.After the better understanding of the working principle of the NER was selected after considering over a simple rule based approach because it enhances contextual understanding. While stemming was considered for dimensionality reduction, it was not employed due to its tendency to distort meanings.

Text Vectorization: The stage after the cleaning of the text is done, the next step to follow is to convert the text into a numerical format so that it can be understood by the machine learning model. Initially, TF-IDF (Term Frequency-Inverse Document Frequency) is applied, which helps in assessing and weighting the terms based on their importance across the dataset. In subsequent steps, more advanced techniques such as Word2Vec embeddings may be explored for better text representation. Alternatives, like GloVe and FastText were also considered but that methods were less optimal when it was compared in capturing the nuanced contextual relationships in news headlines.

5 Level Sentiment Classification: Using an 5-level sentiment classification to capture the nuanced sentiment variations present in news headlines, which a simpler binary or 3-class system might overlook. The dataset distribution shows a balanced spread, with negative (27.37),positive (22.81), and neutral (36.50) sentiments each forming a significant portion, ensuring that no single category dominates the dataset. The inclusion of very negative (6.39) and very positive (5.47) categories allows for a more granular sentiment representation, particularly in political discourse, where tone can range from strongly critical to highly supportive. This refined classification improves sentiment differentiation, leading to better model performance and more insightful election sentiment analysis. There were total 547 preprocessed headlines collected from the different news platforms well it has well balanced with the negative, positive, neutral were the majority while very negative and very positive category would provide an additional granularity This classification ensures a nuanced understanding of sentiment trends.

3.2 Model Building

Feature Engineering: Additional features from the headlines are extracted in order to improve the predictive power of the model and also there are many linguistic features to be kept in mind like adding up the sentiment scores for the headlines, parts-of-speech tags and the syntactic structures, which in return gives out the best insights when we add the model for the prediction of sentiment of the headlines.

Model Selection and Training: The next step of the process involves in selecting most appropriate machine learning algorithm for classification tasks. As discussed in the related works section after going through many previous papers we have found out that Support Vector Machine(SVM) is used for its effectiveness in handling high dimensional data and its ability to find out the optimal hyperplane makes it best for the classification tasks. The labelled dataset, which comprises both positive and negative sentiment categories, is used to train the model. Cross-validation is used to adjust different hyperparameters during training to make sure the model performs well when applied to new data. Following training, the model's performance is assessed using measures such as F1-score, recall, accuracy, and precision. The results produced from the models are examined in order to pinpoint the areas the require improvement, such as overfitting or underfitting of the model, which may affect its capacity for accurate forecasting opinions. In order to classify the sentiment into more granular categories (such as very negative, neutral, and very positive), additional testing and evaluation may be conducted for fine-grained sentiment analysis.

3.3 Optimization:

To increase the accuracy and precision of sentiment classification, extra measures like hyperparameter tuning, the use of sophisticated embeddings like Google News Word2Vec, or model ensembling techniques are investigated if the initial performance is inadequate. This methodology ensures that the data is cleaned, transformed, and modelled effectively, and that the final sentiment analysis model is both accurate and reliable. The steps outlined above are crucial for improving the classification of sentiments in newspaper headlines, particularly for election-related news, which is often complex and nuanced.

Methodology Justification: The selected methodology combines conventional and cutting-edge techniques to guarantee thorough analysis. Due to their inability to capture the intricate semantics of election-related headlines, alternative approaches—such as more straightforward rule-based NER techniques and other embeddings like FastText—were contemplated but never put into practice. This followed methodology has capacity and also efficient in handling the sentiment analysis challenges which was ensured in the section of the pre processing, feature engineering and model optimization.This method works well with both small and large datasets because it strikes a balance between creativity and pragmatism.

4 Design Specification

4.1 Deep Learning Model for Election Prediction

The deep learning models started with an hectic process of collecting the news headlines from multiple news publishing sources from USA because we cannot lose its originality

of its being produced to the people of USA for that we collected the news from multiple credible news sources, such as CNN, Fox News, BBC, Al Jazeera, Hindustan, ABC, Reuters, The New York Times, Forbes, The Guardian, and others. The gathered data undergoes preprocessing, including lemmatization, stopwords removal, and the elimination of small or noisy text fragments. The cleaned text is then vectorized using TF-IDF or advanced embeddings like Google News Word2Vec 300. For sentiment classification, a variety of machine learning and deep learning models are employed:
• Traditional Models: Support Vector Machine (SVM) with multiple kernels, XGBoost, and Random Forest.
• Transformer Models: SVM integrated with BERT and RoBERTa. Before giving out the headlines to the model, the headlines sentiment scores ranging from 1 (bad negative) to 5 (good positive) are analysed manually and passed on to the models. These models output a sentiment score ranging from 1 (bad negative) to 5 (good positive) for each headline. For assuring the nuanced understanding and accurate predictions fine grained analysis is performed .

