Sentimental Analysis on Political Speeches

by

Asif Atiq
18101556
Abrar Shahriar Abeed
18101257
Azher Ahmed Efat
18101027
Armanul Momin
17101281

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
January 2022

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Asif

Asif Atiq

18101556

Abrar Shahriar Abeed

18101257

Azher Ahmed

Azher Ahmed Efat

18101027

Amomin

Armanul Momin

17101281

Approval

The thesis titled "Sentimental Analysis on Political Speeches" submitted by

1. Asif Atiq (18101556)
2. Abrar Shahriar Abeed (18101257)
3. Azher Ahmed Efat (18101027)
4. Armanul Momin (17101281)

Of Fall, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on January, 2022.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam, Phd

Associate Professor
Department of Computer Science and Engineering
BRAC University

Program Coordinator:
(Member)

Md. Golam Rabiul Alam, Phd

Associate Professor
Department of Computer Science and Engineering
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD

Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Politics is an essential part of human society. From the start of human civilization, politics has controlled every human society. Political speeches have had one of the most influential roles in shaping the world. Speeches of the written variety have been etched in history. These sorts of speeches have a great effect on the general people and their actions in the coming few days. With advancing technologies, people from all across the world get to listen to these speeches hence the impact on the listener is increasing on a global scale. We analyzed the performance of different models on our corpus of speeches using sentiment and context analysis and then we compared the results of those models to see the difficulty in analyzing sentiment and context of speeches of country leaders. In our research we have focused on the presidents/prime ministers of the five permanent members of the United Nations Security Council which are France, China, Russia, United Kingdom and United States. Moreover, if left unchecked, a political personnel or party may cause major problems. In many cases there may be a warning sign that the government needs to change their policies and also listen to the people. By classifying the speeches into positive, negative or neutral categories in terms of sentiment and five context categories international, nationalism, development, extremism and others and evaluated the accuracy of our models. By using approaches such as Longformer (RoBERTa based model), TF-IDF with ensemble learning models and LDA topic modeling along with ensemble learning models, we were able to achieve some satisfactory results. We have used a modified Bidirectional Encoder Representations from Transformers (BERT) algorithm which is Longformer and TF-IDF with ensemble learning models for sentiment analysis and an LDA based topic model implemented on ensemble learning models to analyze our speeches for context analysis. We have achieved a 0.67 score on the accuracy of Sentiment and we also achieved a 0.67 accuracy on contexts.

Keywords: Political Speeches, Sentiment Analysis, Context Analysis, LDA Topic Modeling, Longformer, Ensemble Learning.

Dedication

We would like to dedicate our work to parents, who have given us tremendous support and love without which we could have never had the opportunity to do such an incredible research. We would also like to give a very special thanks to our Thesis Supervisor Md. Golam Rabiul Alam for his unwavering support and utmost dedication to help us.

Acknowledgement

Firstly, all praise to the almighty Allah for whom our thesis have been completed without any major interruption.
Secondly, to our Thesis Supervisor Md. Golam Rabiul Alam for his kind support and advice in our work. He helped us whenever we needed help.
Thirdly, we would like to thank Joydhriti Choudhury, Lecturer at BRAC University, for his advice. Fourthly, we would like to thank all the people who helped in data labeling
And finally, to our parents, without their throughout support it may not be possible. With their kind support and prayer we are now on the verge of our graduation.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

$BERT$ Bidirectional Encoder Representations from Transformers.

CAT Category

GB Gradient Boosting

LDA Latent Dirichlet Allocation

MLP Multi Layer Perceptron

NLP Natural Language Processing

$ReLU$ Rectified Linear Unit

SVM Support Vector Machine

$TF-IDF$ Term Frequency-Inverse Document Frequency

UNO United Nations Organization

XGB eXtreme Gradient Boosting

Chapter 1

Introduction

Politics is an intrinsic part of our society. According to Aristotle, politics is to investigate on the basis of the constitutions collected, what makes for good government and what makes for bad government and to identify the factors favorable or unfavorable to the preservation of a constitution [19]. As human beings we depend on skills which allow us to persuade or criticize others. Such types of statements can have a profound effect on a country or its people. In most cases political speeches are controversial. The after effects of such speeches or opinions may be felt years after they were made. Political speeches concern decisions about possible courses of action which are contentious and contested and about which people might reasonably disagree [22]. Sentiment is a thought, opinion or idea based on a feeling about a situation, or a way of thinking about something [21]. In most cases politicians give speeches to increase their support in public. Politicians announce policies, agendas and laws which are meant for the betterment of the people. However, that may not always be the case as speeches or opinions tend to be divisive. In most cases, opposition parties and their supporters tend to have an opposite reaction if the statement is a divisive one. Thus, it is important to keep track of what the intended reaction was and what the reaction received was.

Analyzing the data to understand what sentiment it showed and what context they were made on is an intriguing field to conduct research on. Because in many cases people may jump to conclusion about a topic without knowing it's context and sentiment. Our paper focuses on the steps required to conduct research on the above mentioned ideas which is sentiment analysis and context classification of speeches from political figures. To perform sentiment analysis, we used deep learning algorithms based on BERT architecture. The BERT architecture has the benefit of self-attention layers. However, due the immense size of each speeches, we had to consider models such as Longformer and applied TF-IDF to generate vectors and then train ensemble learning algorithms. For the context classification we used LDA topic modeling to find topics and then we applied ensemble learning models on top of it. By doing so we were able to make a comparative analysis.

1.1 Motivation

Throughout time we have observed wars between countries be it a trade war or otherwise. Decisions made by political leaders or speeches made by them influences a country's own citizens and even the neighbouring countries. Choice of words in what they say can bring about conflicts or prosperity. It is quite possible for the global politics shift just from a single speech delivered by a political figure. To determine whether a political leader is a fit to govern the country, it is of paramount importance to understand their ideologies and how their comments may shift the global politics. Surely, it is not an easy task for a human being and even much more difficult for machines. An early step towards this end is to understand what these political figures deliver to the commoners through their speeches and find a pattern within it. Of course everyday numerous speeches are being delivered in one place or another. Finding a pattern or trends in all of these speeches through human labour is not feasible and so this demands for an automated system - a system that can determine the overall sentiment of the speech - a system that can determine the context a speech is addressing to. And, hence, our research starts from here and it is just one of the early steps of even a larger research.

1.2 Research Problem

Previously politicians used to give speeches in front of gatherings of people or using any news media. Nowadays, in addition to the traditional ways, use of social media to share their thoughts can also be seen. As written speeches are available in on the government websites, we decided to work with the written speeches. The politicians appeal to certain demography of people by conveying certain sentiment in the context of different situations. In this way, they try to preach their agenda and attain their goals. Influencing the audience means changing their sentiment towards any particular topic or group of people. This influence can be either positive or negative. For example, when it comes to winning elections, politicians try to increase their popularity. Therefore, they become aggressive towards their opponents with words which encourages their supporters to not like the opponents too. There are some speeches that changed the world like the Quit India speech given by Mahatma Gandhi on 8 August, 1942 which helped India to gain independence or "I Have a Dream" by Martin Luther King, Jr. which is considered an iconic speech in American history. This speech made a call for freedom and equality and it is considered as a defining moment for the civil rights movement of America. We can also look into the 7^{th} March speech of Sheikh Mujibur Rahman which played a vital role in the independence of Bangladesh. On the other hand, there are speeches that spread hatred towards others and speeches that are based on prosperity and more. Moreover, with the change of society and depending on the situations, the key focus of speeches changes too. As the motive of the speech is to influence people, it is important to identify sentiment values and the context of the speech itself. In this research we intend to answer this question using natural language processing algorithms.

