Classifying Claims from Hindi and English Tweet Texts: ML and DL Tricks

Thesis submitted in partial fulfillment of requirements
For the degree of
Master of Technology in Computer Technology
of
Computer Science and Engineering Department
of
Jadavpur University

by

Pradip Debnath
Registration No. - 154205 of 2020 - 2021
Examination Roll No. - M6TCT23015B

under the supervision of

Assistant Professor (Dr.) Dipankar Das

Department of Computer Science and Engineering
Jadavpur University
Kolkata, West Bengal, India

2023

Certificate from the Supervisor

This is to certify that the work embodied in this thesis entitled **"Classifying Claims from Hindi and English Tweet Texts : ML and DL Tricks"** has been satisfactorily completed by **Pradip Debnath** (Registration Number 154205 of 2020 – 2021; Class Roll No. 002010504040; Examination Roll No. M6TCT23015B. It is a bona-fide piece of work carried out under my supervision and guidance at Jadavpur University, Kolkata for partial fulfillment of the requirements for the awarding of the **Master of Technology in Computer Technology** degree of the Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jadavpur University, during the academic year 2022 – 23.

Asst. Prof. (Dr.) Dipankar Das,
Professor,
Department of Computer Science and Engineering,
Jadavpur University.
(Supervisor)

Forwarded By:

Prof. (Dr.) Nandini Mukherjee,
Head,
Department of Computer Science and Engineering,
Jadavpur University.

Prof. (Dr.) Swaswati Mazumder,
DEAN,
Faculty of Engineering & Technology,
Jadavpur University.

Department of Computer Science and Engineering
Faculty of Engineering And Technology
Jadavpur University, Kolkata - 700 032

Certificate of Approval

This is to certify that the thesis entitled **"Classifying Claims from Hindi and English Tweet Texts: ML and DL Tricks"** is a bona-fide record of work carried out by **Pradip Debnath** (Registration Number 154205 of 2020 – 2021; Class Roll No. 002010504040; Examination Roll No. M6TCT23015B) in partial fulfillment of the requirements for the award of the degree of **Master of Technology in Computer Technology** in the **Department of Computer Science and Engineering, Jadavpur University**, during the period of September 2022 to June 2023. It is understood that by this approval, the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn therein but support the thesis only for the purpose of which it has been submitted.

Examiners:

(Signature of The Examiner)

(Signature of The Supervisor)

Department of Computer Science and Engineering
Faculty of Engineering And Technology
Jadavpur University, Kolkata - 700 032

Declaration of Originality and Compliance of Academic Ethics

I hereby declare that the thesis entitled **"Classifying Claims from Hindi and English Tweet Texts: ML and DL Tricks"** contains a literature survey and original research work by the undersigned candidate, as a part of his degree of **Master of Technology in Computer Technology** in the **Department of Computer Science and Engineering, Jadavpur University**. All information has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: Pradip Debnath

Examination Roll No.: M6TCT23015B

Registration No.: 154205 of 2020 – 2021

Thesis Title: Classifying Claims from Hindi and English Tweet Texts: ML and DL Tricks

---

Signature of the Candidate

ACKNOWLEDGEMENT

I am pleased to express my gratitude and regards towards my Project Guide Dr. Dipankar Das, Asst. Professor, Department of Computer Science and Engineering, Jadavpur University, without whose valuable guidance, inspiration, and attention. I have, pursuing my project would have been impossible.

I am grateful to all the members of **Alpha Lab, Department of Computer Science and Engineering, Jadavpur University**, for cooperating with me in all possible ways and providing me with a good working environment throughout the work.

I express my regards to my parents, my elder brother Krishnendu Das and my friends Debam Saha for bearing with me, helping me and for being a source of constant motivation during the entire term of the work.

I would like to thanks my friends Debasis Ray, Pranay Ghosh, Rana Das for constantly supporting me.

I am so grateful towards my family Late Ashwini Debnath, Late Charubala Debnath, Late Rabindra Nath Das, Lipi Das, Sharannya Das, Champakoli Biswas, Gobinda Biswas, Madhusmita Biswas, Chandra Biswas, Shankar Acharya, Monika Acharya, Madhumita Dey, Manash Dey, Ena Dey, Arjya Dey, Sumitra Mondal, Kishore Mondal, Subhankar Mondal, Bandona Mondal, Sadananda Mondal, Subrata Mondal, Puja Mondal and towards my Parampara family Sushanta Das, Debashri Gangopaddhay, Siddhartha Shankar Ghosh, Piyali Sen, Priyanka Mishra, Alok Dutta, Sangita Mondal.

I would also like to thanks my bosses Anurag More, Ankit Biyani, Navin More, Wassim Shaikh and to my colleagues Hitesh Bhageria, Sumanta Das, Aditya Konar, Rishabh Mundra for constantly supporting me throughout the research work...

Pradip Debnath
MTCT Final Year
Exam Roll No. - M6TCT23015B
Regn. No. - 154205 of 2020 – 2021
Department of Computer Science and Engineering,
Jadavpur University.

Abstract

This thesis, "Classifying Claims from Hindi and English Tweet Texts: ML and DL Tricks," tackles the challenge of classifying claim veracity in multilingual tweet texts using machine learning (ML) and deep learning (DL). We collect and pre-process a comprehensive dataset of Hindi and English tweets, annotated for veracity. Our research explores the performance of ML and DL models, including traditional classifiers and neural networks, in differentiating true and false claims. Additionally, we delve into advanced techniques like transfer learning and ensemble to enhance model accuracy. The results showcase the potential of these "tricks" to improve classification across languages. This study contributes to the vital field of social media information verification and offers insights into the development of effective tools for combating misinformation on platforms like Twitter. This thesis highlights the significance of harnessing both ML and DL techniques, along with the employment of "tricks" to enhance their performance, in addressing the pressing issue of claim classification in multilingual tweet texts. Apart from this, the results on this database using a three classifiers and two models architecture. Three classifiers and two models are used: Logistic Regression, Decision Tree, SVM, LSTM, BERT. Among these three ML classifiers Decision Tree achieved 65% the best accuracy for English and two DL models BERT achieved the best accuracy for Hindi, i.e., 75% .

Keywords: Tweet claim classification, JUTDP, Deep Learning.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 NLP: Natural language processing

NLP[1] a sub-field of computer science, is more specifically the field of “artificial intelligence” (AI), which is concerned with giving computers the ability to grasp spoken and written words in a manner akin to that of human life.

Computational linguistics, which models human language using rules, is combined with statistical, machine learning, and deep learning models in NLP. These kinds of technologies have made it possible for computers to fully ”understand” what is being said or written, including the intentions and mood of the speaker or writer, whether it be in the form of text or audio data. NLP underpins computer systems that translate text between languages, respond to spoken commands, and swiftly summarize massive quantities of text—even in real-time. NLP is undoubtedly already being utilized by you in the form of voice-activated GPS systems, digital assistants, speech-to-text dictation programs, customer service Chatbots, and other consumer conveniences. However, as a means of optimizing business operations, increasing

worker productivity, and streamlining mission-critical business processes, the usage of NLP in corporate solutions is growing.

Since human language is so ambiguous, it is particularly difficult to develop algorithms that can accurately determine the intended meaning of text or speech data. The irregularities in human language that take humans years to learn but that programmers must teach natural language-driven applications to recognize and understand accurately from the beginning if those applications are to be useful include homophones, sarcasm, idioms, metaphors, exceptions to the rules of grammar and usage, and changes in sentence structure.

Many NLP tasks deconstruct human text and voice data to help the computer comprehend the text and speech data it is ingesting. These are only a few of these tasks: -

- **Speech Recognition**[2]: - Speech recognition, often known as speech-to-text, is the process of accurately converting voice input into text. Speech recognition must be used by any program that reacts to voice commands or questions. Speech recognition is particularly challenging because of the way people speak—quickly, slurring words together, with different emphasis and intonation, in numerous 7 dialects, and frequently using incorrect grammar.

- **Part of Speech Tagging**[3]: - Part of speech tagging, also known as grammatical tagging, is the process of determining a word's part of speech based on its usage and context. The word "make" is categorized as both a verb and a noun, as in the lines "I can make a paper airplane" and "What kind of car do you own?"

- **Word sense disambiguation**[4]: - Word sense disambiguation is the process

of selecting a word's meaning from among its possible meanings using semantic analysis to determine which word makes the most sense in the given context. Word sense disambiguation, for example, makes it clear how the verbs "make," "make the grade," "make a bet," and other similar expressions indicate different things.

- **Name Entity Recognition**[5]**: -** By employing named entity recognition, or NEM, words or phrases are recognized as useful entities. NEM recognizes "Kentucky" as a location or "Fred" as a person's name.
- **Co-reference resolution**[6]**: -** Coreference resolution is the process of establishing whether and when two words refer to the same thing. The most common instance is determining who or what a certain pronoun refers to (for instance, "she" = "Marie"), but it may also be necessary to identify a metaphor or idiom that is utilized in the text (for instance, when "bear" refers to a huge, hairy human instead of an animal).

Text classification[7] scans texts for elusive aspects including attitudes, sentiments, sarcasm, confusion, and mistrust. **Natural language generation**, often known as speech-to-text or voice recognition in reverse, is the process of translating structured text input into human language.

1.2 Fake News

Fake News means false or misleading information represented as factual news. These can be present in various forms like written articles, videos, images, audio, and social media posts. Nowadays fake news has become a compelling issue in the growable

digital world due to the contentment of circulating information online. Describe some complete details about fake news below: Types of Fake News:

- **Misinformation:** Wrong or false information shared without malevolent intent. It may result from genuine errors, misconceptions, misleading information, and misunderstandings.
- **Click-bait and Profit:** Fake news generates website traffic, clicks, and ad revenue, making it financially appealing to some publishers.
- **Confirmation Bias:** People are more likely to believe and share information that aligns with their existing beliefs, creating a fertile ground for fake news.
- **Disinformation:** Fake information is dispersed intentionally to deceive or manipulate the public. This technique is generally used for political or ideological agendas.
- **Misinformation:** When sharing information with the intention to hurt and harm, such as sharing private data, or spreading embarrassing information.

1.3 Textual Claim Identification [8]

Textual claim identification is the process of identifying and evaluating claims made within a body of text. Claims are statements or assertions that can be either true or false, and they are often used to persuade or convince the reader or listener of a particular viewpoint or argument. Claim identification is an important task in various fields, including information retrieval, and natural identification:

Definition: Textual claim identification involves automatically identifying statements within an including politics, science, health, and more.

Importance: Importance of Textual Claim identification mentioned below

- **Fact-Checking:** Identifying claims is essential for fact-checking organizations to verify the accuracy of statements made by individuals or organizations.
- **Information Retrieval:** It helps in extracting relevant information from large volumes of text, making it easier for users to find specific claims or information.
- **Argument Analysis:** In argumentation analysis, claim identification is a crucial step in understanding the structure and strength of arguments.

Challenges:

- **Ambiguity:** Some claims can be ambiguous or implicit, making them difficult to identify.
- **Context Dependency:** The interpretation of a claim often depends on the context in which it is presented.
- **Variability:** Claims can vary in terms of their format, structure, and language.

In summary, textual claim identification is a vital task for extracting and assessing claims within text data, with applications ranging from fact-checking to information retrieval and argument analysis. It involves various methods and techniques, including rule-based and machine-learning approaches, and requires careful consideration of the challenges and ethical implications associated with the task.

1.4 Research Challenges

Numerous studies on text classification and its use to identify fake news have been published in a variety of research papers in the past; very little research has been

done on Indian political tweets. There has undoubtedly already been research on the claim classification of political tweets, but those tweets are in English. In order to obtain sentiment scores and better understand the function of the claim classification lexicon-based approach in the detection of fake news, In the past research work, there have been lots of diverse works on claim classification and the role of claim classification in the detection of fake news in various research papers but extremely uncommon work has been noticed on the Indian political tweets. Certainly, there has already been work on claim classification on political tweets however the tweets are in the English language. In our thesis work, we have done our study on Hindi tweets as well as English Tweets to acquire sentiment scores and try to understand what is the role of the claim classification lexicon-based approach behind Fake news detection. As we have seen past study has been done on Indian political tweets but not on Hindi political tweets in that fashion. Here we have an attempt to do a job on claim classification on Indian political Hindi tweets. Related to that claim classification I have discussed an observation based on how much contribution is behind the propagation of fake news, but we studied both Hindi and English tweets for our thesis. As we have seen, previous research on Indian political tweets has been conducted, but not in the same way for Hindi political tweets. Here, we make an effort to do claim classification on tweets written in Hindi on Indian politics. In connection with that claim classification, I have talked about an observation based on how much is contributed to the spread of false information.

1.5 Motivation

In recent years, social media has evolved into a vital platform where anybody may express their thoughts. They are able to upload pictures and videos. It might be Facebook, Twitter, or Reddit, and it could be from the recent past to the present day. Social media is a major source of bogus information, including rumors, news, images, audio, and video. People from various regions and countries are perplexed by how news is now disseminated through social media. In most cases, it is discovered that people are disseminating false information in accordance with the various, either favorable or bad, opinions of the news spreader. We are aware that some of them have put considerable effort into enacting limitations on the dissemination of news and that people have also been made aware of it in previous research articles. But in this case, if we think about technically employing claim classification and natural language processing, we may sometimes impose limits on how far consumer news is distributed. And in order to achieve that aim, we go forward and attempt to implement countless things in various ways, drawing inspiration from these notions.

1.6 Hypothesis

The proliferation of fake news on social media, including Twitter, has prompted the need for effective detection methods. In this hypothesis, we aim to classify the top words and phrases from fake news in multilingual Twitter datasets using a variety of machine learning and deep learning techniques. By identifying these keywords and phrases, we can gain insights into the linguistic patterns commonly associated with fake news across different languages. We hypothesize that by applying a combination of machine learning and deep learning techniques, we can effectively classify the

top words and phrases associated with fake news in multilingual Twitter datasets. Specifically, we aim to achieve the following:

1. **Data Collection and Preprocessing:** Collect multilingual Twitter datasets with labeled fake news samples. Preprocess the data by tokenizing, stemming, and removing stop words.

2. **Feature Extraction[9]:** Employ traditional techniques like TF-IDF and Bag of Words for machine learning models. Utilize word embeddings (e.g., Word2Vec, FastText) for deep learning models.

3. **Machine Learning Models:** Applying classical machine learning algorithms techniques should be able to predict the variables for that data. Text classification uses supervised machine learning namely, Logistic Regression is a great starter algorithm for text-related classification. Conditional Random Fields identify gene and protein, gene prediction. NLP part of Speech (POS) Tagging. NLP Named Entity Recognition (NER). A decision tree is a flowchart-like tree structure that learns a hierarchy of questions from a large training dataset of examples. SVMs use a hyperplane to create a boundary between two classes of data.

4. **Deep Learning Models:** Employ Convolutional Neural Networks (CNNs) to capture local patterns within textual data. Implement Recurrent Neural Networks (RNNs) such as Long Short-Term Memory (LSTM) networks to capture sequential dependencies in text data. Utilize transformers-based models like BERT and RoBERTa for contextual understanding.

5. **Multilingual Approach:** Investigate the effectiveness of machine learning and

deep learning models across different languages. Assess the performance of multilingual models that can handle linguistic variations.

6. **Evaluation:** Evaluate the models using metrics such as accuracy, precision, recall, and F1-score to assess their ability to classify top words, and phrases and classify types of claims from fake news in multilingual Twitter datasets.

7. **Expected Outcomes:** We expect that the condition of machine learning and deep learning techniques will allow it to identify keywords, phrases, and claim labels commonly from fake news in multilingual Twitter datasets. The outcomes may include the following:

    - Identification of common linguistic patterns and keywords in fake news, regardless, regardless of the language of the tweet.
    - Improved performance of deep learning models, particularly transformer-based models, in understanding the context and nuance of multilingual fake news.
    - Insights into which machine learning and deep learning techniques are most effective in identifying top words, phrases, claim labels from fake news in multilingual Twitter data.

    The proposed hypothesis aims to contribute to the development of robust methods for detecting fake news on Twitter across multiple languages. By identifying the top words, phrases, and claim labels associated with fake news, we can improve the accuracy and efficiency of fake news classification systems, ultimately helping to combat the spread of misinformation on social media platforms.

1.7 Objective to our thesis

Our goal in this thesis study is to develop novel forms of text classification lexical models on the foundation of traditional, pre-existing text classification models that are entirely based on many Indian languages. This can provide us with accurate text classification of Hindi tweets and English tweets, as well as Text classification of current socio-political issues in India. It can also help us create a model to identify fake tweets on current political issues, such as the Delhi Riots, which are widely shared on Indian social media.

1.8 Contribution

In this thesis, we used a variety of claim classification techniques to obtain sentiment scores on tweets about the Delhi Riots, Global Warming, Terrorism, Covid-19, Gujarat riots, Adani, Bollywood, Fake News, Narendra Modi, Karnataka Election, Geo Politics, Manipur crisis, Ram Mandir, Cricket, Artificial Intelligence, Rafale Aircraft, Omicron, Covid Vaccination, Narendra Modi(Hindi tweets), Khalistani, NRC and CAA, Farmer Protest, Article 370, Gujrat Riot(Hindi Tweets), Adani(Hindi Tweets), Wrestlers Protest, Assam Flood, Assam Border Dispute, Ram Mandir(Hindi Tweets), Pulwama Attack, Amrit Mahotsav and Indian Krishibill. To get better results, we use Machine learning models like SVM and Deep learning models like LSTM and Bert. We get three different text classification techniques, but also a similar correlation. Then, utilizing tweets about the Delhi riots, we worked on categorizing fake news using text classification tools.

1.9 Dataset

In this study, we mainly used Indian Hindi tweets and English tweets for our thesis work. First, we aggregate only English and Hindi tweets from Twitter and collect the raw tweets. We then added annotations to them. Annotations can have a variety of properties, including Simple, Composite, and Compound. Later on, we'll go into more detail regarding these kinds of datasets and annotations.