4.2 Sentiment Classification Process

The sentiment classification process involves manual categorization of news headlines into five levels:
• 1: Bad Negative
• 2: Negative
• 3: Neutral
• 4: Positive
• 5: Good Positive
The robust ground truth for the models to learn from will be provided by classifying the headlines into the elaborated version and manually verifying after the labelling is complete. After the labelling of the headlines, the cleaned and categorized data is fed into the machine learning pipeline with transformed headlines and the categorized sentiment score as the target variable. include:
1.Preprocessing: Headlines are tokenized, lemmatized, and vectorized using either TF-IDF or Word2Vec embeddings.
2. Model Training: The dataset has been experimented with no of the models such as SVM, XGBoost, and Random Forest with that of the labelled dataset. Transformers like BERT and RoBERT are also finetuned for the sentiment classification with multiple epoch runs.
3. Performance Metrics: As we were evaluating the models' accuracy, precision, recall, and f1 score, we continuously tracked their performance to review and adjust the metrics to enhance the outcomes.
4. Prediction Output: Since each headline was given a sentiment score, this would characterize the tone of the headlines indicating which party is in control, evaluate trends, and forecast which party will win elections.

4.3 Visualization and Insights

Even after going through all the results and performance metrics from the model better understanding is produced when there is better visualization for all the work, so for all the following classification, sentiment distribution trends and their possible influence on

election results are examined using visualization tools. The visualizations are performed with that of the pyPlot and wordcloud etc, but the visualization for the transformer models are performed with the external visualization which is accessed through the Api key and run over that platform wandb (Weights and Biases)and the insights and trends are shown in the report.
• Trends in sentiment towards an party or the candidate over the time are shown. Analysing sentiment across many news sources in comparison.
• Patterns of misclassification and prediction confidence levels. Technologies Tools Used
• Libraries: scikit-learn, NLTK, spaCy (encore websm), SentimentIntensityAnalyzer
• Visualization: wandb with API key for tracking and visualization
• Models: SVM with multiple kernels, XGBoost, Random Forest, BERT, RoBERTa

5 Implementation

5.1 Data Collection

Newspaper headlines are collected with many of the trustworthy sources such as CNN, Fox, Hindustan Times, BBC, ABC, Al Jazeera, The New York Times, Reuters, The Guardian, Forbes, and Washington News which was gathered in the initial data collection phase. The collection of the news headlines was collected in the dates between September 5 to November 5,2024 which was precisely 2 months before the voting poll date which was on November 5,2024. At the initial stages tried collecting it using the News Api, Google news API which generated very less news and failed to collect the targeted news which was focused towards the election and election candidates trump and Harris. Hence the data was collected completely manual from various news sources which is an hectic process which involved many hours of work. This led to the creation of a dataset with few hundreds of headlines that was intended to convey the mood around the election and its main figures.

5.2 Data Preprocessing

After the headlines was collected with appropriate links to the websites containing the news and the news channels names published that news. The gathered news headlines underwent a thorough preparation step. As part of the preprocessing stages, all text was converted to lowercase and special characters, unwanted connecting words were eliminated. SpaCy and NLTK were used for lemmatization, tokenization, and stopword elimination. Named entity recognition (NER) was used to tag important entities like "Donald Trump," "Joe Biden, "kamala Harris," and party names after headlines with fewer than three characters were filtered out. The themes and stakeholders in the headlines were identified with the aid of this procedure and multiple example of the three words has mentioned in the test cases In order to guarantee a high-quality labelled dataset for modelling, sentiment labels were lastly manually assigned on a scale from 1 (poor negative) to 5 (good positive) for supervised learning.

5.3 Feature Engineering

To accurately represent the richness of the textual material, critically assessed feature engineering approaches were used. As an first step, the relevance of each word in the headlines was measured using TF-IDF vectorization. In order to capture semantic linkages, pre-trained word embeddings were also incorporated, such as Google News Word2Vec which has 300 dimensions which would make the words more meaningful when read as sentence and with their capacity to comprehend the context and subtleties of the text, contextual embeddings from BERT and RoBERTa were also used. Features that provided information about the meaning of particular words, occasions, and people were named entities and topic modelling outputs. Strong inputs for the machine learning models were guaranteed by these varied feature representations.

5.4 Model Development

To accurately forecast sentiment scores, the research used a variety of models. The baseline was a Support Vector Machine (SVM) with three different kernels: linear, RBF, and polynomial. For contrast, more sophisticated algorithms like Random Forest and XGBoost were also created. SVM classifiers were fed embeddings produced by the BERT and RoBERTa transformers in order to take advantage of deep learning. The implementation was made easier by libraries like scikit-learn, HuggingFace Transformers, and spaCy. Eventhough we have acheived an accuracy score of 54

5.5 Model Evaluation

In order to deal with the nuances of sentiments displayed in the headlines we have used the SVM's which have an carved niche due to their ability to handle an high dimensional spaces effectively, making them exceptionally suited for the text categorization and also we have researched using multiple variety of models including SVM with different kernels: linear, RBF, and polynomial. For the comparison and clear explanation, more sophisticated algorithms like Random Forest and XGBoost were also created. SVM classifiers were also fed to the embeddings produced by the BERT and RoBERTa transformers in order to take advantage of deep learning. The whole implementation was made easier with using the libraries like scikit-learn, HuggingFace Transformers, and spaCy. Even though we have achieved an accuracy score of 54

5.6 insights and Analysis

Further analysis has been carried out in order to understand further from the modelling part which was to uncover the sentiment patterns within the dataset. Whereas the conservative sites spread more positive attitudes, leftist sources headlines were primarily critical of the republican candidates. Since the political stances of the source often matched with the opinions, media's impact of the people about the politics were made clear in the headlines and showed how the sentiment trends might be used to predict election results.

