Chapter 1

1. **Introduction**

This aspect of the research contains information on the background of study, problem statement, objectives and the organization of the study.

**1.1 Background Of Study**

The State of the Nation Address (SONA) is an annual speech to Parliament given by the President of the Republic of Ghana, covering the economic, social, and financial state of the country, as mandated by Article 67 of the 1992 Constitution of Ghana (Ghana Web, 2024.).The SONA was first implemented under the administration of the second President of the Fourth Republic of Ghana, John Agyekum Kufuor. Records indicate that during his 18 years in office, former President Jerry John Rawlings did not deliver a single State of the Nation Address, despite the constitutional requirement (Ghana Constitution, 2024). Former President John Agyekum Kufuor was the first to adhere strictly to the constitutional mandate, delivering the SONA at the beginning and close of every parliamentary session. Since then, the SONA has become a consistent feature of Ghana's political landscape (Ghana Web, 2024.). Interestingly, the tradition of delivering such an address in person before a joint session of Congress dates back to George Washington, who first fulfilled this presidential duty on January 8, 1790, in New York City(*State of the Union*, 2022).

The SONA warrants critical analysis and scrutiny as it is one of the President’s most significant formal governance and communication events, providing a blueprint for the nation’s developmental agenda for the year (Sikanku, 2022). With how important these speeches are one would think that they are written to perfection without any flaw or deceit behind them but that is not the case. A blog by Julius Yao Petetsi on the Ghana Times website expressed how a State of the Nation Address (SONA) by then-President Nana Addo Dankwah Akufo-Addo did not accurately reflect the true state of the country. He continued to explain how a minority was able to detect instances of deceit in the speech and expressed in a paragraph that in the view of Mr Iddrisu, the success reported by the president in the various sectors of the economy including health, education, roads, security amongst others were just a rehatch of the promises that won him the election(JULIUS YAO PETETSI, 2022). The problem with deceit and false information in these speeches is that it can be very difficult to detect, especially when the information is outdated.

Artificial intelligence is developing at an incredibly fast pace. The potential is enormous and it's hard to see where it will end. Artificial intelligence is based on maths and logic. We know the work processes, but we don't always know how the AI arrives at a particular solution. Therefore, as researchers and society, we must make demands on the use of technology, both in legislation and morally. In Ghana, there is a need for a system that can analyze these SONA speeches by the various presidents, categorize them, and identify instances of deception and false information. In this study, we employ Natural Language Processing (NLP) and machine learning algorithms to perform a classification analysis of the SONA texts. We aim to classify segments of the speeches based on sentiment (positive/negative) and truthfulness (truth/lie), thereby uncovering patterns in political communication. Using NLP techniques such as tokenization, lemmatization, and parts of speech tagging, we pre-process the text data to make it suitable for analysis(Chai, 2023). We then apply various machine learning models, including Support Vector Machines (SVM), Random Forests, and neural networks, to classify the speech segments. This approach not only enhances our understanding of the content and tone of the addresses but also contributes to the broader field of political discourse analysis(Sagar, 2023). By leveraging advanced computational methods, this research aims to shed light on the dynamics of presidential communication in Ghana, providing a data-driven perspective on the effectiveness and transparency of the State of the Nation Addresses.

**1.2 Problem Statement**

In Ghana the analysis of SONA is important to individuals as it gives an overview of the current position and stand of the governance as well as the economic condition of the country which is the reason why most press institutions give a commentary on it when it is delivered, and example being a commentary of the Ghana state of the nation address which was published on the Ghana Investment promotion website(Benjamin A. Alomatu, 2024). Most of the analysis done on SONA are from these online sources or the internet and recent research has proven that the internet isn’t a reliable place for information. The information on the analysis can be twisted to the view or to suit the way the analyser(*LibGuides: Evaluating Information Sources: Should I Trust Internet Sources?*, n.d.).

SONA speeches are lengthy and to be able to understand and grasp which aspects of the speech is true or false a huge amount of research is needed on that single document. Rigorous fact checking and analysis is required which can be time consuming and put significant toll on the individual. In addition to text analysis being time consuming it is also very costly as efficient analysis would require the person who wants the analysis to pay multiple people to work on efficient fact-checking of the information he or she wants to derive from the set of data(Phillip P. Adu, 2022).