1.3 Research Objectives

The research aims to find out the sentiment and context behind the speeches of the presidents/prime ministers/country representatives of five permanent members of the United Nations Security Council. The objectives of this research are:

1. Classification of political speeches into positive, negative and neutral sentiments.

2. Classification of political speeches on the basis of context of the speeches into development, nationalism, extremism, international and others.

Chapter 2

Related Work

In this section, we talk about previous works which are relevant and similar to our work.

The research work [7] focuses on sentiment analysis which performed during the 2016 US presidential election, there was a political homophily phenomenon on Twitter. First of all, they collected the twitter data then they identified political and non-political tweets after then the Both political and non-political tweets were analyzed for sentiment.They then identified each user class (six in total: positive, negative, neutral, whatever, Trump supporter, Hillary supporter) and evaluated political homophily on Twitter.. They compared between homophily in uniplex and multiplex connections. Finally, it was discovered that negative users have the highest amount of homophily and form the most homogeneous societies.

Arabic being the 5^{th} most spoken language around the world, while also being Morphological rich presents a challenge in sentiment analysis, not to mention there is the issue of data sparsity. A sentiment Tree Bank [3] is being made from scratch which is a required resource for the RNN model to be made. Ambiguity was overcome by in-context morphological analysis and disambiguation and other relevant machine learning. During disambiguation and extracting in-context morphological features it was primarily focused on extracting features like predicted diacritics along with stem and lemma . This is used in both syntactic parsing, binarization and later on morphological enrichment. Tokenization is used to create trees that represent untokenized text in its raw form. The nodes in the binary trees are given labels, this task was done via crowdsourcing. Finally, the treebank is being made morphologically enriched.

The authors [15] conducted polarity along with subjectivity sentiment analysis on tweets by taking time as the basis for the dimension of SA. They used word frequency as a measure to look for words which are linked to politicians within a given timeframe. They figured out how to justify computed sentiments along with the highly occurring words that are associated with the topics for every politician. Using the results from experimentation they presupposed that despite the political party working as a platform for selling the personality of that candidate, the accepting candidate or party can add to the winning an election. TextBlob, LDA, SENTIWORDNET, NBC are used here. SENTIWORDNET, result of automatic annotations by each of the synsets11 of WORDNET by labeling them as 'positivity,' 'negativity,' and 'neutrality.' Sentiment analysis of tweets from twitter involves understanding the sentiments from tweets. This is done using different Natural

Language Processing (NLP) techniques. Polarity determines if a tweet conveys a neutral, positive, or negative opinion. Finally, they look at the contributions frequently recurring words make to the importance of the LDA topics. Some results they got are that they believe that tweets during the election day are essential for gaining an insight into the voters mind and this could be an important political insight that can be gained from them. From SA, they figured out what people's attitudes toward the politicians were for a given time period and if these attitudes are factual or opinionated. Results obtained from TextBlob are 2447 (32.93 percent) Positive, 3971 (53.44 percent) neutral and 1012 (13.62 percent) negative respectively. Again, results obtained from SentiWordNet were 2916 (39.25 percent) Positive, 3085 (41.52 percent) neutral and 1429 (19.23 percent) negative respectively.

The paper [5] portraits a primitive approach on sentiment analysis on tweets from twitter. The dataset on which the analysis is being carried out also includes tweets from different languages and dialects which are being translated to English. The data preprocessing takes advantage of two new different sets of dictionaries: emoji dictionary and acronym dictionary. The data preprocessing includes the following steps: removal of emoji, URL, followed by tokenization, removal of non-English word containing tweet, removal of @target, repeating sequence, stop word and finally POS. Then the paper discusses the design of the kernel tree.

Nowadays, social media has become one of the important sources of understanding the political polarity of the citizen and campaigning according to that. In the paper, [9] the researchers used hybrid n-gram models to counter the 'Zero Count problem'. Moreover, to modify the method for maximum likelihood estimate for computing n gram sentiment score, the researchers have used Katz back-off. They have also applied the laplace smoothing for the purpose of creating a classifier to achieve higher accuracy. In the study, they have used Obama-McCain dataset that has been already used in previous studies. The researchers found out that the Unigram model outperforms the n gram model in case of positive sentiment but n gram model outperforms Unigram in case of negative sentiment. However, Hybrid n gram models perform equal to unigram and n gram in case of positive and negative sentiment respectively and combinedly performs better than both with average F measure of 0.80. Moreover, their hybrid n gram model also surpassed the previous study on the same dataset indicating that their goal was achieved. Their research result also relates to the result of the US 2008 presidential election. However, they had to create a custom lexicon for unigram as they could not obtain the unigram lexicon of earlier study but they were able to make a valid comparison regardless of the limitation.

With the increasing number of internet users, social media has become a key platform of political campaigning for the politicians. In the paper, [16] the researchers focused on the sentiment that users showed to the Facebook posts made at the time of the election campaign and wanted to see how accurately the reactions reflected the outcome of the election and the emotion towards the candidates of the State of Mexico in 2017. To do the research, they collected and analyzed 4128 Facebook posts and the reactions of these posts. They classified the reaction based on positive or negative polarization. The sentiment of the Facebook users was measured using the Oh and Kumar Sentiment Index. The sentiment index they got for the political parties was 1.08 for PRI, 2.72 for MORENA, 3.46 for PRD and 1.06 for PAN.

However, the sentiment index does not represent the actual outcome of the election result. Though PRI was third out of four political parties in terms of sentiment index, it won the election with 33.72 percent votes. This difference between the user sentiment and actual result may be a result of the difference between a party's social media strategy vs on-field strategy. A candidate may have used good social strategy and have boosted their posts to make a greater reach but the users who interacted and shared positive reactions might not have the intention to vote for that particular candidate or may have reacted positively for multiple candidates. The online followers of the candidates might not be from the country/region from where the candidates participated. Moreover, the voters who are poor or may not use Facebook very much might have played a key role too. We believe that this research can be improved in future by analyzing the country's political culture and the discussion that users do on their own posts instead of discussion on the candidate's posts in addition to the current approach. Various social media in addition to Facebook can be considered too to improve the accuracy.

The research work [8] proposes IAD sentiment analysis by using word embedding which is basically a feature extracting model. Logistic Regression, Decision Tree, Support Vector Machine, and Naive Bayes are the four supervised learning methods they use.. All of the methods achieved the best performance. But they said that negative samples LR and SVM classifiers give slightly better performance and vice versa relation between DT and NB.

The idea of [4] is that with the help of automatic natural language systems, we need to determine for a given tweet text and a target entity, whether the tweeter supports the provided target, opposes the given target, or believes that neither inference is likely. Previous research has not effectively addressed the connection between sentiment and stance. One of the main reasons for this is the unavailability of a dataset and therefore a new enriched dataset is made. The paper is heavily focused on the dataset from all possible perspectives starting from the dataset requirement to quality control in annotation procedure. For quality control CrowdFlower's gold annotations scheme was used, in which the authors annotated five percent of the data internally. These questions are known as gold questions. The gold questions are intermingled with other questions during crowd annotation, and the annotator was not aware of the gold question and others questions. He/She is immediately notified when he or she answers any gold question incorrectly. If the annotator's accuracy on the gold questions is less than 70 percent then he or she will be denied further annotation. At least eight people responded to each question. Annotation procedures were executed for both stance and sentiment annotation.

