FAKE NEWS DETECTION USING FINE-TUNEDPRETRAINED LANGUAGE MODELS

A PROJECT REPORT
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTSFOR THE AWARD OF THE DEGREE OF

MASTER OF TECHNOLOGYIN
ARTIFICIAL INTELLIGENCE

Submitted by

ADITYA JAIN
2K21/AFI/20

Under the supervision of
DR. PRASHANT GIRIDHAR SHAMBHARKAR
Assistant Professor
DEPARTMENT OF COMPUTER SCIENCE &ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)Bawana Road, Delhi 110042

MAY, 2023

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)Bawana Road, Delhi-110042

CANDIDATE'S DECLARATION

I, ADITYA JAIN, Roll No's – 2k21/AFI/20 student of M.Tech (Department of Computer Science & Engineering),hereby declare that the project Dissertation titled "FAKE NEWS DETECTION USING FINE-TUNED PRETRAINED LANGUAGE MODELS" which is submitted by me to the Department of Computer Science, Delhi Technological University, Delhi in partial fulfilment of the requirement for the award of degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any Degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi ADITYA JAIN

Date: 31.05.2023

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the Project Dissertation titled "FAKE NEWS DETECTION USING FINE-TUNED PRETRAINED LANGUAGE MODELS" which is submitted by ADITYA JAIN, Roll No's – 2K21/AFI/20, DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, Delhi Technological University, Delhi in partial fulfilment of the re- quirement for the award of the degree of Master of Technology, is a record of the project work carried out by the students under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi DR. PRASHANT GIRIDHAR SHAMBHARKAR

Date: 31.05.2023 SUPERVISOR

DEPARTMENT OF COMPUTER ENGINEERING & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

We wish to express our sincerest gratitude to DR PRASHANT GIRIDHAR SHAMBHARKAR and NIKHIL SHARMA Department of Computer Science & Engineering DELHI TECHNOLOGICAL UNIVERSITY (Formerly Delhi College of Engineering) Bawana Road, Delhi 110042 MAY, 2023 for his continuous guidance and mentorship that he provided us during the project. He showed us the path to achieve our targets by explaining all the tasks to be done and explained to us the importance of this project as well as its industrial relevance. He was always ready to help us and clear our doubts regarding any hurdles in this project. Without his constant support and motivation, this project would not have been successful.

Place: Delhi ADITYA JAIN

Date: 31.05.2023

Abstract

Due to lightning growth of fake news the erosion of democratic and public trust, fake news detection and intervention are becoming increasingly important. In order to ensure that the public gets the correct information, various Machine Learning Deep Learming approaches have been developed regardless of whether the news is traditional or non-traditional. The aim of this paper is to identify false news using cutting-edge detection methods i.e. Analysis of fake news taxonomy has been conducted, therefore the purposeis to examine cutting-edge methods for spotting false news and discuss the drawbacksof those methods. Additionally, it discussed credibilityBased, social contextBased, time-Based substance-based detection of false news. In Addition we reviewed the recent work of other academics on the subject and provided a roadmap for future study to aid scholarsin combating the issue related to fake news And In recent years, proliferation of social media websites has resulted in a surge of fake news being created and disseminated for various commercial and political reasons. The use of misleading language in these online articles makes it easy for users to be misled, leading to significant offline consequences. Therefore, crucial steps should be taken to enhance the reliability of information in online social networks.we propose a technique for detecting fake-news that leverages fine-tuned pretrained language models(including BERT, AlBERT, DistilBERT, XLNet RoBERTa). Our approach involves preprocessing the text data, then fine tuning the pretrained models on dataset of labeled fake-real news, and then evaluating their performance on a held-outtest set. Our experiments demonstrate that finetuning pretrained models can effectively detect fake-news with high-accuracy, and can be a promising approach for building auto-mated systems to combat the spread of misinformation.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 MOTIVATION

The proliferation of false news has grown to be a serious problem in politics, the media, and society at large in recent years. Spreading false information has the potential to erode public discourse quality, threaten democratic institutions, and harm faith in reliable news sources. Therefore, precise and trustworthy techniques for identifying bogus news are required. The use of pre-trained language models, which have demonstrated outstanding performance on a number of natural language processing tasks, is one promising strategy for detecting false news. We may use these models' capacity to comprehend intricate linguistic patterns and contextual clues to increase the precision and dependability of false news identification. We want to contribute to the field of fake news identification with this thesis by evaluating the effectiveness of pre-trained language models for detecting bogus news.. On a huge corpus of news stories, we will specifically polish a number of cutting-edge language models and assess how they perform in comparison to more conventional machine learning techniques. We will also look into how different hyperparameters and training approaches affect how well these models function. This study identifies efficient pre-trained language models and honed techniques for false news identification. This might enhance the reliability and accuracy of fake news identification and aid in the creation of more potent techniques for halting the spread of false information. This is a crucial and current issue that should be researched since it might have an effect on many different professions and industries.

1.2 WHAT IS FAKE NEWS?

1.2.1 Definition

News has also gone digital as a result of the internet revolution and content digitalization. However, it also has unfavourable side effects. The problem of spreading false information and influencing people's behaviour has become widespread across society as a result of fake news. Recent examples of false news may be seen in the US presidential election [1] verma et al ,when it was utilised to sway voters' opinions.

With the advent of the technology, people's reliance on News has multiplied in INDIA as well [2] paka et al .Even in the case of digital news, individuals are more likely to pay attentionto a news story featuring a visual component than one that is purely composed of text. The news organisations profit greatly from this concept. News stories with accompanyingvisuals are more widely read and have a wider audience.

Unfortunately, a number of online elements have utilised this concept in a notorious manner. The propagation of the fake news is aided by morphed images, misleading information, and the audience that the false news is intended for.

In this thesis, we experiment the possibility of detecting fake news which has textual and visual information embedded in a single package by applying fake news detection using fine-tuned pretrained language models(including BERT, AlBERT, DistilBERT, XLNet RoBERTa) performed on a FNC dataset containing distinct kind of news.

1.2.2 Fake News Characterization

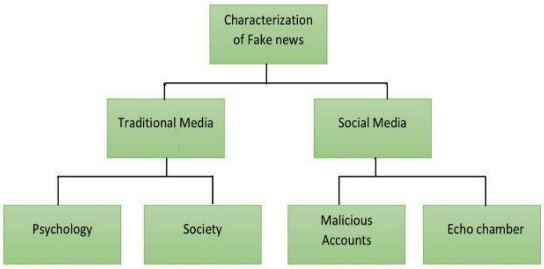

Figure 1.1: Characterization of Fake news

Traditional Media : Media that existed before social media referred as traditional media. Fake news generation and propagation are driven by psychological factors as well as the societal environment in which people are living. This is expressed as the following:

(1) Psychology: Humans are predisposed to believe fake-news due to inbuilt psychological vulnerabilities.An example of this vulnerability is naive-realism, in which people assume only their own interpretation of the world is real, while others' interpretationsare incomplete or misinformed [3] jawet et al . "Desirability bias" is characterized by assimilation of information that is entertaining for the respondent , while "confirmation bias" is characterized by belief only in information that conforms to the perception of the respondent[4] Thota et al. , "Illusion of asymmetric insight" is a vulnerability in which people are convinced they are more knowledgeable than others [5] vishal et al. These psychological factors, therefore, create an environment where fake news can thrive.

(2) Society: There are 2 main factors ie consumers and publishers that help in consuming and generating the information. As a result of biases involved in the processing of certain events, there is a certain degree of distortion in the representation of a news article. Additionally, publishers may have certain expectations or interests that can be seenas short-term utilities, such as financial gains and readership while publisher's reputation and authenticity are long-term utilities[6] Roy et al.

(B) Social Media

False news spreading on social media platforms is mainly caused by malicious accounts echo chamber effects,are outlined below:-

(1) Malicious-Accounts: Various methods are used by malicious accounts to spread fake news, including cyborg accounts, which are created by human being but run, program

influence behalf of certain activities [7] Nasir et al .Their main aim is to either fame or demote an organisation or individual

(2) Echo-Chambers: Connections and made friendships that have allowed people with similar tastes opinions to interact with one another and form a network despite being physically far apart [8] jay et al. For instance, the majority of Facebook users associate with others who have similar opinions and beliefs [9] paka et al. Also the more often a news article is circulated, the more people accept its veracity even if it's based on an unverifiable source [10] varun et al.

Hence, we can say authenticity intent can be used to detect the false news. Verifying information in news report is the definition of authenticity. This could be done to make sure that there are no conspiracies at play and that the news reports haven't been changed. The second element, intent, refers to the motivation behind false information. Whether the news is intended to intentionally mislead the audience.

1.3 FEATURE EXTRACTION

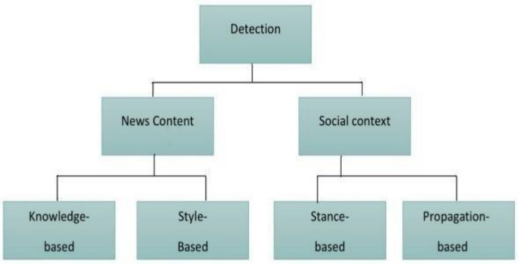

Figure 1.2: Detection Using Fake news

1.3.1 News Content Features

Now that a fake news has been defined and a goal has been established, attention isturningto the characteristics that can be utilised to categorise a false news. According to the general guidelines and news content, false news is made up of four main elements:

1. Source: Information about the news's originator, timestamps, and credibility.
2. Headline: The headline, or opening line, is what draws readers' attention to the news item.
3. Body Text: The news's textual content, or the article's main body.
4. Image-Video: Another component of fake news, which typically combines text with images and

video The four components described above are divided into features. Text and pictures/videos are the most significant components of this. These characteristics are linguistic and vi- sual. The key distinction based on linguistic qualities is the differentiation of language, While false news may be well-altered to create confusion, present the facts incorrectly, or even stray from reality, true news always be more formal. Lexical cues can be used to get over these issues and identify the bogus lingual material.

1.3.2 Social Context Features

Three crucial factors are constantly taken into account while sharing news on social media. User viewpoint, post perspective, and group perspective are these elements.

In the event of a news story, for example, it is possible to assess certain users' activity the meta-data to determine whether a user is at risk. Risks include the potential to propagate incorrect information or to fall into its trap.

In order to offer information on the veracity of news, post aspect is utilised to analyse the meta-data from the different posts. Similar to posts or users, group metadata is analysed by the group aspect to provide key information about that specific group

1.4 NEWS CONTENT MODELS

1.4.1 Knowledge-based models

The model based on news content's main objective is to determine the reality of an article's key assertions in order to verify the news's legitimacy or truthfulness. Assigning truth values to assertions based on their context and external sources is the most important principles of factchecking news content[10].

The following categories can be used to group current fact-checking techniques:

(1) Expert-oriented: It is a process of examining important data documents in order to construct verdicts of claim veracity that relies heavily on human domain experts, such as PolitiFact11 or Snopes12. And expert-oriented factchecking requires a lot of time and mental effort.

(2) Crowd Sourcing: Factchecking based on crowdsourcing uses "wisdom-of-crowds" to enable common people to fluctuate the news content to determine the news' authenticity.For eg,In Fiskkit13, users interact with each other to discuss and annotate specific articles' accuracy.

(3) Computational Oriented : The purpose is to generate an automatic scalable system that helps in classifying true from false claims. Two major problems have beensolved by previous computation-based fact checking methods: the identification of factual claims that are worth checking and the differentiation between factual claims that are not true[11] Each of the strategies discussed above has advantages and disadvantages. When it comes to an expert-oriented approach, employing people can be expensive and time-consuming. The experts might just be slightly knowledgeable in that area, and it might be laborious for them to process all the new information. The majority may not always be correct, and news stories may be written in a way to circumvent fact-checking algorithms. The output of calculation techniques might not always be accurate.

1.4.2 Style-Based Model

A fake news publisher often Influences large consumer communities by disseminating distorted and misleading information, which requires specific writing style in order to appeal persuade a large audience.It is possible to categorize them as follows :

(1) Deception-oriented : This method is based on the Undeutsch Hypothesis in forensic psychology, which is motivated by misleading statements or claims contained in news content. Content analysis tools such as cientific-based Content Analysis [12] CriteriaBased Content Analysis [38] has been developed as forensic tools.

(2) Objectivity-oriented : A lack of objectivity in news content can be indicated by hyperpartisan styles and yellow journalism,which can lead to consumer misinformation.It is common for hyperpartisans to exhibit the behavior in support of specific party, and this behavior is often correlated with the creation of fake news. Linguisticbased features can be used to find the hyperpartisans article[13].

1.5 SOCIAL CONTEXT MODELS

As well as completing and improving News Content Models.As a result, social context models examine significant user social interactions from many angles.These are possible classifications for them:

1.5.1 Stance based

As well as completing and improving News Content Models.As a result, social context models examine significant user social interactions from many angles.These are possible classifications for them:

1.5.2 Propagation based

The links between pertinent social media postings are analysed in order to forecast news credibility[14]. Propagation-based methods rely on elements of sharing that are dependent on propagation, such as social media likes, comments, and the quantity of retweets.