There are various text analysis automated systems however most of the ones available aren’t really focussed on deriving the facts in SONA and the ones available incorporate analysis of information which could be termed as irrelevant in our research(Shashank S. Gupta, 2018). There is a need for a simpler analysis of SONA text speeches and classification to reduce the strain of traditional methods of classification of text speeches(Phillip P. Adu, 2022).

**1.3 Objectives**

**1.3.1 Main Objective**

The primary objective of this research is to develop a machine-learning model that can accurately categorize speech content into distinct segments based on sentiment (positive/negative) and truthfulness (truth/lie).

**1.3.2 Specific Objectives**

This research seeks to:

1. Automate the process of scanning through SONA texts.
2. Using Python and NLP libraries to pre-processed scanned texts this is transforming the speeches so it can be understood by the computer or the model.
3. Perform fact-checking and use pre-process data to perform classification analysis and label the data on whether it is true or false or has neutral elements.

**1.5 Significance of Study**

This study will provide a much simpler alternative to the robust and difficult to use analysis systems. It also provides a specific outcome as it is focussed mainly on identifying deceit or truthfulness in SONA speeches. By locating or identifying instances of deceit and false delivery of information, the study adds to greater transparency in political communication.The research can hold political leaders accountable for the accuracy of their statements, thereby promoting integrity in governance and the nation as a whole.

The research aims to aid individual that will indulge in the analysis of SONA speeches have a system that can reduce fact-checking tasks so they can accurately and efficiently analyse the text without much problems or stess.The study offers a data-driven on the effectiveness and transparency of SONAs, helping government officials, researchers, and the public in understanding the presidents and the information they share to us Ghanaians.

**1.6 Organisation of study**

This study is a five parts study with aspects of the stufy including labelled chapters, from Chapter One to Chapter Five. Chapter one encompases the introduction of the study. This aspect of the study contains the background of the study, problem statement, main as well as the specific objectives, and the significance of the study. Chapter two is where the literature review that backs our study will be. In Chapter three we discuss the steps and the procedures as well as various tools and techniques needed to complete the work. We explain how we build the models and perform text analysis with SONA speeches. We also discuss the data types, and the data collection methods. In chapter four, we discuss the results of our study. In this section of the work we discuss the results we were able to derive from the methodology we implemented in our work and how efficient and effective the model was able to process the text data analyse and classify them. Chapter 5 concludes the study. It contains a summary of the entire work as well as expectations for the future and some recommendations for future works. In addition, all references used in the development of this stufy will be added at the end of the final chapter.

**CHAPTER TWO**

**Literature Review**

**2.0 Introduction**

This chapter is the literature review of this research where definition of concepts, and subtopics as well as related works and some summary of works can be found.The conclusion is the final part of this chapter given an overview of the literature review.

**2.1. Definition of Concepts**

**2.1.1 State of The Nation Address (SONA)**

The State of the Nation is an annual address to Parliament given by the President of the republic of Ghana covering economic, social, and financial state of the country according to Article 67 of the 1992 constitution of Ghana(Ghana Web, n.d.).

**2.1.2 Machine Learning**

According to a webpost on the MIS website Machine learning is a subfield of artificial intelligence, which is broadly defined as the capability of a machine to imitate intelligent human behavior. Artificial intelligence systems are used to perform complex tasks in a way that is similar to how humans solve problems(*Machine Learning, Explained | MIT Sloan*, 2021).

**2.1.3 Natural Language Processing**

Natural language processing (NLP) is the discipline of building machines that can manipulate human language or data that resembles human language in the way that it is written, spoken, and organized. It evolved from computational linguistics, which uses computer science to understand the principles of language, but rather than developing theoretical frameworks, NLP is an engineering discipline that seeks to build technology to accomplish useful tasks. NLP can be divided into two overlapping subfields: natural language understanding (NLU), which focuses on semantic analysis or determining the intended meaning of text, and natural language generation (NLG), which focuses on text generation by a machine(*Natural Language Processing (NLP) [A Complete Guide]*, 2023).

**2.1.4 Text Classification**

Text classification is a machine learning technique that automatically assigns tags or categories to text. Using natural language processing (NLP), text classifiers can analyze and sort text by sentiment, topic, and customer intent – faster and more accurately than humans(*Go-to Guide for Text Classification with Machine Learning*, 2020).