6 Evaluation

RQ1: How do sentiment trends in newspaper headlines leading up to the 2024 US election influence public opinion and voting behavior?

Spikes and Dips in Sentiment: As shown in fig.2 clearly there was clear record of the good level of sentiment trends are visible for the kamala Harris throughout the

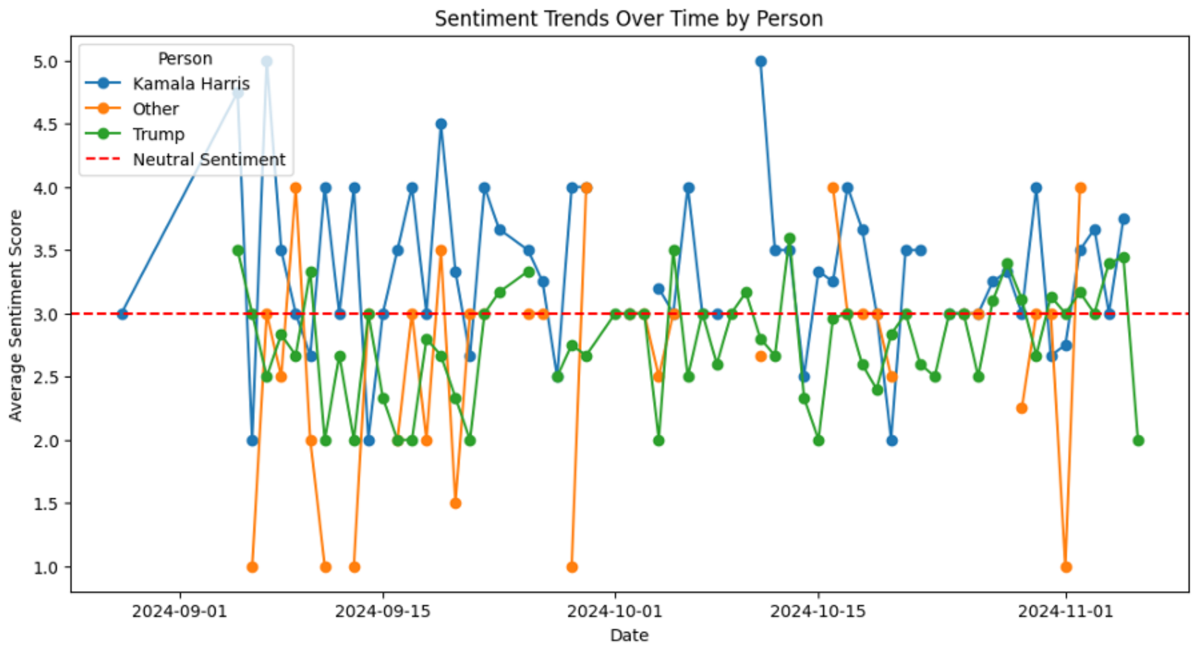

Figure 2: Sentiment Trends over time by person

election campaigns as most of the spikes are seen above the neutral level which is in blue clearly denotes that Harris has been favorites in the newspaper articles with that of the headlines published. As we can see at the beginning of the campaigns ranging in between the timeline of the first 1 month Harris has been above the neutral because of her successful debate in an face off against the trump which was conducted by the ABC News which was aired on September 11, where she proved her worthiness of an being president candidate as she gained trust from the people and heads turned towards her speech and where trump failed to demonstrate his suitability. In the Fig can see lot of green spikes below the neutral which earned trump lot of negative comments, this was one impactful scenario in the campaign period.

Even though trump an candidate of the republican party failed to achieve his popularity with news articles but we can see his involvement in the election with the continuous green line which denotes he never misses to be in the headlines let it be in the positive or negative we can see the continuous record of him being the in the headlines of the news which made people believe that he is an tough contender.

Comparative Trends: After seeing displayed trend comparison in the fig2, we can understand that it is evident that Donald Trump maintains the stable sentiment trend throughout the campaign period which we can see in the chart.Although this stability often fluctuates close to or below the neutral sentiment level.The elements of Trump's strategy of consistently appearing in news articles, albeit with controversial opinions, are made clear by this analysis.As evidenced by the chart's numerous dips below the neutral line, Trump's emotions patterns indicate that, despite his continued prominence in media conversations, this coverage frequently tended toward controversy or criticism.