1.10 Thesis outline

This thesis work has been divided into seven chapters.

Chapter 1has the introductions of the topics we are going to discuss.

Chapter 2 discusses the works that have already been done on the topics.

Chapter 3 discusses how the raw dataset has been collected and prepared to be fed to different models.

Chapter 4 discusses how claim classification has occurred from the Indian Hindi tweets and how many models we have used.

Chapter 5 discusses how our fake news classification model works on current socio-political issues which we extract from Indian tweets.

Chapter 6 discusses the system, experiment, results, and observations.

Chapter 7 discusses the conclusion and future work that can be done additionally in this domain.

Chapter 2

Related Work

2.1 Literature Survey on Claim Features

AR Mahlous et. al.[10] studies show that in March 2020, the World Health Organization declared the COVID-19 outbreak to be a pandemic. Soon afterward, people began sharing millions of posts on social media without considering their reliability and truthfulness. While there has been extensive research on COVID-19 in the English language, there is a lack of research on the subject in Arabic. In this paper, we address the problem of detecting fake news surrounding COVID-19 in Arabic tweets. We collected more than seven million Arabic tweets related to the coronavirus pandemic from January 2020 to August 2020 using the trending hashtags during the time of the pandemic. We relied on two fact-checkers: the France-Press Agency and the Saudi Anti-Rumors Authority to extract a list of keywords related to the misinformation and fake news topics. A small corpus was extracted from the collected tweets and manually annotated into fake or genuine classes. We used a set of features extracted from tweet contents to train a set of machine learning classifiers. The manually anno-

tated corpus was used as a baseline to build a system for automatically detecting fake news from Arabic text. Classification of the manually annotated dataset achieved an F1-score of 87.% using Logistic Regression (LR) as a classifier with the n-gram-level Term Frequency-Inverse Document Frequency (TF-IDF) as a feature, and a 93.3% F1score on the automatically annotated dataset using the same classifier with count vector feature. The introduced system and datasets could help governments, decision-makers, and the public judge the credibility of information published on social media during the COVID-19 pandemic.

Suryavardan et.al.[11] developed FACTIFY is a multi-modal fact verification dataset that aims to combat fake news and expose false claims before they cause harm. The dataset consists of 50K data points, covering news from India and the US, and is the largest multi-modal fact verification dataset available. It includes images, textual claims, reference textual documents, and images labeled with three categories: support, no-evidence, and refute. The dataset serves as a stepping stone for building novel multi-modal fact verification systems and addresses the lack of focus on multi-modal or cross-modal fact verification in existing research efforts and datasets. The majority of the present fact-checking research relies on unimodal techniques and limited annotated datasets, making FACTIFY a valuable resource for the research community.

Helmstetter et. al.[12] introduces an alternative approach for creating a large-scale dataset for tweet classification with minimal user intervention. The approach relies on weak supervision and automatically collects a large-scale, but very noisy, training dataset comprising hundreds of thousands of tweets. The tweets are labeled based on their source, i.e., trustworthy or untrustworthy, and a classifier is trained on this dataset. The classifier is then used for the classification of fake and non-fake tweets,

even though the labels may not be accurate for this new classification target. The results show that despite the unclean and inaccurate dataset, the performance is comparable to using a manually labeled set of tweets. Combining the large-scale noisy dataset with a human-labeled one yields even better results.

Buket Erúahin et. al.[13] presents a classification method for detecting fake accounts on Twitter. The dataset is preprocessed using a supervised discretization technique called Entropy Minimization Discretization (EMD) on numerical features. The results of the Naïve Bayes algorithm are analyzed. The Naïve Bayes learning algorithm uses 16 attributes for classification. Performance metrics such as Accuracy (ACC), F-Measure, and confusion matrix are used to evaluate the algorithm's performance. F-Measure is the harmonic mean of precision and recall, where precision is the ratio of predicted positive items to actual positive items, and recall is the ratio of actual positive items to predicted positive items. Experimental results show the accuracy and confusion matrix results before and after discretization. The paper provides tables with accuracy and F-Measure results for the Naïve Bayes algorithm.

Maryem Heidari et. al.[14] focuses on using the BERT model for fake news detection in the context of the COVID-19 pandemic. Social media platforms are identified as the primary source of information exchange during the pandemic. Social bots are highlighted as one of the main sources of misinformation during the pandemic. The paper aims to investigate whether social bots spread fake and real news with the same ratio as human accounts on social media platforms. The authors use transfer learning and new features to improve the fake news detection model based on the COVID-19 dataset. The paper provides preliminary research results and offers a new perspective on fake news detection and bot detection on online platforms. The paper discusses different types of misinformation, including false, mistaken, misleading, and

inaccurate information. It mentions the use of distance-constrained misinformation combat to detect the spread of false information and increase the spread of true information. The authors highlight the prevalence of conspiracy theories during natural disasters, financial crises, wars, and attacks, such as the conspiracy theories surrounding COVID-19 and its vaccine. The paper emphasizes the importance of detecting and combatting misleading information, which aims to destroy trust in governments, tamper with economies, or gain benefits. Machine learning models, specifically the BERT model, are utilized to provide a new model for fake news detection on online platforms.

2.2 Literature Survey on Claim Classification

Bibek Upadhayay et.al.[15] focuses on the rampant integration of social media into our everyday lives and culture has given rise to faster and easier access to the flow of information than ever in human history. However, the inherently unsupervised nature of social media platforms has also made it easier to spread false information and fake news. Furthermore, the high volume and velocity of information flow in such platforms make manual supervision and control of information propagation infeasible. This paper aims to address this issue by proposing a novel deep-learning approach for the automated detection of false short-text claims on social media. We first introduce Sentimental LIAR, which extends the LIAR dataset of short claims by adding features based on sentiment and emotion analysis of claims. Furthermore, we propose a novel deep learning architecture based on the DistilBERT language model for the classification of claims as genuine or fake. Our results demonstrate that the proposed architecture trained on Sentimental LIAR can achieve an accuracy of 70%, which is

an improvement of 30% over previously reported results for the LIAR benchmark.
William Yang Wang[16] developed LIAR dataset, which is a publicly available dataset for fake news detection, consisting of 12.8K manually labeled short statements collected from POLITIFACT.COM over a decade. It is an order of magnitude larger than previously available datasets of similar type. The dataset includes detailed analysis reports and links to source documents for each case, providing a grounded and natural context for the instances. The dataset can be used for fact-checking research and enables the development of statistical and computational approaches to fake news detection. A hybrid convolutional neural network model that integrates metadata with text has been designed to improve the performance of a text-only deep learning model. The LIAR dataset also facilitates research on stance classification, argument mining, topic modeling, rumor detection, and political NLP research.
Valerio La Gatta et. al.[17] focuses on false and unverified information on social media during major geo-political events threatens the quality of the digital information ecosystem. Several fact-checking organizations have been verifying stories related to the Russia-Ukraine conflict on social media. The paper proposes a methodological framework for automatically identifying false and unsubstantiated claims spreading on Twitter during the conflict. It consists of two models: claim detection and claim retrieval. The models are based on pre-trained language models and are fine-tuned for text classification and information retrieval tasks. The effectiveness of the methodology is validated using 83 verified false claims that spread on Twitter during the first week of the invasion. The experiments show that the proposed methodology outperforms standard baselines for both claim detection and claim retrieval. The models are also tested on a larger Twitter dataset to assess their performance in a real scenario. Although no formal quantitative analysis is performed, manual verification of

a sample of tweets confirms the models' correct predictions. The paper compares the performance of the proposed approach with a sentence-BERT baseline. The results show that the proposed approach achieves better performance in claim retrieval tasks. The data for the analysis is collected from the Russia-Ukraine Conflict-Misinfo Dashboard, which provides a collection of true and false claims verified by fact-checking outlets. The dataset includes English tweets shared during the initial weeks of the invasion.

Mohammed Samdazi et. al.[18] address the lack of fake news datasets in Persian and introduce a new dataset called TAJ, crawled from different news agencies. The authors propose two deep models based on the Bidirectional Encoder Representations from the Transformers model (BERT) for fake news detection. They utilize two different settings of BERT: pool-based representation and sequence representation. In the pool-based representation, a single-layer perceptron (SLP) is connected to BERT for direct embedding-based fake news detection. In the sequence representation, a Convolutional Neural Network (CNN) is used after the BERT's embedding layer to extract additional features based on word collocation. The proposed models are evaluated on the TAJ dataset and two Persian rumor datasets as baselines. The results show that both BERT-SLP and BERT-CNN models outperform previous baselines and traditional machine learning models, with significant improvements in performance. The paper highlights the effectiveness of deep contextualized embedding approaches for fake news detection in Persian.

Arianna D'Ulizia et. al.[19] survey that systematically reviews twenty-seven popular datasets for fake news detection, providing insights into the characteristics of each dataset and comparative analysis among them. The paper presents a Fake News Detection Datasets (FNDD) characterization composed of eleven characteristics ex-

tracted from the surveyed datasets, along with a set of requirements for comparing and building new datasets. The paper analyzes the frequency distribution of the eleven characteristics of the surveyed datasets, such as news domain, application purpose, type of disinformation, language, size, news content type, rating scale, media platform, spontaneity, availability, and extraction period. The paper also provides a summary of the datasets selected for further analysis, along with their main characteristics. The paper highlights the importance of evaluating datasets for fake news detection and identifies a gap in existing surveys that do not detail the characteristics and requirements of these datasets.

S. Dutta et. al.[20] presents a token-level annotation of factual claims in tweets from Indian Twitter, focusing on fine-grained labeling to facilitate better fact-checking systems. The dataset includes tweets in English, Bengali, Hindi, and their code-mixed variants, making it the first of its kind in terms of both labeling scheme and data sources. The paper defines a categorization scheme for the annotated factual claims based on the type of facts represented, highlighting the need for codemixed and multilingual systems for fact-checking in the Indian context. The authors emphasize the importance of selecting minimal spans of meaningful text that express verifiable information for claim identification. The dataset aims to help build effective fact-checking systems, especially in the Indian context and is expected to contribute to the development of novel claim identification schemes.

Chapter 3

Dataset Preparation

In this section, the specifics of creating the JUTDP v1 dataset have been discussed in detail. We have covered data collecting methods, dataset nomenclature, download data from the tweeter, and annotations of raw data in the following subsections.

3.1 Dataset Nomenclature

Our developed dataset has been named JUTDP v1, where JUTDP stands for "Jadavpur University Tweet Dataset Pradip" and v1 stands for "Version 1". Currently, we have developed version 1 of the JUTDP v1 dataset. It is to be noted that the dataset contains 3 different types of claims namely, 'Simple', 'Compound', and 'Composite'. **Table 3.1 provides an illustration of each type of claim for different languages like Hindi and English. Here we used to classify the news as per the given dataset.**

3.2 Collection of raw data

First, we applied on Twitter for a Twitter developer account using all my university details, course details, and thesis work details. After verification of all data, Twitter approved my developer account and provided Twitter API credentials (Consumer key, Consumer secret, access token, access token secret). With the help of Python code, we prepared a code for extracting tweet data with the help of the Tweepy library function from Twitter.

TWEEPY Tweepy is an easy-to-use Python library for accessing the Twitter API. It simplifies the process of authenticating with Twitter, making API requests, and handling responses. Below, we will explain some of the key functions and concepts within the Tweepy library: Once we have configured our Twitter API credentials, the next thing is to build a Bot or pull tweets for claim classification, etc. In our work, we will share complete knowledge but in a simple way to use Tweepy for performing read and write operations, like posting tweets, deleting tweets, retweeting, etc. To start off, first, we need Twitter API credentials. First of all, we opened a Twitter account, then I applied for a Twitter developer account with reference to that account. We are asked for many details like personal details, education details, Why we want developer account access, Google Scholar links, university details, and many other pieces of information. Twitter has a couple of API versions illustrated in the Figure 3.1: V1.1 and V2.0. Following the Tweepy documentation, we shall do the following:

1. Twitter API v1.1 Authentication using OAuth 1.0a - User Context Using Tweepy Class API To perform write operations on Twitter A/c via API, like posting tweets, deleting tweets, following someone, unfollowing someone, liking a tweet, retweeting, etc.

2. Twitter API v2.0 Authenticate using OAuth 2.0 - Bearer Token (App only) Using Tweepy Class client To pull tweets on a specific keyword to perform text analytics

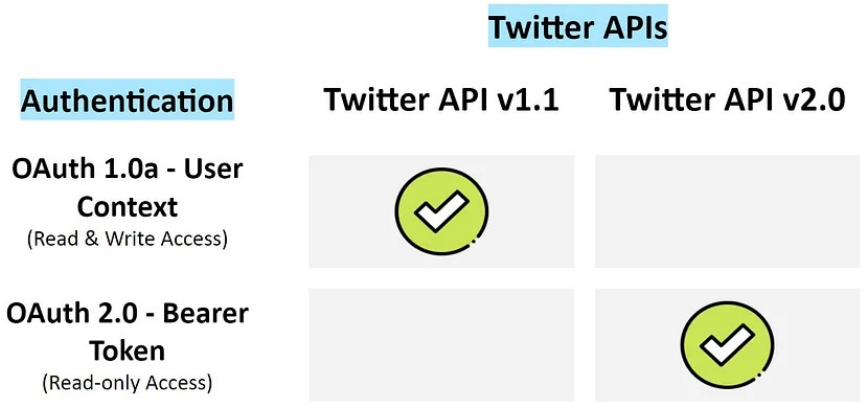

Figure 3.1: Tweeter API Authentication

Our high-level plan of action:

- Pull 7500 tweets for English and 5200 for Hindi from Twitter on a certain trending keyword, like **"#Assamborderdispute, #narendramodi, #krishibill, #covid-19, #covidvaccination, #terrorism, #delhiriots, #gujratriots, #adani, #karnatakelection, #geopolitics, #manipurcrisis, #rammandir, #omicron, #rafaleaircraft, #khalistani, #article370, #wrestlersprotest, #assamflood, #amritmohatsav, # pulwamaattack."**
- Perform claim classification on Tweets, using: Python NLTK library, stopwords, wordtokenize, WordNetLemmatizer, and string.
- Create beautiful visualization to draw inferences

Figure 3.2: Caption

Tweepy Code Walk-through: To get started with Tweepy, as a first step, we need to set up our environment. Here, we are basically installing and then importing tweepy.

Performing write operations:

- Next up, we are first authenticating for Twitter APIv1.1. So, we authenticate using OAuth 1.0a - User Context. Update your API creds below as required.
- Let's start off by making a tweet saying, "Hello World!".
- Now, let's pull some tweets from your home timeline. Home timeline is your Twitter home basically, that has tweets from the people you follow & your own recent tweets.
- Next up, let's get some of the recent tweets posted by you on your account, using the **user_timeline** method.
- When we pull tweets from Twitter, there is additional information that is passed along with the tweet, like the creation date, tweet id, time, location, etc. Let's try printing some of this information here.

- Tweet ID (id_str) is a unique identifier for every tweet posted on the platform. Now, using this Tweet ID, let's try deleting a tweet. For this, first up, We create a tweet.
- Now, using the tweet ID, let's perform the delete operation. We may check on our Twitter account if changes are made.
- We may follow an account with the create_friendship method. For this, we are using Narendra Modi's Twitter handle, which We are not following for now.
- Similar way, you may unfollow as well.

Alright, so this was Twitter API v1.1, with which we performed write operations.

Pulling tweets (read operation): Now let's jump onto the Twitter API v2.0, and try pulling some tweets

- For this, the first step is authentication. We are using OAuth 2.0 this time, which grants us Read-only access. Which is absolutely fine for pulling public tweets.
- Next up, we are pulling recent tweets on the hashtag #elonmusk, with these specific directions, meaning: We don't want retweets, We want tweets in English and Hindi languages, and We want a minimum of 1300 - 1500 results.

Well, these are only some of the many operations Tweepy may perform for us. To try out more Tweepy operations, like retweeting, etc. In code first, we install tweepy library then use Twitter API credentials to validate my developer account. Then use a keyword or name of the topic on the basis that tweet code can extract tweets. We set the number of tweets extracted per run-time of the code is 1000 and write

a unique file name for downloaded CSV files of extracted tweets. Also, we used a function to iterate over tweets containing the keyword.

3.3 Annotation of JUTDP

First, we downloaded multilingual tweets (for English we pulled near about 7600 tweets, and for Hindi 5300 tweets) with the help of tweepy. The format we used to prepare the dataset was developed by DRDO (Defence Research & Development Organization). Here we prepared two datasets (one is for English and another is for Hindi) and the whole process is illustrated in the Figure 3.11

- Tweet ID: I have copied the IDs from the raw dataset in the Tweet ID as illustrated in the Figure 3.3 column. The reason for keeping the IDs is so that I can justify that the tweets we have used to prepare the dataset are actually downloaded from the Twitter portal.

A
ID
1669782054240680000.00
1669781942504380000.00
1669770688868260000.00
1669770453961900000.00
1669770307605850000.00
1669768396219570000.00
1669763683126320000.00
1669757508917900000.00
1669753567782240000.00
1669749493682630000.00
1669742953638550000.00
1669737314669440000.00

Figure 3.3: Figure of Tweet Account ID from JUTDP

- **Tweet:** In the Tweet column we have put tweets directly copied from the raw dataset as illustrated in Figure 3.4 and in Figure 3.5 for both English and Hindi tweets accordingly.