**2.1.5 Text Pre-processing**

Text Processing pertains to the analysis of text data using a programming language such as Python. Text Processing is an essential task in NLP as it helps to clean and transform raw data into a suitable format used for analysis or modelling(GeeksforGeeks, 2024).

**Subtopics**

**2.2.1 Text Classification in State of the Nation Addresses**

Text classification is an important aspect in Natural language (NLP) that involves categorizing text into predefined labels. In the area of State of the Nation addresses, text classification can be used to identify key themes, sentiments, and policy areas discussed in these speeches (Devlin et al., 2019). This process mostly involves several steps of text pre-processing, including tokenization, stemming, and stop-word removal, to prepare the text for classification algorithms (Ehud & Dale, 2020).

**2.2.2 NLP Techniques for Text Pre-processing**

Effective text pre-processing is important for the success of text categorizing tasks. Techniques such as tokenization, which breaks down text into individual words or tokens, and stemming, which reduces words to their root forms, are commonly used in the pre-processing of text before implemented in a model (*Speech and Language Processing*, n.d.) Stop-word removal is another essential step, as it eliminates common words that do not contribute to the meaning of the text (Manning et al., 2008). These pre-processing steps help in reducing the of the text data and improve the performance of classification algorithms.

**2.2.3 Machine Learning Algorithms for Text Classification**

Various machine learning algorithms can be employed for text classification tasks. Traditional algorithms like Naive Bayes, Support Vector Machines (SVM), and decision trees have been widely used due to their simplicity and effectiveness(Kowsari et al., 2019a). However, recent advancements have seen the rise of deep learning models, particularly those based on neural networks, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which have shown superior performance in text classification tasks (Yin et al., 2020).

**2.2.4 State of the Nation Address Analysis Using NLP**

Analyzing State of the Nation addresses using NLP techniques involves extracting meaningful insights from the text. This can include identifying the frequency of key terms, sentiment analysis, and topic modeling to understand the main issues addressed by the speaker (Bhatia et al., 2019). Such analyses can provide valuable information about the political and social priorities of a country at a given time.

**2.2.5 Challenges in Text Classification of State of the Nation Addresses**

One of the primary challenges in text classification of State of the Nation addresses is the complexity and variability of the language used. Political speeches often contain nuanced language, rhetorical devices, and context-specific references that can be difficult for algorithms to interpret accurately (Jurafsky & Martin, 2020). Additionally, the need for large labelled datasets for training machine learning models poses a significant challenge, as manual annotation can be time-consuming and resource-intensive.

**2.2.6 Future Directions in NLP for Political Speech Analysis**

The future of NLP in the analysis of political speeches lies in the integration of more sophisticated models and techniques. Advances in transfer learning, such as the development of pre-trained models like BERT (Bidirectional Encoder Representations from Transformers), have the potential to enhance the accuracy and efficiency of text classification tasks (Devlin et al., n.d.) Moreover, the incorporation of multimodal data, such as combining text with audio and video analysis, could provide a more comprehensive understanding of political speeches (Müller et al., 2020).

**2.3 Review of Related Works or Systems**

A survey conducted by (Li et al., 2022) reviewed the differences and the pros and cons or text classification using Natural language processing as compared to the traditional methods of doing classification. It provides an overview of text classification methods and approaches from 1961 to 2022. And concluded that the more ambitious the classification needed to be performed the more challenging and complex implementation of the model would be. The research provides a comprehensive understanding of other techniques used in the classification of text and the different deep learning models available.

Convolutional neural networks were utilized to develop a sentiment analysis of online products in a reserach by. The aim was analysis to classify online product reviews as positive, negative or neutral. This model had a 92 percent accuracy for categorizing data. Evaluation was done with various methods and it was a nice addition to research utilizing natural language processing methods.

A guide provided on how decision are to be taken before utilizing analysis for true, false or neutral contents by (Barberá et al., 2021) show that two reasonable approaches to corpus selection yield radically different corpora and advocate for the use of keyword searches rather than predefined subject categories provided by news archives. (Barberá et al., 2021) demonstrate the benefits of coding using article segments instead of sentences as units of analysis and show that, given a fixed number of coding, it is better to increase the number of unique documents coded rather than the number of coders for each document. Finally, (Barberá et al., 2021) find that supervised machine learning algorithms outperform dictionaries on a number of criteria.

