|| Volume 8 || Issue 12 || December 2024 || ISSN (Online) 2456-0774
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Analysis of Political Leaders Speech Using NLP and AI: Techniques and Algorithms

Malayaj Kumar[1], Dr. Anuj Kumar Singh[2], Dr. Soumitra Das[3]

Research Scholar, Department of CSE, Shri Jagdishprasad Jhabarmal Tibrewala University, Rajasthan [1]
Research Guide, Shri Jagdishprasad Jhabarmal Tibrewala University, *Rajasthan*[2]
Research Co-Guide, Shri Jagdishprasad Jhabarmal Tibrewala University, Rajasthan[3]
kumarmalayaj@gmail.com, mail2anujji@gmail.com, soumitra_das@yahoo.com

-- ***---

Abstract: - Artificial intelligence, also referred to as AI, is one of the most important developments in this rapidly developing subject, and its application is rising. Among the numerous ways artificial intelligence is changing the world today is by analyzing political statements using sophisticated algorithms. We explore the ways in which artificial intelligence (AI) is transforming the analysis of public speeches and how this science is becoming more and more crucial for political academics, analysts, and strategists. Researchers may find trends in political statements that might otherwise go unreported by using artificial intelligence (AI). Repetition, alliteration, and exaggeration are examples of rhetorical techniques that AI systems can identify. Algorithms are also capable of identifying the political speech's emotions, such as happiness, sorrow, or wrath. By using Artificial Intelligence (AI) and Natural Language Processing (NLP) to political speech analysis, scholars may get profound understanding of ideological tendencies, rhetoric, and mood. A pipeline comprising preprocessing of texts, extracting features, and predictive machine learning or neural network models customized to particular analytic objectives comprise the process.

Keywords: NLP, Deep Learning, Machine Learning, K-Means, CNN, RNN, Decision Trees, SVM

--- ***---

I INTRODUCTION

The field of politics has embraced the use of machine learning (AI), especially in the study of political discourse. Advances in artificial intelligence (AI) technology have made it feasible to transform the analysis of political speeches and provide a more thorough and precise comprehension of the messages that public figures, including politicians, express [1, 2]. The process of natural language processing, or NLP, techniques are a major use of AI in political speech analysis. By analyzing and interpreting human language, these algorithms enable the extraction of significant insights and patterns that could have been missed by more conventional analytic techniques.

NLP is a computational technique that aims to make it possible for computers to process, interpret, and comprehend human languages using a range of methods. Recently, the field of electoral discourse analysis has garnered a lot of scholarly attention [1, 2]. The procedure of sentiment analysis, which evaluates the emotions expressed in the text, is one of the most popular NLP approaches, according to the literature. This method focuses on recognizing and classifying viewpoints in texts in order to determine the author's or speaker's position on a certain subject.

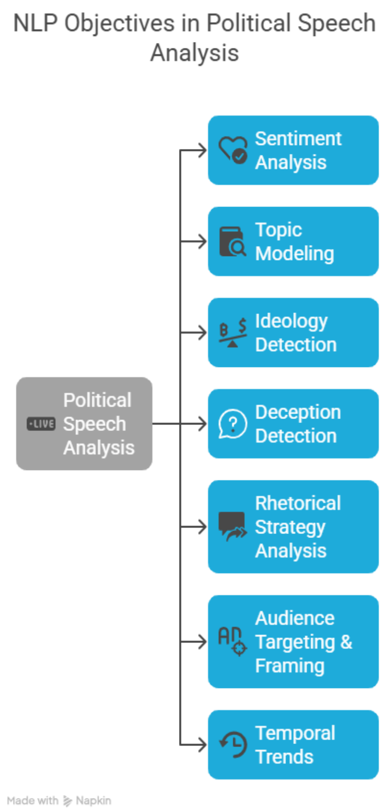

Figure-1: NLP in political leaders' speech analysis

Algorithms for natural language processing can determine the main points and subjects being covered, in addition to the speaker's perspective and voice, by examining the setting and subject matter of political speeches [1, 2]. Sentiment analysis automatically categorizes subjective data as unbiased, adverse, or favorable using natural language processing and artificial intelligence. Though it has special difficulties when it comes to speeches by leaders of government and officials, it is useful for evaluating political speeches, especially in international politics.

Though it suffers with the complex, confusing vocabulary of international statements, sentiment evaluation offers valuable insights into public debate [1, 2]. Irony, ambiguity, and courteous hedging are common in diplomatic discourse, and they may deceive instruments intended for direct communication. Furthermore, historical and cultural factors often play a significant role in diplomatic addresses. These nuances may be difficult for sentiment analysis systems to pick up on. Furthermore, sentiment analysis may overlook the precise objectives, critiques, and variances inside a speech, even while it might provide an overall good or negative score.