Tweet
#IntelBrief: In addition to #AlQaeda & ISK, there are a number of other terrorist groups active in #Afghanistan, such as the TTP, IMU, Islamic Jihadi Group, Jamaat Ansarullah & remnants of the TIP. https://t.co/ebxyydJ30S#Terrorism #Counterterrorism #ISIS #AlShabaab
Afghanistan-based extremists spark terror fears in Central Asia via @NikkeiAsia |#terrorism #extremism #extremists #terror #violence #attack #propaganda#recruitment #highthreat #threat #threats https://t.co/NO6CMqETJo
New Hizbul Mujahideen Terror Camp Supervised By ISI Detected In Pakistan Occupied Kashmir's Solanwali: Reports https://t.co/zYrUZRIRTS #Pakistan #Terrorism #PakistanOccupiedKashmir #Hizbul
Indian police say five foreign militants killed in Kashmir #terrorism https://t.co/8Ebl3vywFB

Figure 3.4: Raw English Tweet from JUTDP

Tweet
2024में किसी की ताकत नहीं जो #नरेंद्र मोदी जी को हरा सके! हर हर मोदी, घर घर मोदी!! जय जय श्री राम
आदरणीय #प्रधानमंत्री श्री #नरेंद्र #मोदी जी के #मनकीबात कार्यक्रम का #102वां एपिसोड इस माह #जगदलपुर #बस्तर में सुना। #मन_की_बात #नरेन्द्र #मोदीजी #नरेंद्र_मोदी #नरेन्द्र__मोदी #नरेंद्रमोदीजी #ManKiBaat #Narendra #Modi #NarendraModi #NarendraModiji #Modiji @narendramodi @PMOIndia https://t.co/ZCP53gbe2P
हिमंता बिस्वा सरमा का ये वक्तव्य आपको #नरेंद्र #मोदी के #भारत के नये काल... https://t.co/Tq7GPswRWF via @YouTube
माननीय प्रधानमंत्री श्री #नरेंद्र मोदी जी के नेतृत्व में #सफलतम 9 वर्ष सेवा, सुशासन और गरीब कल्याण के पूर्ण होने पर #अनुसूचित जनजाति मोर्चा द्वारा #जनजाति जनप्रतिनिधि सम्मेल का आयोजन हुआ । https://t.co/2IyJfzgezi

Figure 3.5: Raw Hindi Tweet from JUTDP

- **Selected Claim:** We manually processed the tweets before putting them in the Selected Claims column, after removing the special characters, the numbers,

URLs, emojis, smileys, and reserved words we have taken that make a meaningful sentence as illustrated in Figure 3.6 and in Figure 3.7 for both English and Hindi tweets accordingly.

Selected Claim
In addition to #AlQaeda & ISK, there are a number of other terrorist groups active in #Afghanistan, such as the TTP, IMU, Islamic Jihadi Group, Jamaat Ansarullah & remnants of the TIP
Afghanistan-based extremists spark terror fears in Central Asia via
New Hizbul Mujahideen Terror Camp Supervised By ISI Detected In Pakistan Occupied Kashmir's Solanwali
Indian police say five foreign militants killed in Kashmir

Figure 3.6: English selected claim for tweets from JUTDP

Selected Claim
2024में किसी की ताकत नहीं जो नरेंद्र मोदी जी को हरा सके!
प्रधानमंत्री श्री #नरेंद्र #मोदी जी के #मनकीबात कार्यक्रम का #102वां एपिसोड इस माह #जगदलपुर #बस्तर में सुना।
हिमंता बिस्वा सरमा का ये वक्तव्य आपको नरेंद्र मोदी के भारत के नये काल
माननीय प्रधानमंत्री श्री नरेंद्र मोदी जी के नेतृत्व में सफलतम 9 वर्ष सेवा, सुशासन और गरीब कल्याण के पूर्ण होने पर अनुसूचित जनजाति मोर्चा द्वारा जनजाति जनप्रतिनिधि सम्मेल का आयोजन हुआ

Figure 3.7: Hindi selected claim for tweets from JUTDP

- **Type of Claim:** In the Type of Claim Column as illustrated in Figure 3.8 we have classified tweets into 3 types, Simple, Compound, and Composite. The

method we applied here to separate the tweets into different classes was developed by DRDO (Defence Research and Development Organisation).

Figure 3.8: Type of Claim in JUTDP

- **Entities:** In the entity column can be illustrated as Figure 3.9 and from Figure 3.10 we select only main phrases or words without any repetition and iterate am, is, are, and, have, full stop, special character, numbers, emojis, smileys, etc., taken from the selected claim column to prepare a basic dictionary.

ID	Tweet	Selected Claim	Type of Claim	Entities
1.66976E+18	@soldiersspeaks Bitter truth, There is no any #democracy exits in #Pakistan Anyone involved in #Corruption should be condemn no matter which ethnic or sect or any party they belong to.#Corruption Should be condemned, Also I strongly condemned #Terrorism . https://t.co/xhUtpgFWyz	soldiersspeaks Bitter truth, There is no any democracy exits in Pakistan Anyone involved in #Corruption should be condemn no matter which ethnic or sect or any party they belong to.#Corruption Should be condemned, Also I strongly condemned #Terrorism	Simple	soldiers, speaks, Bitter, truth, There, democracy, exits, Pakistan, Anyone, involved, Corruption, condemn, ethnic, condemned, Terrorism

Figure 3.9: English Entities from English Tweets

ID	Tweet	Selected Claim	Type of Claim	Entities
1.66743E+18	हमारे ही नहीं विश्व भर में जो अति लोकप्रिय हैं ऐसे महान हस्ती आदरणीय प्रधानमंत्री #नरेन्द्र #मोदी जी को समस्त लाभार्थियों की तरफ से बहुत बहुत धन्यवाद आभार।🙏🙏 जिला #रुद्रप्रयाग #9YearsOfSeva @blsanthosh @VanathiBJP @AshaNautiyal8 @ajaeybjp https://t.co/0UEoyjDTU5	हमारे ही नहीं विश्व भर में जो अति लोकप्रिय हैं ऐसे महान हस्ती आदरणीय प्रधानमंत्री #नरेन्द्र #मोदी जी को समस्त लाभार्थियों की तरफ से बहुत बहुत धन्यवाद आभार।🙏🙏 जिला #रुद्रप्रयाग	Simple	महान हस्ती, आदरणीय, प्रधानमंत्री, नरेन्द्र मोदी, समस्त, लाभार्थियों, जिला रुद्रप्रयाग

Figure 3.10: Hindi Entities from Hindi Tweets

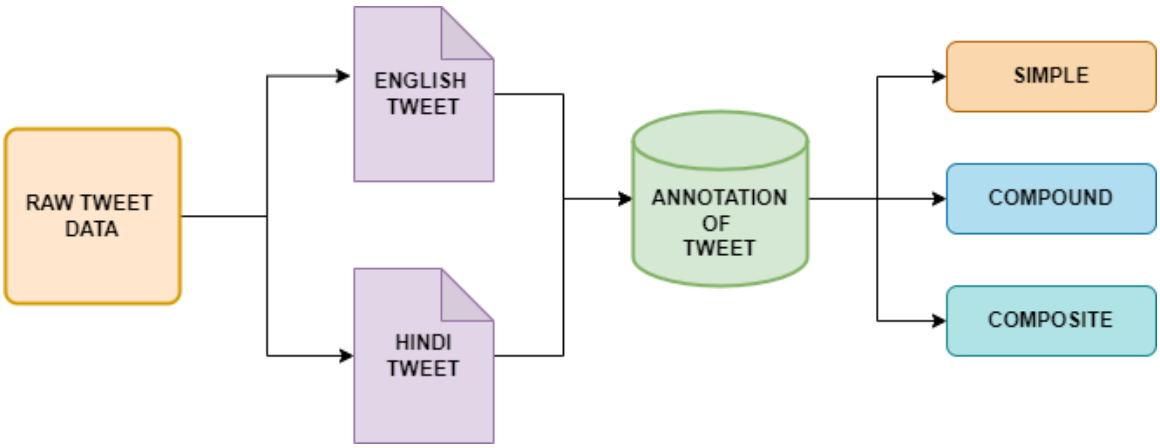

Figure 3.11: Annotation Process of JUTDP

3.4 Types of Tweets in JUTDP

Simple: We studied the tweets to see if there were any links to direct responses, or statements from a renowned person, public figure, any government organizations, or court. Example: court ordered stay order against the application of three Krishi bills, this statement future we can easily find out what is the actual statement and whether this statement is true or false. These types of tweets can be shown in Figure 3.12 and Figure 3.13 for both English and Hindi Tweets from JUTDP.

Compound: After checking tweets we opt for tweets in which two or more than two sentences completely depend on each other or hold a major link with each other. Example: Narendra Modi said In Parliament If China doesn't stop aggression in the border area, there may be a high probability of starting war; After winning the gold medal in Javelin through an honorary post of Captain was given to Neeraj Chopra (Indian Athletic) from Indian BSF force. In the two sentences we cited here as examples, it is seen that if China does not stop the aggression, there may be a war. It means that whether or not there will be a war depends on whether China stops

the aggression. For another example, Neeraj Chopra has been given the honorary post of captain in the BSF as he has won gold medals for the country. If he had not won the gold medal, he would not have got this honorary post. These types of tweets can be shown in Figure 3.12 and Figure 3.13 for both English and Hindi Tweets from JUTDP.

Composite: We selected those tweets whose statements could be traced to reality or a link will surely be found but it will not be very comfortable. For example, if you want to justify a statement given by Lal Bahadur Shastri Ji, the complexity will increase a lot. First of all, he is no longer alive, secondly, we do not have any proper record of it. Another example is when the British gave independence to India and Pakistan it is a very tough and complex matter to verify those statements that have been made about it. The West Bengal CM went to a party meeting and made some statements about the Indian PM, it is not possible to take any action against those statements because she went there as Chief Minister and not as a party worker or it can be said that she did not publish these words from any government organization, or government medium. If anyone wants to take action on the basis of those statements, she can easily say that this is a fake video, so we are unable to prove the actual reality. These types of tweets can be shown in Figure 3.12 and Figure 3.13 for both English

and Hindi Tweets from JUTDP.

ID	Tweet	Selected Claim	Type of Claim	Entities
1.66978E+18	#IntelBrief: In addition to #AlQaeda & ISK, there are a number of other terrorist groups active in #Afghanistan, such as the TTP, IMU, Islamic Jihadi Group, Jamaat Ansarullah & remnants of the TIP. https://t.co/ebxyydJ30S#Terrorism #Counterterrorism #ISIS #AlShabaab	In addition to #AlQaeda & ISK, there are a number of other terrorist groups active in #Afghanistan, such as the TTP, IMU, Islamic Jihadi Group, Jamaat Ansarullah & remnants of the TIP	Composite	AlQaeda, terrorist, groups, active, Afghanistan, Islamic Jihadi Group, Jamaat Ansarullah
1.66978E+18	Afghanistan-based extremists spark terror fears in Central Asia via @NikkeiAsia |#terrorism #extremism #extremists #terror #violence #attack #propaganda#recruitment #highthreat #threat #threats https://t.co/NO6CMqETJo	Afghanistan-based extremists spark terror fears in Central Asia via	Simple	Afghanistan, spark, fears, Central, Asia terrorism, extremism, extremists, terror, violence, attack, propaganda, recruitment, high threat, threat
1.66977E+18	New Hizbul Mujahideen Terror Camp Supervised By ISI Detected In Pakistan Occupied Kashmir's Solanwali: Reports https://t.co/zYrUZRIRTS #Pakistan #Terrorism #PakistanOccupiedKashmir #Hizbul	New Hizbul Mujahideen Terror Camp Supervised By ISI Detected In Pakistan Occupied Kashmir's Solanwali	Simple	New Hizbul Mujahideen, Terror, Camp, Supervised, ISI, Detected, Pakistan Occupied Kashmir, Solanwali, Pakistan,terrorism, Hizbul
1.66977E+18	Indian police say five foreign militants killed in Kashmir #terrorism https://t.co/8Ebl3vywFB	Indian police say five foreign militants killed in Kashmir	Simple	Indian, police, say, five, foreign, militants, killed, in, Kashmir
1.66977E+18	Argentine judge calls for detention of four Lebanese citizens in AMIA bombing probe - wow, almost 30 yrs after the bombing, which killed 85!! #terrorism https://t.co/GVfipyylZl	Argentine judge calls for detention of four Lebanese citizens in AMIA bombing probe - wow, almost 30 yrs after the bombing, which killed 85	Composite	Argentine, judge, calls, for, detention, four, citizens, in, bombing, probe, almost, 30, after, bombing, killed, 85
1.66977E+18	#terrorism : #ISIS In #Africa (Chad, Cameroon, Niger, Democratic Republic of the Congo, Nigeria, Mozambique): Targeting Christians – Killing, Beheading, Murdering Priests And Nuns, Burning Churches, Health Clinics, And Homes https://t.co/loz2j3J1Je	ISIS In Africa (Chad, Cameroon, Niger, Democratic Republic of the Congo, Nigeria, Mozambique): Targeting Christians – Killing, Beheading, Murdering Priests And Nuns, Burning Churches, Health Clinics, And Homes	Compound	ISIS, Africa, Chad, Cameroon, Niger, Democratic Republic Congo, Nigeria, ozambique, Targeting, Christians, Killing, Beheading, Murdering Priests And Nuns, Burning Churches, Health Clinics, And Homes

Figure 3.12: English annotated tweet

ID	Tweet	Selected Claim	Type of Claim	Entities
1.67023E+18	हिमंता बिस्वा सरमा का ये वक्तव्य आपको #नरेंद्र #मोदी के #भारत के नये काल... https://t.co/Tq7GPswRWF via @YouTube	हिमंता बिस्वा सरमा का ये वक्तव्य आपको नरेंद्र मोदी के भारत के नये काल	Simple	हिमंता बिस्वा सरमा, वक्तव्य, आपको, नरेंद्र मोदी, भारत, नये, काल
1.66998E+18	माननीय प्रधानमंत्री श्री #नरेंद्र मोदी जी के नेतृत्व में #सफलतम 9 वर्ष सेवा, सुशासन और गरीब कल्याण के पूर्ण होने पर #अनुसूचित जनजाति मोर्चा द्वारा #जनजाति जनप्रतिनिधि सम्मेल का आयोजन हुआ । https://t.co/2IyJfzgezi	माननीय प्रधानमंत्री श्री नरेंद्र मोदी जी के नेतृत्व में सफलतम 9 वर्ष सेवा, सुशासन और गरीब कल्याण के पूर्ण होने पर अनुसूचित जनजाति मोर्चा द्वारा जनजाति जनप्रतिनिधि सम्मेल का आयोजन हुआ	Simple	माननीय, प्रधानमंत्री, नरेंद्र मोदी, नेतृत्व, सफलतम, 9 वर्ष सेवा, सुशासन, गरीब, कल्याण, पूर्ण, होने, पर, अनुसूचित, जनजाति, जनप्रतिनिधि सम्मेल, आयोजन, हुआ
1.66971E+18	जय हिन्द दोस्तों, #महाराष्ट्र के गौरव #पंतप्रधान #नरेंद्र #मोदी जी ने प्रधानमंत्री #आवास योजना से कई गरीबों को अपना घर मिला ! #आयुष्मान भारत से लगभग आधी आबादी को लाभ मिला ! (जय हिन्द, जय महाराष्ट्र. जय भीम) https://t.co/tX7u1GoMuz	महाराष्ट्र के गौरव पंतप्रधान नरेंद्र मोदी जी ने प्रधानमंत्री आवास योजना से कई गरीबों को अपना घर मिला आयुष्मान भारत से लगभग आधी आबादी को लाभ मिला	Compound	महाराष्ट्र, गौरव, पंतप्रधान, नरेंद्र मोदी, प्रधानमंत्री, आवास, योजना, गरीबों, अपना, घर, मिला, आयुष्मान, भारत, लगभग, आधी, आबादी, लाभ, मिला
1.66907E+18	अब हम तीसरी महाशक्ति 2011प्रधानमंत्री #मनमोहन_सिंह ! अब हम विकासशील देश नहीं 2016 प्रधानमंत्री #नरेन्द्र मोदी के 2 साल के शासन में ! अब हमारा मुक़ाबला नेपाल श्रीलंका पाकिस्तान नाइजीरिया से चल रहा है #भुखमरी बेरोजगारी में बस एक बन्दा यही है जो देश को कटोरा देगा ? https://t.co/tdge7fAl0W https://t.co/DypXLUsq0y	अब हम तीसरी महाशक्ति 2011प्रधानमंत्री #मनमोहन_सिंह ! अब हम विकासशील देश नहीं 2016 प्रधानमंत्री #नरेन्द्र मोदी के 2 साल के शासन में ! अब हमारा मुक़ाबला नेपाल श्रीलंका पाकिस्तान नाइजीरिया से चल रहा है #भुखमरी बेरोजगारी में बस एक बन्दा यही है जो देश को कटोरा देगा ?	Compound	तीसरी, महाशक्ति, 2011, प्रधानमंत्री, मनमोहन_सिंह, विकासशील, 2016, नरेन्द्र मोदी, 2 साल, शासन, मुक़ाबला नेपाल, श्रीलंका, पाकिस्तान, नाइजीरिया, चल, भुखमरी, बेरोजगारी, बन्दा, कटोरा, देगा
1.67071E+18	प्रधानमंत्री मोदी की अमेरिका यात्रा से पहले चीन की बड़ी प्रतिक्रिया, तारीफ के साथ भारत को दी नसीहतें #pmmodiusvisit #china #india #america #xijinping #narendramodi #joebiden #चीन #अमेरिका #भारत #शीजिनपिंग #जोबाइडेन #नरेंद्रमोदी https://t.co/vLRdugDAqa	प्रधानमंत्री मोदी की अमेरिका यात्रा से पहले चीन की बड़ी प्रतिक्रिया, तारीफ के साथ भारत को दी नसीहतें	Composite	प्रधानमंत्री, मोदी, अमेरिका, यात्रा, पहले, चीन, बड़ी, प्रतिक्रिया, तारीफ, साथ, भारत, नसीहतें