A research paper written by (Kowsari et al., 2019b) expressed the growing level of text document in the world and how there is a need for the automation of these files in by utilizing Natural language processing. (Kowsari et al., 2019b) also emphasized the significance of text pre-processing before feeding text data into any model as well as the various techniques utilized to complete text classifications.

A paper by Francis Adoma acheampong on text-based emotion detection investigated deep learning techniques for recognizing emotion expressed in text information or data through various social media platforms namely Facebook and twitter as well as customer reviews and chat conversation(Acheampong et al., 2020). The research centred around transformer-based approaches like BERT, and how such models have achieved state of the art performance in emotion recognition tasks. The study which focused on emotion classification metrics had issues with cross-domain generation and interpretability of emotion detection(Acheampong et al., 2020).

Multilingual text classification model for social media analysis was developed based on deep learning and the aim was to accurately categorize social media posts in multiple languages. The project involved the use of pre-processed texts and a trained deep neural network architecture. It achieved a high classification accuracy of 90% across multiple tests and was evaluated using standard metrics such as precision and recall as well as f-1 score(Zade & Ajani, 2022).

(Salminen et al., 2020)designed and implemented a machine learining system for automatically detecting hate speech and toxic content in online forums and other social media platform and utilized support vector machines which was the supervised model used after data preprocessing had taken place. It got an F-1 score of 0.90 compared to traditional machine learnig approaches and was a huge addition to research in the field of machine learning. However the study also had challenges deleting speeches which were unbiased(Salminen et al., 2020).

The Automated fact-checking research by (Thorne et al., 2018) which seeked to unvestigate the use of natural language processing and machine learning techniques for automatically verifying the factual accuracy of claims and statements utilized a created database or dataset which contained claims and associated evidence then proceded to incorporate this into a network based model. The research played a pivotal role in combating misinformation particularly form online sources(Thorne et al., 2018).

Learning towards conversational AI: The paper discusses the evolution of open-domain dialogue systems, highlighting their ability to handle unrestricted conversation topics compared to task-oriented systems. As well as the frameworks utilized in bringing dialogue models to life. It also emphasised how dialogue systems should be informative and controllerble(Fu et al., 2022).