The use of NLP techniques to analyze emotions has grown significantly in the last few years. The area of sentiment evaluation in diplomatic relations is expanding.Artificial intelligence (AI) is transforming the political discourse landscape by providing real-time analysis of political debates, campaign rhetoric, and media coverage [1–4]. Because there is so much information accessible on so many platforms, people are finding it harder and harder to distinguish reality from fantasy. AI-driven tools that offer real-time verification of facts, analysis of emotions, and bias identification enable more openness in political communication.

Additionally, AI is now able to customize political analysis based on an individual's media consumption habits, political preferences, and hobbies. By using computer learning and the processing of natural languages (NLP), artificially intelligent systems (AI) can tailor content to help consumers understand stories about politics and their implications. This article looks at the primary technologies utilized in the analysis of political discourse using artificial intelligence.

II ANALYSIS OF POLITICAL SPEECHES

A key instrument for electoral interaction, speeches about politics help to articulate policy viewpoints, rally support, and shape the public's views. Throughout the globe, politicians use addresses to influence and convince audiences at home and abroad in addition to providing information [1, 2]. These remarks often mirror the political leaders' and their parties' philosophies, agendas, and methods of speech. Scholars may find patterns in language usage, emotional tone, and thematic emphasis by methodically examining them; these patterns provide important insights into public discourse and political conduct.

Massive analysis of governmental texts are now more practical because to developments in machine learning and natural language processing (NLP). Scholars may look at the superficial characteristics and deeper implications of political speeches using methods like sentiment evaluation, modeling topics, and discourse evaluation. This kind of study may identify disinformation or abuse, identify changes in political narratives, and even predict political tactics [1, 2]. Using a combination of qualitative and quantitative techniques, the current research attempts to investigate the linguistic and affective aspects of political speeches in order to analyze politician's rhetoric in various settings.

AI-powered algorithms can analyze political addresses, debates, and election statements with very high accuracy. These systems work by examining large volumes of text, searching for patterns, and verifying assertions with reliable sources. There are many reasons why AI technologies are crucial for political evaluation [1, 2].

1. Instantaneous Fact-Checking: Artificial intelligence systems cross-reference political statements with reputable databases and sources to ensure they are accurate. In only a few seconds, some advanced AI systems can recognize fraudulent claims and provide context.

2. Mood Analysis: AI can evaluate the emotional content and tone of political discourse, helping to gauge popular sentiment toward politicians, policies, and events.

3. Prejudice Detection: By analyzing speech patterns, word choices, and news coverage, AI may identify ideological biases in political speeches and media narratives. This makes it possible for individuals to have a more objective perspective.

4. Voice Recognition and Transcriptions: AI-powered speech recognition methods are used to convert political speeches and live debates into text. This allows for the examination of language, rhetorical tactics, and message trends across time.

5. Trends and Story Analysis: AI can predict the direction of stories across several media channels, track the evolution of policy discussions, and spot emerging political trends.