The research work [10] conducted an experiment to see how the appearance of political hashtags on social media can influence people's reactions. They basically conducted 2*3 factorial experiment where political hashtags are included and excluded and there are three types of comments: firstly positive comment, secondly negative comment and lastly no comment. Finally, they came to know that when a hashtag is used, the particular news gets more attention. And that news became more partisan and controversial.

Authors [13] talk about predicting the trends of the real stock market by using multiple machine learning algorithms. These are the analysis of public sentiment and political situation. Here, by using 10 different types of machine learning algorithms on their dataset, from Naïve Bayes to MLP, they used a variety of advanced technology and tested the results. Next, they divided their task into three stages being the SA, political situation analysis and finding the stock interdependence analysis. Some results they got from the sentiment analysis were using the sentiment attributes for stock market prediction increases accuracy of ML algorithms by 0-3 percent. Next, from initial prediction, the most accurate prediction can be gained on day 7 to find stock market trends. Moreover, from the day of initial prediction, day 1 and 2 are the least effective days. SMO was the best in performance for running both of SA approaches. Thus, it is the best for predicting stocks using sentiment analysis. LWL performed well in differing positive and negative trend classes. For stock prediction via analysing political situations, MLP and DT have shown highest accuracy, which is an accuracy of 75.38 percent (achieved on day 5). However, most of the classifiers showed lower accuracy on day 1. Thus, a political event will cause the minimum effect on day 1. Political events can have a significant (about 20 percent) impact on the KSE stock market's stock values.

The transformer [18] is one of the state-of-the-art transactional models capable of solving machine translation and other NLP problems at a much faster rate than CNN, RNN, GRU and LSTM. It is composed of five parts. There is a 6-layer encoding and decoding layer through which the data passes through. Self-attention is an important part of this architecture. Self-attention is very helpful as it has 3 advantages. Firstly, the total computational layer complexity is less. Next, the number of computations can be parallelized. Finally, the pathway between long range dependencies in a network is significantly shorter. They used 8 NVIDIA P100 GPUS in one machine to run this model. Moreover, the datasets are encoded and tokenized by byte-pair encoding. They also used the Adam optimizer as an optimizer. At last, a residual drop rate is also applied here. Here, they exhibited the Transformer which is the first sequence transduction model. It is based on attention, which replaces the recurrent layers that used in encoder-decoder along with multi-headed self-attention layers. In comparison to architectures based on CNN or RNN, Transformers can be trained faster. They achieved a new state of the art in WMT 2014 English-to-German translation and in WMT 2014 English-to-French for French translation. In the English to French translations, their model outperforms all other previous entries.

Here in the research work [6] they applied their data set which contains 2500 positive and 2500 negative comments to know what percentage of a nation's citizens welcome refugees inside the country(approve) and what percentage deny(disapprove). They took 20 percent of the data as testing data, applied SVM and got 79 percent accuracy. They have also tried with the naïve bayes algorithm but could not achieve satisfactory outcomes.

In the paper, [17] the researchers tried to find out the relationship between moral-emotional behavior and political ideologies. In the research they ensured variety in the data by analyzing data from the United States of America and Japan which are different in terms of political and government cultures. The researchers have used Frequency Analysis and Regression Analysis in their study. In terms of the emotional factors, the study suggests that emotion has little to do regarding positional or political ideologies in the United States.. However, in Japan, the positional factor matters more than the ideological factor in the case of emotional state. Similar results were also found in the case of moral aspects. The results suggest that the relationship between moral-emotional behavior and ideologies may vary depending on the political culture and government structure of a country.

In this study [14], the authors show that political party representatives can be characterized using data sciences along with their evolution. Tweets of political parties are a good way to guess the emotional state of the party as well. They started with frequency analysis which helped in extracting each party's distinct ideological bubble, then they moved on to a more sophisticated analysis. This was done to quantify the evolution of Sentiment, that is negative and positive as a response to certain political events. Finally, identifying party ideology adhering is possible with the help of predictive AI tools by analyzing each individual's tweet. They were able to train an AI capable of recognizing political affiliations of a tweet. Moreover, the origin of a tweet can be predicted with the AI as well. It had a precision of range of 71-75 percent. They were also able to find left or right political alignment with a precision of 90 percent .

In the paper [2] discusses the system level learning technique which is partially-supervised for emotion classification. The classification task is subdivided into four main tasks: feature selection, clustering, profile generation and classification. Among these, the first three are speaker independent and unsupervised whereas the last one is speaker dependent and supervised. Therefore, this makes the approach semi-supervised. According to this paper, AHC (Agglomerative Hierarchical Clustering) is used. AHC avoids problems that are often common in other clustering techniques like K-means or GMM-EM. Also AHC being a bottom-up process is algorithmically faster relative to other top-to-bottom clustering techniques. The key idea for AHC is to regard each data as a cluster. The cluster starts to merge on a closest-pair basis which continues until the stopping condition is satisfied. The stopping condition needs to be specified beforehand which is basically a number - n, number of clusters.

This paper [20], discusses opinion mining of Movie Reviews-Based Applications using CART (Clustering and Regression Trees) with the help of ROCK (Robust Hierarchical Clustering Algorithm). The CART technique makes use of a link to gather information on connections between two points, with the goal of deciding which points should be merged into a single group. ROCK recognizes clusters with the help of agglomerative progressive grouping calculations. ROCK uses various levelled bunching techniques which is a strategy for organizing information focuses. These information focuses are put in a bunch or collection with similar properties, whereas information focuses in partitioned groupings are distinct.

This paper discussed [1] about the combination of both LDA and SVM properties to make a new classifier with a huge margin. Under the SVM framework, the proposed SVM/LDA formulation may be considered as an extension of a conventional SVM classifier that integrates global data properties .They proved that the suggested new technique can be implemented with current SVM model and also the suggested method might be expanded to semi supervised learning. They get a very good output for example Twonorm dataset (SVM/LDA = 0.983 , SVM = 0.958 , LDA = 0.974).On the downside, only linear decision boundaries can be implemented using the SVM/LDA formulation proposed in this study.

The paper [12] focuses on to propose a model by using the pretrained transformer which is considered as the base classifier to select tough training sets to fine tune and achieves the profits of both increasing ensemble in NPL tasks and pretraining language knowledge. They present the Boosting-BERT model, which incorporates multi-class boosting into BERT. They test their model on the GLUE dataset and three prominent Chinese NLU benchmarks. They try to prove that this is the first study to show that boosting, rather than bagging or stacking, may be utilized to improve BERT performance. In addition, their results demonstrate that BoostingBERT consistently beats bagging BERT. They compare between two approaches weight privacy vs weights sharing. The findings of the experiment reveal that BoostingBERT beats the strong BERT baseline on all tests and that it is beneficial in a variety of NLP tasks.

To address the limitation of 512 token limits of BERT modeel, in the paper [11], Longformer model was introduced. It comprises of an attention mechanism that is capable of scaling linearly with sequence length. They introduce the Longformer with an attention mechanism that scales linearly with sequence length. Their key idea is to substitute the original self-attention mechanism with a new one that combines local windowed attention with a task specific global attention. This new self-attention strategy scales linearly with increasing input sequence.