Whereas, on the other hand Kamala Harris shows greater sentimental volatility. Even if her sentiment scores show some peaks and dips, the majority of the surges tend to be above the neutral threshold, especially during significant events like her victory over

Trump in the September 11 debate. This variation suggests that, in contrast to Trump's more consistent (but unfavorable) coverage, Harris's campaign generated a greater emotional response, both good and negative.

The highlights of the stories from the various media sources are shown in the comparative trend pattern that is shown in Figure 2. While Harri's varying but largely positive sentiment suggests dynamic and that made an impactful coverage which shows the public Harri's clear idea, Trump's consistent but less favorable campaigns demonstrate how strongly he is involved with the media scrutiny and divisive opinions about his candidacy.

RQ2: Can advanced NLP techniques, such as BERT and RoBERTa, significantly improve sentiment analysis accuracy compared to traditional methods (e.g., TF-IDF and SVM) in the context of political news?

In this research question we have applied multiple traditional models not only the SVM or Deep learning models which includes the transformer models such as BERT or RoBERTa, but also experimented with many other contrast models in order to see its performance with our dataset, the following results are produced by the models in the sentiment analysis of the news headlines such as Support Vector Machines(SVM), Logistic Regression, Random Forest, and transformer based models like BERT and RoBERTa.

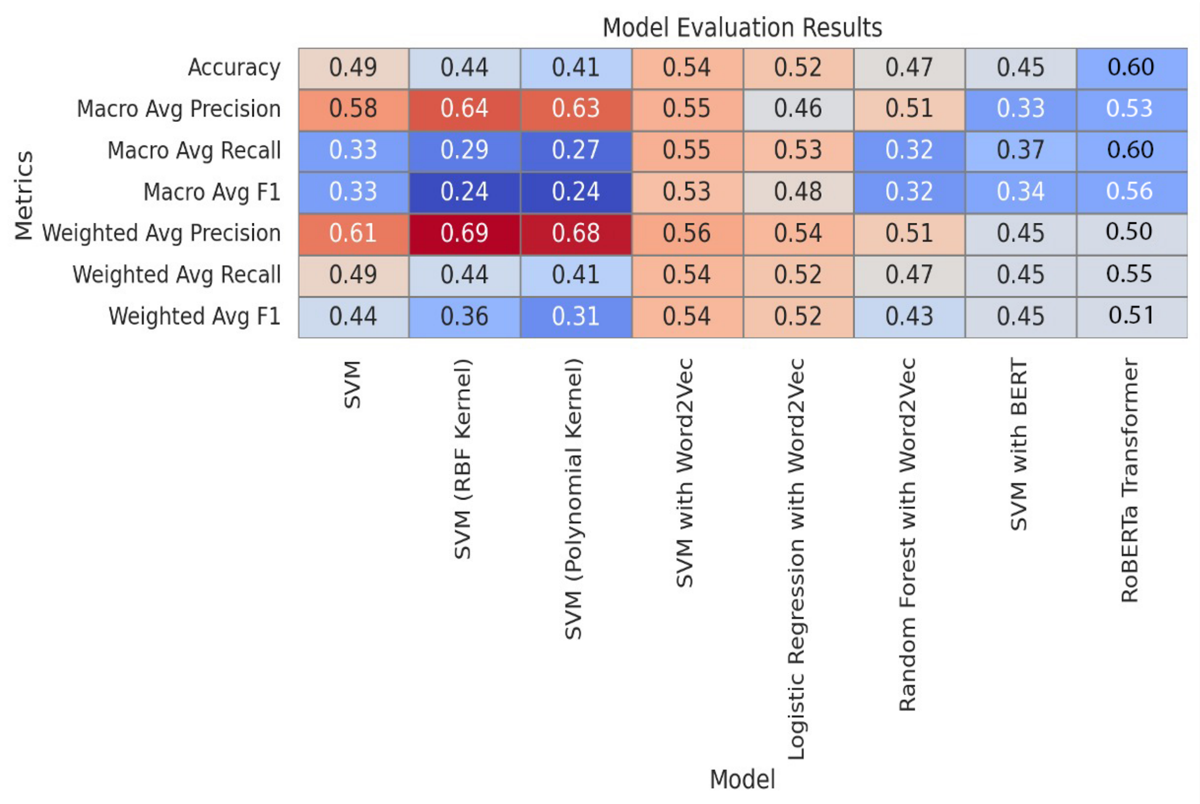

Figure 3: All Models Evaluation Results

1. SVM Model

• Accuracy: 49

• Macro Average: Precision = 0.58, Recall = 0.33, F1-Score = 0.33
• Weighted Average: Precision = 0.61, Recall = 0.49, F1-Score = 0.44

The research started with using the SVM with the an linear kernel which offers an baseline performance with moderate precision, but the recall value and F1 scores are significantly lower, that denotes that the model is struggling with predicting the both positive and negative classes effectively.