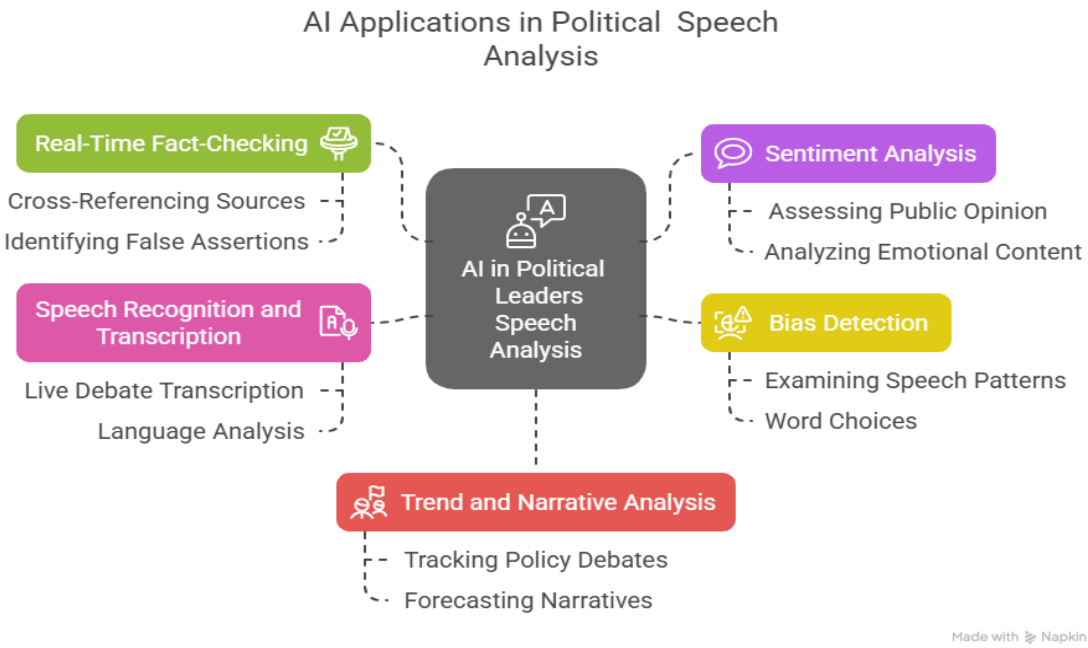

Figure-2: AI in Political Leaders Speech Analysis

III TEXT PREPROCESING AND FEATURE EXTRACTION

One of the first steps in the computer evaluation of public statements is text preparation. Distortion in raw voice records, such as grammar, unique characters, empty words, and irregular layout, may make analysis difficult. Preprocessing methods including lowercasing, stopping word elimination, splitting or a nutshell and a tokenization are often used to remedy this. By standardizing the text and distilling it to its most basic words, these procedures enhance the efficiency of ensuing natural language processing (NLP) activities. To maintain crucial context, domain-specific modifications may sometimes be required, such as maintaining political phrases, party names, or policy keywords.

The unstructured text is transformed into structured data representations using feature extraction techniques once it has been cleaned and standardized. The incorporation of words (e.g., Word2Vec, GloVe, or BERT-based encoding), Bag-of-Words (BoW), and Term Frequency- Inverse Document Frequency (TF- are typical methods. These techniques aid in capturing the vocabulary, setting, and regularity of elected leaders' usage. Emotion scores, identified entities, part-of-speech tags, and grammar patterns are examples of sophisticated characteristics that enhance the analysis. In addition to improving the interpretability of political speech, efficient preprocessing and feature extraction make it possible to use machine learning algorithms for analyzing sentiment, modeling of topics, and political categorization with greater accuracy.

The following are specific procedures for preprocessing political speech:

1. Tokenization

Typically, token consist of phrases, keywords, or phrases [5]. Tokenization is the is the act of breaking down a text into smaller parts known as tokens. In the field of speech analysis, the most frequent method of tokenization is word-level tokenization. As an example, the phrase "We are going to fight for justice!" has the form ["You", "may", "battle", "over", "equality", "!"]. By isolating meaningful units, tokenization makes it possible to do analysis on them either on their own or in context. Because it lays the groundwork for all of the future preprocessing operations, accurate tokenization is an absolutely necessary first step.

2. Lowercasing

For the purpose of ensuring that the same term is not handled differently owing to capitalization, it is necessary to convert all of the letters in the text to lowercase [5]. As an example, the phrases "democratic" and "democracy" would be considered to be distinct concepts in any other context. Despite sacrificing the semantic content, lowercasing makes analysis easier and minimizes the amount of the lexicon.

3. Stopword Removal

Frequent stop phrases like or, exists, and so on have little lexical weight. Eliminating stops helps analyze deeper words. Politics evaluation of speech requires care since regularly eliminated phrases like not and never may have high emotion or rhetorical value.

4. Punctuation and Special Character Removal

When they have context (e.g., "?" indicates a rhetorically question), quotation marks, hash tags, and acronyms are usually omitted [5]. Cleansing text reduces noise and ensures formatting.

5. Lemmatization

Lemmatization is is the process of breaking down words into their most basic or lexicon form, known as a word. Lemmatization takes into account a word's structure and context, as opposed to stemming, which only removes word ends. For instance, operate is the only word that remains after "operating," "operated," and "operates." This standardization helps to more precisely recognize speech patterns and enhances the matching of similar phrases.

6. Stemming

Although it uses a rule-based method that could not result in real words, stemming likewise lowers words to their basic form. a democracy and "democratic," for example, may both become "democ." Although stem is quicker, it is less precise and may not work well with complex political terminology. Thus, lemmas is often the method of choice for discourse analysis.

7. Part-of-Speech (POS) Tagging

POS tagging involves labeling each token with its grammatical category (e.g., noun, verb, adjective). This information can be used to analyze the syntactic structure of a speech or to extract features such as noun-phrase density, verb usage patterns, or adjective-based sentiment.

8. Named Entity Recognition (NER)

NER is the procedure of finding correct titles and classifying them as things such as dates, places , individuals , or organisations To track allusions to political figures, nations, events, and organizations, NER is very helpful in political speech analysis.

9. N-gram Extraction

N-grams are n-word patterns that are ongoing. When it comes to collecting popular collocations or rhetorical statements, bigrams (two-word phrases) and trigrams (three-word phrases) are helpful (e.g., "human rights", "make in India"). N-gram models are useful for both subjective and quantitative studies because they maintain the contextual links connecting words.

In order to prepare raw political expression text for study using mathematical methods, natural language processing models, or artificial intelligence algorithms, these preparatory processing stages convert it into an organized form. The study objectives—whether emphasizing a target audience, logical style, political stance, or emotive tone—determine which of these stages to use and in what mixture.