Chapter 2

LITERATURE REVIEW

2.1 INTRODUCTION

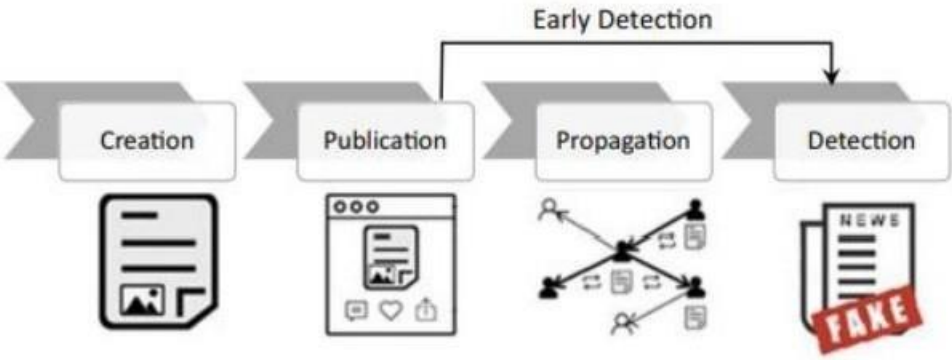

Figure 2.1: Importance of Fake news Detection[7]

The term ”fake news” has been used by politicians and media personalities. According to most definitions, fake news refers to fabricated or untrue stories about political or non-political issues that are meant to deceive or mislead. In addition to traditional media like newspapers, television non-traditional media, including social media and websites, can also spread these types of stories. Often, these fake news stories focus on the political landscape and are used to tarnish the reputations of individuals or organizations, or to manipulate public discourse in favor of the news distributors’ interests. Some of these fake news stories aim to gain popularity and sensationalism among the masses, posing a serious threat to free expression and debate. Recent advancements made it easily possible to understand human behavior through various emerging technologies. In ancient times, the idea of machine-human contact was only an imagination. The ability to manage a territory and contribute to the development of a civilization is based on the co-operation of groupsof people. To regulate these activities, the state’s government establishes regulations that society must follow. Various emerging technologies can assist the government in ensuring the security of its policies and its citizens[15]. Fake news is information that is false or misleading and is presented as real news[16]. During US elections in 2016, the phrase ”fake-news” introduced. As a result, Facebook, Twitter, and Google all took efforts to combat fake information.As online news portals has dramatically increased, making it harder to tell if anything is true or false.

As Internet access has improved, Twitter is easy to access and quickly disseminates news, it has become popular to spread and receive news.However, it becomes very difficult to differentiate between legitimate the fake news on these platforms.False news can be propagated unintentionally by wellmeaning individuals without rigorous verification[17].Proliferation of fake news and false data has escalated, potentially causing major issues in society. As an example malicious people circulate false news about a business which badly effect the company's reputation.It has been noted that the links between pol- itics and misleading information predate the advent of ”false news”. Although research and its outcome have demonstrated that a link exists between the two. Historically, media organizations, particularly those affiliated with political parties, have emphasized one-sided views and unreliable material [18].

In modern era, politicians members of the media have been using term ”fake news” commonly. According to most definitions, fake news refers to fabricated or untrue stories about political or non-political issues intended to mislead or to deceive someone.In addi- tion to traditional news media like print newspapers and television, non-traditional media such as social media websites can disseminate these types of stories.These news, stories, or articles primarily focused on the political landscape to tarnish the reputations of indi- viduals or organizations, change the public discourse to suit the interests of these news distributors.Moreover, some of these fake news aim to gain popularity and sensationalism among the masses, which poses a serious threat to free expression and debate.Human behavior can be understood through several emerging technologies. In Ancient time, thinking of machine-human contact was just the Imagination only[19].This makes us feel connected to our planet and makes it possible for us to live in a big civilization.It is theability of a group of people to manage a territory that contributes to the development of this civilization.To monitor these activities, that state's government establishes regula- tions that society must follow .Various emerging technologies could assist the government in establishing security of the policies and the humans. Then, various ML DL methodswere demonstrated as potential tools for predicting the future [20]. As the new Technolo- gies are rising, each area has been advanced significantly. Hence government used these new emerging technologies for the betterment of the humans.In text categorization, texts are categorized as well as tagged based on their content.There are a wide range of appli- cations for intent identification in NLP[21].Based on it, text analyzers detect and identify content in medical documents various other publications, resulting in predefined labels or classifications for each document.Classifiers determine which categories of textual content should be classified, but they must check the consistency of all the inputs[4].The pro- cess of extracting information from text is called text mining and utilizing it for various tasks.Textual materials can be taken with the use of IE technology. As Internet access has improved, Twitter is easy to access and quickly disseminates news, it has become popular to spread and receive news.However, it becomes very difficult to differentiate between legitimate the fake news on these platforms.False news can be propagated unintentionally by well-meaning individuals without rigorous verification[22].Proliferation of fake news and false data has escalated, potentially causing major issues in society.

As an example malicious people circulate false news about a business which badly effect the company’s reputation.It has been noted that the links between politics and misleading information predate the advent of ”false news”. Although research and its outcome have demonstrated that a link exists between the two. Historically, media organizations, particularly those affiliated with political parties, have emphasized one-sided views and unreliable material [23].

2.2 REVIEW OF EXISTING MODELS

Table 2.1 An overview of all models and their performance

Author	Year	Description	Dataset	Method Used	performance
Verma et al.[24]	2022	An approach to detecting fake news that embeds words over linguistic features	WELFake	CV &TF-IDF	Acc=0.967
Paka et al. [25]	2022	Using Cross-SEAN to Spot Fake News Based on COVID-19	CTF	Attention & Cross-Stitch Mechanis,RoBERT	F1-Score=0.953 , Acc=0.954
Nasir et al. [26]	2021	Detecting Fake News Using a Hybrid RNN-CNN	ISO, FA-KES	CNN,RNN	Mean F1-Score = 0.79, Mean Accuracy = 0.79.
Baruah et al. [27]	2021	BERT's use to identify fake news spreaders	Microblogs	BERT	Acc =0.690
Anoop et al. [28]	2020	Emotional Intelligence Enhances Health Fake News Detection	HWB	AdaBoost	Acc = 0.965
Del Vicario et al.[29]	2020	The Early Warning System for Potential Misinformation Targets: Polarization and Fake News	CLEF 2019	CNN	Acc=0.918
Jawa et al. [30]	2019	Using BERT to detect fake news	FNC	BERT	F1-Score=0.746
Roy et al. [31]	2019	Incorporating deep learning into the detection of fake news	LIAR	BiLSTM & CNN	F1-Score = 0.43
Thota et al. [32]	2018	Use of TF-IDF & DNN for Stance Detection	FNC	NN	Acc=0.942
Granik & Mesyura [33]	2018	Use of NB to Spot Fake News	Postsof Facebook	NB	Acc=0.74
Horne and Adali[34]	2017	Fake news contains a lot of information in the headline and text body and is more	Silverman’s Buzzfeed Political News	SVM	Acc = 0.77

		like comedy than true news.			
Long et al. [35]	2017	Detecting Anti-fake news using attention-based approaches	LIAR	Attention-Mechanism & LSTM	Acc=0.415
Rashkin et al[36]	2016	Examining Language used in Political Fact-checking	Fact Checking	Maxent, LSTM	F1-Score = 0.56
Popat et al. [37]	2016	Evaluation of the Reliability of Textual Claims on the Internet.	Snopes	LR	Acc = 0.72
Rubin et al.[38]	2015	Detection of Fake news using satirical cues	News Articles	SVM and TF-IDF	F1-Score=0.87
AlRubaian et al.[39]	2015	Microblog Credibility Analysis Model with Multiple Stages	Ad-hoc from Twitter	NB	F1-Score= 0.926, Acc = 0.903
Ma et al. [40]	2014	Social context-based rumour detection	Microblogs	SVM	Acc =0.871, F1-Score = 0.87
Mukherjee et al.[41]	2014	Detecting Fake News Using Style-Oriented	Mechanical Turk & Yelp	SVM	F1-Score = 0.857, Acc=0.861,
Kwon et al.[42]	2013	the use of several features, such as linguistic, structural, and temporal, to identify fake news	Microblogs	RF	F1-Score=0.878 Acc=0.897,

2.3 RELATED WORK

Related work can subdivided into following:(1) Detection with Conventional ML (2) Detection based on DL, (3) Detection Based on Pretrained language models, and (4) Comparison studies

2.3.1 Detection with conventional ML

There are a few ways to automate the finding of fake news using traditional ML approaches. It is suggested that linguistic-based criteria, such as total words, characters per word, large word frequency, phrase frequency, i.e., ”n-gram” and bag-of-words techniques, and parts-of-speech (POS) tags, may be helpful for spotting false information[7].The use of part-of-speech tags and content-related n-grams alone did not solve a classification test.With rule categories (lexicalized, non-lexicalized, parent nodes, etc.), deception detection accuracy of 85%-91% was reached [8]. compared to bi-gram TF-IDF, however, as- sessed PCFG characteristics to see if they were successful at spotting bogus news [9]Since there may be a link between the sentiment of the title and the sort of content, sentiment analysis has also been proposed as a method for spotting fraud in various study papers..On Twitter, Cliche describes using n-grams to detect sarcasm using tweets tagged as sarcas- tic. Also, he used sentiment analysis and topic detection in order to improve prediction accuracy by identifying words that tend to be repeated in tweets.

(1) Logistic Regression

A popular statistical technique for binary classification is logistic regression, which uses a logistic function to represent the likelihood of an event occurring. In logistic regression, the projected probability is produced by applying the logistic function to the log chances after a linear model has been trained to predict the log odds of an event. LR is statistical method for analyse data. LR is used to determine the likelihood that a class member will exist. Consequently, LR is the best choice when the target class is categorical [10].By estimating probabilities with the use of a logistic function, ore independent variables. The use of LR yields fruitful results for binary classification. The probability values are forecastusing the sigmoid function. It creates a map from values ranging from 0-1.

(2) Decision Tree

Significant is the supervised learning algorithm DT. To perform a range of tasks, such Random Forest or Gradient Boosting, researchers frequently utilise tree-based ensemble models. The decision tree provides a clear

explanation of how decisions and predictions are made. However, because of its poorlearner, it could do badly on little datasets.

(3) Random Forest

The RF ensemble learning model is based on trees and predicts by integrating the predictions of numerous weak learners. By using various bootstrap samples, RF employs bagging for training decision-trees. These bootstrap samples are created using replacement subsampling of the training dataset, where the sample size is equal to the training dataset size [7]. The biggest difficulty in creating a decision tree in an RF is identifying the properties of the root-node at each level. The procedure called attribute selection.

(4) Multinomial naive bayes

NB is a supervised algorithm that determines the class of a fresh sample using the conditional probability theorem [6]. By using the training dataset, NB calculates conditionalprobability of vector class. Each vector's probability value is computed, and then new vector class is determined based on those conditional-probabilities. For text-based classification issues, NB is frequently utilised REGRESSION

(5) Gradient Boosting Machine(GBM)

GBM is a collection of artificial intelligence classifiers that joins a number of ineffective classifiers to create a potent learning model [9]. Decision trees are frequently utilised for performing gradient boosting. GBM requires a lot of time and money because each tree is developed separately. GBM strengthens the probability approximating correct learning (PAC) learning algorithm. On raw data, PAC produces notable results. GBM effectively handles missing values.

2.3.2 Detection based on DL

It's been demonstrated that CNN can outperform other traditional ML models.A comprehensive analysis of linguistic features was performed[10] .Three layers of hierarchical attention networks (word, phrase, and headline of a news story) were postulated by the author[11].A model was developed to record text, response, and source attributes as a re- sult of user behaviour.Explainability is one of the most crucial characteristics of false news identification, according to a recent article by Shu et al. Authors created an explainable, checkable sentence-comment co-attention subnetwork that takes advantage of the con- tents of user comments in order to detect bogus news.The authors [12] claim that the identification of false news using a multimodal variational autoencoder used a bi-modal variational autoencoder with a binary classifier. According [10], fake news spreads via social networks in patterns, and the relationships between spreaders are studied,therefore proposed a hyrid approach [13], and they used node2vec to extract user characteristics.

(1) Neural Network

Both natural and artificial neural networks act in the same manner. They attempt to mimic how the human brain functions and would behave in certain circumstances if left alone (Fig. 3.1). This was motivated by the brain's neuronal network. In biological terms, a brain recognises signals given from various dendritic groups, and if the signalfrom the dendrites is strong enough to cause the brain to respond, signal flow then moves ahead. Human brain neurons are separate from other neurons in the system and are not interconnected. This phenomenon, made possible by synaptic gaps and the two differentneurons, does not occur until the connections between an axon from one neuron and dendrites linked together from the others are activated.

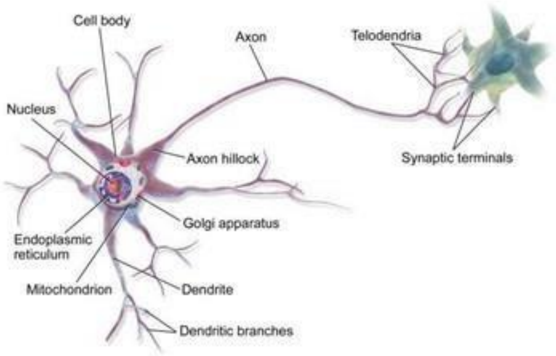

Figure 2.2: Neural Network in human brain[44]

Let’s say that we want an output layer for a job that requires classification across multiple labels (3 labels). The first probable solution that pops into our heads would resemble the neural network topology seen in Figure 2.2.

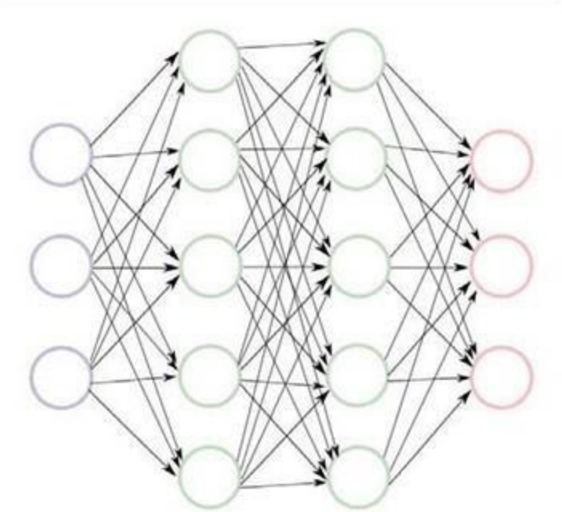

Figure 2.3: Neural networks having 2 Hidden layers 3 output labels[47]

(2) Recurrent Neural Network

RNN is yet another subset of ANN. To read the many inputs in these neural networks, cells or subnetworks are repeatedly used throughout the system. This arrangement is depicted similarly in Figure 2.4.