Figure 3.13: Hindi annotated tweet

3.5 JUTDP Statistics

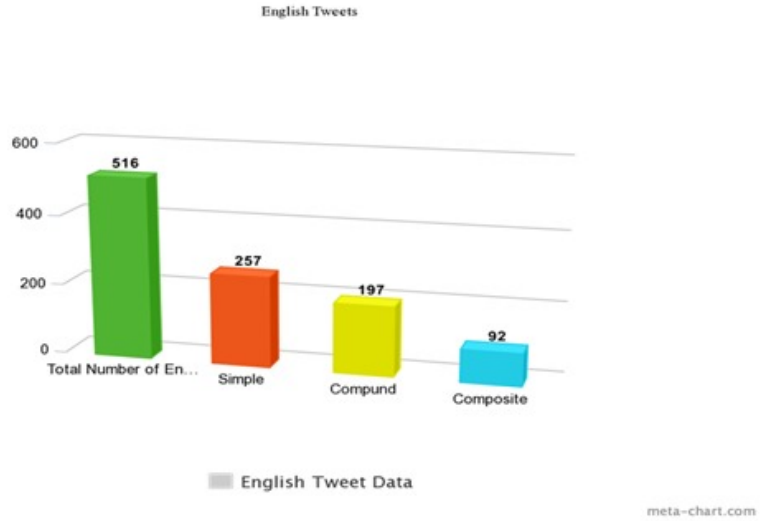

Figure 3.14: Bar Chart Representation Of Numbers of English Tweet Classes in JUTDP

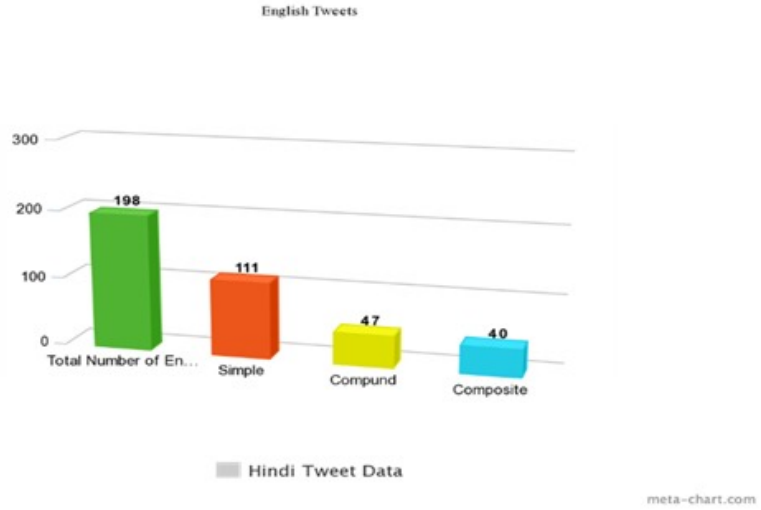

Figure 3.15: Bar Chart Representation Of Numbers of Hindi Tweet Classes in JUTDP

Chapter 4

Feature Extraction

4.1 Word Level

Feature extraction from a "Tweet" column in a tweet dataset at the word level involves converting the text data into numerical features that can be used for machine learning and deep learning or natural language processing tasks.

- **Tokenization:** Break the text in the "Tweet" column into individual words or tokens. This can be done using libraries like NLTK or spaCy.
- **Stop Word Removal:** Remove common stop words (e.g., "and," "the," "in,") as they may not provide useful information for analysis.
- **Stemming or Lemmatization:** Reduce words to their base or root form to standardize variations of words (e.g., "running" to "run").
- **Count Vectorization:** Create a document-term matrix where each row represents a tweet and each column represents a unique word. The cell values indicate the frequency of each word in a tweet.

- **TF-IDF Vectorization:** Instead of raw word counts, use TF-IDF (Term Frequency-Inverse Document Frequency) to weigh words based on their importance in the dataset.
- **N-grams:** Consider including word n-grams (e.g., bigrams, trigrams) to capture word combinations and phrases.
- **Feature Scaling:** Normalize or scale the feature values to ensure consistent ranges.
- **Feature Selection:** Depending on the dataset and analysis goals, you may want to select a subset of the most relevant features to reduce dimensionality.
- **Word Embeddings:** You can also use pre-trained word embeddings like Word2Vec, GloVe, or FastText to represent words as dense vectors.
- **Custom Feature:** You can create custom features based on domain-specific knowledge or analysis requirements.
- **Vectorized Data:**The result of these steps will be a numerical representation of the "Tweet" column for each tweet, which can be used for various machine learning and deep learning tasks, such as classification or clustering.

tweet	top_word1	top_word2	top_word3	top_word4	top_word5
extremist spark terror fear central asia via nikkeiasia terror extrem extremist terror violenc attack propaganda recruit highthreat threat threat http	extremist	terror	threat	fear	highthreat
new hizbul mujahideen terror camp supervis isi detect pakistan occupi kashmir solanwali report http pakistan terror pakistanoccupiedkashmir hizbul	hizbul	pakistan	terror	mujahideen	pakistanoccupiedkashmir
indian polic say five foreign milit kill kashmir terror http	kashmir	milit	five	polic	foreign
soldierspeak bitter truth ani democraci exit pakistan anyon involv corrupt condemn matter ethnic sect ani parti belong corrupt condemn also strongli condemn terror http	condemn	corrupt	ani	belong	bitter
us continu proclaim afghanistan hub terrorist activ continu allow problem grow easili gather terror problem go significantli expand potenti spread central asian state car	problem	continu	asian	easili	hub
ronfilipkowski trump say end war ukrain less hour stop ukrain bomb russia happi http elect gop maga putin terror fbi cia doj dh china north korea iran cuba venezuela nicaragua http	ukrain	cuba	fbi	gop	nicaragua
claxton right yr investig minneapoli polic depart long thi shoulda higher prioriti mcquad make excus imagin terror amp kill happen yr jdbalart msnbc doj blm georgefloyd	yr	claxton	excus	georgefloyd	imagin

Figure 4.1: English Top 5 Words Illustration

Tweet	top_word1	top_word2	top_word3	top_word4	top_word5
2024में किसी की ताकत नहीं जो #नरेंद्र मोदी जी को हरा सके! हर हर मोदी, घर घर मोदी!! जय जय श्री राम	2024	किसी	ताकत	नहीं	नरेंद्र मोदी
आदरणीय #प्रधानमंत्री श्री #नरेंद्र #मोदी जी के #मनकीबात कार्यक्रम का #102वां एपिसोड इस माह #जगदलपुर #बस्तर में सुना। #मन_की_बात #नरेन्द्र #मोदीजी #नरेंद्र_मोदी #नरेन्द्र__मोदी #नरेंद्रमोदीजी #ManKiBaat #Narendra #Modi #NarendraModi #NarendraModiji #Modiji @narendramodi @PMOIndia https://t.co/ZCP53gbe2P	प्रधानमंत्री	मनकीबात	जगदलपुर	एपिसोड	कार्यक्रम
हिमंता बिस्वा सरमा का ये वक्तव्य आपको #नरेंद्र #मोदी के #भारत के नये काल... https://t.co/Tq7GPswRWF via @YouTube	काल	भारत	वक्तव्य	बिस्वा	हिमंता
माननीय प्रधानमंत्री श्री #नरेंद्र मोदी जी के नेतृत्व में #सफलतम 9 वर्ष सेवा, सुशासन और गरीब कल्याण के पूर्ण होने पर #अनुसूचित जनजाति मोर्चा द्वारा #जनजाति जनप्रतिनिधि सम्मेल का आयोजन हुआ । https://t.co/2IyJfzgezi	सफलतम	कल्याण	प्रधानमंत्री	mission2024	जनप्रतिनिधि
जय हिन्द दोस्तों, #महाराष्ट्र के गौरव #पंतप्रधान #नरेंद्र #मोदी जी ने प्रधानमंत्री #आवास योजना से कई ग़रीबों को अपना घर मिला ! #आयुष्मान भारत से लगभग आधी आबादी को लाभ मिला ! (जय हिन्द, जय महाराष्ट्र. जय भीम) https://t.co/tX7u1GoMuz	महाराष्ट्र	आवास	आयुष्मान	विकासशील	नाइजीरिया
अब हम तीसरी महाशक्ति 2011प्रधानमंत्री #मनमोहन_सिंह ! अब हम विकासशील देश नहीं 2016 प्रधानमंत्री #नरेन्द्र मोदी के 2 साल के शासन में ! अब हमारा मुक़ाबला नेपाल श्रीलंका पाकिस्तान नाइजीरिया से चल रहा है #भुखमरी बेरोजगारी में बस एक बन्दा यही है जो देश को कटोरा देगा ? https://t.co/tdge7fAl0W https://t.co/DypXLUsq0y	मनमोहन_सिंह	पाकिस्तान	राजनाथ सिंह	श्रीलंका	अमित शाह
@ramesh_ya1 प्रधानमंत्री #नरेंद्र मोदी जी गृहमंत्री #अमित शाह जी और रक्षा मंत्री #राजनाथ सिंह जी #Ahir Regiment का हक और अधिकार रखते हैं हम इसलिए हम यह लड़ाई लड़ते रहेंगे जब तक आप सेना से जातिवाद खत्म नहीं कर देंगे। #अहिर रेजीमेंट हक है हमारा #Ahir Regiment Demand https://t.co/FpCBK4SXcZ	अमेरिका	वॉशिंगटन	नरेन्द्रमोदी	प्रेमानंदतिगा	स्वावलंबी

Figure 4.2: Hindi Top 5 Words Illustration

4.2 Entity Level

Perform feature extraction from the "Tweet" column in a tweet dataset at the entity level.

- **Tokenization:** Tokenize the text in the "Tweet" column into individual words or phrases to break the text into manageable units.
- **Named Entity Recognition (NER):** Use NER techniques to identify and classify entities such as people, organizations, locations, dates, and more.

Tweet	Name Entity Tagger
#IntelBrief: In addition to #AlQaeda & ISK, there are a number of other terrorist groups active in #Afghanistan, such as the TTP, IMU, Islamic Jihadi Group, Jamaat Ansarullah & remnants of the TIP. https://t.co/ebxyydJ30S#Terrorism #Counterterrorism #ISIS #AlShabaab	[('#AlQaeda &', 'ORG'), ('ISK', 'ORG'), ('Afghanistan', 'GPE'), ('TTP', 'ORG'), ('IMU', 'ORG'), ('Islamic Jihadi Group', 'ORG'), ('Jamaat Ansarullah &', 'ORG'), ('#Counterterrorism #ISIS', 'MONEY'), ('AlShabaab', 'PERSON')]
Afghanistan-based extremists spark terror fears in Central Asia via @NikkeiAsia |#terrorism #extremism #extremists #terror #violence #attack #propaganda#recruitment #highthreat #threat #threats https://t.co/NO6CMqETJo	[('Afghanistan', 'GPE'), ('Central Asia', 'LOC'), ('@NikkeiAsia', 'GPE'), ('#extremism #', 'MONEY'), ('#', 'CARDINAL'), ('#', 'CARDINAL'), ('#', 'CARDINAL'), ('#', 'CARDINAL'), ('#', 'CARDINAL'), ('#', 'CARDINAL')]

Figure 4.3: Entity Tagger on English Tweets

Tweet	Entity
आदरणीय #प्रधानमंत्री श्री #नरेंद्र #मोदी जी के #मनकीबात कार्यक्रम का #102वां एपिसोड इस माह #जगदलपुर #बस्तर में सुना। #मन_की_बात #नरेन्द्र #मोदीजी #नरेंद्र_मोदी #नरेन्द्र__मोदी #नरेंद्रमोदीजी #ManKiBaat #Narendra #Modi #NarendraModi #NarendraModiji #Modiji @narendramodi @PMOIndia https://t.co/ZCP53gbe2P	आ
आदरणीय #प्रधानमंत्री श्री #नरेंद्र #मोदी जी के #मनकीबात कार्यक्रम का #102वां एपिसोड इस माह #जगदलपुर #बस्तर में सुना। #मन_की_बात #नरेन्द्र #मोदीजी #नरेंद्र_मोदी #नरेन्द्र__मोदी #नरेंद्रमोदीजी #ManKiBaat #Narendra #Modi #NarendraModi #NarendraModiji #Modiji @narendramodi @PMOIndia https://t.co/ZCP53gbe2P	##र

Figure 4.4: Entity Tagger for Hindi Tweets

- **Entity Frequency:** Count the frequency of each entity type within the "Tweet" column. This will help you understand which types of entities are mentioned most often.
- **Entity Co-occurrence:** Analyze which entities tend to co-occur frequently. This can reveal relationships or patterns in the data.
- **Entity Features:** Extract features for each identified entity, such as the number of mentions, positions within the text, and context words around the entity.
- **Claims classification:** Perform claim classification on the entities to determine whether they are associated with positive, negative, or neutral sentiment.
- **Word Embeddings:** Utilize word embeddings like Word2Vec or GloVe to represent entities as dense vectors, which can capture semantic relationships between entities.
- **Named Entity Disambiguation:** Resolve entity mentions to canonical forms to deal with variations and ambiguities in entity names.
- **Entity-based Clustering:** Group similar entities together based on their features or embeddings to identify clusters of related entities.
- **Visualization:** Visualize the extracted entity-level features to gain insights and make it easier to interpret the results.

Using these steps we have extracted meaningful features at the entity level from the "Tweet" column in our tweet dataset, allowing us to perform various analyses and gain valuable insights.

Chapter 5

Classification Framework

It's important to note that handling claims in tweet texts, especially across different languages, can be challenging due to the informal and noisy nature of tweets. Classifying claims from Hindi & English tweet texts can be a complex task, but it can be accomplished using various frameworks and techniques. Here is a description of the classification framework for this purpose:

- **Data Collection and Preprocessing:** Gather a diverse dataset of tweet texts in both Hindi and English that contain claims or statements. Preprocess the data, including tasks such as tokenization, lowercasing, removing stopwords, and special character handling.
- **Data Labeling:** Annotate the dataset with labels that represent the nature of the claims. For example, labels are multiclass-like, Simple, Compound, Composite (fact-based, opinion, news, misinformation, etc.).
- **Model Selection:** Choose a suitable classification model based on the dataset type and complexity. Here we used three classifiers namely Logistic Regression, Decision Tree, and SVM, and two models namely BERT, and LSTM.

- **Training:** Split the dataset into training and testing for model training. Train the selected model on the training data, using the extracted features and labeled claims.
- **Multilingual Handling:** For a multilingual classification task, we need to consider language-specific preprocessing and feature extraction techniques for Hindi and English text. Bert's base model can be beneficial, as they can handle multiple languages simultaneously.

5.1 Logistic Regression

Logistic Regression[21] is a statistical model used for binary classification, which involves categorizing data into one of two possible outcomes. It's a fundamental tool in machine learning and statistics. This model works by applying the logistic function, also known as the sigmoid function, to a linear combination of input features, resulting in an output between 0 and 1. This output represents the probability of the data point belonging to the positive class. If the probability is above a certain threshold (usually 0.5), the data point is classified as the positive class; otherwise, it's classified as the negative class. Logistic Regression is interpretable and efficient, making it a popular choice for problems like spam detection, medical diagnosis, and credit risk assessment. It's especially valuable when you want to understand the impact of individual features on the classification decision. Despite its name, logistic regression is not used for regression tasks (predicting continuous values) but for classification tasks (predicting discrete labels).

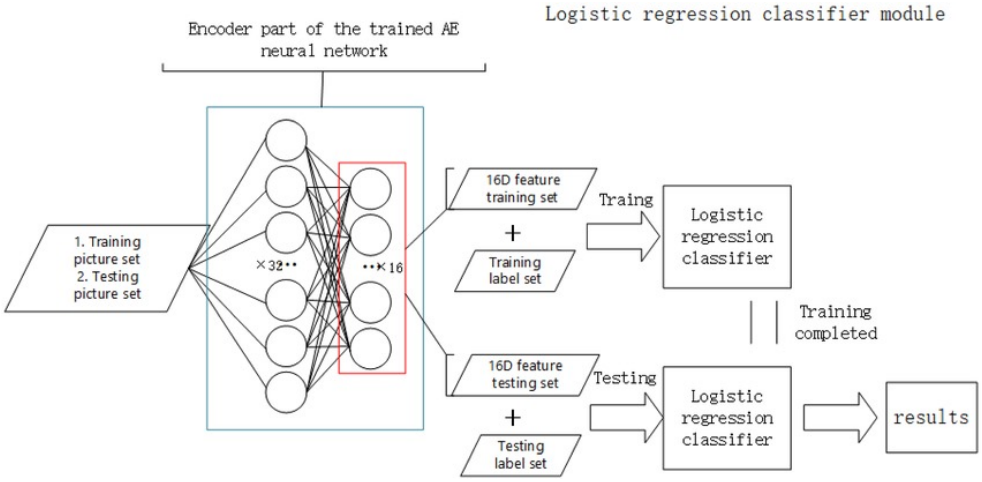

Figure 5.1: Logistic Regression

5.2 Decision Tree

A Decision Tree[22] is a powerful and intuitive machine-learning algorithm used for both classification and regression tasks. It resembles a tree-like structure, where each internal node represents a decision based on a specific feature, each branch corresponds to an outcome of that decision, and each leaf node denotes the final prediction or value. The algorithm recursively partitions the data into subsets, selecting the most informative features at each node to minimize impurity or maximize information gain. Decision Trees are highly interpretable, making them valuable for understanding the decision-making process of the model. However, they are prone to overfitting, which can be mitigated through techniques like pruning. They are widely employed in various domains, including finance, healthcare, and marketing, due to their simplicity, transparency, and ability to handle both categorical and continuous data. Decision Trees serve as the foundation for ensemble methods like Random Forests and Gradient Boosting, which enhance predictive performance by combining multiple trees to

mitigate individual tree biases and errors.