2. SVM with RBF Kernel

Accuracy: 44
• Macro Average: Precision = 0.64, Recall = 0.29, F1-Score = 0.27
• Weighted Average: Precision = 0.69, Recall = 0.44, F1-Score = 0.36

After the fall of precision we moved to the SVM with the Radial Basis Function(RBF) which performed worse than the previous linear kernel, we can see an notable decrease in the accuracy. While only the precision value was improved, recall and F1 scores dropped further, suggesting the SVM with RBF kernel cannot able to generalize well with the data.

3. SVM with Polynomial Kernel

• Accuracy: 41
• Macro Average: Precision = 0.63, Recall = 0.27, F1-Score = 0.24
• Weighted Average: Precision = 0.68, Recall = 0.41, F1-Score = 0.31 Then tried with using an another polynomial kernel which once again showed an poor performance in terms of accuracy and recall. The precision of the model is higher than the recall, but this indicates that the models is not detecting all the relevant instances, which in return resulting in an higher false-negative rate.

4. SVM with Word2Vec Embeddings

• Accuracy: 54
• Macro Average: Precision = 0.55, Recall = 0.55, F1-Score = 0.53
• Weighted Average: Precision = 0.56, Recall = 0.54, F1-Score = 0.54 After the installation of the 1.6gb file which contains of google 300 dimensions, the Word2Vec embeddings combined with SVM was applied which led to an increase in the accuracy of the predicting model and balance in the precision, recall and F1-score. This combination gives an clear understanding of how the inclusion of the pre-trained word embeddings from Word2Vec helps capture the semantic relationships, which could lead to better feature representation and generalization.

5. Logistic Regression with Word2Vec

• Accuracy: 52
• Macro Average: Precision = 0.46, Recall = 0.53, F1-Score = 0.48
• Weighted Average: Precision = 0.54, Recall = 0.52, F1-Score = 0.52 After the integration of the SVM with word2Vec we tried using the logistic regression with the Word2vec embeddings which resulted moderately well, though the recall is slightly better than the precision. The overall F1-score indicates there is some room for improvement, particularly in the precision- recall trade-off.

6. Random Forest with Word2Vec

• Accuracy: 47
• Macro Average: Precision = 0.51, Recall = 0.32, F1-Score = 0.32
• Weighted Average: Precision = 0.51, Recall = 0.47, F1-Score = 0.43 For the surprise that logistic regression performed moderately well with the logistic regression we tried integrating Random forest with the Word2Vec embeddings which did not perform well, showing poor recall and F1-scores, which gives an understanding of the model struggles

with correctly identifying positive and negative sentiment in the dataset.

7. SVM with BERT Embeddings

• Accuracy: 45

• Macro Average: Precision = 0.33, Recall = 0.37, F1-Score = 0.34

• Weighted Average: Precision = 0.45, Recall = 0.45, F1-Score = 0.45 After integrating the data with the models we tried dealing the data with the transformer pipelines in order see the maximum potential and make use of the deep learning concepts hence we used it with the BERT embeddings with SVM which further did not improve in performance over the results produced by the models of SVM. Precision and recall are still suboptimal, and F1 score further suggests that model struggled with classifying the sentiments correctly. The use of BERT embeddings with SVM did not significantly improve performance over other SVM models. Precision and recall are still suboptimal, and the F1-score further suggests that the model struggled with classifying the sentiment correctly.

8. RoBERTa Transformer Model (5 Epochs)

Epoch 1:

•Training Loss: 1.62, Validation Loss: 1.26,

•Accuracy: 51.81

•Precision: 0.43, Recall: 0.52, F1-Score: 0.43

Epoch 2:

•Training Loss: 1.52, Validation Loss: 1.21,

•Accuracy: 52.73

•Precision: 0.36, Recall: 0.53, F1-Score: 0.43

Epoch 3:

•Training Loss: 1.43, Validation Loss: 1.19,

•Accuracy: 52.73

•Precision: 0.36, Recall: 0.53, F1-Score: 0.43

Epoch 4:

•Training Loss: 1.13, Validation Loss: 1.10,

•Accuracy: 60.91

•Precision: 0.53, Recall: 0.61, F1-Score: 0.57

Epoch 5:

•Training Loss: 1.13, Validation Loss: 1.17,

•Accuracy: 55.45

•Precision: 0.51, Recall: 0.55, F1-Score: 0.52

The results produced from the RoBERTa model demonstrated an steady reduction in training loss when its run over the epochs and achieved an highest accuracy of over 60.91 percent in its fourth epoch run. There was also an notable improvement in the F1-score at this stage which indicates an better balance between the precision and recall of the model.The performance dropped during the 5-epoch run, which is the only negative aspect. This likely indicates overfitting or the need for additional hyperparameter fine-tuning.