IV FEATURE EXTRACTION IN POLITICAL SPEECHES

Following preliminary processing, feature extraction—the method for converting unorganized text data into organized numerical representations—is a crucial next step in political speech analysis. Classification, sentiment analysis, and topic modeling are among the downstream tasks made possible by these attributes. Some common methods in this field are as follows [9, 10, 11]:

1. Bag of Words (BoW)

BoW is a basic technique that creates a vector of word counts by calculating the frequency of each word in the text [9]. This method is straightforward and efficient for determining if a phrase is present or absent, but it disregards the arrangement of words and contextual. BoW treats the terms "the nation shall rising" and "rising the nation will" equally, despite the fact that they may have significantly different meanings. A common baseline for text categorization tasks is BoW [4].

2. Term Frequency-Inverse Document Frequency (TF-IDF)

BoW is a simple method that computes the occurrence of every word in the text to produce an array of word counts [9]. This approach is simple and effective for identifying if an expression is present or missing, but it ignores contextual information and word order. Even though "the nation shall rising" and "rising the nation will" may signify very different things, BoW regards both identically. BoW [4] is a widely used baseline for text classification tasks.

3. Word Embeddings (Word2Vec, GloVe, FastText)

Word embeds are thick vector representations of words that preserve semantic links. Models may comprehend word similarity and relatedness because embeddings take into account the context in which words occur, unlike BoW or The TF-ID [10]. For instance, the vectors for "government" and "administration" will be comparable. By learning these associations from huge corpora, methods such as Word2Vec [10], GloVe, and FastText allow for deeper feature extraction from political speeches.

4. Contextual Embeddings (BERT, RoBERTa)

By creating distinct vector illustrations for the same word based on how it appears in a phrase, embedded contexts go one step further. Models like as BERT and RoBERTa are perfect for examining the complex nature of political language since they combine transformers topologies and deep learning to capture subtleties in language. For tasks like attitude analysis, position identification, and sarcastic recognition in political speech, these models do very well [11].

V CONVENTIONAL MACHINE LEARNING ALGORITHMS

In the study of election-related speeches, conventional techniques for machine learning are essential for tasks including sentiment analysis, categorization, grouping, and subject exploration. The efficacy, comprehension, and accuracy of these algorithms in processing structured feature data extracted from voice text make them popular.

1. Naive Bayes

The theorem of Bayes serves as the foundation for the stochastic classifier naive Bayes, which assumes feature neutrality. Its effectiveness and ease of use allow it to serve very well in language classification tasks, including sentiment or subject categorization in political conversations, notwithstanding this reducing assumption [12].

2. Support Vector Machines (SVM)

SVMs are strong models for learning under supervision that perform well in features spaces with many dimensions. They are especially helpful for differentiating various political ideas, attitudes, or mood in speeches because they provide the best hyperplanes to divide data into distinct groups [13].

3. Logistic Regression

One linear equation for estimating the likelihood of categorical outcomes is logistic regression. Its interpretability makes it popular for situations involving binary or multiclass categorization. This makes it appropriate for delicate political situations when it is essential to comprehend the logic of the concept [14].

4. Decision Trees

To produce a decision structure like a tree, decision trees recursively divide data according to feature values. They are appropriate for exploratory examination of political rhetoric and speech patterns of reasoning due to their transparency and intuitive character [10].

5. Random Forests

Random Forests are ensemble techniques that prevent overfitting and increase classification accuracy by combining many decision trees. When dealing with complicated patterns in political speech datasets, including different speaker groups or mixed rhetorical styles, this method works very well [15].

6. K-Means Clustering

K-Means is an unsupervised technique that uses feature similarity to divide talks into k different groups. In vast collections of political writings, it is useful for locating speaker parts or subject clusters [16].

7. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN can find clusters of any size and form since it bases its cluster identification on the density of data points. With no need to pre-specify the number of clusters, it is helpful in identifying specialized or fringe issues in political speech corpora [17].

Table 1: Traditional Machine Learning Models: Method, Opportunities and Challenges [12-17]

Method	Opportunities	Challenges	Examples/Case Studies
Naive Bayes	✓ Fast, easy to implement	✓ Assumes feature independence	✓ Sentiment analysis of political tweets
SVM	✓ Handles high-dimensional data well	✓ Computational cost, needs tuning	✓ Political stance detection
Logistic Regression	✓ Interpretable model	✓ Limited to linear relationships	✓ Predicting policy support from speech content
Decision Trees	✓ Transparent, rule-based classification	✓ Prone to overfitting	✓ Classification of rhetorical devices
Random Forests	✓ Reduces overfitting, good accuracy	✓ Less interpretable than single trees	✓ Detecting hate speech in political discourse
K-Means	✓ Simple and fast clustering	✓ Sensitive to initial conditions	✓ Grouping political leaders by topic usage
DBSCAN	✓ Handles noise and arbitrary	✓ Parameter sensitivity	✓ Unsupervised discovery of

	cluster shapes		discourse communities

VI DEEP LEARNING ALGORITHMS

Neural Networks with Recurrence (RNNs):

Because they can recognize intricate patterns in data, especially in sequential and contextual tasks, deep learning models are effective tools for studying political leaders' statements. Various model types provide unique benefits based on the speech's characteristics and the analysis's objectives [19]. Because recurrent neural networks (RNNs) are made to process inputs sequentially, they are ideal for jobs where word order is important. But long-term dependencies are a problem for RNNs, which means they have trouble relating early speech segments to later material. This makes them less useful for evaluating lengthier political speeches, as comprehension of the overall message often depends on context from the outset.