**2.4 Summary of Related Topics/Research**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Project**  **Name** | **Citation** | **Main**  **Objectives** | **Methodologies** | **Results** | **Evaluation**  **On Metrics** | **Conclusion** |
| Sentiment Analysis of Online Product Reviews | (L-, 2023) | Develop a sentiment analysis model to classify online product reviews as positive, negative, or neutral. | preprocessed the text data, and trained a deep learning-based sentiment classification model using convolutional neural networks (CNNs). | CNN-based sentiment analysis model achieved an accuracy of 92% in positive, negative, and neutral categories. | The model was evaluated using standard classification metrics such as accuracy, precision, recall, and F1-score. | valuable insights for businesses to understand customer perceptions and improve their products or services. |
| Fake News Detection using Machine Learning | (Raja & Raj, 2022) | Design and implement a machine learning-based system to detect fake news in social media posts. | Collected a dataset of real and fake news articles, extracted textual features using natural language processing techniques, and trained various machine learning models | The random forest classifier achieved the highest accuracy of distinguishing fake news from real news. | The models were evaluated using accuracy, precision, recall, and F1-score metrics. | Machine learning-based fake news detection can be an effective tool to combat the spread of misinformation on social media platforms. |
| Automated Abstractive Text Summarization using Deep Learning | (Karuna et al., 2023) | The objective is to construct an abstractive text summarizer using deep learning | The algorithm which is been used here is the Long Short Term Memory model (LSTM) which is a type of RNN model. | Summerized text accurately | The models were evaluated using accuracy, precision, recall | Performed assigned tasks allocated to it. |
| An intent recognition pipeline for conversational AI | (Chandrakala et al., 2024) | Investigate the state-of-the-art in intent detection techniques for conversational AI systems, such as chatbots and virtual assistants. | review on intent detection approaches, including rule-based, machine learning, and deep learning methods, and their applications in various domains. | The survey identified the key challenges, techniques, and performance metrics in intent detection for conversational AI, and highlighted emerging trends and future research directions. | The review was based on a systematic analysis of the existing literature, with a focus on the strengths, limitations, and comparative performance of the discussed approaches. | Intent detection is a crucial component of conversational AI, and continued research in this area can lead to more natural and intelligent human-computer interactions. |
| Automated Text Summarization using Deep Learning | (Barberá et al., 2021) | Develop a deep learning-based model for generating abstractive summaries of text documents. | Employed a sequence-to-sequence neural network architecture, specifically a transformer-based model, to learn the mapping from input text to concise summaries. | The deep learning-based text summarization model outperformed traditional extractive and abstractive summarization techniques in terms of ROUGE scores and human evaluation. | The summarization model was evaluated using standard metrics like ROUGE. | Accomplised Research goal. |
| Emotion Recognition from Text using Deep Learning | (Kowsari et al., 2019b) | Investigate the use of deep learning techniques for recognizing emotions in text data. | The study reviewed the existing literature on emotion recognition from text, covering various deep learning architectures (e.g., recurrent neural networks, transformers) | The review found that deep learning models, particularly transformer-based approaches like BERT, have achieved state-of-the-art performance in emotion recognition tasks. | The reviewed studies were evaluated based on standard emotion classification metrics, such as accuracy, F1-score, and area under the curve (AUC). | Emotion recognition from text using deep learning has numerous applications in customer experience management, mental health monitoring, and human-computer interaction |
| Multilingual Text Classification for Social Media Analysis | (Acheampong et al., 2020) | accurately categorize social media posts in multiple languages | The project involved the use of pre-processed texts and a trained deep neural network architecture | It achieved a high classification accuracy of 90% across multiple tests | The models were evaluated using accuracy, precision, recall, and F1-score metrics. | Achieved focus of study and was fairly accurate. |
| Automated Hate Speech Detection in Online Communities | (Salminen et al., 2020) | Design and implement a machine learning-based system for automatically detecting hate speech and toxic content | The researchers collected a dataset of user-generated content from various online platforms, annotated the data for the presence of hate speech, and trained supervised learning models | The deep learning-based hate speech detection models achieved superior performance compared to traditional machine learning approaches | F1-score, and area under the receiver operating characteristic (ROC) curve. | Achieved focus study. |
| Automated Fact-Checking using Natural Language Processing | (Thorne et al., 2018) | Use of NLP and machine learning techniques for automatically verifying the factual accuracy of claims or statements | Created the FEVER (Fact Extraction and Verification) dataset and then developed neural network-based models to classify the claims as supported | FEVER dataset, with the best-performing systems reaching an F1-score of over 0.65. | The fact-checking models were evaluated using standard metrics like accuracy, precision, recall, and F1-score, as well as task-specific measures like FEVER score. | The fact-checking models were evaluated using standard metrics like accuracy, precision, recall, and F1-score, as well as task-specific measures like FEVER score. |
| Learning towards conversational AI: A survey | (Chandrakala et al., 2024) | he goal of the paper is to address dialogue act (DA) classification in domain-independent conversations. | Retrieval based methods, Generation based methods and Hybrid based methods | Informative | Word-overlap measures, embedding based measures and metrics based questions | Research aimed on open  Was up for more scrutiny |

**2.5 Summary of Literature review**

The literature review has highlighted the various concepts in NLP and Machine learning as well as highlighted similar works and how they have been implemented. Existing research provides a solid understanding of the importance of text classification and analysis. The table gives a brief overview on what to expect when dealing with works or research similar to the owns in the table. This chapter allows or shows how important Natural Language Processing is and how it is prevailing very rapidly in out now economy or society. The systems implemented in most of the projects although prevailed were not perfect which goes to show that there is much needed improvement in how we handle NLP and machine learning systems and with time there should be better data retrival, classification and databases as well as models which will make the implementation of robust systems easier.

**Chapter 3**

**3.0** **Introduction**

The third chapter of this project lays down the methodology implemented in order to achieve the aims of this project as well as the objectives which were stated earlier the previous chapters. This chapter will involve a detailed documentation of all the steps involved in text analysis of the SONA speeches as well as how we intend to implement the NLP and Machine learning models into our project. This section of the document will contain sections like data collection methos, system Architecture, system components, algorithms, operational methods and the summary of the documentation which will be provided at the end of this chapter.

**3.1 Research Design**

**3.1.1 Research Philosophy**

The research philosophy adopted for this study is positivism. This approach is suitable as it emphasizes the use of quantitative methods which is objective measurements and the statistical, mathematical, or numerical analysis of data collected through polls, questionnaires, and surveys, or by manipulating pre-existing statistical data using computational techniques. This research relies on researching and finding truths or lies in the text speeches using this philosophy will be of great help to the development of the project. The goal is to produce results that are reliable, simple, and structured.