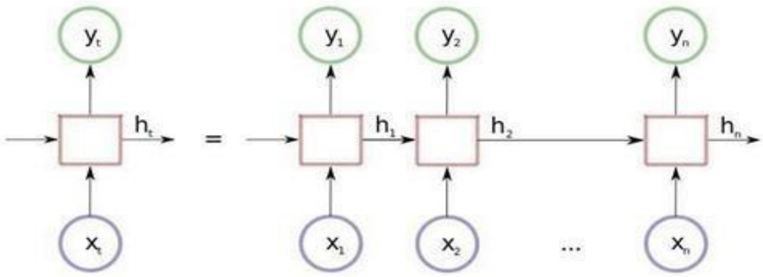

Figure 2.4: RNN Architecture[52]

These systems are especially designed with a sequential data handling architecture to manage inputs that are not given sequentially to neural network systems. To be supplied into a network cell in the order of their individual numbers, the sequential data is broken up into smaller bits of the same or different sizes. Recurrent neural networks have their own constraints in collecting extended streams of data and their relationships, despite being designed to cope with and act appropriately on the sequence of data presented. Asa result, an LSTM network has been developed to address this system’s shortcomings. This LSTM model is a modified RNN that includes additional characteristics like a gating mechanism.

(3) LSTM

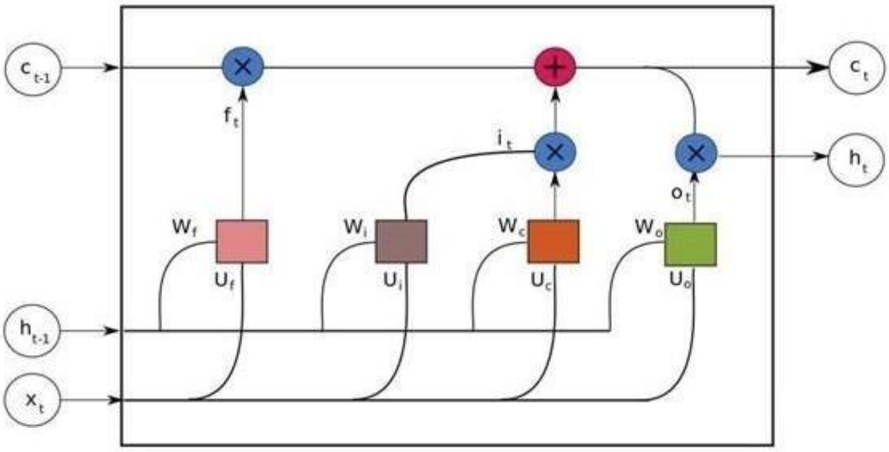

Figure 2.5: Internal structure of LSTM[48]

Due to its capacity to address the limits of RNN, this LSTM model has advanced neural network systems at a time when the current RNN systems had their own restrictions on how they processed data. Three more gates are added to a network cell to give the basic

components of memory. Actually, the input stream is read and stored in memory as a buffer, which is updated each time a network cell does so.

(4)Bi-directional LSTM

Bi-Directional LSTM (BLSTM) is a updated version of LSTM that can process the input sequence in both forward and backward directions. This allows the model to capture not only the contextual information from the past but also from the future, which can improve the accuracy of the predictions.

BLSTMs consist of two separate LSTMs that process the input sequence in opposite directions. The outputs of the two LSTMs are then concatenated to obtain the final output sequence.

The basic LSTM cell contains three gates: input gate, forget gate, and output gate,in addition to the memory cell. These gates regulate the flow of information into andout of the cell. The input gate determines how much of the input information should be stored in the cell, the forget gate decides which information to discard from the cell, andthe output gate controls the amount of information to be outputted from the cell.

The addition of the memory gate in LSTM allows model to either selectively remember or forget informations from previous time steps, thus providing a mechanism for long-term memory.

BLSTM is a popular option for text classification tasks like false news detection since it has generally demonstrated positive results in a number of natural language processing tasks, including sentiment analysis, named entity identification.

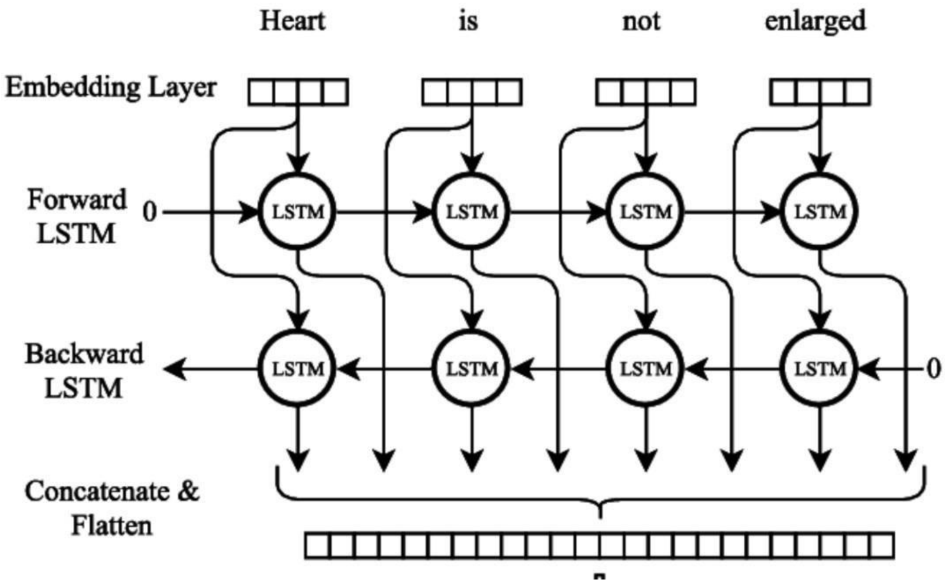

Figure 2.6: Structure of BLSTM[49]

2.3.3 Detection Based on Pretrained language models

There are number of natural language tasks, including text-classification, are being addressed with highly advanced pre-trained language models[14].However, the use of these methods have been limited to few studies.Hence, relationship between the headlines news body was analyzed for detection of false news [15].BERT's deep contextualization improves the F-score by 0.14 in comparison to previous models.Using BERT and RNN together,[12] a hybrid architecture that tackles the problem of fake news .Using a semi-supervised pseudo-label datasets [16]

2.3.4 Comparison studies

In contrast to most existing research, very few studies are conducted to compare many approaches to detect fake news independently on different datasets.According to[16], dataset 'LIAR' was used to comparing the performance of LR,CNN,Bi-LSTM SVM.In addition to analyzing different datasets (e.g., SVM, CNN, LSTM).A number of ML models (in- cluding k-NN, AdaBoost, Decision-Trees, Naive-Bayes, and Bagging) were evaluated to detect the fake news on different available datasets [17].

2.4 Challenges and their Solutions

Fake news has been a concern since the printing press was first in 1439.This section examines methods and approaches for spotting false news as well as the challenges they might provide along with their solution.

(1) Logistic Regression

Challenges:

1. Logistic Regression models may not capture the complex relationships between different parts of the input.

2. Logistic Regression models may not handle imbalanced data well

3. Logistic Regression models may suffer from the problem of multicollinearity, where some of the input features may be highly correlated with each other.

Solutions:

1. Using techniques such as feature engineering or dimensionality reduction can help capture the complex relationships between different parts of the input. Additionally, using hybrid models may help

2. Balancing the data and enhancing model performance may be achieved by employing strategies like oversampling or undersampling. AUC and F1 score are two additional evaluation measures that can help one understand how well a model performs when faced with unbalanced data.

3. Using techniques such as regularization can help mitigate the problem of multicollinearity

(2) Random forest

Challenges:

1. Random Forest models may overfit on the training data.

2. Random Forest models may not capture the complex relationships between different parts of the input.

3. Random Forest models could struggle to deal with data that is unbalanced, such as when one class has much more samples than the other.

Solutions:

1. Preventing overfitting and enhancing the model's generalisation performance may be accomplished by using approaches like cross-validation, early halting, or regularisation.

2. Using techniques such as feature engineering or dimensionality reduction can help capture the complex relationships between different parts of the input. Additionally, using hybrid models may help.

3. Balancing the data and enhancing model performance may be achieved by employing strategies like oversampling or undersampling. AUC and F1 score are two additional evaluation measures that can help one understand how well a model performs when faced with unbalanced data.

(3) Support Vector machine

Challenges:

1. SVMs may not handle imbalanced data well
2. SVMs generally not able to capture the complex relationships of the input data.
3. SVMs may not perform well, when input data is not linearly*separable.*

Solutions:

1. Balancing the data and enhancing model performance may be achieved by employing strategies like oversampling or undersampling. AUC and F1 score are two additional evaluation measures that can help one understand how well a model performs when faced with unbalanced data.

2. Using techniques such as feature engineering or dimensionality reduction can help capture the complex relationships between different parts of the input. Additionally, using hybrid models may help

3. Using techniques such as kernel tricks or non-linear SVMs can help handle non-linearly separable data

(4) Naive bayes

Challenges:

1. Naive Bayes models assume independence between input features,
2. Naive Bayes models may not handle imbalanced data well.
3. Naive Bayes models could be unable to depict the intricate connections between the input's many components.

Solutions:

1. Using techniques such as N-gram models, sentiment analysis, or word embeddings can help capture the complex relationships between different parts of the input. Additionally, using hybrid models may help

2. Balancing the data and enhancing model performance may be achieved by employing strategies like oversampling or undersampling. AUC and F1 score are two additional evaluation measures that can help one understand how well a model performs when faced with unbalanced data.

3. Using techniques such as feature engineering or dimensionality reduction can help capture the complex relationships between different parts of the input.

(5) Decision

TreeChallenges:

1. When the tree depth is not adequately managed, trees may be vulnerable to overfitting.

2. Decision Trees may not capture the complex relationships between different parts of the input.

3. Decision Trees may not be able to handle high-dimensional or noisy data effectively.

Solutions:

1. Overfitting can be avoided by employing strategies like trimming, regularisation, or early halting. Additionally, the model's performance may be enhanced and made more resistant to data noise by applying ensemble approaches like Random Forests or Gradient Boosted Decision Trees.

2. Using techniques such as feature engineering or dimensionality reduction can help capture the complex relationships between different parts of the input. Additionally, using hybrid models may help.

3. Using techniques such as feature engineering or dimensionality reduction can help capture the complex relationships between different parts of the input.

(6) RNN

Challenges:

1. It may be challenging to train the model efficiently since RNNs are susceptible to disappearing or bursting gradients.

2. It's possible that RNNs can't detect long-term relationships in the data.

3. Variable-length input sequences may be difficult for RNNs to handle well.

Solutions:

1. The disappearing or inflating gradient issue can be reduced by employing methods like gradient trimming, batch normalisation, or weight initialization. Additionally, utilising other RNN architectures like LSTMs or Gated Recurrent Units (GRUs) helps increase the model's stability during training.

2. To help identify long-term relationships in the data, approaches like hierarchical RNN (HiRNN), attention mechanisms, or memory networks can be used. The model's performance may be further enhanced by employing strategies like transfer learning or hybrid models, which integrate RNNs with other machine learning models.

3. To successfully deal with variable-length input sequences, methods like padding, masking, or dynamic input length can be used. The model's effectiveness can also be increased by employing strategies like bucketing or sorting input sequences by length.

(7) LSTM

Challenges:

1. LSTMs could overfit easily and struggle to generalise to new data.
2. To train efficiently, LSTMs can need a lot of labelled data.
3. It’s possible that LSTMs will miss the intricate connections between the input’s many components.

Solutions:

1. Overfitting can be avoided by employing strategies like regularisation, early halting, or dropout. Additionally, the model’s performance may be enhanced and made more resistant to data noise by applying ensemble approaches like stacking or bagging.
2. By utilising unlabeled data, training requirements can be decreased by utilising techniques like unsupervised pre-training, transfer learning, or data augmentation. Additionally, the computational cost of the model may be decreased without compromising speed by adopting methods like pruning or quantization.
3. Memory networks, hierarchical LSTM (HiLSTM), and attention mechanisms are a few examples of approaches that may be used to record complicated interactions between various input components. The model’s performance may also be enhanced by employing strategies like transfer learning or hybrid models, which integrate LSTMs with other machine learning models.

(8) Bi-directional

LSTMChallenges:

1. Bi-LSTMs may be computationally costly and may need a lot of labelled data totrain well.
2. Bi-LSTMs could overfit easily and struggle to generalise to new data.
3. It could be challenging to decipher Bi-LSTMs and understand how they makedecisions.

Solutions:

1. The quantity of labelled data needed for training can be decreased by utilising approaches like unsupervised pre-training, transfer learning, or data enhancement. Additionally, the computational cost of the model may be decreased without compromising speed by adopting methods like pruning or quantization.Using techniques such as regularization, early stopping, or dropout can help prevent overfitting.

2. Decision-making process of the model may be better understood by using methods like attention processes, gradient-weighted class activation mapping. Additionally, employing strategies like adversarial training or input perturbation might enhance the model's resilience.

3. Methods like gradient-weighted class activation mapping, saliency maps, and attention processes can provide light on the decision-making process of the model. The model's robustness may also be improved by using techniques like adversarial training or input perturbation.

2.4 DATASETS AVAILABLE

Testing different detection techniques with datasets is a very important part of research. corpora's usefulness depends on a number of factors, namely the following: give false and true instance of the news and the process should be clearly transparent i.e no heterogeneity should be there. Hence, we have analysis a series of datasets.