Chapter 3

Data set Preparation and Methodology

3.1 Dataset Creation

We have built a custom dataset for the purpose of our research. We have built the dataset by using the speeches of permanent members of the United Nations Security Council. These members include, The United States of America, The United Kingdoms, Russia, China and France. These countries play a vital role in global politics and decision making. As a result, any statements made by these countries are integral for understanding the global political sentiment. In the case of the USA, UK, France and Russia we have collected speeches of the presidents or prime ministers. For China, we had to collect speeches from the minister of foreign affairs as well as the prime minister and president since there was no exact website that contained only presidential speeches. All the speeches were collected from official websites of the government. We have only considered speeches that contain only one speaker so that the sentiment of the speaker is the only focus in these speeches. As a result, there is no chance of multiple sentiments from the speech. We also omitted speeches that are just short messages as they are not proper speeches but rather greetings to the general people. We picked the speeches which followed these criteria, resulting in the data set that we built. As we have collected these speeches based on our own judgement, there may be some biases as we have omitted some speeches which we felt were not relevant to our research.

We also recorded the name of the speaker, date, the country, the designation of the speaker, the headlines and speech link along with the speech transcripts. For France we used the google translator to translate the speeches into English as the speeches were initially available in French. As a result, there may be some context lost in translation. We have collected 3091 speeches from five different countries across multiple years. We manually scrapped the data from four of the websites as they were not suitable for code implemented scrapping. The UK speeches were collected directly from the website by using a scraping program. It took us about two months to collect all the speeches of our corpus. Some of the description of our collected speeches are mentioned in the table 3.1

Country	Head of the Government	Number Of Speeches	Year	Speaker	Main Website Link
Russia	President	557	2000-2021	Vladimir Putin, Dmitry Medvedev	http://en.kremlin.ru
US	President	989	2012-2021	Barack Obama, Donald Trump, Joe Biden	https://obamawhitehouse.archives.gov/ https://trumpwhitehouse.archives.gov/ https://www.whitehouse.gov/
China	President	278	2003-2021	Wang Yi Xi Jinping Li Keqiang Hu Jintao	https://www.fmprc.gov.cn
France	President	1081	2014-2021	Emmanuel Macron, François Hollande	https://www.elysee.fr
United Kingdom	Prime Minister	186	2012-2019	Boris Johnson, Theresa May, David Cameron	https://www.gov.uk

Table 3.1: Data Set Features

3.2 Data Set Annotations

After collecting the data, we decided to label each speech. Each speech is labeled into two criteria. One is based on sentiment and the other is based context. For the the annotation part, the speeches were given to three different annotators at first. They all annotated each speeches based on their interpretation. The three annotators gave their respective labels and then we compared their labels. After comparing the labels, we assigned the label which had most votes as that label of that speech. In case of a tie, we sent it to further one person to review the speech. The labels of speech were then once again reviewed by us. All four team mates participated in annotating. We also outsourced one label of the three labels here as well. This way we were able to check the three labels by at least two of our team mates. However, we were not able to cross check with any expert to validate our data set. As we have done our labeling based on annotators interpretation, there might be some unwanted bias in our dataset.It took us four months to label the entire data set.

3.3 Data Pre Processing:

Pre processing is a very integral part of NLP based ML techniques. The preliminary steps like removing stop words were carried out. Moreover, nouns, special characters, extra spaces, alphanumeric characters and slashes were also removed. We have also lemmatized the speeches and made all the letters lowercase. Additionally numerical values were erased since the main content of the data set are speeches of large political figures from which we are trying to classify sentiments and contexts where numerical values like dates are less likely to contribute any sentimental value. This also reduces the size of vector matrices and optimizes the feature extraction. As we are dealing with entire speeches, the token length of each speech is quite large. The highest token size of a speech is 2800 tokens. While the average token ranged from 600 to 700 in token size.

3.4 Algorithms Used

As we are evaluating, the speeches are in terms of two different categories. For that purpose we need to implement different algorithms for each case. For Sentiment analysis, we went with Longformer Algorithm and TF-IDF vectorizer along with different Ensemble Learning Algorithms. Ensemble learning algorithms include XGB Classifier, CATBoost classifier, Gradient Boosting classifier. For evaluating the context of a certain speech, we opted to use LDA Topic Modeling along with different Ensemble Learning Algorithms only. The learning algorithms are XGB Classifier, CATBoost Classifier, Gradient Boosting classifier.

3.4.1 TF-IDF Vectorizer:

To use Ensemble Learning models for sentiment analysis, we chose TF-IDF vectorizer for feature extraction. TF-IDF (Term-Frequency and Inverse Document Frequency) harbors the key idea that the TF-IDF value of a word is positively proportional to the number of occurrences in the entire document (TF) and inversely proportional to number of documents the word appears (IDF).

3.4.2 LDA Topic Modeling:

Topic modeling is a statistical modeling for finding the abstract topics that are present in a corpus of documents. LDA topic modeling is used to identify various topics in a document. It works with very basic two assumption. One is that a document is a mixture of topics and another is that a topic is a mixture of words. However, LDA does not label any topic nor does it label the document with any particular topic. Instead, it provides a document – topic frequency matrix and a topic – word frequency matrix.

3.4.3 Ensemble learning:

Ensemble Learning combines the decision from multiple algorithms to improve the overall performance. There are various advanced ensemble learning techniques like stacking, bagging, blending and boosting. We have used Gradient Boosting, XGBoost and CatBoost algorithm in our research. Boosting algorithm is a sequential process where each subsequent model attempts to correct the errors of previous model.

3.4.4 Longformer :

Before understanding longformer it is necessary to understand BERT. Bi-Directional Encoder Representation from Transformers(BERT) has achieved breakthrough in NLP tasks. BERT is a transformer based deep learning model or simply to put it is a stack of trained transformer encoders. Whereas transformers are basically a pair of encoder and decoders. For example in a language translating transformer an input sequence of English words can be passed as an input and output will be a sequence of words but in a different language while retaining the semantic values. What essentially occurs is that the encoder portion of the transformer takes the input sequence and encodes them into a different format of data which is decoded by the decoder for prediction purposes. Since the architecture of the encoder in BERT is very close to that of the original Transformer models, BERT is often called a Transformer based model. BERT takes advantage of the self-attention mechanism within the encoder. Primitive machine learning techniques are not usually capable of inferring words in relation to other words. However, the self-attention mechanism is capable of doing so. What BERT essentially does is that it produces word piece embeddings or vector representations of a word. This leads to the other advantage BERT has over other word embeddings or vector generating. BERT is context-aware. For example one word can have different meanings based on the other words in a sentence it is surrounded by. Vector generating models like Word2Vec would produce exactly the same embeddings for a word anywhere irrespective of context or in what sentences a particular word is used for. As a result, non-context-aware models like Word2Vec would not be able to learn the difference in meaning of the words like ''nails" between "finger nails" and "hammer nails". However, BERT is context-aware meaning the embeddings produced for a word is influenced by its surrounding words, hence, the embeddings will not exactly be the same at every instance making them dynamic resulting in an effective learning. Another matter of importance is how BERT reads the input sequence. Traditional directional models like RNNs which read input sequences from left to right or right to left are often exhaustive in terms of time complexity and consume a large amount of time in the training phase. However, BERT takes input sequences all at once. This saves a lot of computational power and enables training the model with a large amount of data using the time saved by reading the sequence all at once.