The above evaluation highlights the experimental results obtained from various machine learning models for sentiment analysis of newspaper headlines related to the 2024 U.S. election. The results reveal incremental improvements in the RoBERTa model's performance compared to earlier experiments, but challenges remain in achieving consistently high scores across all metrics. These findings will be analyzed critically to identify areas of strength, weaknesses, and opportunities for further improvement in sentiment classification models. From both practical and academic perspectives, these insights offer

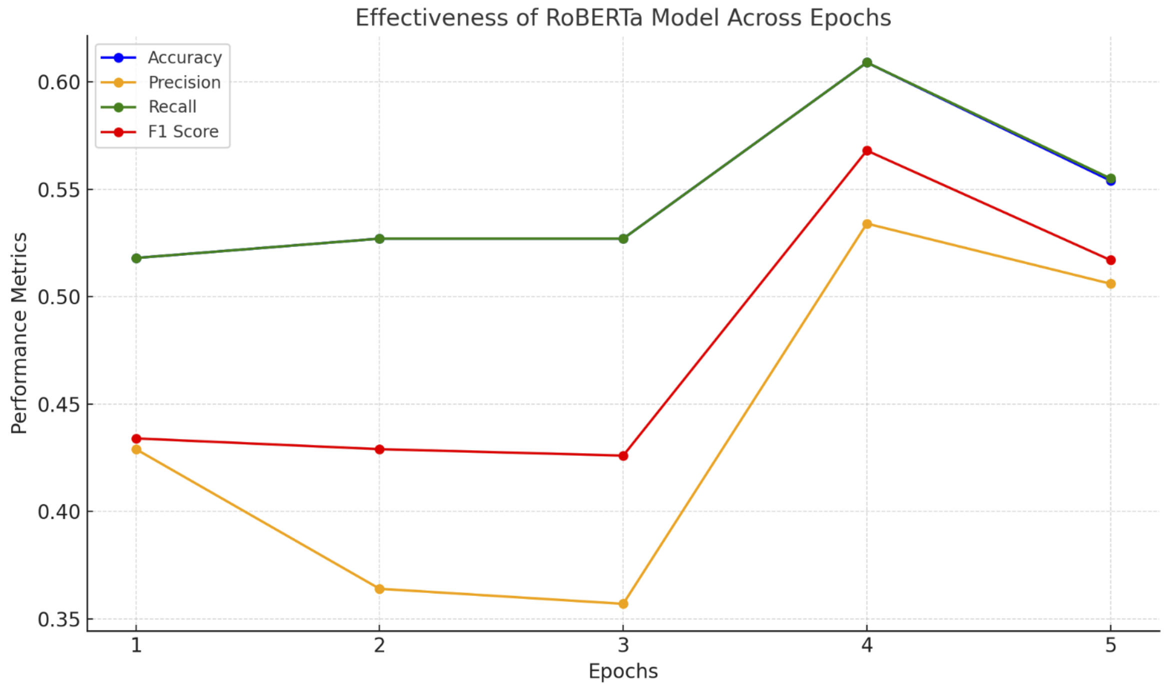

Figure 4: Effectiveness of Roberta Model

guidance on refining models to better address the research question.

Model Performance Evaluation

This research assessment of several models gave thorough grasp of the sentiment analysis of the capabilities for the news paper headlines about the 2024 US election. In spite of experimenting with multiple models which includes deep learning models, the SVM with Word2Vec, was the best performing classical model with an accuracy of 54. This proves that using the pre trained word embeddings would greatly increase the model's capacity to capture the semantic links in the text and which further increase the feature representation. Using accuracies ranging from 41 to 49 and lower recall values, SVM using linear, RBF, and polynomial kernels fared poorly in comparison, indicating that these kernels had trouble efficiently generalizing across classes.

Though its precision-recall trade-off was less advantageous than SVM with Word2Vec, Logistic Regression with Word2Vec nevertheless demonstrated respectable performance (accuracy: 52). When it comes to handling the complex sentiment of the distribution in the dataset once again it is proved that Random Forest performs poorly even associated with word2vec models which achieved an accuracy of only 47.

Among the transformer based models, RoBERTa showed potential for improvement in deep learning approaches, but the results weren't up to the mark because the deep learning models which has the potential to deal with the larger datasets which is spread over all the criteria. The RoBERTa gave us an accuracy of 60.91 percentage with the total of 5 epoch gave an good result at the 4 epoch after three epochs of run. Even after the slight reduction in the training loss, the model struggled to generalize due to limited training data and insufficient fine- tuning. Similarly, SVM with BERT embeddings underperformed (accuracy: 45), emphasizing the need for advanced fine-tuning strategies and larger labelled datasets to fully exploit the power of transformer-based models.

In conclusion, while the traditional practises for the NLP like SVM with Word2Vec gave a robust performance, where as the advanced transformer like BERT and RoBERTAa with further optimization and more training data would surpass the traditional methods in this domain. The above mentions findings addresses the RQ2, demonstrating that while advanced NLP models have potential, traditional models remain competitive for sentiment analysis in resource-constrained settings.