Table 2.2 Datasets Available for Fake News Detection

Dataset	Year	Description	Instances	Language
Spanish-v2 [24]	2022	Dataset containing data of science, sports, politics & society	572 articles	Spanish
DANFEVER [25]	2022	Dataset containing data of Wikipedia, Den Store Danske	6,407 claims	Danish
FakeCovid[26]	2021	Dataset containing data of COVID-19	12,805 articles	40 languages
FA-KES [27]	2021	Dataset containing data of Syrian War	804 articles	English
Breaking![28]	2020	Dataset containing data of 2016 US election	679 articles	English
Rumor-anomaly [29]	2020	Dataset containing data of Politics, Fraud & Scam, Science, etc.	1,022 threads	Syrian War
FakeNewsAMT [30]	2019	Dataset containing data of Politics, Technology, Education & Business	480 articles	English
FakeNewsNet [31]	2018	Dataset containing data of Celebrity & Politics.	23,921 news	English

Media_Weibo [32]	2017	Dataset containing data of Sina community management	9,528 posts	Chinese
Constrant[33]	2017	Dataset containing data of COVID-19 with politifact & snopes.	10,700 posts	French, Italian, English
Arabic_Stance [34]	2016	Dataset containing data of Arab-related news	4,547 claims	Arabic
PHEME [35]	2015	Dataset containing data of Society, Politics using Crowdsourcing	330 threads	English

Chapter 3

PROPOSED WORK

3.1 DATASET CHARACTERISTICS

The scarcity or a limited access to properly managed multimedia information is a concern. To avoid this, we use standard dataset which helps in testing our model to detect fake news. The FNC-1 (Fake News Challenge) dataset was created to facilitate research in the detection of fake news. The dataset contains a collection of news articles, along with their headlines and the bodies of the articles, and labels indicating whether the articles are fakeor real. The FNC-1 dataset contains approximately 50,000 news articles from a variety of sources, including both mainstream and alternative media. The dataset was developed based on the dataset used for the 2017 FNC challenge, which aimed to develop automated systems that can detect fake news. The FNC-1 dataset consists of three main components: the headline of the news article, the body of the news article, and the label indicating whether the article is fake or real. The dataset is divided into training, development, and test sets. Researchers can use the FNC-1 dataset to develop and evaluate algorithms andmodels for the detection of fake news. The dataset has been used in various studies andcompetitions related to fake news detection, and it continues to be a valuable resource for researchers in this field.

(1) Creation: A group of academics from several universities, including the School of Computer Science college Victoria, the University of Illinois at Chicago, and the Qatar Computing Research Institute, produced the FNC-1 dataset. By compiling news stories from multiple sources and manually classifying them as false or authentic, the dataset was built.

(2) Size: The FNC-1 dataset, which is broken up into training, development, and test sets, has about 50,000 news items. 35,000 articles make up the training set, 2,500 go into the development set, and 2,500 go into the test set.

(3) Labeling : The dataset classifies each news item as either true or fraudulent. The labels were given based on the fact-checking conducted by expert journalists and professional fact-checkers from variousmedia outlets.

(4) Challenges: The FNC-1 dataset has been utilized in a number of tasks and contests involving the identification of false news. For instance, the FNC-1 wasemployed in the 2017 false News Challenge, which sought to create algorithms that could more reliably detect false news stories than people.

(5) Data format : The tab-separated value (TSV) format for the FNC-1 dataset is available. The TSV file's columns include details like the article ID, the article headline, the article content, and the article label. Each row represents a news story.

(6) Limitations: The FNC-1 dataset has a number of limitations that researchers should be aware of. For instance, the dataset only contains English-language news items, and a small panel of fact-checkers and journalism experts choose what labels to assign to the articles. Additionally, only a small percentage of specialized types of false news, such propaganda and satire, are included in the dataset.

”We began our analysis by examining the FNC-1 dataset, which consists of 27,585 news articles labeled as either ’fake’ or ’real’. We observed that the dataset contained some quality issues, such as duplicate articles and missing labels. To address these issues, we removed duplicates and articles with missing labels, resulting in a total of 26,721 articles. Next, we applied several filtering criteria to the dataset to remove non-news articles and low-quality articles. Specifically, we removed any articles that were not in English, as we did not have the resources to translate them. We also removed any articles with a credibility score of less than 0.5, which was determined by the authors of the dataset based on a manual assessment of each article's credibility. After filtering the dataset, we performed additional preprocessing steps to clean and normalize the text data. We removed stop words, punctuations, and special characters, and performed stemming to reduce inflectional forms of words to their root form. These steps helped to reduce the noise in the text data and improve the accuracy of our analysis. We encountered some challenges during the filtering and preprocessing stage, particularly with respect to class imbalance in the dataset. To address this issue, we employed oversampling techniques to create synthetic samples of the minority class (fake news) to balance the dataset. After applying all of our filtering and preprocessing techniques, we arrived at a final filtereddataset consisting of 23,826 articles, with 11,913 fake news articles and 11,913 real news articles. This dataset was used for our subsequent analysis and modeling. Overall, the dataset filtering and preprocessing steps were crucial for improving the quality of our data and preparing it for analysis. The criteria and methods we used were based on best practices in the field of fake news detection, and were chosen to ensure the accuracy and reliability of our findings.”

Our analysis of the FNC-1 dataset has revealed important information on the features of false news and how it spreads across various platforms. To investigate the dataset and find significant patterns and trends, we employed a variety of data exploration approaches, including data visualization, statistical analysis, and natural language processing. According to our study, the FNC-1 dataset contains approximately 27,000 news stories with a fairly even mix of bogus and legitimate news. Additionally, we discovered that the collection includes news stories written in a variety of languages, with English being the most common. We found that fake news items frequently center on political subjects, and the bulk of them are posted on social media websites like Twitter and Facebook. During the data exploration process, we ran across various restrictions, such as missing data and data quality concerns, which might have had an impact on the accuracy of our research.

	id	title	text	la
0	0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	
1	1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	
2	2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	
3	3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	
4	4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	
...	...	...	...	
72027	4490	State Department says it can't find emails fro...	The State Department told the Republican Natio...	
72028	8062	The 'P' in PBS Should Stand for 'Plutocratic' ...	The 'P' in PBS Should Stand for 'Plutocratic' ...	
72029	8622	Anti-Trump Protesters Are Tools of the Oligarc...	Anti-Trump Protesters Are Tools of the Oligar...	
72030	4021	In Ethiopia, Obama seeks progress on peace, se...	ADDIS ABABA, Ethiopia —President Obama convene...	
72031	4330	Jeb Bush Is Suddenly Attacking Trump. Here's W...	Jeb Bush Is Suddenly Attacking Trump. Here's W...	

72032 rows × 4 columns

Figure 3.1: FNC Dataset

However, we were able to get around these issues and discover some insightful information about the dataset by using a variety of data exploration approaches. Overall, the FNC-1 dataset investigation we conducted has given us a strong platform on which to build our research and modeling. Researchers and decision-makers seeking to stop the spread of fake news will find the knowledge we gained from our investigation to be useful. Moving forward, we advise that research concentrate on gathering more varied and thorough datasets to enhance the precision of false news detection programs.