The main component of the dataset that we are working with are speeches from leading political figures. These speeches are often long and often have more than thousands of words in them. Lengthy speeches are not the issue here. The issue is the token limits in BERT models. Most of the BERT models that we have come across can accept only up to a max of 512 tokens. Therefore, in order to fit the long speeches a rigorous text preprocessing has been carried out to remove unnecessary characters and words. However, even after text preprocessing it is found that the highest number of words in a speech is around 2800 words. A naive approach to adjust this is by truncating speeches and keeping only the first 512 words of the speech. The problem with this approach is that the portion that may help decide the overall sentiment of the speech may not be present within the first 512 words of all speeches. Fortunately, the Longformer model has a larger token limit - 4096 tokens. The highest number of words in our speeches is around 2800 which is well within the range of Longformer's token limit. Also Longformer is based on RoBERTa, and so it is expected to have a good training accuracy. For our research purpose, all things considered, opting for Longformer over models seems to be the go to decision.

3.5 Proposed Models:

As we are evaluating, the speeches are in terms of two different categories. For that purpose we need to implement different algorithms for each case. For Sentiment analysis, we went with our modified Longformer Algorithm and TF-IDF vectorizer along with different Ensemble Learning Algorithms. Ensemble learning algorithms include XGB Classifier, CATBoost classifier, Gradient Boosting classifier. For evaluating the context of a certain speech, we opted to use LDA Topic Modeling along with different ensemble Learning Algorithms only. The learning algorithms are XGB Classifier, CATBoost Classifier, Gradient Boosting classifier. The three models are;-

1) Longformer: The longformer model, is a RoBERTa based model. We have used a model comprising of the following parts. Figure 3.1 shows our proposed model

a. Longformer layer: We used the base Longformer model to encode the data and generate a vector matrix. This layer takes to input and makes sure the dimensions of the all the speeches are equal. We added a maximum padding of 2500 so that the speech doesn't go over this boundary. This ensures that every speech is of equal length.

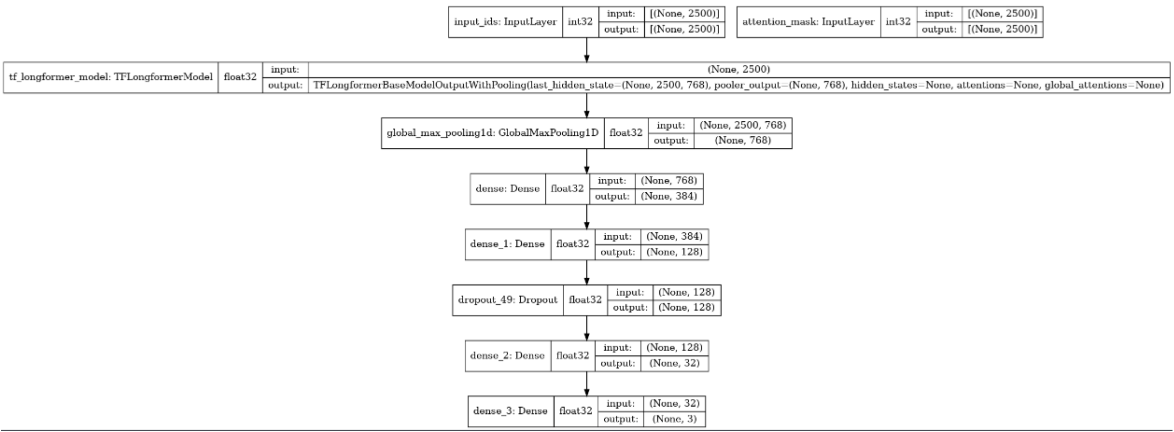

Figure 3.1: Longformer Model

b. Pooling Layer: The pooling layer is used to reduce the dimensions of the embeddings generated by the Longformer layer. And so it reduces both the computational load and memory strain. We used a layer of max pooling which takes the largest values of the feature map.

c. Dense Layer: The fully connected layers or dense layers take input from its previous layers, so that they can compute which layer is more likely to match a particular class. It assigning weights to specific features, using which necessary important features are brought to the forefront while others that do not help in this endeavor are excused. We have used 4 dense layers, which takes the takes inputs from the previous pooling layer and also previous dense layers as well and uses it to see find features which help in classification of sentiment.

d. Activation Function: We used ReLU and softmax activation function in the dense layers. These functions have yielded the best results on our model.

$$relu(x_i) = max(0, x_i) \tag{3.1}$$

$$softmax(x_i) = \frac{exp(x_i)}{\sum_j exp(x_j)} \tag{3.2}$$

ReLU is a piecewise function linear function that shows output if the input is a positive in nature, otherwise the output will be zero. We used the ReLU function on in the four dense layers and we used softmax on the last layer. In the equations, x the value of a number. In softmax, value lies between 0 and 1. This allows for normalization of output of network to a probability distribution for any output that was specified.

2. TF-IDF with Ensamble Learning: To use Ensemble Learning models for sentiment analysis, we chose TF-IDF vectorizer for feature extraction. For our feature extraction purposes with better accuracy, we opted for unigram, bi-gram, and tri-gram in our vector matrices. Usually during extraction of features, in the text vectorization process, order of the words are lost. To resolve this, N-Grams are implemented, however, the vector matrix becomes too large as a result which is computationally inefficient. Therefore, we drop both high and low frequency N-Grams to reduce complexity. This gives rise to another issue that certain keywords of great importance but less occurring might also get removed from the vector matrices. This issue is addressed in TF-IDF and, therefore, our reason for opting to this approach to extract features. We then applied boosting ensemble learning algorithm. We have used the Gradient Boosting, XGBoosting and CATboosting algorithms for our research. Boosting algorithm is a sequential process which means each subsequent model will make attempts to correct the errors seen in previous model.

3. LDA with Ensemble Learning: For the Context analysis part, we used a LDA topic modeling along with a boosting ensemble learning algorithm approach. A speech contains many hidden topics in the subtexts and to categorized the speech with a label, we have to identify those topics first. Thus, LDA is used. Moreover, the LDA along with ensemble learning approach is more effective in getting more accurate results due the topic modeling finding the actual context far better as compared to the other machine learning models. As a result, we implemented the LDA and boosting algorithm approach. We used the coherence scores and perplexity scores to generate an idea of how many topics we need to generate. We applied ensemble learning algorithm on top of LDA to run train data and test the predicted values of the labels.
Figure 3.2 shows a bird's eye view of our working steps.