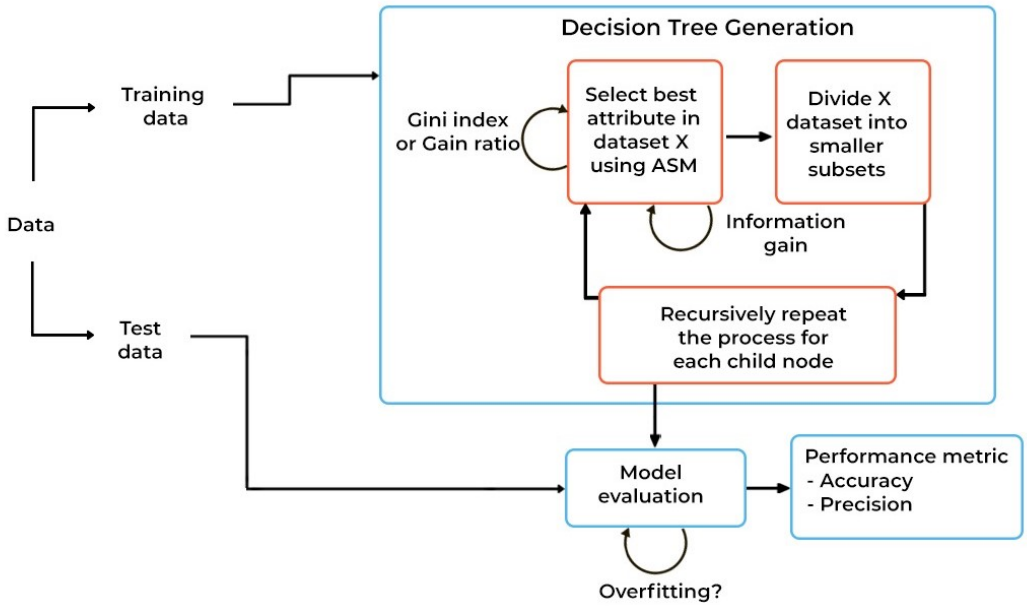

Figure 5.2: Decision Tree

5.3 SVM Classifier

Support Vector Machine (SVM)[23] is a powerful supervised machine learning algorithm used for classification and regression tasks. It works by finding the optimal hyperplane that best separates data into different classes or predicts a target variable with the maximum margin of separation, making it effective in high-dimensional spaces. SVM aims to minimize classification errors while maximizing the distance between the decision boundary and the nearest data points, known as support vectors. This margin ensures robust generalization to new, unseen data. SVM can handle linear and nonlinear data by employing different kernel functions, such as linear, polynomial, and radial basis functions, which transform the input data into higher-dimensional spaces. SVM is widely used in various domains, including image

recognition, text classification, and bioinformatics, due to its versatility and strong theoretical foundation. However, SVM can be sensitive to parameter settings, and its computational complexity increases with large datasets. Still, it remains a valuable tool in the machine learning toolbox for its ability to handle complex decision boundaries and provide accurate predictions.

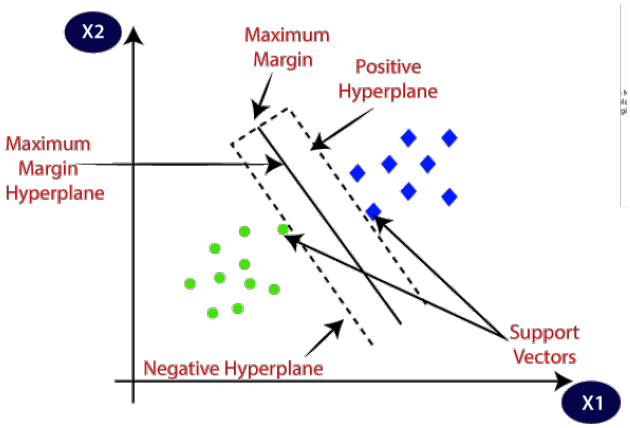

Figure 5.3: Support Vector Machine Classifier

5.4 LSTM - Long Short Term Memory

Long Short-Term Memory[24] (LSTM) is a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem in traditional RNNs. LSTMs are widely used in various applications, particularly in natural language processing and time series analysis. What sets LSTMs apart is their ability to capture long-term dependencies in sequential data. They achieve this by incorporating a memory cell that can store and retrieve information over extended sequences. This cell is equipped with three critical gates: an input gate, an output gate, and a forget gate. The input gate controls the flow of new information into the cell, the forget gate manages the removal of unnecessary information, and the output gate regulates

the information to be passed to the next time step. This gating mechanism enables LSTMs to maintain information over longer periods, making them especially effective in tasks like language modeling, machine translation, and speech recognition. The network's capacity to handle both short-term and long-term dependencies has made LSTMs an integral component of many deep-learning applications where sequential data analysis is essential.

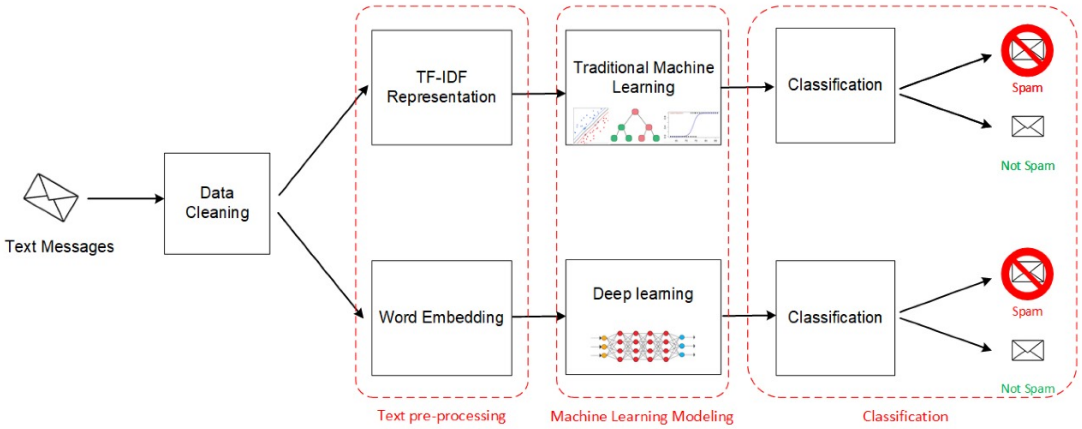

Figure 5.4: LSTM - Long Short Term Memory

5.5 BERT - Bidirectional Encoder Representations From Transformers

BERT[25], or Bidirectional Encoder Representations from Transformers, is a groundbreaking natural language processing model introduced by Google in 2018. Unlike previous language models that read text in a unidirectional manner, BERT revolutionized the field by understanding context bidirectionally. It pre-trains vast corpora of text, learning to predict missing words by considering both the preceding and following words. This bidirectional comprehension enables BERT to capture intricate

nuances and dependencies within language. Fine-tuned for specific tasks, such as claims classification, question answering, or language translation, BERT has set new performance benchmarks in various natural language understanding tasks. Its success has made it a cornerstone in modern NLP, and it has paved the way for subsequent transformer-based models. BERT's contextual understanding and transfer learning capabilities have been leveraged in applications ranging from chatbots and virtual assistants to content recommendation systems, greatly enhancing the way computers interpret and generate human language.

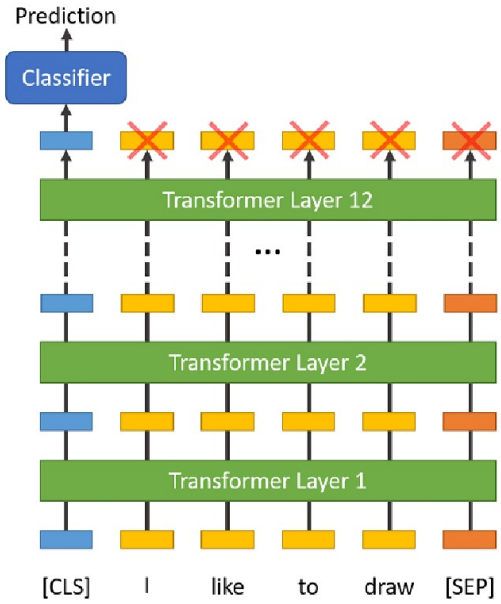

Figure 5.5: BERT - Bidirectional Encoder Representations From Transformers

5.6 Tweet Level and Claim Level Analysis

Tweet-level and claim-level analysis are two approaches used to examine and evaluate information shared on social media platforms, especially in the context of news, rumors, or the impact of statements made on social media. Here's an overview of

each:

- **Tweet-Level Analysis:** Tweet-level analysis focuses on individual tweets or social media posts as the primary unit of analysis. Each tweet is examined in isolation to determine its content, context, and potential implications. **Key aspects of tweet-level analysis include Content:** Analyzing the text, images, videos, or links shared in the tweet. **Source:** Identifying the user who posted the tweet and assessing their credibility or reputation. **Language and Tone:** Evaluating the number of likes, tweets, and comments to gauge the tweets and reach. **Engagement:** Examining the number of likes, retweets, and comments to gauge the tweet's impact and reach. **Timing:** Considering when the tweet was posted in relation to significant events or developments. Tweet-level analysis helps with misleading or false information, as well as the potential for virality and its impact on public opinion.

  In the thesis we use various methods to analyze tweets on multilanguage (english and hindi). First we get raw tweets using tweepy from Twitter portal based on our chosen topics (Delhi Riots, Global Warming, Terrorism, Covid-19, Gujarat riots, Adani, Bollywood, Fake News, Narendra Modi, Karnataka Election, Geo Politics, Manipur crisis, Ram Mandir, Cricket, Artificial Intelligence, Rafale Aircraft, Omicron, Covid Vaccination, Narendra Modi(Hindi tweets), Khalistani, NRC & CAA, Farmer Protest, Article 370, Gujrat Riot(Hindi Tweets), Adani(Hindi Tweets), Wrestlers Protest, Assam Flood, Assam Border Dispute, Ram Mandir(Hindi Tweets), Pulwama Attack, Amrit Mahotsav and Indian Krishibill). We used those tweets for which later we can verify its reality like Narendra Modi discussed about Krishi bills in parliament. If we wish to check the reality of this tweet, we can do this easily.

- **Claim-Level Analysis:** Claim-level analysis takes a broader perspective by examining the veracity of specific claims or statements made across multiple tweets or social media posts. The approach seeks to assess the accuracy and reliability of the information being disseminated. **Key aspects of claim-level analysis include: Identifying Claims:** Identifying specific assertions or claims made by various users on social media. **Fact-checking:** Conduct fact-checking to determine the accuracy of these claims by cross-referencing with credible sources or evidence. **Spread and Influence:** Tracking how these claims propagate across social media platforms and their potential impact on public precipitation or behavior. **Debunking:** If a claim is found to be false, provide evidence and explanations to debunk it and correct misinformation. Claim-level analysis is especially valuable for addressing rumors, conspiracy theories, and misinformation by systematically examining the truthfulness of specific statements.

  Both tweet-level and claim-level analyses are essential tools in the realm of digital media literacy and the fight against misinformation and disinformation. They help individuals, researchers, and organizations better understand the information landscape on social media, enabling them to make informed decisions and counter false narratives effectively.

Chapter 6

Experimental Result

The present effort includes the development of English and Hindi-annotated tweet datasets for claim classification. This dataset is called JUTDP and it is created keeping the DRDO's text classification techniques in mind. Five different models are trained and evaluated on the dataset. These models namely, Logistic Regression, Decision Tree, SVM (Support Vector Machine), LSTM (Long Short-Term Memory), and BERT (Bidirectional Encoder Representations from Transformers).

6.1 Evaluation Metrics

Evaluation metrics [26] are used to measure the quality of the model. One of the most important topics in machine learning and deep learning is how to evaluate models. We have used different models to evaluate our dataset. To analyze these models, various metrics are used. Metrics, such as precision-recall, and F1-scores can be applied to evaluate classification-based methods. Five models namely, Logistic Regression, Decision Tree, SVM (Support Vector Machine), LSTM (Long Short-Term Memory), and BERT (Bidirectional Encoder Representations from Transformers) are

trained and tested on this dataset. We have achieved 49% accuracy on the Logistic Regression model, while the Decision Tree has achieved an accuracy of 46%, SVM (Support Vector Machine) has achieved an accuracy of 62%, LSTM (Long Short-Term Memory) has achieved an accuracy of 48%, BERT (Bidirectional Encoder Representations from Transformers) has achieved an accuracy of 52% for the English language. We have achieved 57% accuracy on the Logistic Regression model, while Decision Tree has achieved an accuracy of 65%, SVM (Support Vector Machine) has achieved an accuracy of 65%, LSTM (Long Short-Term Memory) has achieved an accuracy of 60%, BERT (Bidirectional Encoder Representations from Transformers) has achieved an accuracy of 75% for the Hindi language. In the following subsections, we have defined the metrics that are used for the determination of the classification accuracy of the model. Here, we have shown the main components of evaluation metrics which are obtained by the five state-of-the-art.

- **Accuracy** Accuracy is only a measurement of how frequently the classifier makes accurate predictions. The ratio of the number of accurate forecasts to the total number of predictions is one way to qualify accuracy.

$$Accuracy = (TP + TN)/(TP + TN + FP + EN) \tag{6.1}$$

- **Precision** The precision of a prediction is measured by calculating the number of positive observations that can be anticipated. A low percentage of false positives is indicative of high accuracy.

  1. **Ture Positive (TP):** When an outcome is positive and it is also predicted as positive by the model, it is called True positive.
  2. **True Negative (TN):** When an outcome is negative and it is also predicted

as negative by the model, it is called True negative.

3. **False Positive (FP):** If the number of outcomes that are negative is predicted as positive, it is called false positive. These errors are also called Type 1 Errors.
4. **False Negative (FN):** When the number of outcomes that are positive is predicted as negative, it is called a false negative. These errors are also called Type 2 Errors.

$$Precision = TP/(TP + FP) \tag{6.2}$$

- **Recall** Recall calculates the ability of a classifier to find positive observations in the database. The greater the number of false negatives the model predicts, the lower the recall becomes.

$$Recall = TP/(TP + FN) \tag{6.3}$$

- **F1-Score** It is a single metric that combines both Precision and Recall. The performance of the model improves with increasing F1 scores. The range of F1-score is [0, 1]. The weighted average of recall and accuracy is the F1 score. Both the precision and recall must be high for the classifier to have a high F1 score. This metric solely rewards classifiers with comparable recall and accuracy. The F1 score is a measurement that takes into account both accuracy and recall. The harmonic mean is defined as the simple weighted average of accuracy and recall. Using P for precision and R for recall, we can represent the F1 score as:

$$F1 = 2PR/(P + R) \tag{6.4}$$

Confusion Matrix The performance of the classification models for a certain set of test data is evaluated using a matrix called the confusion matrix. It can be calculated after the real values of the test data are known. Although the matrix itself is simple to understand, some of the terminology may be confusing. It is sometimes referred to as an error matrix as it displays the faults in the model performance in the form of a matrix.

6.2 Results

The outcomes of the three machine learning models and 2 deep learning models have been examined in this section. The outcomes of the foundational learners have been displayed graphically. When it comes to identifying certain claims, certain models have been observed to be more accurate than others. In addition, a report detailing the accuracy of the classifications has been provided as a classification report, and a confusion matrix is also given.