Both Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) networks

Long Short-Term Memory (LSTM) networks and Gated Recurrent Units (GRUs) were created to overcome this constraint. In order to assist them recall long-range relationships in the data, these models use gating techniques, which enable them to store relevant information across longer sequences [20]. Long, intricate political speeches with meaning that often changes over time are much easier for LSTMs and GRUs to handle. These models work especially well for sentiment analysis or topic modeling in political debate, when it's critical to comprehend changes in themes or tone.

Detecting Local Patterns using Convolutional Neural Networks (CNNs):

Convolutional neural networks, or CNNs, on the other hand, are usually used in image processing, however they have been modified for text analysis. CNNs use convolutional filters on the input sequence to detect local patterns in text, such key n-grams or brief sentences. Because of this, they may be used to extract characteristics from brief political statements that may be suggestive of political attitude or position [19–21]. CNNs are good at identifying local dependencies, but they are not as good at handling long-term context as transformers or LSTMs. Nonetheless, they may provide a more thorough study when paired with other models, such as LSTMs.

Transformers' Power: RoBERTa, T5, and BERT:

Transformers, which use attention processes to analyze complete sequences at once, have raised the bar for natural language processing. Examples of these models are BERT, RoBERTa, and T5. In contrast to RNNs, which analyze words individually, transformers are able to take into account the whole context of a speech or phrase simultaneously [22, 23], which enables them to pick up on subtle meaning subtleties and long-range relationships. This makes them particularly effective for applications like political speech sentiment analysis, topic recognition, and ideological categorization. The ability of BERT and its variations to comprehend the underlying themes, tones, and intents in political debate is at the cutting edge of performance.

Speech Analysis and Simulation Using Generative Models

It is also important to analyze political speeches using generative models such as GPT-2 and GPT-3. These models are intended to produce content that is logical and appropriate for the context when given input prompts [19–24]. For activities like creating speech drafts, simulating political debate, and summarizing speeches, they are very helpful. Generative algorithms may also help forecast possible speeches or replies by comprehending the words and rhetorical devices employed in political speeches. Furthermore, they may be used to examine the persuasive methods employed by political figures, providing valuable perspectives on their rhetorical approaches.

Table 2: Deep Learning Models [19-24]

Method	Opportunities	Challenges	Examples/Case Studies
RNN	✓ Models sequential data	✓ Poor long-term memory	✓ Sentence-level speech analysis
LSTM/GRU	✓ Captures long-range dependencies	✓ Slower training than CNNs	✓ Temporal emotion trends in political campaigns
CNN	✓ Extracts local features efficiently	✓ Limited sequential understanding	✓ Phrase-level sentiment detection
BERT/Transformers	✓ Rich context awareness, pre-trained models	✓ Resource-intensive	✓ Named entity recognition, argument structure analysis
GPT-2, GPT-3	✓ Text generation and simulation	✓ Bias in generation, data requirement	✓ Simulating political statements or summaries

VII SENTIMENT AND EMOTION ANALYSIS

Comprehension of the feelings and rhetorical tools used in political leaders' speeches requires a thorough comprehension of sentiment and emotion analysis. In order to shape ideology, mobilize support, or affect public opinion, political speech often uses intricate emotional appeals. From rule-based approaches to sophisticated deep learning techniques, NLP provides a variety of tools and models to identify and decipher these emotional signals [23–25].

VADER (Valence Aware Dictionary for Sentiment Reasoning) is a popular analysis of sentiment tool. Specifically created to assess sentiment in political speech and social media, VADER is a rule-based approach. In addition to positive, neutral, and negative evaluations for a particular paragraph, it offers a compound sentiment score. Campaign slogans and argument soundbites are examples of brief, informal, and emotionally charged comments that VADER is especially good at assessing.

TextBlob, a lexicon-based model that rates text's subjectivity and polarity (from -1 to 1), is another well-liked tool. Despite its simplicity and ease of use, TextBlob is best suited for generic sentiment detection and could miss the complex emotional tone of more complex political discourse [23–25]. Nonetheless, it works well for initial analysis or identifying sentiment trends in big datasets.

DeepMoji provides a potent substitute for a more profound emotional comprehension. This model can identify subtle emotional signals like sarcasm, empathy, or wrath since it was trained on millions of tweets with emoji labels. It is especially helpful in determining the underlying emotional tactics politicians use to engage their audiences because of its capacity to distinguish subtle emotional states.