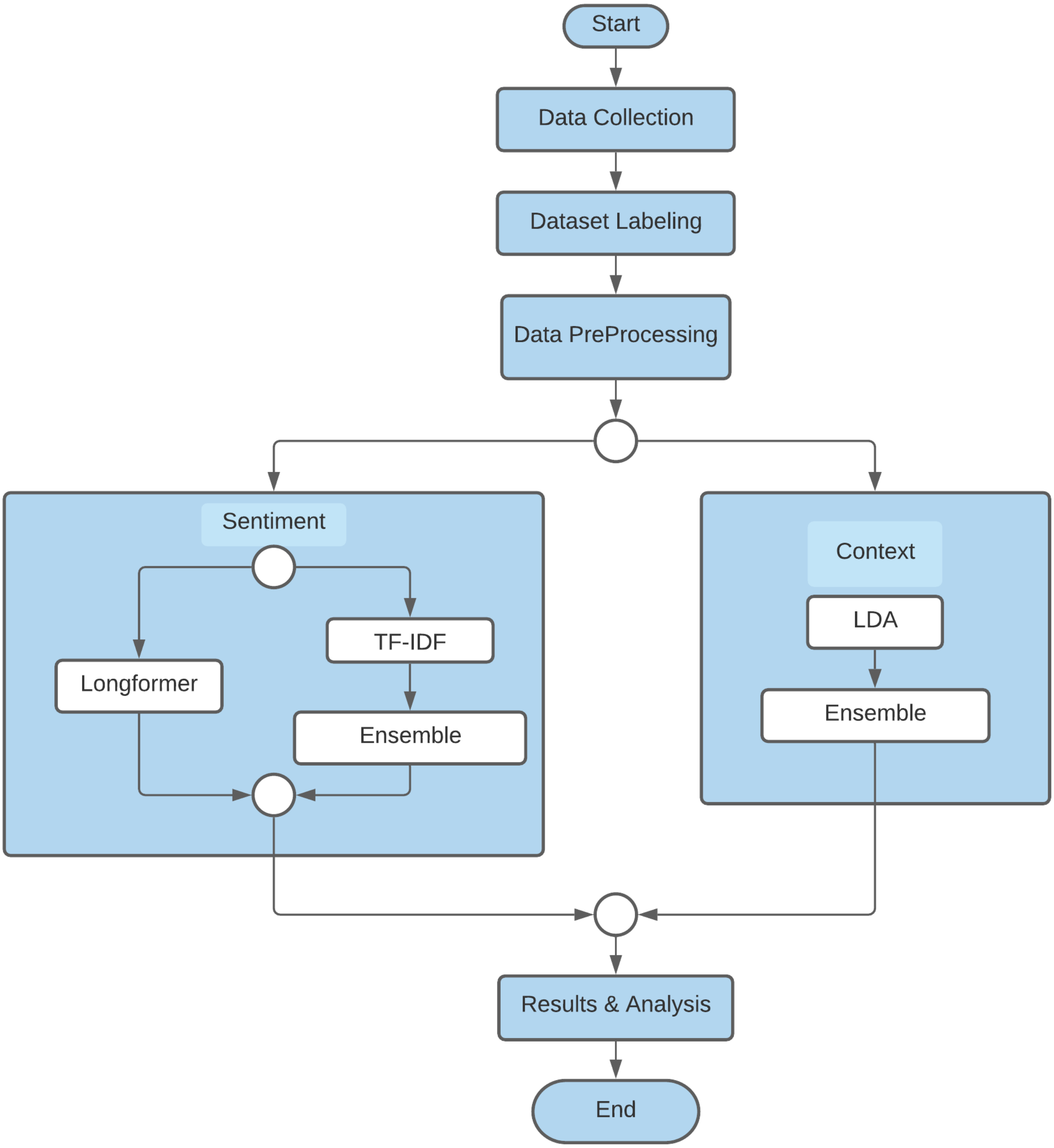

Figure 3.2: Work Flow Diagram

Chapter 4

Implementation and Result Analysis

4.1 Implementation of Sentiment Analysis Models:

For sentiment analysis, we tried a couple of different approaches. One of them was the modified BERT model Longformer and the other is TF-IDF vectorizer along with ensemble learning algorithms. We also tried using Baseline Algorithms such as the SVM, Naive Bayes and Random Forest. We found that ensemble learning algorithms are better equipped than the baseline algorithms. We also tried using the LDA model instead of TF-IDF along with the ensemble learning algorithm. However, they yielded a 0.60 score so we discarded them. In this section, we will talk about the results which we have obtained from testing our data set on our political speeches data set. It is divided into two parts. Here our evaluating factor was accuracy. The accuracy can be calculated by the following formula

$$Accuracy = \frac{\sum_{n}^{j} B^i}{n} \tag{4.1}$$

Here, n denotes specimen count and B^i is a function that returns 1 or 0 for the i^{th} specimen.

TF-IDF model: We set the following parameters for TF-IDF Vectorizer in our experiment. minimum difference = 0.1 max difference= 0.8 ngram range = (1,3)

Ensemble Learning: We used a multitude of boosting techniques. The boosting algorithms work by trying to correct the wrong predicted values of the previous models. Basically by combining the predictions, we can get a better result than traditional ML models. We used XGB Classifier, CATBoost Classifier and Gradient Boosting classifier on our dataset. The parameters for the different algorithms are given in the tables 4.1

The results obtained from the TF-IDF and boosting and Longformer algorithm is given in the table 4.2

Here we see that the results produced from the Longformer and the ensemble learning algorithms along with the TF-IDF Vector model have yielded scores between 0.6 and

Algorithm	Random State	Learning Rate
XGB Classifier	1	0.01
Gradient Boosting Classifier	1	0.01
CAT Boost Classifier	Default	Default

Table 4.1: Parameters of the algorithms

Algorithm	**Longformer**	Gradient Boosting Classifier	XGB Classifier	CAT Boost Classifier
Result	**0.6638**	**0.6649**	**0.6746**	**0.6734**

Table 4.2: Results of sentiment analysis

0.7. Here, we understand that the ensemble learning algorithms are out performing the Longformer model. Due to huge token length of each sentence, Longformer struggles to find the accurate results. Normally we label each individual sentence. But in this case the model can not give any output due to the fact that we labeled the entire speech in place of a certain sentence. This results in yielding poor results compared to the boosting algorithms. They do not have any limitations in regards to token length as the TF-IDF removes that complexity. Hence, we see the Longformer model struggling.

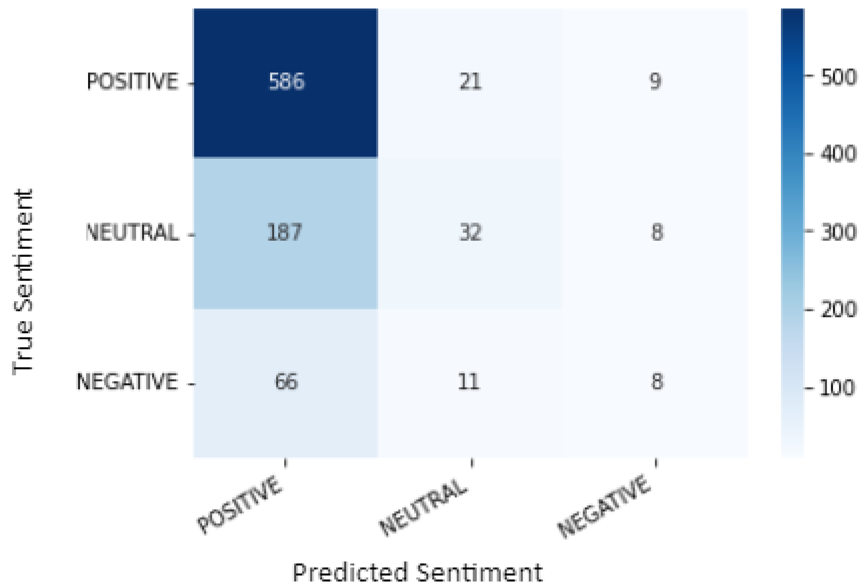

Figure 4.1: XGB Sentiment Confusion Matrix

We only showed the confusion matrix for XGB on top of TF-IDF because this is showing us the best results. Looking at the confusion matrix in figure 4.1. We see that the entire model is training speeches for only positive sentiment properly. The other two sentiments were not trained for properly and that resulting in the lower accuracy. The reason behind this might be due to different subtext of a speech carrying different sentiment making it difficult to draw an overall sentiment value for the entire speech. Furthermore, during data annotations, annotators had background knowledge pertaining to the speeches which is not the case for the machine learning models. Additionally, class imbalance with positive sentiment labeled speeches outnumbering the others. thus, the data set may need to be under sampled to show better testing results. The count of the original speeches were 2051 for positive speeches, 757 speeches and 283 for negative speeches. Thus by testing on 0.3 percent of the model has yielded poor results due to class imbalance.