- **Results of Machine Learning Classifiers on our Dataset JUTDP on English Tweets**

Tweet Class	Precision	Recall	F1-Score
Composite	0.00	0.00	0.00
Compound	0.44	0.11	0.18
Simple	0.49	0.94	0.65
Accuracy		0.49	
Macro Avg.	0.31	0.35	0.28
Weighted Avg.	0.53	0.49	0.37

Table 6.1: **Classification Report from Logistic Regression Classifier on English(Tweet Column)**

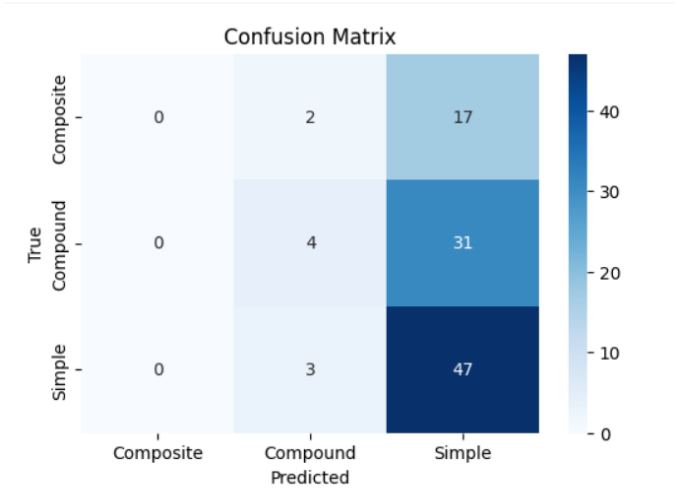

Figure 6.1: **Confusion Matrix from Logistic Regression Classifier on English(Tweet Column)**

Tweet Class	Precision	Recall	F1-Score
Composite	0.23	0.16	0.19
Compound	0.40	0.40	0.40
Simple	0.55	0.62	0.58
Accuracy		0.46	
Macro Avg.	0.39	0.39	0.39
Weighted Avg.	0.44	0.46	0.45

Table 6.2: **Classification Report from Decision Tree Classifier on English (Tweet Column)**

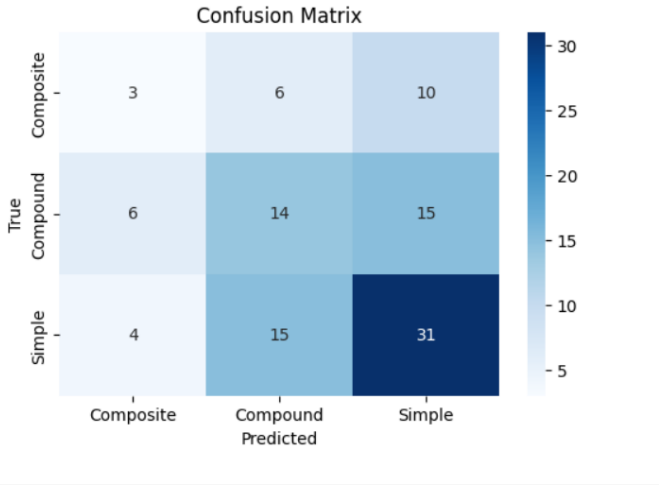

Figure 6.2: **Confusion Matrix from Decision Tree Classifier on English (Tweet Column)**

Tweet Class	Precision	Recall	F1-Score
Composite	0.25	0.05	0.09
Compound	0.50	0.54	0.52
Simple	0.71	0.88	0.89
Accuracy		0.62	
Macro Avg.	0.49	0.49	0.46
Weighted Avg.	0.56	0.62	0.57

Table 6.3: **Classification Report from SVM Classifier on English (Tweet Column)**

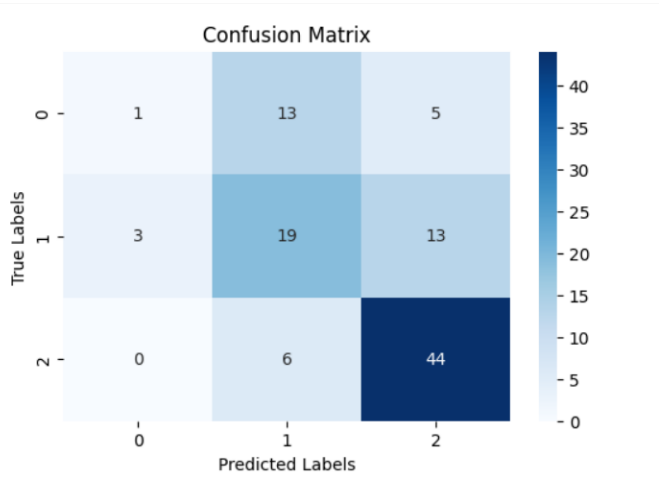

Figure 6.3: **Confusion Matrix from SVM Classifier on English (Tweet Column)**

Tweet Class	Precision	Recall	F1-Score
Composite	0.00	0.00	0.00
Compound	0.17	0.03	0.03
Simple	0.48	0.94	0.94
Accuracy		0.46	
Macro Avg.	0.22	0.32	0.23
Weighted Avg.	0.29	0.46	0.32

Table 6.4: **Classification Report from Logistic Regression Classifier on English (Selected Claim Column)**

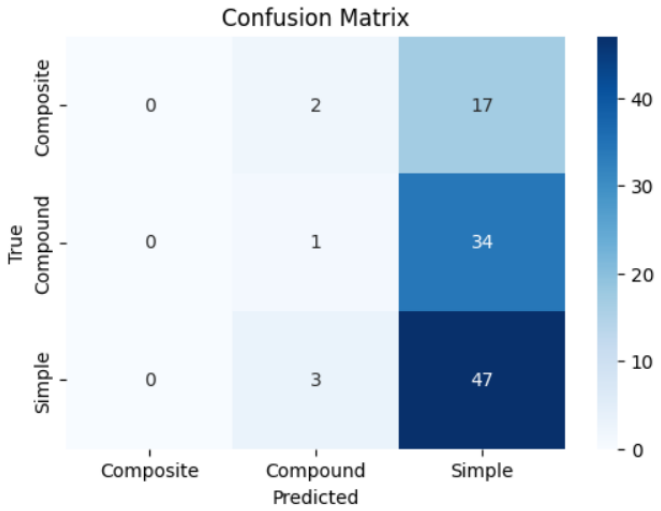

Figure 6.4: **Confusion Matrix from Logistic Regression Classifier on English(Selected Claim Column)**

Tweet Class	Precision	Recall	F1-Score
Composite	0.12	0.05	0.07
Compound	0.42	0.31	0.36
Simple	0.60	0.84	0.70
Accuracy		0.52	
Macro Avg.	0.38	0.40	0.38
Weighted Avg.	0.45	0.52	0.47

Table 6.5: **Classification Report from Decision Tree Classifier on English (Selected Claim Column)**

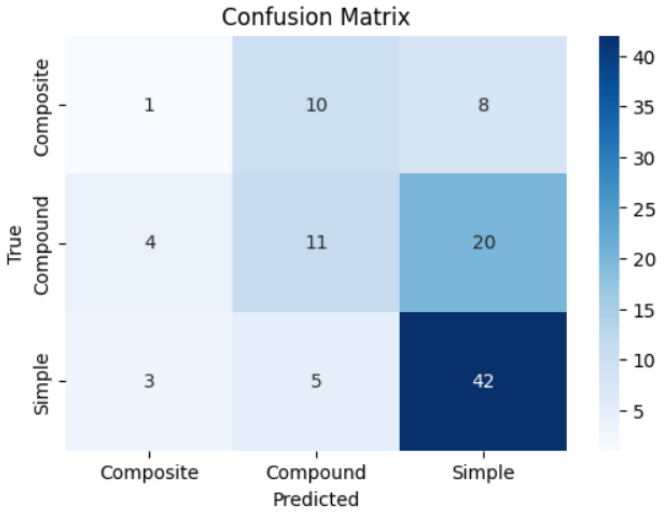

Figure 6.5: **Confusion Matrix from Logistic Decision Tree on English(Selected Claim Column)**

Tweet Class	Precision	Recall	F1-Score
Composite	0.40	0.21	0.28
Compound	0.54	0.43	0.48
Simple	0.65	0.86	0.74
Accuracy		0.60	
Macro Avg.	0.53	0.50	0.50
Weighted Avg.	0.57	0.60	0.57

Table 6.6: **Classification Report from SVM Classifier on English (Selected Claim Column)**

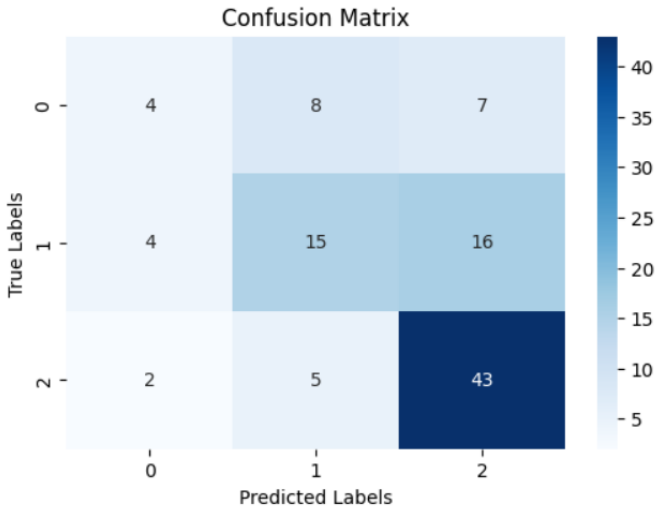

Figure 6.6: **Confusion Matrix from SVM classifier on English(Selected Claim Column)**

- **Results of Machine Learning Classifiers on our Dataset JUTDP on Hindi Tweets**

Tweet Class	Precision	Recall	F1-Score
Composite	0.00	0.00	0.00
Compound	1.00	0.18	0.31
Simple	0.58	0.95	0.72
Accuracy		0.57	
Macro Avg.	0.53	0.38	0.34
Weighted Avg.	0.60	0.57	0.48

Table 6.7: **Classification Report from Logistic Regression Classifier on Hindi (Tweet Column)**

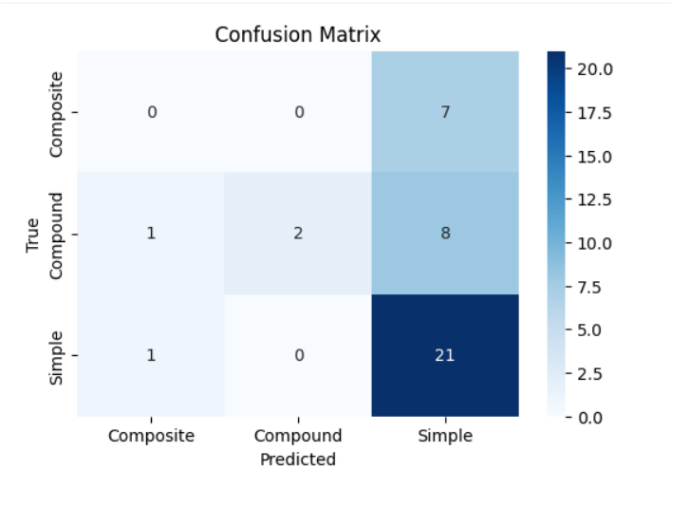

Figure 6.7: **Confusion Matrix from Logistic Regression Classifier on Hindi Tweets(Tweet Column)**

Tweet Class	Precision	Recall	F1-Score
Composite	0.33	0.29	0.31
Compound	0.50	0.36	0.42
Simple	0.77	0.91	0.83
Accuracy		0.65	
Macro Avg.	0.53	0.52	0.52
Weighted Avg.	0.62	0.65	0.63

Table 6.8: **Classification Report from Decision Tree Classifier on Hindi Tweets (Tweet Column)**

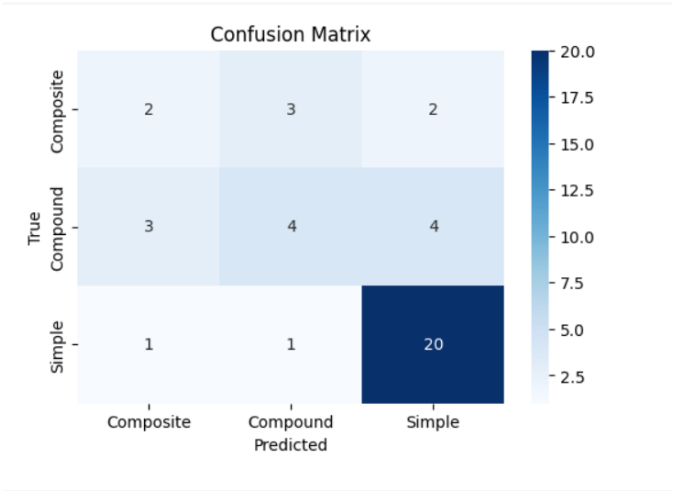

Figure 6.8: **Confusion Matrix from Decision Tree Classifier on Hindi Tweets (Tweet Column)**

Tweet Class	Precision	Recall	F1-Score
Composite	0.33	0.29	0.31
Compound	0.50	0.36	0.42
Simple	0.77	0.91	0.83
Accuracy		0.65	
Macro Avg.	0.53	0.52	0.52
Weighted Avg.	0.62	0.65	0.63

Table 6.9: **Classification Report from SVM Classifier on Hindi Tweets (Tweet Column)**

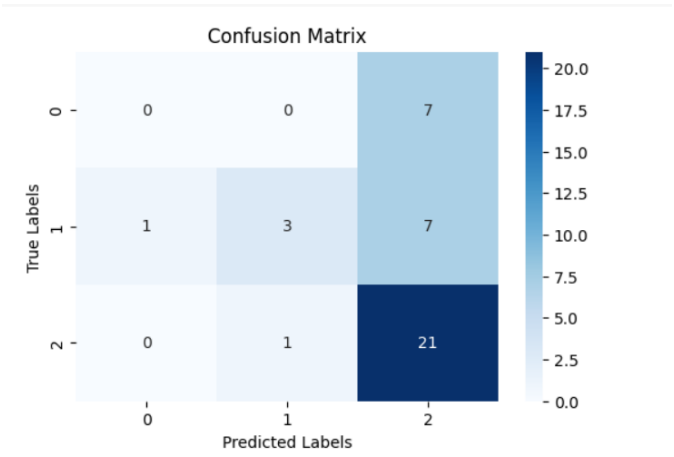

Figure 6.9: **Confusion Matrix from SVM Classifier on Hindi Tweets (Tweet Column)**

Tweet Class	Precision	Recall	F1-Score
Composite	1.00	0.14	0.25
Compound	1.00	0.18	0.31
Simple	0.59	1.00	0.75
Accuracy		0.62	
Macro Avg.	0.86	0.44	0.43
Weighted Avg.	0.78	0.62	0.54

Table 6.10: **Classification Report from Logistic Regression Classifier on Hindi Tweets (Selected Claim Column)**

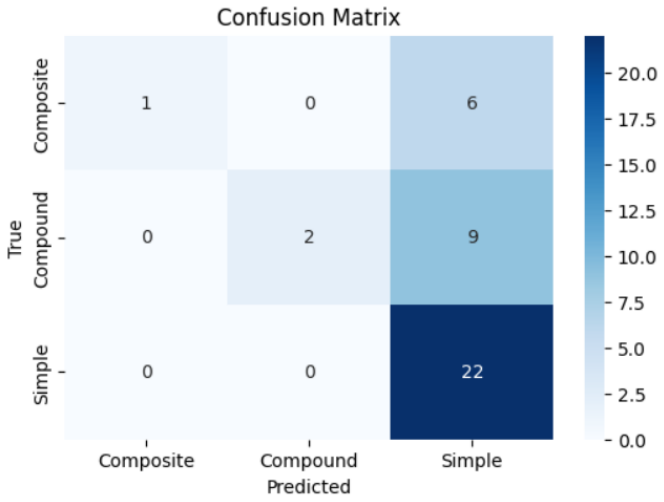

Figure 6.10: **Confusion Matrix from Logistic Regression Classifier on Hindi Tweets (Selected Claim Column)**

Tweet Class	Precision	Recall	F1-Score
Composite	0.33	0.29	0.31
Compound	0.44	0.18	0.25
Simple	0.62	0.82	0.71
Accuracy		0.55	
Macro Avg.	0.45	0.43	0.42
Weighted Avg.	0.51	0.55	0.51

Table 6.11: **Classification Report from Decision Tree Classifier on Hindi Tweets (Selected Claim Column)**

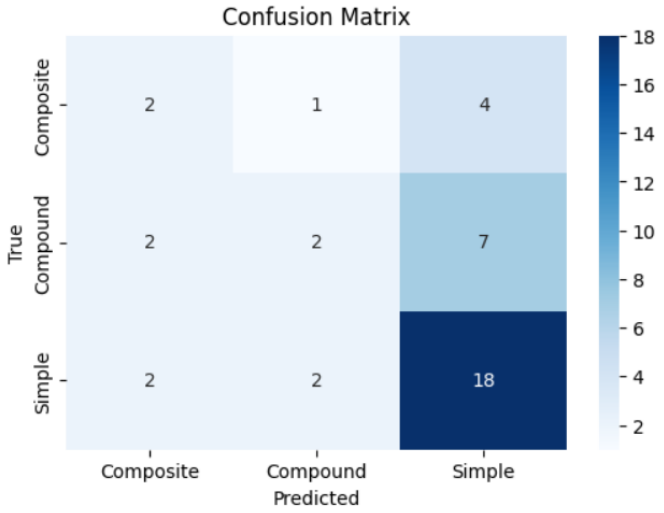

Figure 6.11: **Confusion Matrix from Decision Tree Classifier on Hindi Tweets (Selected Claim Column)**

Tweet Class	Precision	Recall	F1-Score
Composite	0.00	0.00	0.00
Compound	0.67	0.36	0.47
Simple	0.58	0.86	0.69
Accuracy		0.57	
Macro Avg.	0.41	0.41	0.39
Weighted Avg.	0.50	0.57	0.51

Table 6.12: **Classification Report from SVM Classifier on Hindi Tweets (Selected Claim Column)**

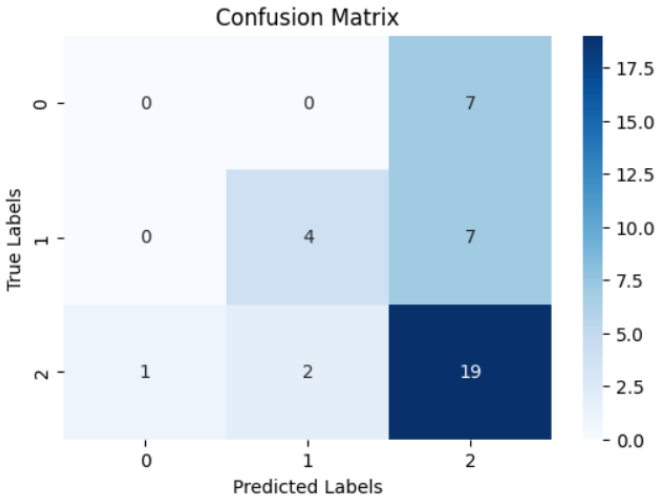

Figure 6.12: **Confusion Matrix from SVM Classifier on Hindi Tweets (Selected Claim Column)**

- **Results of Deep Learning Models on our Dataset JUTDP on English Tweets**

Tweet Class	Precision	Recall	F1-Score
Composite	0.00	0.00	0.00
Compound	0.00	0.00	0.00
Simple	0.48	1.00	0.65
Accuracy	0.48		
Macro Avg.	0.16	0.33	0.22
Weighted Avg.	0.23	0.48	0.31