The most advanced methods for identifying complex emotions in political discourse have lately surfaced: BERT-based emotion classifiers [25]. These models use the BERT architecture's extensive contextual awareness to pinpoint nuanced emotional emotions present in political debate. BERT-based classifiers, in contrast to rule-based models, are able to recognize emotion in context and differentiate between minute changes in tone, such worry vs indignation or hope against optimism. This degree of emotional detail is especially helpful for monitoring how a leader's tone changes throughout a speech or during a campaign.

All things considered, sentiment and emotion analysis using these instruments offers important new perspectives on the emotional content of political discourse. Analysts and researchers may get a more comprehensive knowledge of the persuasive and emotional power present in political speeches by integrating rule-based, lexicon-based, and deep learning methodologies.

Table 3: Sentiment and Emotion Analysis Methods [23-25]

Method	Opportunities	Challenges	Examples/Case Studies
VADER	✓ Real-time, interpretable sentiment	✓ Surface-level analysis	✓ Monitoring live political debates
TextBlob	✓ Easy integration and fast results	✓ Less accurate than deep models	✓ Quick polarity checks on social media posts
DeepMoji	✓ Rich emotional insights	✓ Trained on emoji data, domain mismatch	✓ Emotion detection in campaign speeches
BERT Classifier	✓ Nuanced emotion classification	✓ Needs labeled emotional data	✓ Anger/fear detection in nationalist rhetoric

VIII DISCOURSE AND TOPIC ANALYSIS

Finding major themes, rhetorical patterns, and important policy concerns in a vast corpus of political conversation is made possible by topic modeling and summarization, two crucial methods in political speech analysis [11, 22]. These techniques assist in distilling long and intricate speeches into succinct summaries, which facilitates the tracking of changing political agendas and the comparison of discourse across leaders or parties.

Latent Dirichlet Allocation (LDA) is one of the most used models for topic modeling. A probabilistic model called LDA finds word clusters that regularly occur together to provide interpretable subjects. It is assumed that every text has a variety of themes, each of which is a word distribution. LDA generates subjects that are quite interpretable and accessible by humans, but it works on the premise that words are independent, which may restrict its capacity to grasp more complex semantic links, particularly in complex political jargon.

More recent models such as BERTopic have been created to overcome this constraint [11, 22]. BERTopic groups texts that are semantically related and extracts relevant topics by combining clustering techniques with transformer-based embeddings (from BERT, for example). In contrast to classical LDA, this method enables the model to take use of the contextual richness acquired by transformers, producing more logical and complex topics. When evaluating political speeches, where word meaning and context are crucial to comprehending ideological framing and policy purpose, BERTopic is particularly helpful.

Graph-based options are provided by TextRank and LexRank for keyword extraction and summarization. Using a ranking algorithm modeled after Google's PageRank, these systems construct networks of words or phrases and assign a value [11, 22]. By finding words that commonly appear together and form connections across the text, TextRank is especially good at extracting important phrases from speeches. In contrast, LexRank emphasizes similarity at the sentence level and is often used in extractive summarizing, which selects important phrases from the source text to create a summary. Because both approaches are unsupervised and somewhat light, they may be used for extensive political text analysis.

Transformer-based models like T5 (Text-to-Text Transfer Transformer) and PEGASUS provide strong abstractive summarizing solutions for increasingly complex summary jobs. Instead than only extracting passages from the original text, these models create new sentences that convey its meaning. T5 is a flexible model that transforms all NLP tasks into a text-to-text format, but PEGASUS is tailored to produce summaries that are human-like and particularly made for summarizing jobs. These methods may provide succinct, logical summaries of political speeches that preserve the main points of the speech while demythologizing intricate language for simpler study or public distribution.

To conclude, topic modeling and summarization tools are essential to political speech analysis, ranging from sophisticated transformer models to more conventional LDA and graph-based techniques [11, 22]. They make it possible for scholars to discern new themes, draw insightful conclusions from vast amounts of material, and effectively convey the main idea of political discourse.

Table 4: Discourse and Topic Analysis Methods [11, 22]

Method	Opportunities	Challenges	Examples/Case Studies
LDA	✓ Interpretable topic modeling	✓ Assumes word independence	✓ Identifying themes in parliamentary debates
BERTopic	✓ Uses contextual embedding for better coherence	✓ Computationally expensive	✓ Grouping political narratives
TextRank/LexRank	✓ No training required, graph-based summarization	✓ May miss deeper semantics	✓ Extracting main points from press conferences
PEGASUS/T5	✓ Generates fluent summaries	✓ Requires large GPU and training data	✓ Abstract summaries of political addresses

IX COMPARISON OF ALGORITHMS

Numerous algorithms with distinct advantages and disadvantages are used in the area of political speech analysis [1–5]. Traditional methods such as TF-IDF and Bag of Words (BoW) provide comprehension and ease for text representation, which makes them appropriate for baseline analysis. Nevertheless, these approaches disregard word order and semantic significance. Word2Vec and BERT, on the other hand, provide richer language representation via situational embedding. In political texts, where tone and rhetoric have a big impact on meaning, BERT is particularly useful since it catches the complex interactions between words in context.