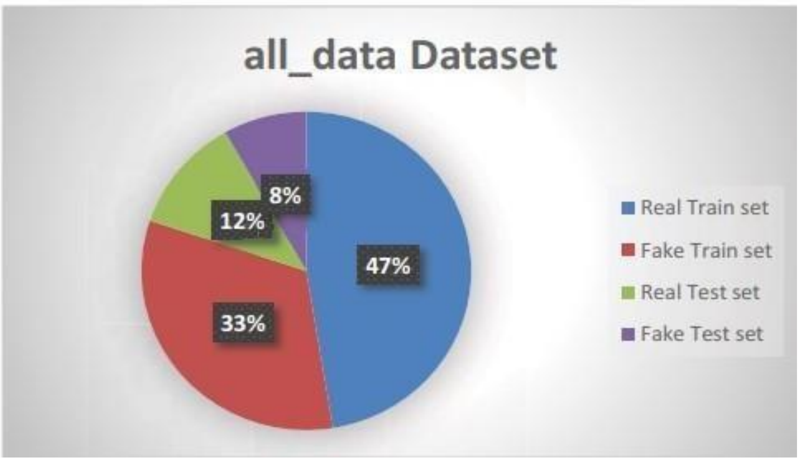

Figure 3.2: Test Train Split after Pre-processing

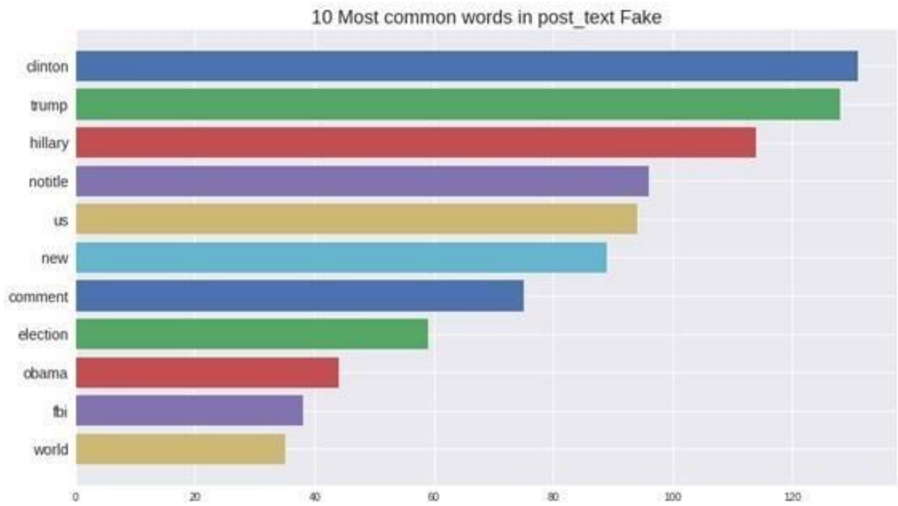

Figure 3.3: 10 most common words in postText with fake label

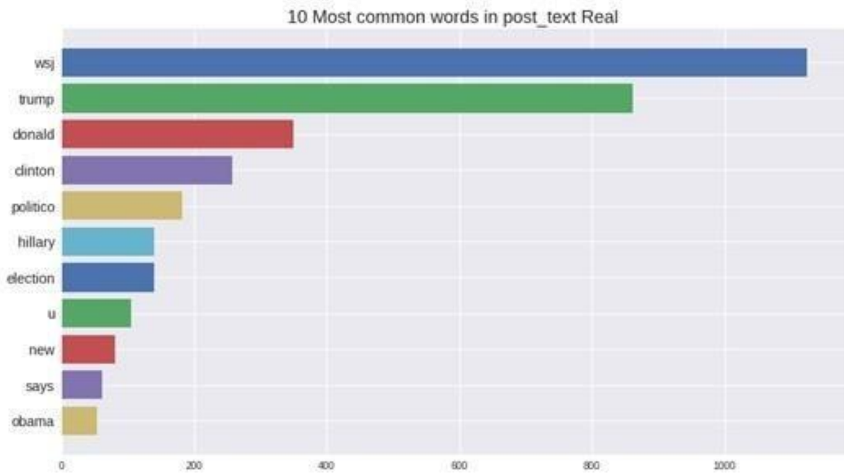

Figure 3.4: 10 most common words in postText with real labelx

Figure 3.5: Word cloud for fake and real post

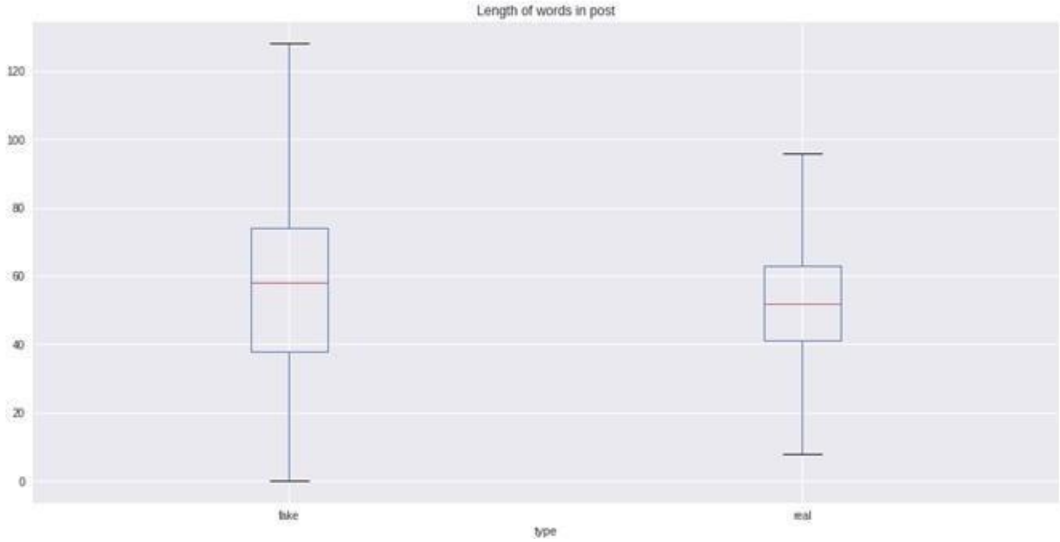

Figure 3.6: Length of words in post for fake and real

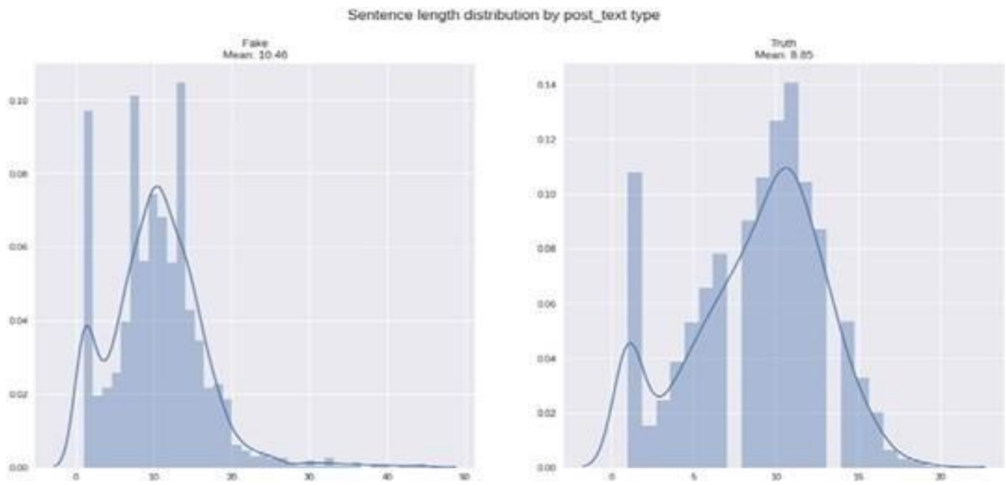

Figure 3.7: Sentence length for fake and real post

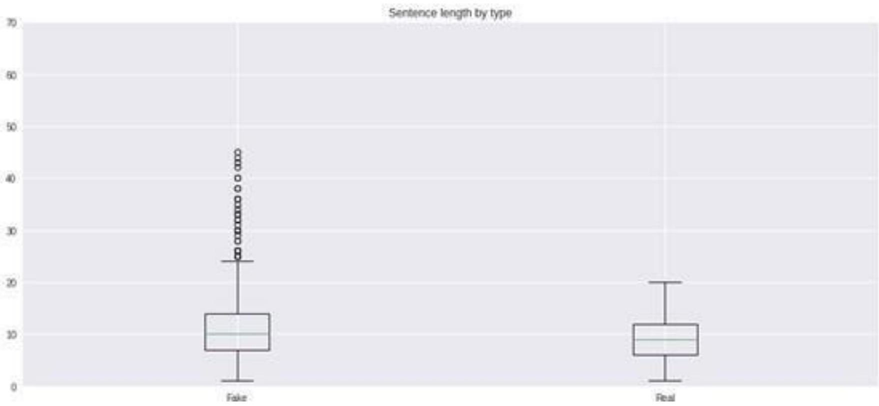

Figure 3.8: Box plot for Sentence length for fake and Real post

3.2 Procedural Flowchart

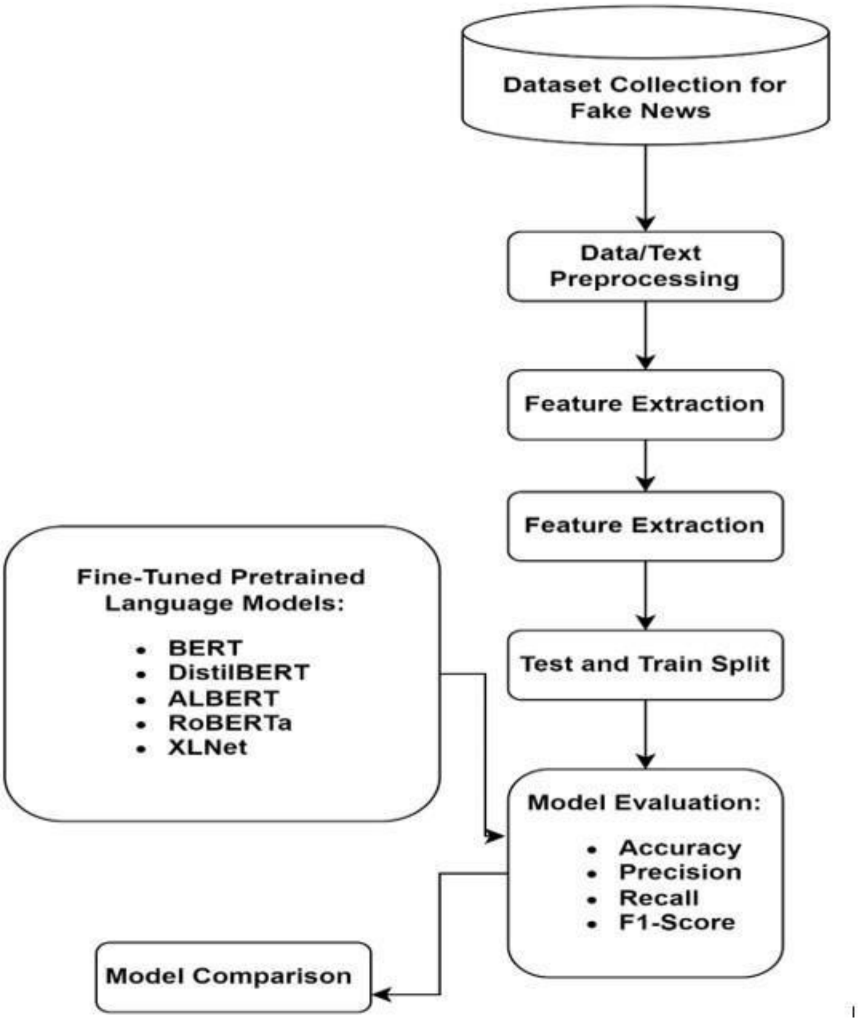

Figure 3.9: Procedural Flowchart

The FNC dataset, which produced for the competition with the goal of developing AI systems that can automatically recognise false news stories, basically it is a collection of news articles with their accompanying labels.The training set and test set of news articles, which together total 24,833 news articles, are used to assess how well machine-learning models perform.The FNC dataset has received a lot of attention. It has been used to assess models built using several methodologies, such as neural networks, ensemble models, and linguistic characteristics. The effectiveness of various models on task of false news identification has also been implemented compared using the FNC dataset.

3.3 PRE-PROCESSING

The method of translating data to something that a machine might comprehend is called preprocessing. Some of the main ways of preprocessing is filtering out unnecessary information. Data preprocessing is very important part of any project as we would not want unnecessary information and other unimportant features occupying space and affecting the performance of ML DL models. Then data is supposed to go through various modifications like the elimination of the stop words, tokenization, discarding of punctuations, and sentence segmentation beforehand. Then it can be represented using the n-gram model. A function was defined to remove the punctuations as well as lowercasing the letters in the Document.

3.4 FEATURE EXTRACTION

The Bag of Words approach is used to create a group of vectors that reflect the occurrences of words in the document to start the feature extraction process. These vectors keep track of how frequently each word appears in the text.When the original raw data is drastically different and cannot be used for machine learning models, feature extraction is frequently done. The raw data is then given the desired shape. The technique of extracting additional, more accurate features from raw data that capture the bulk ofits important information is known as feature extraction. On real-world ML challenges, therefore we must extract the relevant characteristics from the raw data. The TF-IDF vectorizer is one of the important the famous feature extraction techniques we employ.

3.5 TF-IDF VECTORIZER

TF-IDF weight are frequently used in text mining and retrieval of important information. Using versions of the tf-idf weighting algorithm, search engines routinely assess and rank the relevance of documents given a query. Using this weight, we can calculate word's importance in a collection. While a word's frequency in the corpus cancels out the importance bump brought on by its frequency in the text, it still has an impact on the data set's measure of word importance. It is reported to have given better performancefor the tasks related to textual data. The Tf-Idf Vectorizer has an advantage over other Vectorizers in that it considers both the frequency of a word in a particular document and its frequency in general. This gives the model more data to work with and makes it simpler to distinguish between words that contribute more to the classification task than words that contribute less to the task. We have used the TfIdfVectorizer module of the SkLearn module for implementation purposes.

3.6 BUILDING THE MODELS

Transfer learning and optimization techniques are thought to perform better than popular ML models. It is crucial to fully know the features of fake news and the way how it is spreading in the world. The current study makes a contribution in this field by offering a model built using cutting-edge techniques that illustrates the benefit of using pretrained language models to identify false news. The pretrained Language model has to be modified for better outcomes.

3.6.1 Fine-tuning the BERT Model

To illustrate our conclusions, we employed a variety of visualizations, including word clouds and frequency distributions. For instance, the word clouds for legitimate news stories showed a greater frequency of neutral terms like ”government,” ”policy,” and ”report,” but the word clouds for false news articles showed a higher frequency of emotionally charged words like ”scandal,” ”fraud,” and ”crisis.” Overall, our linguistic study identified a number of key language characteristics that distinguish between phony and authentic news stories. These results offer important new insights into the linguistic traits of false news and are consistent with other studies in the area of fake news identification. Our findings have significant ramifications for the creation of algorithms for false news identification that are more precise and efficient.

To illustrate our conclusions, we employed a variety of visualizations, including word clouds and frequency distributions. For instance, the word clouds for legitimate news stories showed a greater frequency of neutral terms like ”government,” ”policy,” and ”report,” but the word clouds for false news articles showed a higher frequency of emotionally charged words like ”scandal,” ”fraud,” and ”crisis.” Overall, our linguistic study identified a number of key language characteristics that distinguish between phony and authentic news stories. These results offer important new insights into the linguistic traits of false news and are consistent with other studies in the area of fake news identification. Our findings have significant ramifications for the creation of algorithms for false news identification that are more precise and efficient.

Table 3.1 Parameters and their Values for the training phase of the BERT model

Parameter Name	Parameter Value
Optimizer	Adam
Learning Rate	2.00E-05
Loss	bce
Metrics	Accuracy
Epochs	5
Batch Size	16
Validation Split	10%

3.6.2 Fine-tuning the DistilBERT Model

For our challenge on fake news identification, we employed the DistilBERT model that has been previously trained. Using a learning rate of 2e-5, a batch size of 16, and a maximum sequence length of 512 on our dataset, we enhanced the DistilBERT model.

We developed the model with the AdamW optimizer and trained it with a weight decay of 0.01 across three epochs.2019 saw the launch of DistilBERT, a smaller and quicker variant of BERT. While being more computationally effective, it is meant to function similarly to BERT. With fewer parameters, this model is trained to mimic nature of larger model. Due to faster inference times and less demand- ing resource requirements, the smaller model is able to perform similarly to the larger model.On several benchmark datasets, DistilBERT has been demonstrated to perform aswell as BERT while being 40

Table 3.2 Parameters and their Values for the training phase of the DistilBERT model

Parameter Name	Parameter Value
Optimizer	Adam
Learning Rate	2.00E-05
Loss	bce
Metrics	Accuracy
Epochs	5
Batch Size	16
Validation Split	10%

3.6.3 Fine-tuning the ALBERT Model

For our fake news identification job, we started with the ALBERT model that had already been trained. Using a lr of 2e-5, a max sequence length of 512 on our dataset, batch size of 16 we fine-tuned the ALBERT model. We used the Adam optimizer to train the model over four iterations, with a weight decay of 0.01 We included an attention mechanism that concentrates on the instructive portions of the input text in terms to enhance the overall performance. To calculate the attention scores for each input token, we employed the self attention-mechanism put forward in transformers model. Weighted sum of embeddings, which represents the informative portions of the attention scores, was then obtained by multiplying the attention scores by the associated token embeddings.

Table 3.3 Parameters and their Values for the training phase of the ALBERT model

Parameter Name	Parameter Value
Optimizer	Adam
Learning Rate	2.00E-05
Loss	bce
Metrics	Accuracy
Epochs	5
Batch Size	16
Validation Split	10%

3.6.4 Fine-tuning the RoBERTa Model

Our attempt to identify bogus news was based on the pre-trained RoBERTa model. We improved the RoBERTa model on our dataset with a maximum sequence length of 512, a batch size of 16, and a learning rate of 2e-5. With a weight decay of 0.01, we trained the model across three epochs using the AdamW optimizer. language features can help fake news identification work better when combined. Model excellent accuracy and F1- score demonstrate that it has the precision and memory necessary to discern between authentic and false news. The effectiveness of a model was further enhanced by the addition of linguistic variables, indicating that linguistic analysis may offer extra details for spotting bogus news.Quality of the linguistic features and the model's applicability to other domains, however, may have an impact on how well the model performs

Table 3.4 Parameters and their Values for the training phase of the RoBERTa model

Parameter Name	Parameter Value
Optimizer	Adam
Learning Rate	2.00E-05
Loss	bce
Metrics	Accuracy
Epochs	5
Batch Size	16
Validation Split	10%

3.6.5 Fine-tuning the XLNet Model

We started with the XLNet model that had already been trained for the false news detection test. Using 16 batches, a 2e-5 learning rate, and a 512-byte maximum for sequence length, we fine-tuned the XLNet model for our dataset. With a weight decay of 0.01, we trained the model using the Adam optimizer over five iterations.BERT and XLNet both use the Transformers architecture. Contrary to BERT's masked language modeling aim, XLNet uses a permutation language modelling objective. Instead than only utilizing the disguised tokens for this goal, the model is trained to forecast the probability distribution over all feasible permutations of the input tokens. By using this method, the model is able to recognise dependencies between all of the input tokens rather than just the masked tokens. As a result, XLNet is better able to handle tasks like document classification and question answering that call for reasoning over longer sequences.