Table 6.13: **Classification Report from LSTM on English Tweets (Tweet Column)**

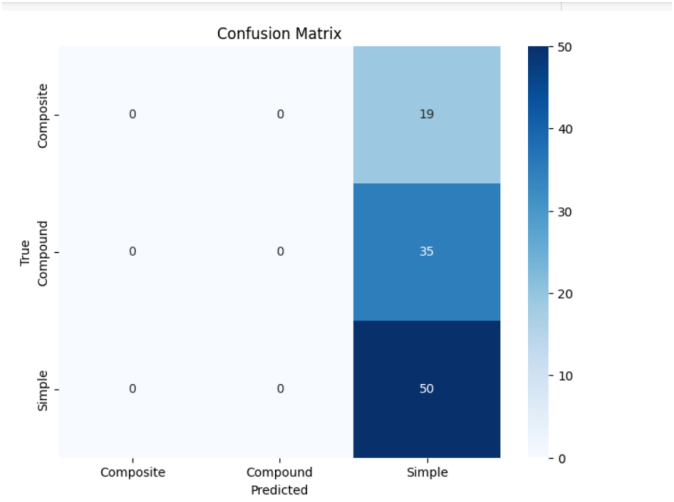

Figure 6.13: **Confusion Matrix from LSTM on English Tweets (Tweet Column)**

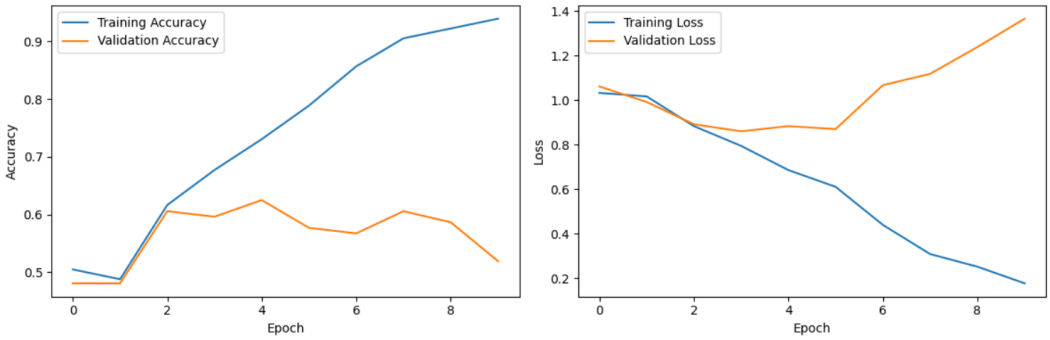

Figure 6.14: **Training-Validation Accuracy and Loss Graph obtained from LSTM Model on English Tweet Column**

Tweet Class	Precision	Recall	F1-Score
Composite	0.29	0.42	0.34
Compound	0.54	0.20	0.29
Simple	0.52	0.78	0.69
Accuracy		0.52	
Macro Avg.	0.48	0.47	0.44
Weighted Avg.	0.53	0.52	0.49

Table 6.14: **Classification Report from BERT on English Tweets (Tweet Column)**

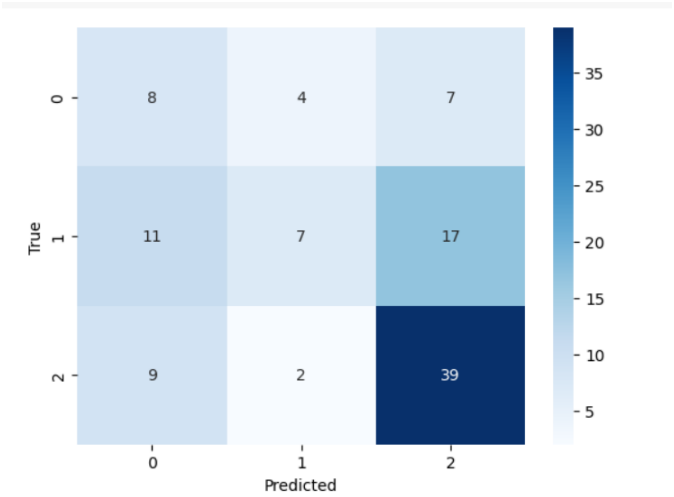

Figure 6.15: **Confusion Matrix from BERT on English Tweets (Tweet Column)**

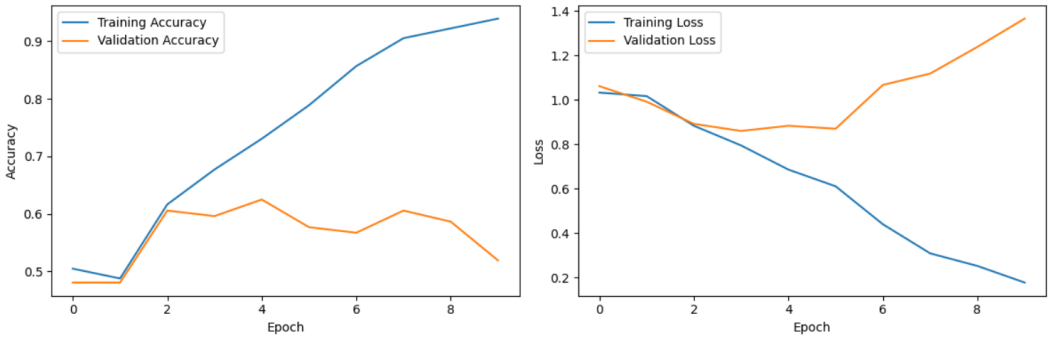

Figure 6.16: **Training-Validation Accuracy and Loss Graph obtained from BERT Model on English Tweet Column**

Tweet Class	Precision	Recall	F1-Score
Composite	0.00	0.00	0.00
Compound	0.40	0.71	0.52
Simple	0.67	0.56	0.61
Accuracy		0.51	
Macro Avg.	0.36	0.42	0.37
Weighted Avg.	0.46	0.51	0.47

Table 6.15: **Classification Report from LSTM on English Tweets (Selected Claim Column)**

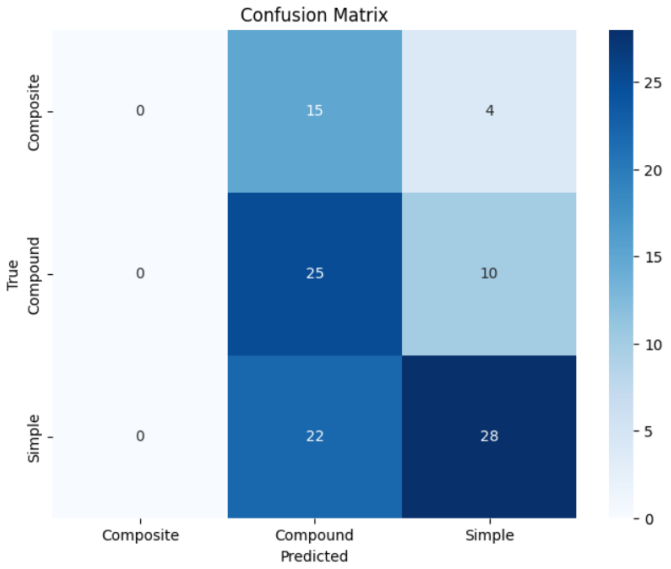

Figure 6.17: **Confusion Matrix from LSTM on English Tweets (Selected Claim Column)**

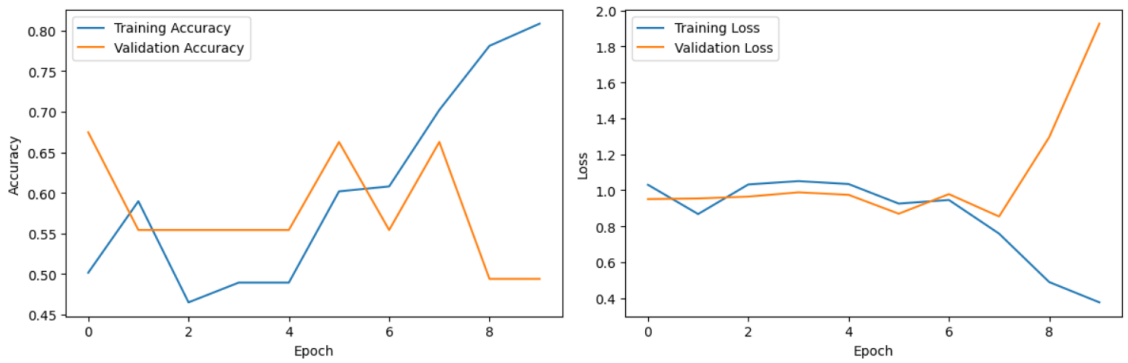

Figure 6.18: **Training-Validation Accuracy and Loss Graph obtained from LSTM Model on English Selected Claim Column**

Tweet Class	Precision	Recall	F1-Score
Composite	0.28	0.53	0.36
Compound	0.64	0.46	0.53
Simple	0.84	0.72	0.77
Accuracy		0.60	
Macro Avg.	0.58	0.57	0.56
Weighted Avg.	0.67	0.60	0.62

Table 6.16: **Classification Report from BERT on English Tweets (Selected Claim Column)**

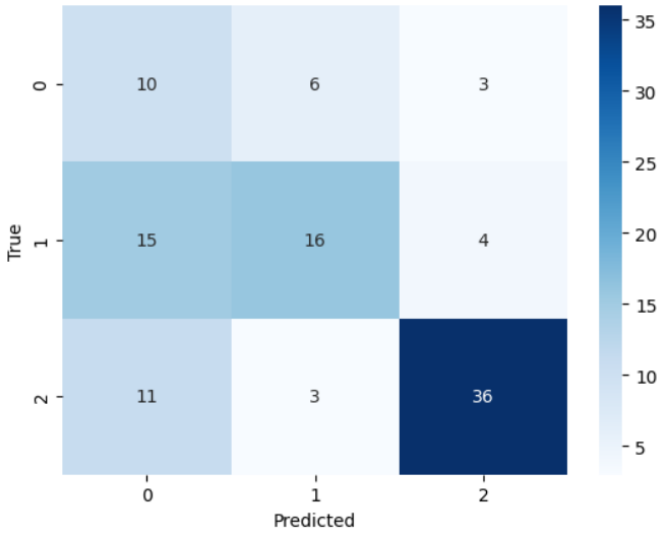

Figure 6.19: **Confusion Matrix from BERT on English Tweets (Selected Claim Column)**

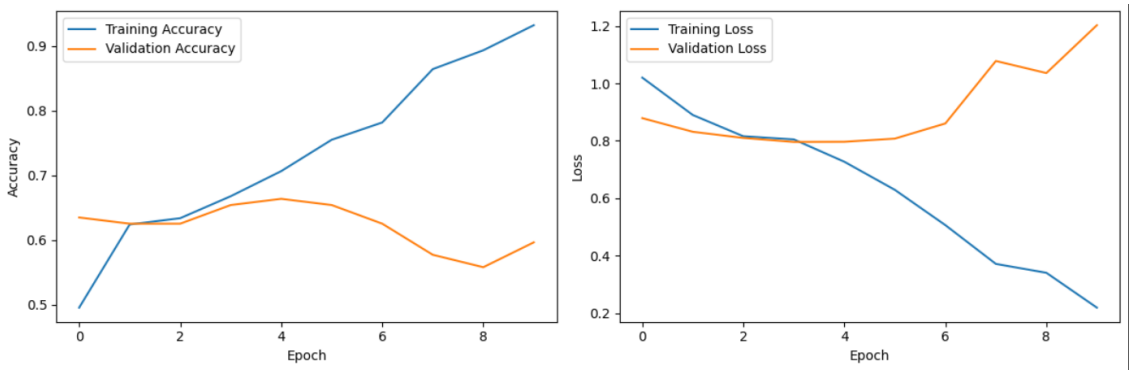

Figure 6.20: **Training-Validation Accuracy and Loss Graph obtained from BERT Model on English Selected Claim Column**

- **Results of Deep Learning Models on our Dataset JUTDP on Hindi Tweets**

Tweet Class	Precision	Recall	F1-Score
Composite	0.00	0.00	0.00
Compound	0.44	1.00	0.61
Simple	0.87	0.59	0.70
Accuracy		0.60	
Macro Avg.	0.44	0.53	0.44
Weighted Avg.	0.60	0.60	0.55

Table 6.17: **Classification Report from LSTM on Hindi Tweets (Tweet Column)**

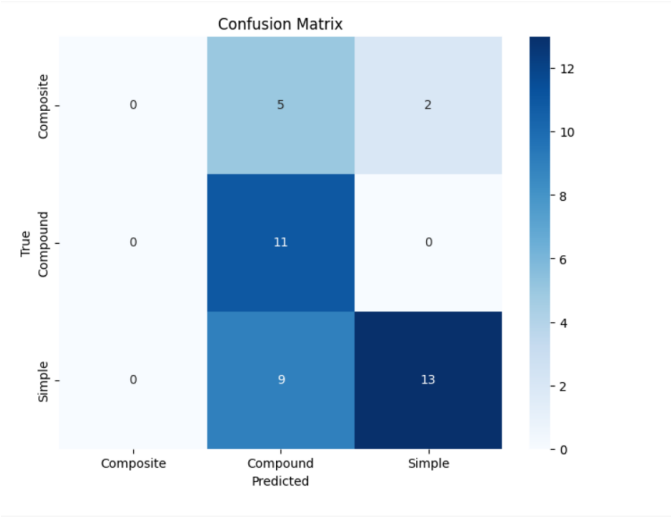

Figure 6.21: **Confusion Matrix from LSTM on Hindi Tweets (Tweet Column)**

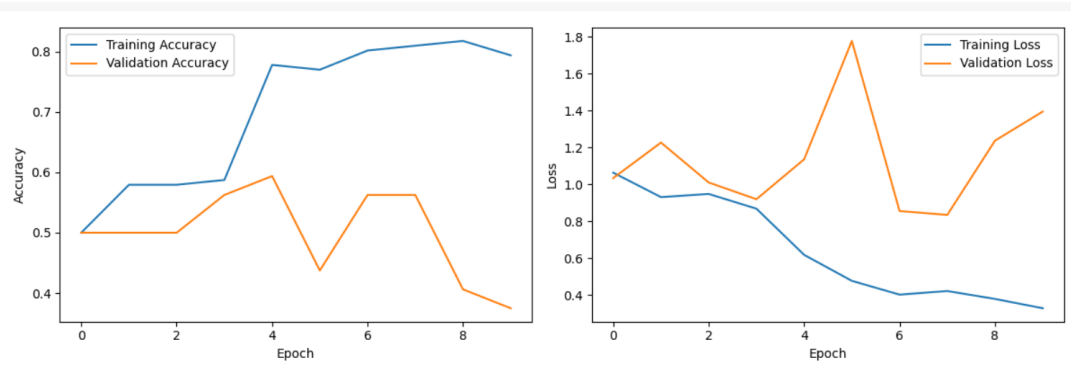

Figure 6.22: **Training-Validation Accuracy and Loss Graph obtained from LSTM Model on Hindi Tweet Column**

Tweet Class	Precision	Recall	F1-Score
Composite	0.50	0.14	0.22
Compound	0.82	0.82	0.82
Simple	0.74	0.91	0.82
Accuracy		0.75	
Macro Avg.	0.69	0.62	0.62
Weighted Avg.	0.72	0.75	0.71

Table 6.18: **Classification Report from BERT on Hindi Tweets (Tweet Column)**

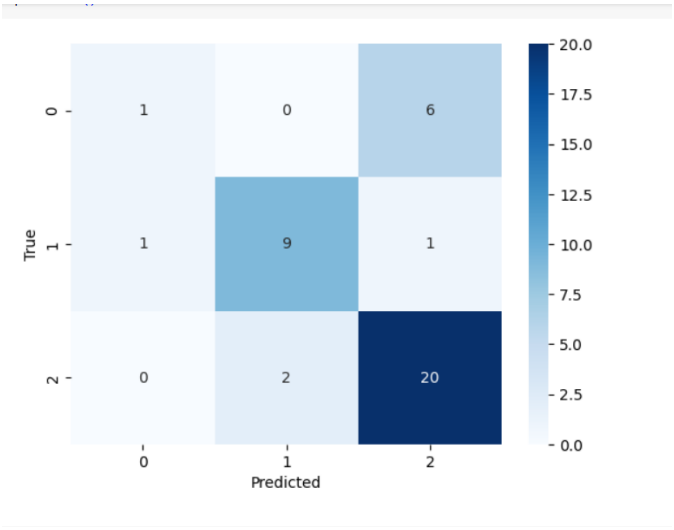

Figure 6.23: **Confusion Matrix from BERT on Hindi Tweets (Tweet Column)**

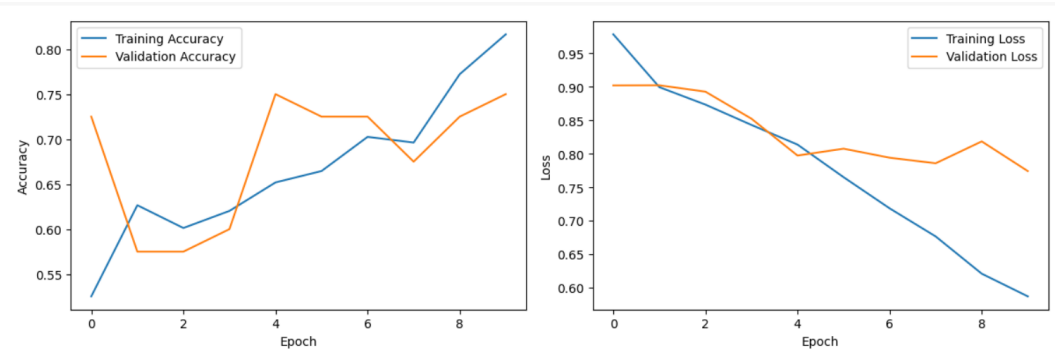

Figure 6.24: **Training-Validation Accuracy and Loss Graph obtained from BERT Model on Hindi Tweet Column**

Tweet Class	Precision	Recall	F1-Score
Composite	0.50	0.14	0.22
Compound	0.82	0.82	0.82
Simple	0.74	0.91	0.82
Accuracy		0.75	
Macro Avg.	0.69	0.62	0.62
Weighted Avg.	0.72	0.75	0.71

Table 6.19: **Classification Report from LSTM on Hindi Tweets (Selected Claim Column)**