RQ3: What is the impact of named entity recognition NER on the performance of sentiment analysis models for political news?
The inclusion of the Named Entity Recognition(NER) played an important role in our project which includes improvement of the results produced from the models as prediction has possible with the help of the (NER) Initially, the sentiment analysis models struggled to determine the specific subject or target of a sentiment in a headline. This lack of specificity reduced the model's effectiveness in predicting sentiment trends for political candidates. After integrating NER using the encore websm model from spaCy, the research could accurately identify entities such as "Donald Trump," "Kamala Harris," and other relevant political figures. By combining NER with the Sentiment Intensity Analyzer (SIA), each headline was associated with a sentiment score linked to its tagged entities. This enabled us to:
1.Isolate Entity-Specific Sentiments: Headlines were no longer generalized instead, sentiment was tied to specific political figures, making the analysis more precise.
2.Identify Entity-Specific Sentiment Trends: As it is shown in the fig.5 we can track how the sentiment trends for the candidates of the US election over time.

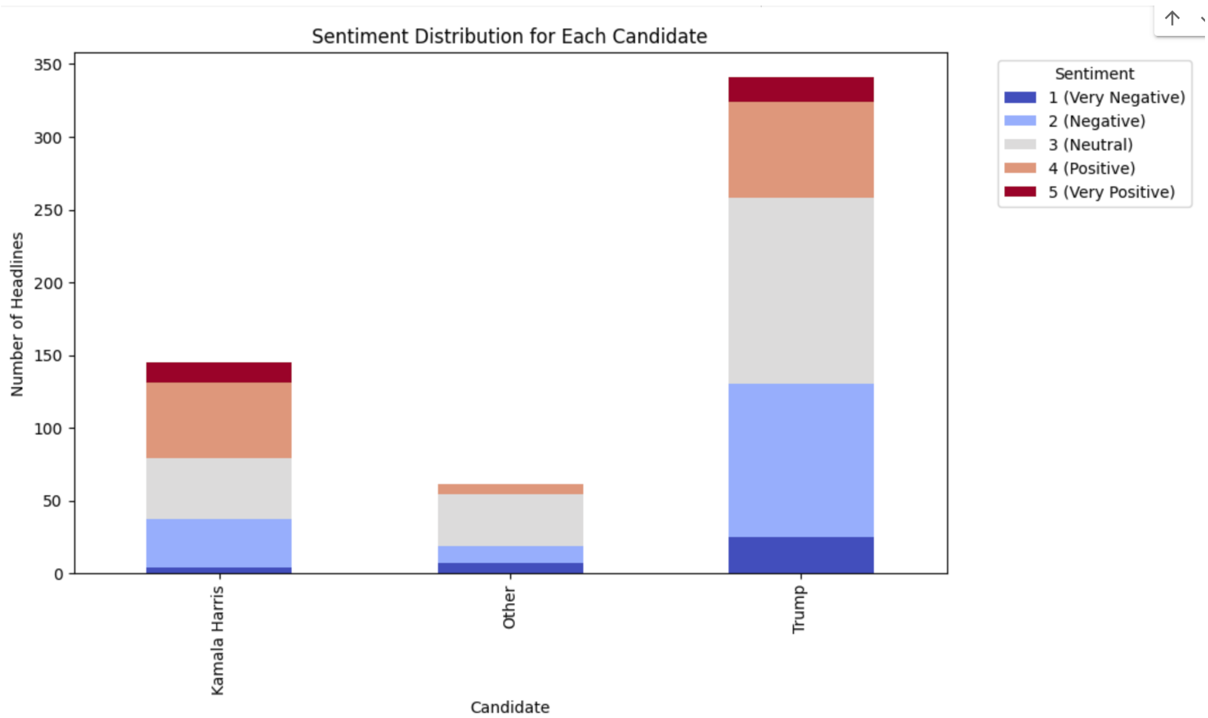

Figure 5: Sentiment Distribution of Each candidate

7 Conclusion and Future Work

Even though this research has advanced significantly, there are still a number of areas that could be improved in the future.First, adding news articles from different sources was enough to study the news headlines, but a bigger dataset would greatly increase the models' generalizability and robustness. Further integration of additional sources of the headlines from various sources such as the social media, blogs and online news archives. Further increasing the time period of collection of the news articles for more number of months before the polling day wouldn't let any important actions take over this time frame. Further fine graining of sentiment intensity not only(e.g., as very positive , positive, neutral, negative, very negative) but also finding the emotion behind headlines such as anger, fear, joy, sadness would an extra feature and would bring more insights towards handling the tone behind the headlines. Involvement of XAI techniques to understand the reasoning behind the model predictions is an way of achieving the advanced modelling.The research on which phrases have the greatest influence on the tone of headlines would benefit from this feature development.

This research aimed to predict the winning party in the 2024 U.S. election through sentiment analysis of newspaper headlines. The research successfully achieved in collecting, preprocessing and analysation of the different kinds of data from reputable news article sources, while achieving an moderate evaluation accuracy of 60.9 percent at its peak with RoBERTa. This paved the way for the partial achievement in the way of leading to ultimate predictive goal. However the overall performance ,including precision and F1-score are assessed in the each stage and fine tuned to bring the best out of the model's performed and also this research laid an strong foundation in sentiment analysis, everaging state-of-the-art models and advanced visualization tools. I believe that this research laid an groundwork for future enhancements to achieve higher predictive accuracy and deeper insights.