Algorithms such as Naive Bayes and SVM are useful for classifying sentiment and topics in classification, clustering, sentiment analysis, and emotion analysis; however, they need labeled data. Once again, BERT's extensive contextual awareness allows it to provide excellent accuracy. Latent themes in unstructured political speech data may be found using unsupervised algorithms like K-Means, DBSCAN, and BERTopic for topic modeling and clustering. BERTopic's use of transformer embeddings makes it unique, but compared to more straightforward models like LDA, it may be computationally demanding and less interpretable. Sentiment techniques like as VADER and TextBlob are simple and lightweight, but they could overlook subtle emotional signals that are better captured by models like DeepMoji and BERT-based emotion classifiers.

Graph-based algorithms such as TextRank are effective for extractive summarizing, keyword extraction, and model interpretability, whereas transformer-based models like PEGASUS and T5 provide abstractive summaries that are more human-like [11]. These are helpful for keeping the main points of long political speeches while cutting them down. However, they could overlook key rhetorical devices or oversimplify. Finally, in politically sensitive situations where openness is critical, tools like as LIME and SHAP are useful for providing an explanation of model predictions. Their implementation may result in increased processing cost, despite the fact that they boost confidence in AI systems. Overall, the work at hand, the intended balance between efficiency and interpretability, and the complexity of the political language under analysis all influence the algorithm selection [22].

X CONCLUSION

Political speech analysis using Natural Language Processing (NLP) provides a full range of tools for comprehending the thematic substance, mood, emotion, and structure of political discourse. Each algorithm is essential to deciphering the intricate and calculated language usage of political leaders, ranging from simple text representation methods like TF-IDF and Word2Vec to sophisticated contextual models like BERT and GPT. While subject modeling and summarizing techniques reduce speeches to their essential points, sentiment and emotion analysis tools assist in revealing the underlying emotional tone. This allows for effective comparison and interpretation across leaders, parties, and historical periods. To accommodate a variety of language difficulties and data sources, the combination of rule-based and deep learning methodologies also offers flexibility and depth in analysis. Modern transformer-based models are superior at capturing context and complex meanings, even if they come at a larger computing cost than older techniques, which are simpler and easier to grasp. Particularly in politically delicate situations, these insights may be comprehended and relied upon thanks to model explainability tools like LIME and SHAP. All things considered, NLP enables scholars, analysts, and decision-makers to acquire data-driven understandings of political communication, facilitating better decision-making, media analysis, and public debate. As these technologies develop further, they offer even more clarity and accuracy in interpreting the language of influence and leadership.

REFERENCES

[1] D. Jurafsky and J. H. Martin, “Speech and Language Processing”, 3rd ed., Pearson, 2022.

[2] Y. Goldberg, "A Primer on Neural Network Models for Natural Language Processing," J. Artif. Intell. Res., vol. 57, pp. 345–420, 2016.

[3] Gaikwad, V. S. ., Shivaji Deore, S. ., Poddar, G. M. ., V. Patil, R. ., Sandeep Hirolikar, D. ., Pravin Borawake, M. ., & Swarnkar, S. K. . (2024), “Unveiling Market Dynamics through Machine Learning: Strategic Insights and Analysis”, International Journal of Intelligent Systems and Applications in Engineering, 12(14s), 388–397, 2024

[4] Suwarna J, Kirti N. Mahajan, Yogesh B. Pawar, Yogita D. Bhise, Bharti Jagdale, Rajendra V. Patil, “Plant Growth Analysis using IoT and Reinforcement Learning Techniques for Controlled Environment”, Advances in Nonlinear Variational Inequalities, 27(3), pp. 706–715, 2024

[5] S. Bird, E. Klein, and E. Loper, “Natural Language Processing with Python”, O’Reilly, 2009.

[6] Rajendra V. Patil, Dr. Renu Aggarwal ”Comprehensive Review on Image Segmentation Applications”, Sci.Int.(Lahore), 35(5), pp. 573-579, Sep. 2023

[7] Rajendra V. Patil., Dr. Renu Aggarwal, “Edge Information based Seed Placement Guidance to Single Seeded Region Growing Algorithm”, International Journal of Intelligent Systems and Applications in Engineering, 12(12s), 753–759, 2024

[8] Rajendra V. Patil, Dr. Renu Aggarwal, Govind M. Poddar, M. Bhowmik. Manohar K. Patil, “Embedded Integration Strategy to Image Segmentation Using Canny Edge and K-Means Algorithm”, International Journal of Intelligent Systems and Applications in Engineering, 12(13s), 01–08, 2024 Retrieved from

[9] C. D. Manning, P. Raghavan, and H. Schütze, “Introduction to Information Retrieval”, Cambridge University Press, 2008.

[10] T. Mikolov et al., "Distributed Representations of Words and Phrases," in Proc. NIPS, 2013.