Table 3.5 Parameters and their Values for the training phase of the XLNet model

Parameter Name	Parameter Value
Optimizer	Adam
Learning Rate	2.00E-05
Loss	bce
Metrics	Accuracy
Epochs	5
Batch Size	16
Validation Split	10%

Chapter 4

RESULTS & DISCUSSIONS

4.1 Results for Fine-Tuned Language Models

Our Model aims to detect fake news using Fine-Tuned Pretrained language Models. The outcomes of our research show how well-tuned pretrained language models can identify bogus news. We tested five different finetuned language models including BERT, Al-BERT, DistilBERT, XLNet RoBERTa. The assessment results showed that the fake news detection capabilities of BERT, DistilBERT, AlBERT, RoBERTa, and XLnet wereall extremely effective, with precision scores ranging from 0.92 to 0.95, recall scores rang- ing from 0.92 to 0.97, and F1-scores ranging from 0.93 to 0.95. RoBERTa XLnet had the highest recall rating of 0.97, while BERT received the highest precision rating of
0.94. DistilBERT achieved the highest F1-score, which was 0.95 and matched BERT's F1-score. BERT received the highest accuracy score (0.939), closely followed by RoBERTa and XLnet, both of which had ratings of 0.931 and 0.930. Therefore, we can say choicing which language model may depends on the particular needs of the application, such as realtime detection or maximizing the detection of all fake news.Overall, our work shows that refined pretrained language models can accurately and effectively identify actual false news, and it offers us a potential method for creating automated systems. A healthy and educated society may be maintained by using these fine-tuned pretrained algorithms to dramatically increase the efficacy of false news identification. So, we achieved the highest accuracy of 93.9% on the FNC Dataset after Fine-tuning the BERT pretrained language Model followed by and XLnet 93.6% RoBERTa 93.1%

Table 4.1 Using the official competition metric classes ,we assessed our Fine-Tuned Language models on the FNC Dataset

Model	Accuracy	Precision	Recall	F1-Score
BERT	0.939	0.94	0.96	0.96
DistilBERT	0.936	0.95	0.95	0.95
AlBERT	0.929	0.93	0.96	0.94
RoBERTa	0.931	0.92	0.97	0.95
XLnet	0.930	0.93	0.97	0.95

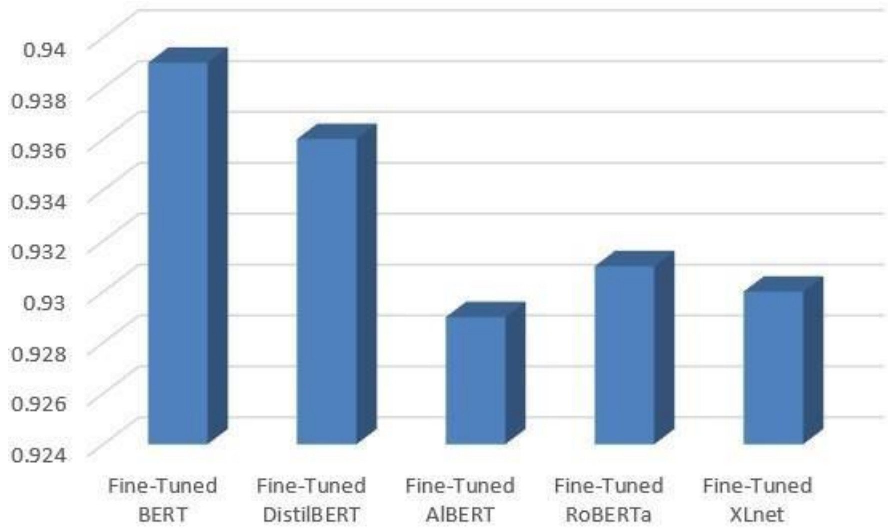

Figure 4.1: Accuracy of the Fine-Tuned Language Models

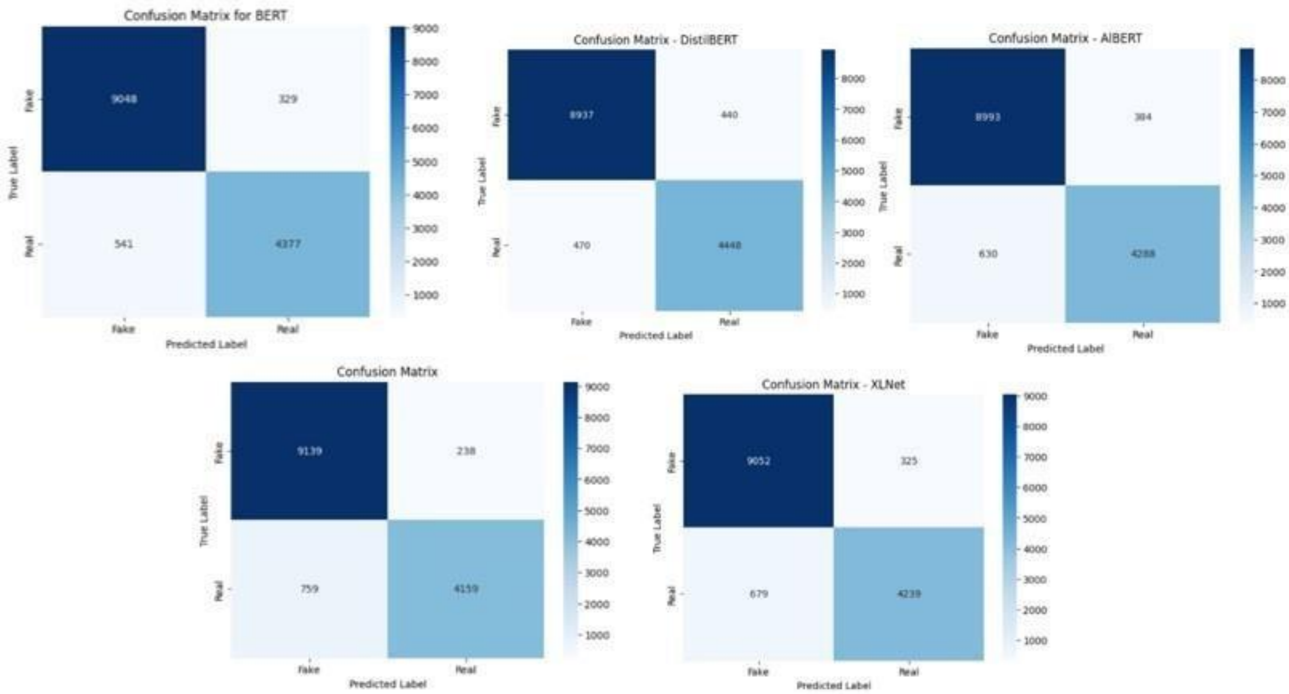

Figure 4.2: Confusion Matrix for Fine-tuned Pretrained Language Models on the FNC dataset

4.2 Results for Machine Learning Models

Additionally, we used machine learning techniques on the FNC dataset. We compare the performance of six machine learning models, including Logistic Regression, Decision Tree, Gradient Boosting, Random Forest, Multinomial Naive Bayes, and Passive Aggressive, using a FNC dataset in the section on machine learning algorithms. The findings show that Logistic Regression and Passive Aggressive models both had accuracy ratings of 91%, which were the highest. The accuracy of the Decision Tree model was 82.5%, while that of the Gradient Boosting and Multinomial Naive Bayes models was 83.8%

and 88.8%, respectively. With a score of 79.8%, the Random Forest model had the least accuracy.The models with the greatest precision scores, both with 86%, are the Logistic Regression and Random Forest models, according to the precision ratings. High accuracy scores (89% and 91%, respectively) were also achieved by the Multinomial Naive Bayes and Passive Aggressive models. The accuracy ratings for the Gradient Boosting and Decision Tree models were 82% and 85%, respectively.The Logistic Regression model obtained a recall score of 90%, while the Passive Aggressive model had the best score of 91%. The recall score for the Multinomial Naive Bayes model was 86% while the recall scores for the Decision Tree, Gradient Boosting, and Random Forest models were 78%, 79%, and 71%, respectively. The F1-score, which combines accuracy and recall, offers a final way to assess the performance of each model. The Logistic Regression model came in second with an F1-score of 90 %, closely trailed by the Passive Aggressive model with an F1-scoreof 91%. The decision tree, gradient boosting, and random forest models got F1-scores of79%, 81%, and 73%, respectively, while the multinomial naive bayes model had an F1-score of 87%. Overall, these assessment measures shed light on each model's advantagesand disadvantages and can assist future model selection and improvement efforts.

Table 4.2 Using the official competition metric classes ,we assessed Machine Learning Models on the FNC Dataset

Model	Accuracy	Precision	Recall	F1-Score
LogisticRegression	0.914	0.91	0.90	0.90
DecisionTree	0.825	0.82	0.78	0.79
GradientBoosting	0.838	0.85	0.79	0.81
RandomForest	0.798	0.86	0.71	0.73
MultinomialNB	0.888	0.89	0.86	0.87
PassiveAggresive	0.912	0.91	0.91	0.91

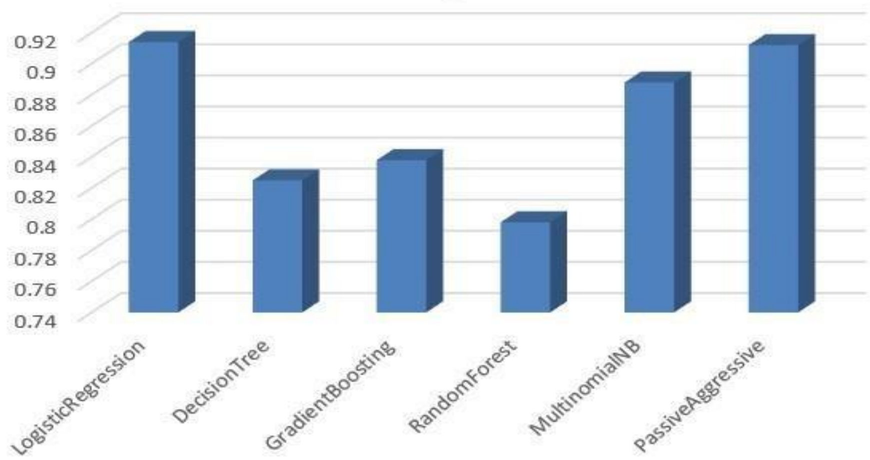

Figure 4.3: Accuracy of the Machine Learning Models

F

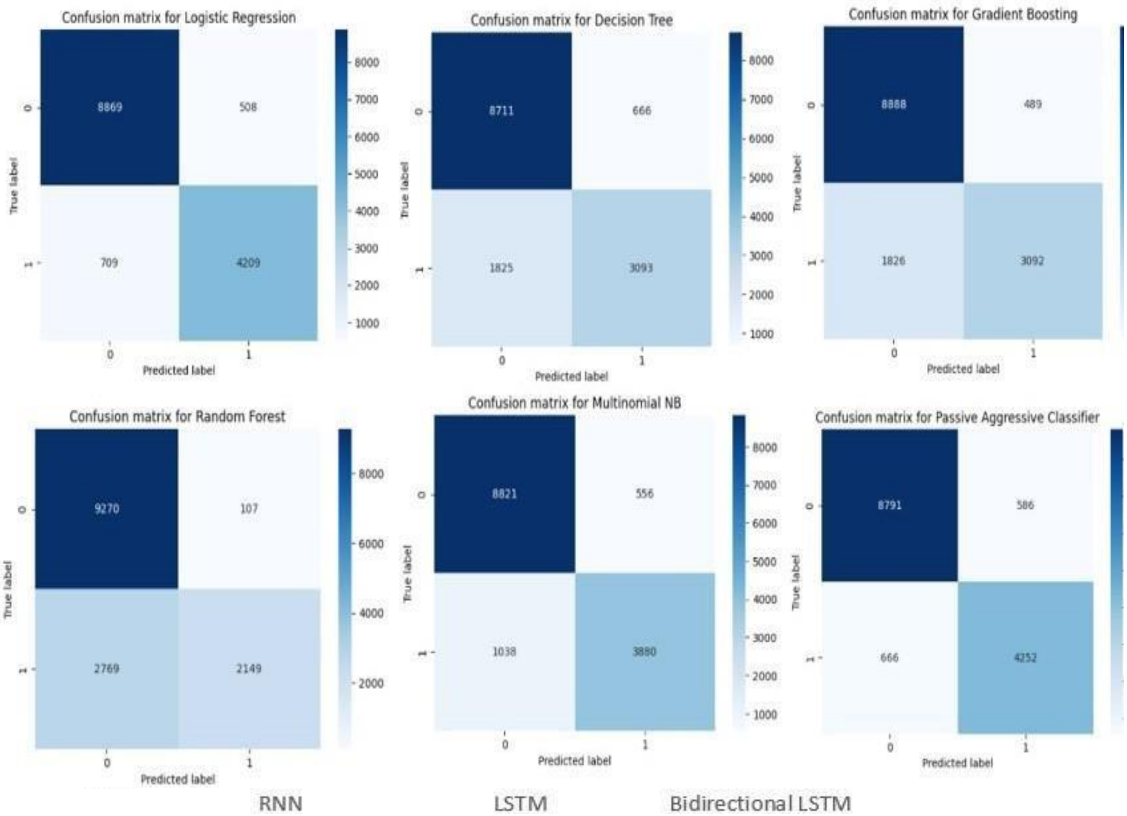

Figure 4.4: Accuracy of the Machine Learning Models

4.3 Results for Deep Learning Models

On the FNC dataset, we also used Deep Learning methods. On a FNC dataset,we analyze the performance of three deep learning models, including RNN, LSTM, and Bidirectional LSTM. According to the findings, the LSTM model had an accuracy scoreof 90.8%, while the Bidirectional LSTM model had the best accuracy score of 91.4%. The RNN model’s accuracy score of 79.1% was the lowest.The RNN model received a precision
score of 79%, while the LSTM and Bidirectional LSTM models both obtained the greatest precision values of 91%.The RNN model had a recall score of 79%, while the Bidirectional LSTM and LSTM models both obtained the best results, both having a recall score of 90%.Finally, both the LSTM and Bidirectional LSTM models scored 91% on the F1-score, which takes into account both precision and recall. The F1-score for the RNN model was 80%. In conclusion, the Bidirectional LSTM model outperformed the LSTM and Bidirectional LSTM models in terms of accuracy, precision, recall, and F1-scores. All assessment measures yielded lower ratings for the RNN model.

These findings offer insightful guidance for choosing the appropriate model for this specific dataset. Overall, these assessment measures shed light on each model's advantages and disadvantages and can assist direct future model selection and improvement efforts.