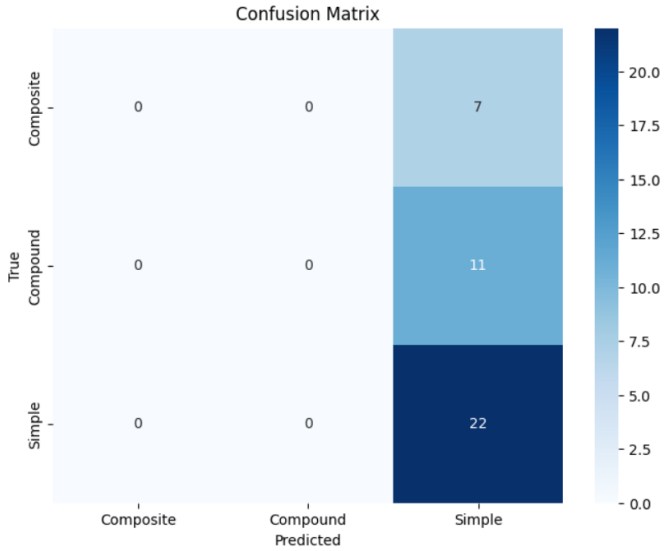

Figure 6.25: **Confusion Matrix from LSTM on Hindi Tweets (Selected Claim Column)**

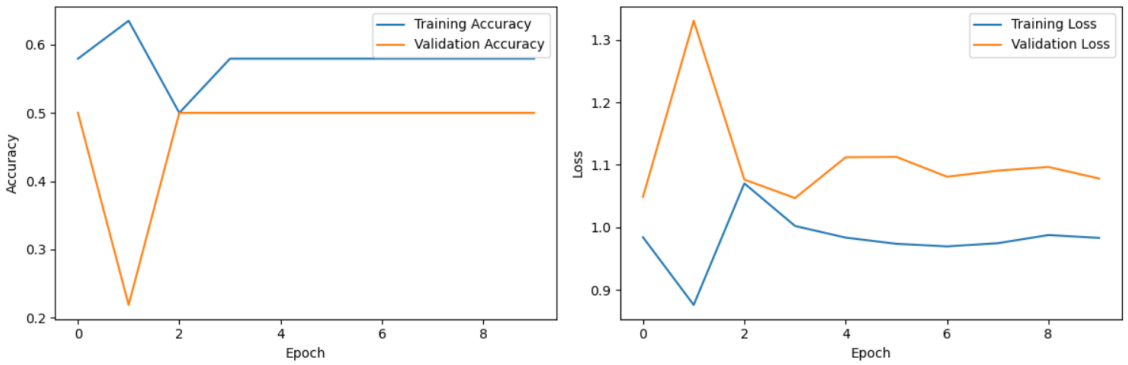

Figure 6.26: **Training-Validation Accuracy and Loss Graph obtained from LSTM Model on Hindi Selected Claim Column**

Tweet Class	Precision	Recall	F1-Score
Composite	0.50	0.14	0.22
Compound	0.82	0.82	0.82
Simple	0.74	0.91	0.82
Accuracy	0.75		
Macro Avg.	0.69	0.62	0.62
Weighted Avg.	0.72	0.75	0.71

Table 6.20: **Classification Report from BERT on Hindi Tweets (Selected Claim Column)**

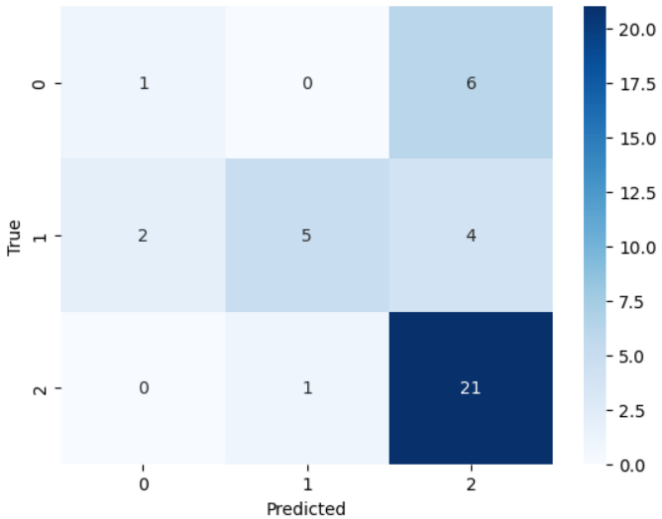

Figure 6.27: **Confusion Matrix from BERT on Hindi Tweets (Selected Claim Column)**

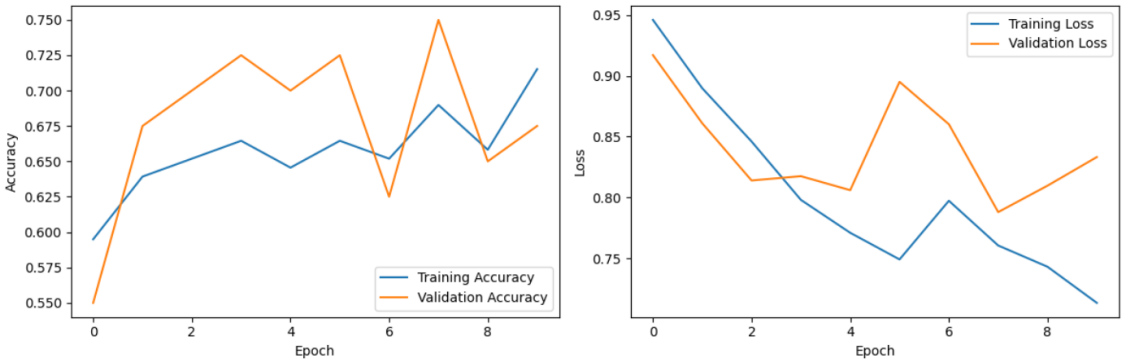

Figure 6.28: **Training-Validation Accuracy and Loss Graph obtained from BERT Model on Hindi Selected Claim Column**

6.3 Result Comparison from Five State-of-the-art

Tweet Level Analysis						
Classifiers	Hindi			English		
	Prec. (%)	Rec. (%)	F-score (%)	Prec. (%)	Rec. (%)	F-score (%)
Logistic Regression	58	95	72	49	94	65
Decision Tree	77	91	83	55	62	58
SVM	60	95	74	71	88	79
BERT	74	91	82	62	78	69
LSTM	87	59	70	48	100	65

Claim Level Analysis						
Classifiers	Hindi			English		
	Prec. (%)	Rec. (%)	F-score (%)	Prec. (%)	Rec. (%)	F-score (%)
Logistic Regression	59	100	75	48	94	64
Decision Tree	62	82	71	60	84	70
SVM	58	86	69	65	86	74
BERT	68	95	79	84	72	77
LSTM	55	100	71	67	56	61

Figure 6.29: **Result comparison between Tweet Level and Claim Level Analysis from our five state-of-the-art.**

6.4 Discussion

Here, we use five states of the arts. That means we used three classifiers and two models. Using three M/L classifiers Support Vector Machine accuracy of 62%, Lo-

gistic regression accuracy of 49%, and Decision tree accuracy of 46% for the English language we achieved a maximum accuracy of 62% through the SVM classifier and achieved the lowest accuracy of 46% with the Decision Tree; Decisions tree accuracy 65%, Support Vector Machine accuracy 60%, Logistic Regression accuracy 57% for Hind language we achieve maximum accuracy 65% through Decision Tree classifier and achieved lowest accuracy 57% with Logistic Regression and two D/L models BERT accuracy 52%, LSTM accuracy 49% for the English language we achieve maximum accuracy 52% through BERT model and achieved lowest accuracy 49% with LSTM; BERT accuracy 75%, LSTM accuracy 60% for the Hindi language we achieve maximum accuracy 75% through BERT model and achieved lowest accuracy 60% with LSTM.

Machine Learning Classifiers	Accuracy(%)
Support Vector Machine	62
Logistic Regression	49
Decision Tree	46

Table 6.21: Accuracy achievement by Machine Learning Classifiers for English Tweets from our Dataset JUTDP

Machine Learning Classifiers	Accuracy(%)
Decision Tree	65
Support Vector Machine	60
Logistic Regression	57

Table 6.22: Accuracy achievement by Machine Learning Classifiers for Hindi Tweets from our Dataset JUTDP

Deep Learning Models	Accuracy(%)
BERT	52
LSTM	49

Table 6.23: Accuracy achievement by Deep Learning Models for English Tweets from our Dataset JUTDP

Deep Learning Models	Accuracy(%)
BERT	75
LSTM	60

Table 6.24: Accuracy achievement by Deep Learning Models for Hindi Tweets from our Dataset JUTDP

After studying all these accuracy scores we understand that in natural language processing for text classification Deep learning models give us better results than Machine learning classifiers. Deep learning models study a full sentence in one iteration but machine learning classifiers study a word in one iteration. So, D/L works first and classifies proper sentiment then M/L. That is also a primary reason for receiving good accuracy from deep learning models instead of machine learning classifiers.

6.5 Miss-classified Tweets for Several Models

A miss-classified tweet refers to a situation where an automated system or algorithm, such as a claims classification tool, categorizes a tweet incorrectly. This could happen when a tweet's content is not accurately assigned to the intended category or sentiment. For example, if a tweet expressing positive sentiment is mistakenly classified as negative, or vice versa, it would be considered a miss-classified tweet. Misclassification can occur due to various factors, including the complexity of lan-

guage, sarcasm, irony, or the limitations of the classification model used.

```
True Label: Compound, Predicted Label: Simple
Tweet:  @INCIndia #Adani's projects stand as beacons of progress and prosperity, transforming local communit

True Label: Compound, Predicted Label: Composite
Tweet: The #geopolitics of #MiddleEast is in a churn &amp;amp; old assumptions are no longer valid in assessing

True Label: Composite, Predicted Label: Compound
Tweet: #Venezuela's Executive Vice President, @delcyrodriguezv received the MoS for International Cooperatio

True Label: Composite, Predicted Label: Simple
Tweet: Covid test kits were being purchased by countries in 2017 .#COVID19 #plandemic #pandemic #governments

True Label: Compound, Predicted Label: Simple
Tweet: And just think yesterday #CovidIsNotOver was #Trending and 🤡🌎 still upset there’s people walking a
```

Figure 6.30: Miss-Classified Tweets for several models

Chapter 7

Conclusions

In conclusion, this thesis provides a comprehensive exploration of claim classification in Hindi and English tweet texts, demonstrating the potential of ML and DL techniques in addressing the challenges of multilingual, informal, and dynamic social media content. As social media continues to evolve, the need for robust and adaptable claim classification models will persist, making this area of research both timely and important. The future work suggested here offers a roadmap for further advancement in this critical domain. In this thesis, we have delved into the challenging task of classifying claims from Hindi and English tweet texts, leveraging a combination of machine learning (ML) and deep learning (DL) techniques. The ever-increasing prominence of social media as a platform for expressing opinions and making claims has necessitated the development of automated tools for assessing the veracity and nature of these claims. This research represents a significant step towards addressing this need, as it tackles the specific challenges posed by multilingual content, informal language, and the dynamic nature of social media discourse. Throughout our studies and investigation, we have accomplished several key objectives:- **Data Col-**

lection and Preprocessing: We collected a diverse dataset of Hindi and English tweets, which were subsequently preprocessed to remove noise, standardize text, and prepare the data for modeling. **Feature Engineering:** We explored various linguistic and content-based features to represent tweet texts, enabling our models to capture the nuanced characteristics of claims effectively. **Machine Learning Models:** We employed traditional ML algorithms, such as SVM, Logistic Regression, and Decision Tree, to establish a strong baseline for claim classification. These models demonstrated competitive performance, emphasizing the importance of feature engineering and data preprocessing. **Deep Learning Models:** We also harnessed the power of deep learning with convolutional neural networks (CNN) and recurrent neural networks (RNN). These models, particularly Bidirectional LSTMs, BERT base, outperformed the traditional ML approaches, highlighting the ability of DL to capture intricate patterns in text data. **Multilingual Aspect:** Our research tackled the multilingual aspect of the problem by considering both Hindi and English, a feature of great significance given the global nature of social media. Our findings underscored the necessity of language-specific models for optimal performance. **Evaluation Metrics:** We employed various evaluation metrics, including accuracy, precision, recall, F1-score, and confusion matrices, to gauge the performance of our models comprehensively.

7.1 Future Work

While this thesis makes substantial contributions to the field of claim classification in multilingual social media text, there are several avenues for future research and improvement:

- **Multimodal Approaches:** Extend the study to incorporate image and video data from tweets, as these modalities can provide valuable context and information for claim classification.
- **Cross-lingual Models**: Investigate the development of cross-lingual models that can effectively handle claim classification in multiple languages without language-specific model training.
- **Real-time Classification:** Adapt the models for real-time claim classification to address the dynamic nature of social media and the need for rapid response to emerging claims.
- **Fine-grained Classification:** Explore the possibility of fine-grained claim classification, which can distinguish between different types of claims (e.g., factual, opinion, controversial, etc.).
- **User and Context Analysis:** Consider incorporating user-profiles and contextual information to improve the accuracy of classification by accounting for user credibility and historical tweet behavior.
- **Ethical and Bias Considerations:** Examine the ethical implications and potential biases in automated claim classification and develop strategies to mitigate biases and promote fairness.
- **Scalability and Generalization:** Evaluate the scalability of the models to handle large-scale social media data and generalize the approaches to other languages and platforms.

References

[1] KR1442 Chowdhary and KR Chowdhary. "Natural language processing". In: *Fundamentals of artificial intelligence* (2020), pp. 603–649.

[2] Santosh K Gaikwad, Bharti W Gawali, and Pravin Yannawar. "A review on speech recognition technique". In: *International Journal of Computer Applications* 10.3 (2010), pp. 16–24.

[3] Atro Voutilainen. *Part-of-speech tagging*. Vol. 219. The Oxford handbook of computational linguistics, 2003.

[4] Roberto Navigli. "Word sense disambiguation: A survey". In: *ACM computing surveys (CSUR)* 41.2 (2009), pp. 1–69.

[5] Alireza Mansouri, Lilly Suriani Affendey, and Ali Mamat. "Named entity recognition approaches". In: *International Journal of Computer Science and Network Security* 8.2 (2008), pp. 339–344.

[6] Heeyoung Lee et al. "Stanford's multi-pass sieve coreference resolution system at the conll-2011 shared task". In: *Proceedings of the fifteenth conference on computational natural language learning: Shared task*. 2011, pp. 28–34.

[7] Kamran Kowsari et al. "Text classification algorithms: A survey". In: *Information* 10.4 (2019), p. 150.

[8] Mihai Dusmanu, Elena Cabrio, and Serena Villata. "Argument mining on Twitter: Arguments, facts and sources". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2017, pp. 2317–2322.

[9] Isabelle Guyon and André Elisseeff. "An introduction to feature extraction". In: *Feature extraction: foundations and applications*. Springer, 2006, pp. 1–25.

[10] Ahmed Redha Mahlous and Ali Al-Laith. "Fake news detection in Arabic tweets during the COVID-19 pandemic". In: *International Journal of Advanced Computer Science and Applications* 12.6 (2021), pp. 778–788.

[11] S Suryavardan et al. "Factify 2: A multimodal fake news and satire news dataset". In: *arXiv preprint arXiv:2304.03897* (2023).

[12] Stefan Helmstetter and Heiko Paulheim. “Collecting a large scale dataset for classifying fake news tweets using weak supervision”. In: *Future Internet* 13.5 (2021), p. 114.

[13] Buket Erşahin et al. “Twitter fake account detection”. In: *2017 International Conference on Computer Science and Engineering (UBMK)*. IEEE. 2017, pp. 388–392.

[14] Maryam Heidari et al. “Bert model for fake news detection based on social bot activities in the covid-19 pandemic”. In: *2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE. 2021, pp. 0103–0109.

[15] Bibek Upadhayay and Vahid Behzadan. “Sentimental liar: Extended corpus and deep learning models for fake claim classification”. In: *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*. IEEE. 2020, pp. 1–6.

[16] William Yang Wang. “” liar, liar pants on fire”: A new benchmark dataset for fake news detection”. In: *arXiv preprint arXiv:1705.00648* (2017).

[17] Valerio La Gatta et al. “Retrieving false claims on Twitter during the Russia-Ukraine conflict”. In: *arXiv preprint arXiv:2303.10121* (2023).

[18] Mohammadreza Samadi, Maryam Mousavian, and Saeedeh Momtazi. “Persian fake news detection: Neural representation and classification at word and text levels”. In: *Transactions on Asian and Low-Resource Language Information Processing* 21.1 (2021), pp. 1–11.

[19] Arianna D’Ulizia et al. “Fake news detection: a survey of evaluation datasets”. In: *PeerJ Computer Science* 7 (2021), e518.

[20] Subhabrata Dutta et al. “A Multilingual Dataset for Identification of Factual Claims in Indian Twitter”. In: *Proceedings of the 14th Annual Meeting of the Forum for Information Retrieval Evaluation*. 2022, pp. 88–92.

[21] Michael P LaValley. “Logistic regression”. In: *Circulation* 117.18 (2008), pp. 2395–2399.

[22] Anthony J Myles et al. “An introduction to decision tree modeling”. In: *Journal of Chemometrics: A Journal of the Chemometrics Society* 18.6 (2004), pp. 275–285.

[23] Thorsten Joachims. *Making large-scale SVM learning practical*. Tech. rep. Technical report, 1998.

[24] Yong Yu et al. “A review of recurrent neural networks: LSTM cells and network architectures”. In: *Neural computation* 31.7 (2019), pp. 1235–1270.

[25] Jacob Devlin et al. “Bert: Pre-training of deep bidirectional transformers for language understanding”. In: *arXiv preprint arXiv:1810.04805* (2018).

[26] Mohammad Hossin and Md Nasir Sulaiman. “A review on evaluation metrics for data classification evaluations”. In: *International journal of data mining & knowledge management process* 5.2 (2015), p. 1.