[11] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, 2019.

[12] A. McCallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," in Proc. AAAI Workshop, 1998.

[13] T. Joachims, "Text Categorization with Support Vector Machines," in Proc. ECML, 1998.

[14] I. Guyon et al., "Design of Experiments for the NIPS 2003 Variable Selection Benchmark," in Adv. Neural Inf. Process. Syst., 2003.

[15] M. T. Ribeiro et al., "Why Should I Trust You?: Explaining the Predictions of Any Classifier," in Proc. KDD, 2016.

[16] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in Proc. NIPS, 2017.

[17] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in Proc. 5th Berkeley Symp. Math. Statist. Prob., 1967.

[18] M. Ester et al., "A Density-Based Algorithm for Discovering Clusters," in Proc. KDD, 1996.

[19] Y. Bengio et al., "Learning Long-Term Dependencies with Gradient Descent is Difficult," IEEE Trans. Neural Netw., vol. 5, no. 2, 1994.

[20] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, 1997.

[21] Y. Kim, "Convolutional Neural Networks for Sentence Classification," in Proc. EMNLP, 2014.

[22] A. Radford et al., "Language Models are Few-Shot Learners," OpenAI, 2020.

[23] C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis," in Proc. ICWSM, 2014.

[24] S. Loria, "TextBlob: Simplified Text Processing," [Online]. Available: https://textblob.readthedocs.io/.

[25] B. Felbo et al., "Using Millions of Emoji Occurrences to Learn Any-domain Representations for Detecting Sentiment, Emotion and Sarcasm," in *Proc. EMNLP*, 2017.

[26] Rajendra V. Patil, Dr. Renu Aggarwal, Dr. Shailesh Shivaji Deore, “Edge Segmentation based Illumination Invariant Feature Detector Phase Congreuncy”, 2024 5th International conference on Mobile Computing and Sustainable Informatics (ICMCSI), pp. 91-96, Jan. 2024

[27] K. C. Jondhale , R. V. Patil, "Edge Based technique to estimate Number of clusters in K-means Color Image Segmentation", *3rd IEEE International Conference on Computer Science and Info Tech, Chegndu , China, 2010*

[28] R. V. Patil, K. C. Jondhale,"Content Based Image Retrieval based on Phase congruency Via Log Gabor Wavelet Filtres", Proceedings of ICCVGIVP 2009, Nagpur, pp 84-85, 13th –14th March 2009.

[29] Sagar V. Joshi, Rajendra V. Patil, Manoj Tarambale, Balkrishna K Patil, Vinit Khetani, Yatin Gandhi, "Stochastic Processes in the Analysis of Electrical Load Forecasting", Advances in Nonlinear Variational Inequalities, 28(1), pp. 15-29, 2025.

[30] Gaikwad, V. S. ., Shivaji Deore, S. ., Poddar, G. M. ., V. Patil, R. ., Sandeep Hirolikar, D. ., Pravin Borawake, M. ., & Swarnkar, S. K. . (2024), "Unveiling Market Dynamics through Machine Learning: Strategic Insights and Analysis", International Journal of Intelligent Systems and Applications in Engineering, 12(14s), 388–397. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/4675

[31] Suwarna J, Kirti N. Mahajan, Yogesh B. Pawar, Yogita D. Bhise, Bharti Jagdale, Rajendra V. Patil, "Plant Growth Analysis using IoT and Reinforcement Learning Techniques for Controlled Environment", Advances in Nonlinear Variational Inequalities, 27(3), pp. 706–715, 2024

[32] Sangeeta Kakarwal, Rahul Mapari, Pranoti Prashant Mane, Madhuri Pravin Borawake, Dhanraj R. Dhotre, Rajendra V. Patil, "A Novel Approach for Detection, Segmentation, and Classification of Brain Tumors in MRI Images Using Neural Network and Special C Means Fuzzy Clustering Techniques", Advances in Nonlinear Variational Inequalities, 27(3), pp. 837–872, 2024.

[33] Tarambale, M. ., Naik, K. ., Patil, R. M. ., Patil, R. V. ., Deore, S. S. ., & Bhowmik, M., "Detecting Fraudulent Patterns: Real-Time Identification using Machine Learning", International Journal of Intelligent Systems and Applications in Engineering, 12(14s), 650, 2024

[34] Govind M. Poddar, Rajendra V. Patill, Satish Kumar N, "Approaches to handle Data Imbalance Problem in Predictive Machine Learning Models: A Comprehensive Review", *Int J Intell Syst Appl Eng*, vol. 12, no. 21s, pp. 841–856, Mar. 2024.

[35] Suwarna J, Kirti N. Mahajan, Yogesh B. Pawar, Yogita D. Bhise, Bharti Jagdale, Rajendra V. Patil, "Plant Growth Analysis using IoT and Reinforcement Learning Techniques for Controlled Environment", Advances in Nonlinear Variational Inequalities, 27(3), pp. 706–715, 2024.