4.2 Implementation of Context Analysis Models:

For the Context analysis part, we used a LDA topic modeling along with a boosting ensemble learning algorithm approach. The reason behind using such an approach is due to the fact that LDA topic modeling gives different topics which work far better than traditional algorithms like SVM, Naive Bayes, Random Forest etc. Moreover, the LDA along with ensemble learning approach is more effective in getting more accurate results due the topic modeling finding the actual context far better as compared to the other machine learning models. As a result, we implemented the LDA and boositing algorithm approach.

LDA Topic Modeling: We used the coherence scores and perplexity scores to generate an idea of how many topics we need to generate. Normally the coherence score is determined as the place where the curve begins to flat out. But in our case even after generating a coherence score till 350 topics, we saw that the curve does not flat out. As there is no clear plateau, we tested the model using from 50 to 100 topics and found 56 topics to yield the best results. The parameters of our LDA Topic Model are as follows:- No. of topics = 56 Random States = 42 Passes = 10 Iterations = 10 Coherence Score: 0.48070 Perplexity Score: -16.2095

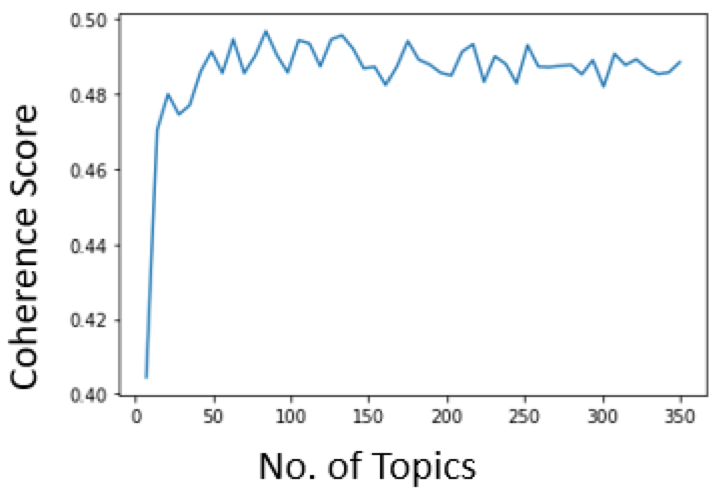

Figure 4.2: Coherence Score Graph

Here are some topic generated using LDA which can be mapped with the categories we have used for context analysing

(2, '0.030*"terrorist" + 0.018*"attack" + 0.017*"country" + 0.014*"also" + 0.014*"terrorism" + 0.014*"act" + 0.013*"security" + 0.013*"support" + 0.012*"fight" +

Algorithm	Random State	Learning Rate
XGB Classifier	1	0.01
Gradient Boosting Classifier	1	0.01
CAT Boost Classifier	Default	Default

Table 4.3: Parameters of the algorithms

0.012*"people"') =extremism

(43, '0.047*"job" + 0.045*"business" + 0.036*"company" + 0.032*"economy" + 0.026*"investment" + 0.020*"new" + 0.017*"make" + 0.015*"world" + 0.013*"create" + 0.012*"year"') = development

(46, '0.216*"europe" + 0.129*"european" + 0.030*"europeanunion" + 0.027*"ukrainian" + 0.022*"council" + 0.017*"border" + 0.015*"continent" + 0.014*"agreement" + 0.013*"policy" + 0.013*"trade"') = international

(36, '0.037*"world" + 0.034*"people" + 0.026*"war" + 0.020*"nation" + 0.016*"peace" + 0.013*"history" + 0.011*"human" + 0.011*"freedom" + 0.011*"today" + 0.010*"justice"') = nationalism

(19, '0.014*"work" + 0.011*"issue" + 0.011*"system" + 0.010*"state" + 0.010*"government" + 0.009*"lawenforcement" + 0.009*"problem" + 0.009*"justice" + 0.008*"criminal" + 0.008*"anybody"') = others.

Now, as we want to use supervised learning algorithm thus we created a feature matrix using the topic vector. We passed the feature vector to our ensemble learning algorithms to train our model.

Ensemble Learning Boosting algorithms: We used a multitude of boosting techniques. The boosting algorithms work by trying to correct the wrong predicted values of the previous model. Basically by combining the predictions, we can get a better result than traditional ML models. We used XGB Classifier, CATBoost Classifier and Gradient Boosting classifier on our dataset. The parameters for the different algorithms are given in the table 4.3

After training our model on our dataset, we see that we get a score of 0.6713 on the LDA CAT Boost Classifier, we get a score of 0.6261 on our XGB Classifier model and finally we get an accuracy of 0.6315 on the Gradient boosting algorithm. This shows that while all the ensemble boosting algorithms are performing between 0.60 to 0.70 score, the score of the CAT Boost Classifier is clearly the highest. The CAT Boost Classifier is very good at handling categorical values. Normally one hot encoding occurs when too many labels are introduced. This results in the exponential increase which the other boosting algorithms are incapable of handling. On the other hand, XGBoost is a highly predictive power gradient boosting technique. As a result it can result in overfitting. To overcome this, XGBoost employs the help of regularization. However in our case, as there are many labels, it becomes necessary to slightly overfit the model as opposed to regularizing it. Thus XGBoost gives the lowest possible accuracy when compared to the other two boosting algorithms. In theory, Gradient boosting algorithms should produce a good accuracy. However, what actually is happening is due to too many labels, Gradient boosting algorithm struggles to go above the 0.65 score threshold. As the Regression trees are used as base learners, Gradient Boosting is more suitable for regression problems than multi label classification problems.

The results obtained from the LDA and boosting algorithm are given in the table 4.4

Algorithm	Gradient Boosting Classifier	XGB Classifier	CAT Boost Classifier
Result	0.6315	0.6261	0.6713