Table 4.3 Using the official competition metric classes ,we assessed Deep Learning Models on the FNC Dataset

Model	Accuracy	Precision	Recall	F1-Score
RNN	0.791	0.79	0.79	0.80
LSTM	0.908	0.91	0.91	0.91
Bidirectional LSTM	0.914	0.91	0.90	0.90

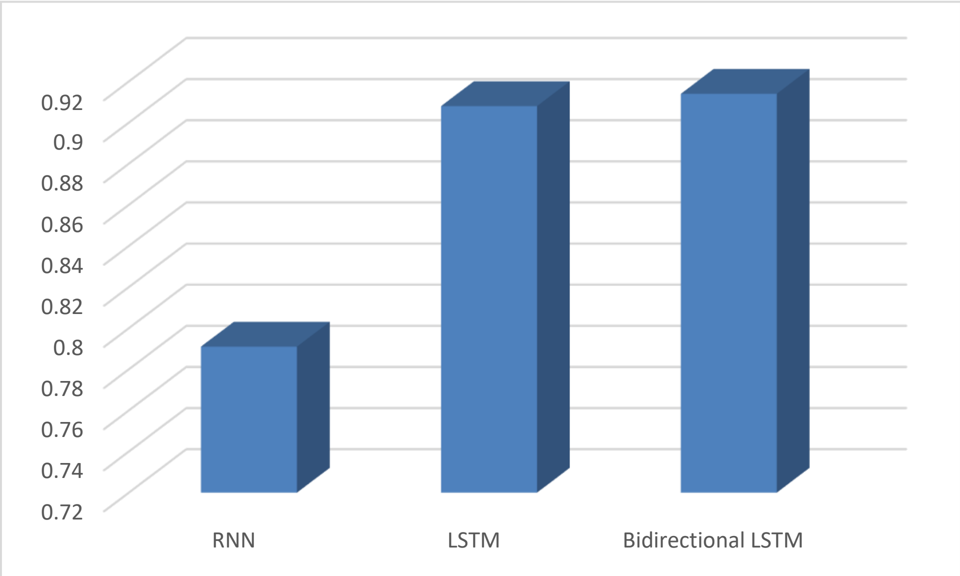

Figure 4.5: Accuracy of the Deep Learning Models

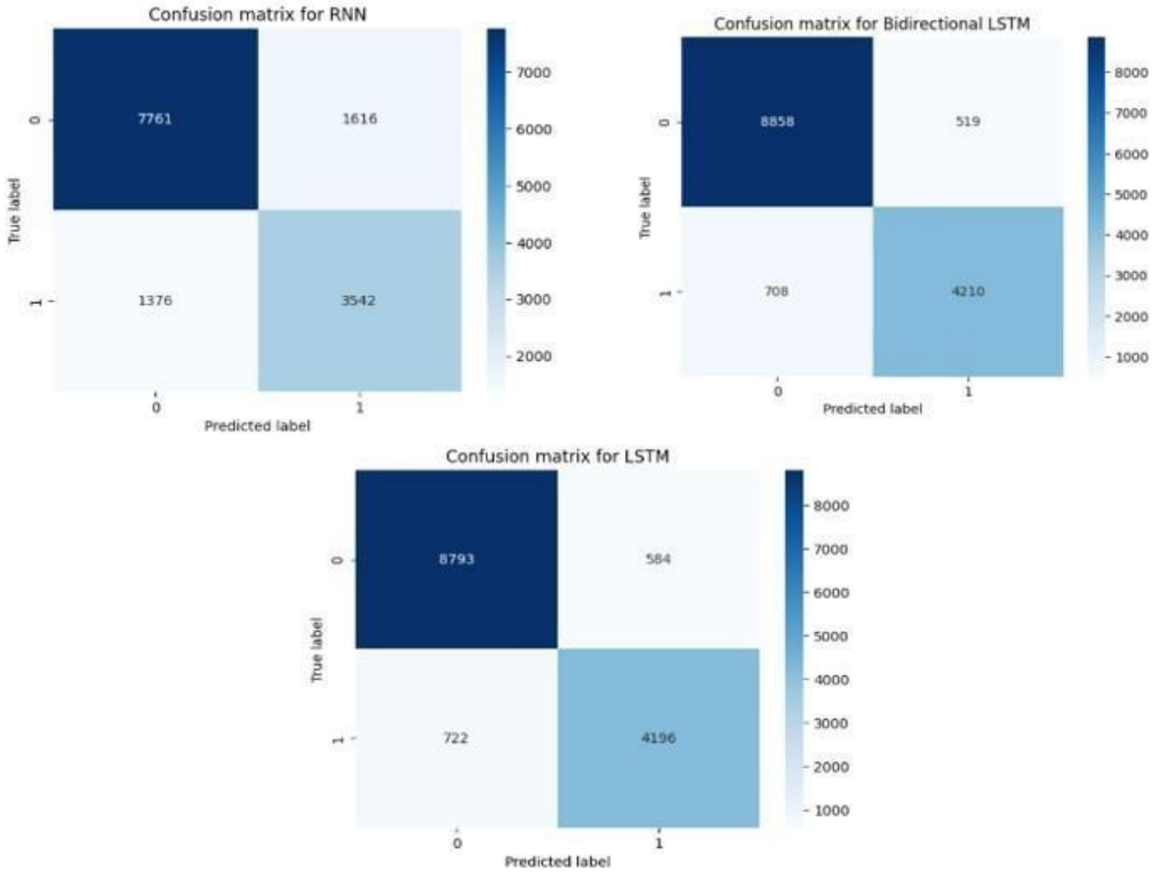

Figure 4.6: Confusion Matrix for Deep Learning Models on the FNC dataset

Chapter 5

CONCLUSION & FUTURE SCOPE

5.1 CONCLUSION

The use of refined pretrained language models can be a workable tactic for extremely accurate false news identification. DistilBERT, AlBERT, RoBERTa, and XLNet models performed well, according to the study's findings, with F1-scores between 0.93 and 0.95 and accuracy between 0.929 and 0.936. The accuracy of the BERT model was 0.939, and its F1-score was 0.95, both excellent results. These results show that pretrained models may be effectively optimized to construct automated systems to prevent the spread of erroneous information. After some fine-tuning, the models must be trained, and the performance of the models may be improved by increasing the amount of the training dataset.According to the findings, DistilBERT, a smaller, quicker model, might be a viable substitute for BERT since it offers equivalent performance while utilizing less computing power.According to our research, pretrained language models that have been optimized can aid in halting the spread of false information on social media platforms. But itis essential to keep in mind that the quality of the labelled data used for the models' training has a significant impact on how well they function.Therefore, future research should focus on developing labelled datasets that are more accurate and thorough for false news identification.

Therefore, we can say choicing which language model may depends on the particular needs of the application, such as realtime detection or maximizing the detection of all fake news.Overall, our work shows that refined pretrained language models can accurately and effectively identify actual false news, and it offers us a potential method for creating automated systems. A healthy and educated society may be maintained by using these fine-tuned pretrained algorithms to dramatically increase the efficacy of false news identification. So, we achieved highest accuracy of 93.9% after using Fine-tuned BERT pretrained language Model followed by XLNet 93.6% RoBERTa 93.1%.

Overall, this study shows the possibility for detecting false news using finetuningof the pretrained language models and offers a direction for further study in this area Researchers and professionals who are interested in developing a deeper understanding devise measures to combat its propagation generation are invited to take part. Data scientists, artificial intelligence experts, journalists, information scientists, psychologists, and social scientists will benefit from this thesis. As a result, this thesis attempted to concentrate on the features of fake news and the methods used to create it. Researchers are trying to understand the principles of those origins of false news because of the difficulties in recognizing it. Future researchers will benefit from the comparative examination of open problems in this area.

5.2 FUTURE SCOPE

Future research can be conducted in a variety of connected topics.For instance, used to identify ”clickbaits,” which are websites that entice viewers to click by utilizing catchy and elegant content.Based on the difference between news title and content, it is beneficial to identify spam in social networks that is disseminated by people, groups of users, or social bots.

The unresolved issues in research are:

(1) Datasets: Fake news in various forms may be found in most public repositories. Focusing the study objectives that encompass all news data kinds is likewise an open challenge.

(2) Source validation: No researchers have looked into the origin of the false information.

(3) Credibility evaluation: The dissemination and knowledge-based elements of a chain of fake news stories written by the same (or different) writers are not examined.

In order to increase our accuracy, we will eventually incorporate more explicit features based on textual material or user profile data. We will test our approach on more publicly accessible fake news datasets as well.

REFERENCES

[1] R. D. Abdiansyah, D. Mutiara, S. P. Sumedha, and N. Hanafiah, "Effective methods for fake news detection: a systematic literature review," in 2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI), vol. 1, pp. 278–283, Jakarta, Indonesia, 2021.

[2] O. Abu Arqoub, A. Abdulateef Elega, B. Efe Özad, H. Dwikat, and F. Adedamola Oloyede, "Mapping the scholarship of fake news research: a systematic review," Journalism Practice, vol. 16, no. 1, pp. 56–86, 2022

[3] M. O. Ahmad, J. Markkula, and M. Oivo, "Factors affecting egovernment adoption in Pakistan: a citizen's perspective," TransformingGovernment: People, Process and Policy, vol. 7, no. 2, pp. 225–239, 2013.

[4] M. Albahar and J. Almalki, "Deepfakes: threats and countermeasures systematic review," Journal of Theoretical and Applied Information Technology, vol. 97, no. 22, pp. 3242–3250, 2019

[5] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, and J. Vilares, "Sentiment analysis for fake news detection," Electron, vol. 10, no. 11, 2021.

[6] R. Azad, B. Mohammed, R. Mahmud, L. Zrar, and S. Sdiq, "Fake news detection in low-resourced languages 'Kurdish language' using machine learning algorithms," Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no.6, pp. 4219–4225, 2021.

[7] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22–36.

[8] Feng, S., Banerjee, R., & Choi, Y. (2012). Syntactic stylometry for deception detection. In Proceedings of the 50th annual meeting of the association for computational linguistics: short papers- volume 2 (pp. 171–175).

[9] Gilda, S. (2017). Evaluating machine learning algorithms for fake news detection. In Research and development (scored), 2017 IEEE 15th student conference on (pp. 110–115).

[10] Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., & Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In Proceedings of the 2017 conference on empiricalmethods in natural language processing

[11] Singhania, S., Fernandez, N., & Rao, S. (2017). 3HAN: A deep neural network for fake news detection. In International conference on neural information processing (pp. 572–581).

[12] Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). Mvae: Multimodal variational autoencoder for fake news detection.In The world wide web conference (pp. 2915–2921).

[13] Hamdi, T., Slimi, H., Bounhas, I., & Slimani, Y. (2020). A hybrid approach for fake news detection in Twitter based on user features and graph embedding. In International conference on distributed computing and internet technology (pp. 266–280).

[14] Tenney, I., Das, D., & Pavlick, E. (2019). BERT Rediscovers the classical NLP pipeline.

[15] Jwa, H., Oh, D., Park, K., Kang, J. M., & Lim, H. (2019). exBAKE: Automatic fake news detection model based on Bidirectional Encoder Representations from Transformers (BERT).Applied Sciences, 9(19), 4062.

[16] Shu, K., Cui, L., Wang, S., Lee, D., & Liu, H. (2019). defend: Explainable fake news detection. In Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining (pp. 395–405)

[17] Gravanis, G., Vakali, A., Diamantaras, K., & Karadais, P. (2019). Behind the cues: A benchmarking study forfake newsdetection. Expert Systems with Applications, 128, 201–213.

[18] Robinson, R. J., Keltner, D., Ward, A., & Ross, L. (1995). Actual versus assumed differences in construal: "naive realism" in intergroup perception and conflict. Journal of Personality and Social Psychology, 68(3), 404

[19] Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. Review of generalpsychology, 2(2)

[20] Pronin, E., Kruger, J., Savtisky, K., & Ross, L. (2001). You don't know me, but i know you: The illusion of asymmetric insight.Journal of Personality and Social Psychology, 81(4), 639.

[21] Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A.(2016). The rise of social bots. Communicationsof the ACM, 59(7).

[22] Quattrociocchi, W., Scala, A., & Sunstein, C. R. (2016). Echochambers on facebook. Available at SSRN2795110

[23] Paul, C., & Matthews, M. (2016). The russian "firehose of falsehood"propaganda model (pp. 2–7). RandCorporation. Polignano

[24] Juan-Pablo Posadas-Durán, Helena Gómez-Adorno, Grigori Sidorov, and Jesús Jaime Moreno Escobar. 2019. Detection of fake news in a new corpus for the Spanish language. Journal of Intelligent & Fuzzy Systems 36, 5 (2019), 4869–4876.

[25] Jeppe Nørregaard and Leon Derczynski. 2021. DanFEVER: claim verification dataset for Danish. In Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa). Linköping University Electronic Press, Sweden, Reykjavik, Iceland (Online), 422– 428.

[26] Gautam Kishore Shahi and Durgesh Nandini. 2020. FakeCovid – A Multilingual Cross-domain Fact Check News Dataset for COVID-19. In Workshop Proceedings of the 14th International AAAI Conference on Web and Social Media.

[27] Fatima K Abu Salem, Roaa Al Feel, Shady Elbassuoni, Mohamad Jaber, and May Farah. 2019. Fa-kes: A fake news dataset around the syrian war. In Proceedings of the International AAAI Conference on Web and SocialMedia, Vol. 13. 573–582.

[28] Archita Pathak and Rohini Srihari. 2019. BREAKING! Presenting Fake News Corpus for Automated Fact Checking. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop. 357–362.

[29] Nguyen Thanh Tam, Matthias Weidlich, Bolong Zheng, Hongzhi Yin, Nguyen Quoc Viet Hung, and Bela Stantic. 2019. From anomaly detection to rumour detection using data streams of social platforms. Proceedings of the VLDB Endowment 12, 9 (2019), 1016–1029.

[30] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic Detection of Fake News. In Proceedings of the 27th International Conference on Computational Linguistics. 3391–3401.

[31] Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media. Big Data 8, 3 (2020), 171–188.