**3.1.2 Research Type**

This research is deductive and quantitative in nature as it begins with a hypothesis about the ability to classify SONA statements as fulfilled or unfulfilled promises based on text analysis. The study primarily employs quantitative methods to analyse and classify the text data, making it a quantitative research type.

**3.1.3 Sampling Strategy**

In this study, the population consists of all existing SONA speeches. Since the aim of the research is to gather all Ghanaian SONAs in existence there is no need for sampling of the speeches. This approach ensures comprehensive coverage and avoids sampling errors which may occur. Accuracy is a big benefit since analysing the entire population removes sampling errors and provides a complete picture of the data.

**3.1.4** **Data Collection**

This is the first and one of the most crucial part of the project. The task to be completed in this section is to collect all SONA from the year 2000 to the current year at the time of writing this documentation 2024, these speeches which are to be collected are necessary for text analysis.

**3.1.5** **Data Collection Methods**

The Sona speeches are going to be collected across various government and news websites namely Ghanaweb, GhanaToday, CityFmOnline , Wikipedia and the largest database coming from the official parliament of Ghana Website.

**3.1.6** **Data Requirements**

There is a need for data collected to be strictly text as the objective of the project is to do analysis on only text data and label or classify it. Also, data collected will be pdfs which will then be converted to csv for model training.

**3.2 System Architecture**

**3.2.1** **Data Labelling**

Collecting of all SONA data is the easy part of the project, here comes the hard part. Here out job will be to scan through the collected SONA documents and identify claims in the speeches and get corresponding year and speaker then proceed to create a csv file that will house all data so it can be used to train the model.

**3.2.2** **Data Pre-processing**

After the data has been collected there is a need for us to convert the SONA claims in the csv file into a format that will make the computer understand the claims and provide effective and efficient classification. We will be using pre-processing techniques like

**3.2.3 Feature Extraction Module**

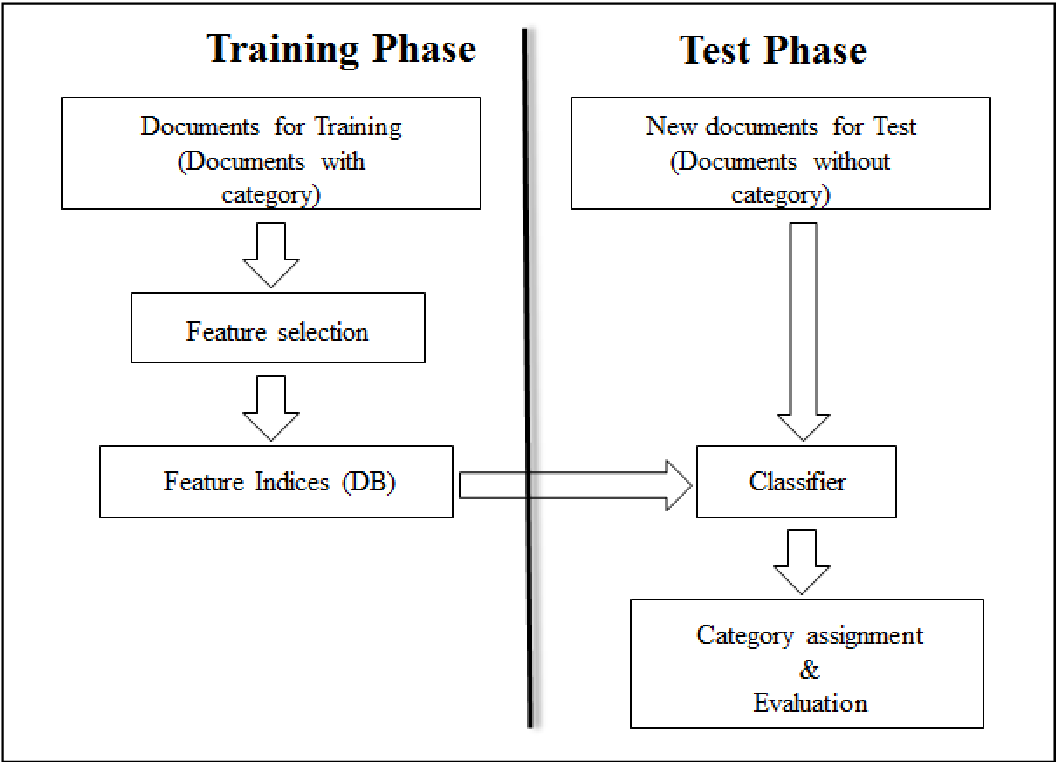
Raw text data cannot be used by the computer, it is important to convert the text data into numerical values so that it can be used by the computer. Feature extraction converts the text into a structured numerical format that algorithms can process. Properly extracted features can significantly improve the accuracy and efficiency of the classification model.