Table 4.4: Results of Context Analysis

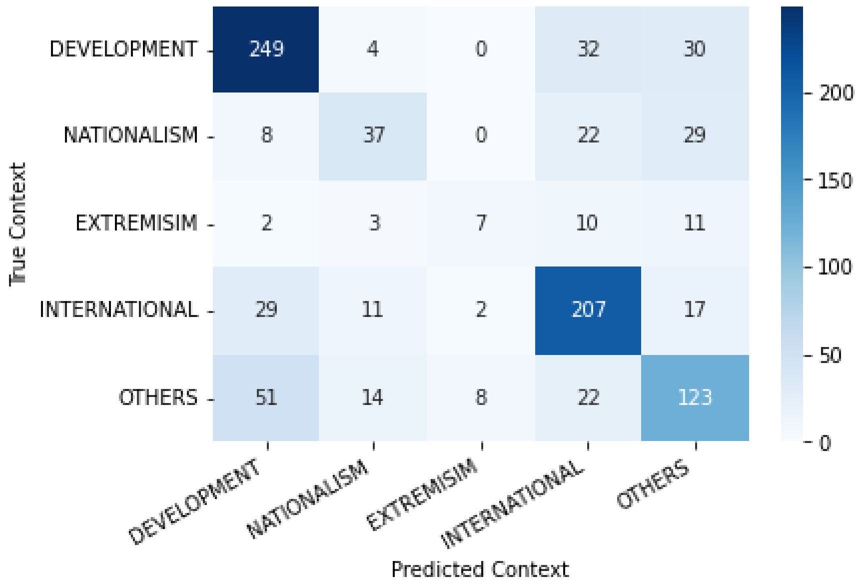

Figure 4.3: CAT Boosting Context Confusion Matrix

Looking at the confusion Matrix in figure 4.3, we see that the score of development is highest. The main reason for that is that the number of speeches labeled as development is the highest. As a result, the model was able to learn development contexts better. It can also be said that the development speeches are commonly confused with speeches which convey the international context. The main reason for such cases occurring is due to the fact that different international events relating to economy, climate, sports, education etc. occur together. So development gets confused with international speeches more when compared to the other speeches. Next, looking at the Nationalism context, the models performed quite poorly, the main reason for this occurring is due the low no. of nationalism speeches. Furthermore, in many cases, nationalism talks about defending the country against alien aggression. Moreover, many speeches use start off nationalism contexts but actually end up with other contexts such as military, memorials, social events and holidays etc. resulting in the score of accuracy dropping. The speeches which promote extremism performed the worst. The main reason behind it is the fact that extremism has the lowest speech count in the corpus. As a result, the data set confuses it with topics such as international and others. A prime reason for such a dismal score is due to the fact that politicians tend to mask extremism while talking about countries rather than their own countries. Terrorism, Hate speeches and inciteful comments are very often hidden cleverly by the politicians as they do not want to cause a diplomatic issue. As a result they tend to hide their true feelings by hiding them within speeches addressing international terrorism and in other issues. Next, international Speeches are identified quite well. It is the second largest count of speeches in the entire corpus. Just like the development case, international is confused commonly with development the most. The reason is the same as before, international summits such as the UNO, Covid-19, International sports are very similar to development topics. Hence, they are very easily confused. Finally, topics that do not relate to any of the above mentioned topics are classified as others. The other category covers a much wider domain in the real world as compared to the other cases. As a result it can be confused with development, international speeches quite easily. The other category of speeches are the average case here. Meaning they are in between the two extremes of context categories. In many cases, the other category can be confused with development due the fact that many politicians talk about development topics while talking about topics like military, events etc. As a result the algorithms get confused about the results and may sometimes predict development instead. Hence the results in the confusion matrices are like this. The total label count is development 1049, international 885, others 728 nationalism 319 and extremism 110.Thus nationalism and extremism under performed here.

Chapter 5

Conclusion

By conducting sentiment analysis and context analysis, we were able to find out that the CAT Boost Classifier with LDA topic modeling performs better to predict the context. However, while analysing the sentiment, we found that the model struggles a bit to predict correctly. This might be because of the speech frequency variation between positive, negative and neutral. As the president/prime ministers of a country usually talk positively from the perspective of their own country, that could be a reason for the large amount of positive speeches. However, working on the data set more to reduce data variance and using different transformer models like Big Bird might increase the efficiency more.

Bibliography

[1] T. Xiong and V. Cherkassky, "A combined SVM and LDA approach for classification," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005*, Montreal, QC, Canada: IEEE, 2006.

[2] E. Provost, K. Han, S. Lee, and S. Narayanan, "A cluster-profile representation of emotion using agglomerative hierarchical clustering"," pp. 797–800, 2010.

[3] H. H. N. H. K. B. S. W. E. -H. Ramy, "A sentiment treebank and morphologically enriched recursive deep models for effective sentiment analysis in arabic," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 16, no. 4, pp. 1–21, 2017.

[4] P. S. S. K. Saif, "Stance and sentiment in tweets," *ACM Transactions on Internet Technology*, vol. 17, no. 3, pp. 1–23, 2017.

[5] R. H. P. Siddu and P. Algur, "Sentiment analysis by identifying the speaker's polarity in twitter data," in *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*, 2017, pp. 1–5.

[6] H. A. Chowdhury, T. A. Nibir, and M. S. Islam, "Sentiment analysis of comments on rohingya movement with support vector machine," Mar. 2018. arXiv: 1803.08790 [cs.IR].

[7] M.-N. Josemar and A. Caetano, "Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 american presidential election," *Journal of Internet Services and Applications*, no. 18, 2018.

[8] A. Alnawas and N. Arici, "Sentiment analysis of iraqi arabic dialect on facebook based on distributed representations of documents," en, *ACM trans. Asian low-resour. lang. inf. process.*, vol. 18, no. 3, pp. 1–17, Jul. 2019.

[9] J. Awwalu, A. A. Bakar, and M. R. Yaakub, "Hybrid n-gram model using naıve bayes for classification of political sentiments on twitter," en, *Neural Comput. Appl.*, vol. 31, no. 12, pp. 9207–9220, Dec. 2019.

[10] M. M. Eugeniaha Rho, "Hashtag burnout? a control experiment investigating how political hashtags shape reactions to news content," *Proceedings of the ACM on Human-Computer Interaction*, vol. 3, pp. 1–25, 2019.

[11] I. Beltagy, M. E. Peters, and A. Cohan, "Longformer: The Long-Document transformer," Apr. 2020. arXiv: 2004.05150 [cs.CL].

[12] T. Huang, Q. She, and J. Zhang, "BoostingBERT:Integrating multi-class boosting into BERT for NLP tasks," Sep. 2020. arXiv: 2009.05959 [cs.CL].

[13] W. Khan, U. Malik, M. A. Ghazanfar, M. A. Azam, K. H. Alyoubi, and A. S. Alfakeeh, "Predicting stock market trends using machine learning algorithms via public sentiment and political situation analysis," en, *Soft Comput.*, vol. 24, no. 15, pp. 11 019–11 043, Aug. 2020.

[14] V. S. Miguel and G. Folgado, *Exploring the political pulse of a country*. 2020.

[15] I. Onyenwe, S. Nwagbo, N. Mbeledogu, and E. Onyedinma, "The impact of political party/candidate on the election results from a sentiment analysis perspective using #anambradecides2017 tweets," en, *Soc. Netw. Anal. Min.*, vol. 10, no. 1, Dec. 2020.

[16] R. Sandoval-Almazan and D. Valle-Cruz, "Sentiment analysis of facebook users reacting to political campaign posts," en, *Digit. Gov.: Res. Pract.*, vol. 1, no. 2, pp. 1–13, Apr. 2020.

[17] H. Takikawa and T. Sakamoto, "The moral–emotional foundations of political discourse: A comparative analysis of the speech records of the U.S. and the japanese legislatures," en, *Qual. Quant.*, vol. 54, no. 2, pp. 547–566, Apr. 2020.

[18] P. Verma, "Attention is all you need? good embeddings with statistics are enough:large scale audio understanding without transformers/ convolutions/ BERTs/ mixers/ attention/ RNNs or," Oct. 2021. arXiv: 2110.03183 `[cs.SD]`.

[19] "Aristotle - political theory," in *Encyclopedia Britannica*.

[20] M. K. Av, "Sentiment analysis using robust hierarchical clustering algorithm for opinion mining on movie Reviews-Based applications"," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8,

[21] *Sentiment*, en, https://dictionary.cambridge.org/dictionary/english/sentiment, Accessed: 2022-1-16.

[22] *What makes a speech political?* en, https://www.psa.ac.uk/what-makes-speech-political, Accessed: 2022-1-16.