8 References

[1].F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer,R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay,"Scikit-learn: Machine learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[2]. C. D. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. J. Bethard, and D. McClosky, "The Stanford CoreNLPnatural language processing toolkit," in Association for Computational Linguistics (ACL) System Demonstrations, 2014, pp. 55–60. [Online]. Available: http://www.aclweb.org/anthology/P/P14/P14-5010
[3]. D. Dor, "On newspaper headlines as relevance optimizers," Journal of Pragmatics, vol. 35, no. 5, pp. 695– 721, 2003. [Online].
[4]. A. Agarwal, V. Sharma, G. Sikka, and R. Dhir, "Opinion mining of news headlines using sentiwordnet,"in Colossal Data Analysis and Networking (CDAN), Symposium on. IEEE, 2016, pp. 1–5.
[5]. K. Yang, Y. Cai, D. Huang, J. Li, Z. Zhou, and X. Lei, "An effective hybrid model for opinion mining and sentiment analysis," in Big Data and Smart Computing (BigComp), 2017 IEEE International Conference on. IEEE, 2017, pp. 465–466.
[6].Jemai, F.; Hayouni, M.; Baccar, S. Sentiment Analysis Using Machine Learning Al-

gorithms. In Proceedings of the 2021 International Wireless Communications and Mobile Computing (IWCMC), Harbin, China, 28 June–2 July 2020; pp. 775–779.
[7]. S. Rana and A. Singh, "Comparative analysis of sentiment orientation using svm and naive bayes techniques," in Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on.IEEE, 2016, pp. 106–111.
[8].R. K. Bakshi, N. Kaur, R. Kaur, and G. Kaur, "Opinion mining and sentiment analysis," in Computing for Sustainable Global Development (INDIACom), 2016 3rd International Conference on. IEEE, 2016, pp.452–455.
[9].S. Aroju, V. Bhargavi, S. Namburi et al., "Opinion mining on Indian newspaper quotations," in Signal Processing and Integrated Networks (SPIN), 2016 3rd International Conference on. IEEE, 2016, pp.749–752. s
[11].Makhmudah, U.; Bukhori, S.; Putra, J.A.; Yudha, B.A.B. Sentiment Analysis Of Indonesian Homosexual Tweets Using Support Vector Machine Method. In Proceedings of the 2019 International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), Jember, Indonesia, 16–17 October 2019; pp. 183–186.
[12]. P. Arora, D. Virmani, and P. Kulkarni, "An approach for big data to evolve the auspicious information from cross-domains," International Journal of Electrical and Computer Engineering (IJECE), vol. 7,no. 2, pp. 967–974, 2017.
[13]. R. Susanti, T. Djatna, and W. A. Kusuma, "Twitter's sentiment analysis on gsm services using multi-nominal naive bayes." Telkomnika, vol. 15, no. 3, 2017.
[14]. M F Rozi1, I Mukhlash1, Soetrisno1, and M Kimura, Opinion mining on book review using CNN-L2-SVMalgorithm
[15]. . Gao, H. Xu, and J. Wang, "A rule-based approach to emotion cause detection for chinese micro-blogs,"Expert Systems with Applications, vol. 42, no. 9, pp. 4517–4528, 2015.
[16]. [D. Chen and C. Manning, "A fast and accurate dependency parser using neural networks," in Proceedingsof the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 740–75
[17]. Jotheeswaran and Y. Kumaraswamy, "Opinion mining using decision tree based feature selectionthrough manhattan hierarchical cluster measure." Journal of Theoretical Applied Information Technology, vol. 58, no. 1, 2013
[18]. H. Akkineni, P. Lakshmi, and B. V. Babu, "Online crowds opinion-mining it to analyze current trend:A review," International Journal of Electrical and Computer Engineering (IJECE), vol. 5, no. 5, pp.1180–1187, 2015.
[19]. Sohangir, S., Wang, D., Pomeranets, A., Khoshgoftaar, T. M. (2018). Big data: deep learn-ing for financial sentiment analysis. Journal of Big Data, 5(1), 3. https://doi.org/10.1186/s40537-017-0111-6.20.
[20] Kriti Nemkul.(2024), Tribhuvan University Journal39(1):124-137 Use of Bidirectional Encoder Representations from Transformers (BERT) and Robustly Optimized Bert Pretraining Approach (RoBERTa) for Nepali News Classification
[21]. Yuxin Jiao, Li Zhao(2024), Real-Time Extraction of News Events Based on BERT Model, Volume 09, No.03, 2024.
[22]. Ramazan Mengi, Hritik Ghorpade, Arjun Kakade(2023), Fine-tuning T5 and RoBERTa Models for Enhanced Text Summarization and Sentiment Analysis.