**3.2.4 Machine Learning Classification Module**

Classification models can learn patterns from labelled training data, this will be useful for classifying fulfilled and unfulfilled promises in the SONA texts and then predict the category for new, unseen statements. The model likely to be implemented will be Support Vector Machines or Logistics regression.

**3.2.5 Evaluation and Refinement Module**

Evaluates model performance and refines the models based on feedback and performance metrics.

****

Workflow

[PDF Files] -> [Claim Extraction] -> [Dataset labelling] -> [Pre-processing] -> [Feature Extraction]

-> [Model Training] -> [Model Evaluation] -> [Prediction]

**3.3 System Components**

**3.3.1 Backend**

The system is first and foremost a backend project. Handling large amounts of text SONA speeches, utilizing machine learning and Natural language processing and creating a database storage are all primary features of the backend. Since this project will be utilizing all this aspects in the work its safe to say that the project is a backend project.

**3.3.3 Algorithms**

The algorithms used in this project are:

Logistics regression, Random Forest Classifiers and Decision Tree Classifier.

**3.4. Operational Methods**

**1. Fact-Checking and Accountability:**

The text classification model can be used to systematically analyze new SONA speeches and identify promises that are classified as "fulfilled" or "unfulfilled".This information can be used to hold the government accountable and track their performance in delivering on their stated commitments.The insights can be shared with the public, media, and other stakeholders to increase transparency and democratic oversight.

**2. Policy Evaluation and Policymaking:**

The analysis of past SONA speeches can provide valuable insights into the government's policy priorities, focus areas, and the evolution of their agenda over time.These insights can inform the evaluation of existing policies and the development of new policies that better address the needs and concerns of the citizens.Policymakers can use the findings to align their agenda and resource allocation with the promises and commitments made in the SONA.

**3. Citizen Engagement and Awareness:**

The project outcomes can be used to create interactive visualizations and reports that help citizens better understand the government's performance and the fulfillment of promises.This can empower citizens to engage more actively in the political process and hold their representatives accountable.The analysis can also be used to educate the public on the importance of the SONA and its role in the democratic process.

**4. Academic Research and Teaching:**

The dataset of SONA speeches and the associated analysis can be a valuable resource for academic researchers studying political discourse, rhetoric, and the relationship between promises and outcomes.The project can be used as a case study in courses on political science, public policy, data analysis, and natural language processing.Researchers can build upon the existing work to explore new research questions and expand the scope of the analysis.

**5. Media and Journalism:**

Journalists and media outlets can leverage the project's findings to enhance their reporting on the government's performance and the fulfillment of promises.The analysis can be used to fact-check claims made in the SONA and provide a more objective and data-driven perspective on the government's achievements.Media organizations can collaborate with the project team to integrate the analysis into their coverage and amplify its impact.

**3.5 Limitations Of Project**

The project faces several challenges and limitations. Due to the largeness of the document or SONA text accuracy may be a problem and since the text is also supposed to be pre-processed some valuable information may be lost in the process as well. Also, data that will be used to train model is information available online and may be untrue. The scalability and processing speed of the system are also important considerations, as analysing and classifying the text of multiple SONA speeches over time can be computationally expensive. There may also be ethical and legal considerations around the use of SONA speeches, which could be considered sensitive or confidential government information. The interpretability and explain ability of the machine learning models used for text classification are also important considerations, as complex models may struggle to provide transparent and interpretable explanations for their predictions. Although the model will be useful now further on along the line it can’t be determined if it will retain its usefulness.

**3.6 Conclusion**

This project elaborates on all the aspects and process involved in producing a classification analysis utilizing natural language processes and machine learning algorithms. It also explains how the project is going to be used therefore speaking on its importance as well as the various stages the data goes through to bring about a useful project.