[32] Zhiwei Jin, Juan Cao, Han Guo, Yongdong Zhang, and Jiebo Luo. 2017. Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In Proceedings of the 25th ACM international conference on Multimedia. 795–816.

[33] Parth Patwa, Shivam Sharma, Srinivas PYKL, Vineeth Guptha, Gitanjali Kumari, Md Shad Akhtar, Asif Ekbal, Amitava Das, and Tanmoy Chakraborty. 2020. Fighting an Infodemic: COVID-19 Fake News Dataset. arXiv preprint arXiv:2011.03327 (2020).

[34] Jude Khouja. 2020. Stance Prediction and Claim Verification: An Arabic Perspective. In Proceedings of the Third Workshop on Fact Extraction and VERification (FEVER). Association for Computational Linguistics,
Online, 8–17. https://doi.org/10.18653/v1/2020.fever-1.2

[35] Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. PloS one 11, 3 (2016), e0150989.

[36] Andreas Vlachos and Sebastian Riedel. Fact checking: Task definition and dataset construction. ACL’14.

[37] Naeemul Hassan, Chengkai Li, and Mark Tremayne. Detecting check-worthy factual claims in presidential debates. In CIKM’15.

[38] Aldert Vrij. Criteria-based content analysis: A qualitative review of the first 37 studies. Psychology, Public Policy, and Law, 11(1):3, 2005.
[39] Tony Lesce. Scan: Deception detection by scientific content analysis. Law and Order, 38(8):3–6.

[40] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. A stylometric inquiry into hyperpartisanand fake news. arXiv preprint arXiv:1702.05638, 2017

[41] Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. ACM Transactions on Internet Technology (TOIT), 17(3):26, 2017

[42] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. News verification by exploiting conflicting social viewpoints in microblogs. In AAAI’16.

[43] I. Ahmad, M. A. Alqarni, A. A. Almazroi, and A. Tariq, “Experimental evaluation of

clickbait detection using machine learning models,” IASC-Intelligent Automation & Soft Computing, vol. 26, no. 4, pp. 1335–1344, 2020.

[44] C. J. Hutto and E. Gilbert, “Vader: a parsimonious rule-based model for sentiment analysis of social media text,” in Proceedings of the international AAAI conference on web and social media, vol. 4no. 3, pp. 216–225, Ann Arbor, Michigan USA, 2014.

[45] M. A. Alonso, D. Vilares, C. Gómez-Rodríguez, and J. Vilares, “Sentiment analysis for fake news detection,” Electron, vol. 10, no.11, 2021.

[46] K. Anoop, M. P. Gangan, and V. L. Lajish, “Leveraging heterogeneous data for fake news detection,” in Linking and mining heterogeneous and multi-view data, pp. 229–264, Springer, Cham,2019.

[47] S. Shahnawaz and P. Astya, “Sentiment analysis: approaches and open issues,” in 2017 International Conference on computing,Communication and automation (ICCCA), vol. 2017-Janua, pp. 154–158, Greater Noida, India, 2017.

[48] D. Berrar, “‘Bayes’ theorem and naive Bayes classifier,” Encyclopedia of Bioinformatics and Computational Biology, vol. 1– 3,no. January 2018, pp. 403–412, 2019.

[49] S. L. Ting, W. H. Ip, and A. H. C. Tsang,“Is Naïve bayes a good classifier for document classification?,” International Journal of Software Engineering and Its Applications, vol. 5, no. 3, pp. 37–46, 2021.

[50] K. P. Murphy, “Naive Bayes classi fiers generative classi fiers, ” Bernoulli, vol. 4701, no. October, pp. 1 –8, 2007.

[51] S. Karthika and N. Sairam, “A Naïve Bayesian classi fier for educational quali fication, ” indian journal of science and technology, vol. 8, no. 16, 2021.

[52] W. Dai, G. R. Xue, Q. Yang, and Y. Yu, “Transferring naive Bayes classi fiers for text classi fication, ” -

[53] L. Zhang, W. Zhou, and L. Jiao, “Wavelet support vector machine, ” IEEE Transactions on Systems, Man, and Cybernetics— Part B, vol. 34, no. 1, pp. 34–39, 2004.

[54] R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. 72022, Article ID 3498123, 17 pages, 2022.

[55] S. Khan, S. Hakak, N. Deepa, B. Prabadevi, K. Dev, and S. Trelova, “Detecting COVID-19-related fake news using feature extraction,” Frontiers in Public Health, vol. 9, no. January, pp. 1–9, 2021.

[56] O. D. Apuke and B. Omar,“Fake news and COVID-19: modelling the predictors of fake news sharing among social media users,” Telematics and Informatics, vol. 56, article 101475, 2021.

[57] M. Bhogade, B. Deore, A. Sharma, O. Sonawane, and M. Singh, “A review paper on fake news detection,”International Journal of Advance Scientific Research and Engineering Trends, vol. 6, no. 5, pp. 94–96, 2021.

LIST OF PUBLICATIONS

[1] A. Jain and P.G. Shambharkar "Analysis on Fake News,Manipulation & Detection Techniques: A Review ," communicated and accepted at 5th IEEE International Conference on Advances in Computing, Communication Control and Networking (ICAC3N-23).

[2] A. Jain and P.G. Shambharkar "Fake News Detection Using Fine-Tuned Pretrained Language models," communicated and accepted at 5th IEEE International Conference on Advances in Computing, Communication Control and Networking (ICAC3N-23).

Acceptance Notification 5th IEEE ICAC3N-23 & Registration: Paper ID 367 Inbox ×

Microsoft CMT <email@msr-cmt.org>
to me

17 May 2023, 00:49 (13 days ago)

Dear Aditya Jain,
Delhi Technological University

Greetings from ICAC3N-23 ...!!!

Congratulations....!!!!!

On behalf of the 5th ICAC3N-23 Program Committee, we are delighted to inform you that the submission of "Paper ID- 367 " titled " Fake News Detection Using Fine-Tuned Pretrained Language models " has been accepted for presentation and further publication with IEEE at the ICAC3N- 23 subject to incorporate the reviewers and editors comments in your final paper. All accepted papers will be submitted to IEEE for inclusion into conference proceedings to be published on IEEE Xplore Digital Library.

For early registration benefit please complete your registration by clicking on the following Link: https://forms.gle/8e6RzNbho7CphnYN7 on or before 31 May 2023.

Registration fee details are available @ https://icac3n.in/register.
https://drive.google.com/file/d/1RGu6i4eGUI5B07zOfgRmDRfyjOhOhiC6/view?usp=share_link

You can also pay the registration fee by the UPI. (UPI id - icac3n@ybl) or follow the link below for QR code:
https://drive.google.com/file/d/1Ry-sF0apvy_0zUjM8lNW02lZgWAsE0YD/view?usp=sharing

You must incorporate following comments in your final paper submitted at the time of registration for consideration of publication with IEEE:

Reviewers Comments:
The theme and title of the article "Fake News Detection Using Fine-Tuned Pretrained Language models" is suitable and appropriate for publication.

Acceptance Notification 5th IEEE ICAC3N-23 & Registration: Paper ID 78 Inbox ×

Microsoft CMT <email@msr-cmt.org> Tue, 16 May, 22:36

to me

Dear Aditya Jain,
Delhi Technological University

Greetings from ICAC3N-23 ...!!!

Congratulations....!!!!!

On behalf of the 5th ICAC3N-23 Program Committee, we are delighted to inform you that the submission of "Paper ID- 78 " titled " Analysis on Fake News,Manipulation & Detection Techniques: A Review " has been accepted for presentation and further publication with IEEE at the ICAC3N- 23 subject to incorporate the reviewers and editors comments in your final paper. All accepted papers will be submitted to IEEE for inclusion into conference proceedings to be published on IEEE Xplore Digital Library.

For early registration benefit please complete your registration by clicking on the following Link: https://forms.gle/8e6RzNbho7CphnYN7 on or before 31 May 2023.

Registration fee details are available @ https://icac3n.in/register.
https://drive.google.com/file/d/1RGu6i4eGUI5B07zOfgRmDRfyjOhOhiC6/view?usp=share_link

You can also pay the registration fee by the UPI. (UPI id - icac3n@ybl) or follow the link below for QR code:
https://drive.google.com/file/d/1Ry-sF0apvy_0zUjM8lNW02lZgWAsE0YD/view?usp=sharing

You must incorporate following comments in your final paper submitted at the time of registration for consideration of publication with IEEE:

Reviewers Comments:
The theme and title of the article "Analysis on Fake News,Manipulation & Detection Techniques: A Review" is suitable and appropriate for publication.

Paytm

Payment Successful

₹7,000

Rupees Seven Thousand Only

To: Manjeet Singh

icac3n@ybl

From: Aditya Jain

ICICI Bank - 8629

UPI Ref. No: 3150342 **38764**

06:38 pm, 30 May 2023

POWERED BY UPI UNIFIED PAYMENTS INTERFACE

ICAC3N-21: 3rd IEEE International Conference on Advances in Computing, Communication Control and Networking

Galgotias College of Engineering & Technology (GCET),Greater Noida U.P. (India) -201306

Greater Noida, India, December 17-18, 2021

Conference website	https://icac3n-21.in/
Submission link	https://easychair.org/conferences/?conf=icac3n21

Topics: evolutionary computing big data machine learning networking

About ICAC3N-21:

3rd IEEE International Conference on Advances in Computing, Communication Control and Networking (ICAC3N-21) will be held during **December 17-18, 2021** in **Galgotias College of Engineering and Technology, Greater Noida, India**. The conference is an international forum which aims to bring together leading academician, researchers and research scholars to exchange and share their experiences and hard-earned technological advancements about all aspects of based on their research related to Computing, Communication Control & Networking. We invite all leading researchers, engineers and scientists in the domain of interest from around the world. We warmly welcome all authors to submit your research papers to ICAC3N-21, and share the valuable experiences with the scientist and scholars around the world.

IEEE Conference Record No. #53548

Publication and Indexing

ICAC3N-21 is indexed in SCOPUS and Google Scholar. All registered & presented papers are available on IEEEXplore DigitalLibrary.

PAPER NAME

FAKE NEWS DETECTION USING FINE-TUNED PRETRAINED LANGUAGE MODELS

AUTHOR

Aditya Jain

WORD COUNT

12945 Words

CHARACTER COUNT

70098 Characters

PAGE COUNT

44 Pages

FILE SIZE

1.0MB

SUBMISSION DATE

May 30, 2023 5:17 PM GMT+5:30

REPORT DATE

May 30, 2023 5:18 PM GMT+5:30

9% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 5% Internet database
- 4% Publications database
- Crossref database
- Crossref Posted Content database
- 7% Submitted Works database

Excluded from Similarity Report

- Bibliographic material
- Cited material

Gmail **Aditya Jain <adityajain9560@gmail.com>**

Registration Confirmation, 5th IEEE ICAC3N-23, Paper ID 78

1 message

Microsoft CMT <email@msr-cmt.org> 31 May 2023 at 01:45
Reply-To: Vishnu Sharma <vishnu.sharma@galgotiacollege.edu>
To: Aditya Jain <adityajain9560@gmail.com>

Dear Aditya Jain,
Delhi Technological University

Greetings from ICAC3N-23 ...!!! Thanks for Completing your registration...!!

Paper ID- "78 “
Paper Title- “ Analysis on Fake News,Manipulation & Detection Techniques: A Review “

This email is to confirm that you have successfully completed your registration for your accepted paper at ICAC3N-2023. We have received your registration and payment details. Further, your submitted documents will be checked minutely and if any action will be required at your end you will be informed separately via email.

For further updated regarding conference please keep visiting conference website www.icac3n.in or for any query please write us at icac3n23@gmail.com.

Regards:
Organizing committee
ICAC3N - 2023

Note:
1. Transfer of E-copyright to IEEE and Presenting paper in conference is compulsory for publication of paper in IEEE. (For this you will be informed separately via email well before conference)
2. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled and paper will be rejected and the authors will be responsible for any consequences. Plagiarism must be less then 20% (checked through Turnitin). However, author will be given sufficient and fair chance to reduce the plagiarism.
3. Change in paper title, name of authors or affiliation of authors is not allowed now.
4. Violation of any of the above point may lead to cancellation of registration.
5. Registration fee once paid is non-refundable.

Gmail **Aditya Jain <adityajain9560@gmail.com>**

Registration Confirmation, 5th IEEE ICAC3N-23, Paper ID 367

1 message

Microsoft CMT <email@msr-cmt.org> 31 May 2023 at 01:45
Reply-To: Vishnu Sharma <vishnu.sharma@galgotiacollege.edu>
To: Aditya Jain <adityajain9560@gmail.com>

Dear Aditya Jain,
Delhi Technological University

Greetings from ICAC3N-23 ...!!! Thanks for Completing your registration...!!

Paper ID- "367 "
Paper Title- " Fake News Detection Using Fine-Tuned Pretrained Language models "

This email is to confirm that you have successfully completed your registration for your accepted paper at ICAC3N-2023. We have received your registration and payment details. Further, your submitted documents will be checked minutely and if any action will be required at your end you will be informed separately via email.

For further updated regarding conference please keep visiting conference website www.icac3n.in or for any query please write us at icac3n23@gmail.com.

Regards:
Organizing committee
ICAC3N – 2023

Note:
1. Transfer of E-copyright to IEEE and Presenting paper in conference is compulsory for publication of paper in IEEE. (For this you will be informed separately via email well before conference)
2. If plagiarism is found at any stage in your accepted paper, the registration will be cancelled and paper will be rejected and the authors will be responsible for any consequences. Plagiarism must be less then 20% (checked through Turnitin). However, author will be given sufficient and fair chance to reduce the plagiarism.
3. Change in paper title, name of authors or affiliation of authors is not allowed now.
4. Violation of any of the above point may lead to cancellation of registration.
5. Registration fee once paid is non-refundable.