A MULTI-DOMAIN FAKE NEWS DETECTION MODEL USING BERT AND DistilBERT

A Project Report

Submitted by

Ms. SREETHU P R

REG NO : TKM20MEAI13

SEMESTER : IV

In partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY
IN
Mechanical Engineering (Artificial Intelligence)

Under the guidance of
Prof. SUMOD SUNDAR

Thangal Kunju Musaliar College of Engineering
Kerala

JULY 2022

DECLARATION

I undersigned hereby declare that the project report "A multi-domain fake news detection model using BERT and DistilBERT", submitted for partial fulfillment of the requirements for the award of degree of Master of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Prof. Sumod Sundar. This submission represents my ideas in my own words and where ideas or words of others have been included, I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Place: Kollam
Date: SREETHU P R

Thangal Kunju Musaliar College of Engineering
Centre for Artificial Intelligence

C E R T I F I C A T E

This is to certify that, this report titled ***A MULTI-DOMAIN FAKE NEWS DETECTION MODEL USING BERT AND DistilBERT*** is a bonafide record of the **Project** presented by **SREETHU P R (TKM20MEAI13)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **M.Tech in Mechanical Engineering (Artificial Intelligence)** in **APJ Abdul Kalam Technological University** .

Project coordinator & Internal Supervisor

Prof. Sumod Sundar
Assistant Professor
Centre for Artificial Intelligence

Head of the Department

Dr. Imthias Ahamed
Professor & HOD
Centre for Artificial Intelligence

Internal Examiner

External Examiner

ACKNOWLEDGEMENT

A successful project is a fruitful culmination of efforts by many people, some directly involved and some others indirectly, by providing support and encouragement. Firstly I would like to thank the almighty for giving me the wisdom and grace for making my project a successful one. I thank him for steering me to the shore of fulfillment under his protective wings

I express my sincere gratitude to **Dr. T A Shahul Hameed**, Principal of TKMCE, and **Dr. Imthias Ahamed**, Professor and Head of the Department, Centre for Artificial Intelligence, TKMCE, for their constant support and encouragement throughout the project work.

I would like to express my heartfelt thanks to our project coordinator cum my internal supervisor **Prof. Sumod Sundar**, Assistant Professor, Centre for Artificial Intelligence, TKMCE, for his expert guidance and cooperation. With a profound sense of gratitude, I would like to thank **Dr.Santhi Natarajan**, Honorary Professor, for her immense encouragement. I would like to express my gratitude to **Mr. Rajeev Azhuvath**, Mentor (TCS Mentor ID:120914), Tata Consultancy Services (TCS), for his expert guidance, and cooperation. I also extend my thanks to the entire faculty and staff members of the Centre for AI, TKMCE, who has encouraged me throughout this work.

I also express my thanks to my loving parents, husband, brother, sister and friends, for their support and encouragement in the successful completion of this project work.

SREETHU P R

Abstract

The popularity and usage of online media platforms are increasing day by day and the dissemination of data is rapidly raised. The rise of social networks has accelerated the dissemination of rumors, satires, and false information, increase in the distribution of fake news. So the identification of such news as real or fake is an important task in digital life. The fake news may be on different domains such as political domain, entertainment domain, sports domain, etc. Various studies regarding machine learning and deep learning algorithms are found in the literature. Generalizing a learning model by identifying patterns in a text will help to differentiate fake news from the real one. Fake news detection using BERT and LSTM techniques is the most competitive study happening now. A model is proposed using BERT and DistliBERT to detect fake news on multiple domains and the performance is compared with Naive Bayes, Decision Tree, Random Forest, Logistic Regression and SVM classifiers. It is evaluated using the datasets: the Twitter dataset, ISOT dataset, LIAR dataset, and Kaggle dataset. BERT is a widely used pre-trained transformer model for various Natural Language Processing applications. Pre-training and Fine tuning are the two tasks carried out by BERT. Pre-training includes named Masked Language Model (MLM) and Next Sentence Prediction (NSP), these are train on simultaneously. The pre-training task improves the performance of BERT model. BERT is an encoder stack, so the outputs are some vectors. The output vectors are given to a fully connected layer. The number of neurons in the layer should be equal to the number of tokens in the vocabulary. Softmax activation is used to convert a word vector to a distribution. DistilBERT is a distilled model of BERT used to reduce the training time and memory size. The BERT model obtained an accuracy of 94.8%, 100% and 99.89% on the Twitter, ISOT, and the Kaggle datasets respectively; DistilBERT obtained an accuracy of 78.68% on the LIAR dataset.

Contents

List of Figures

List of Tables

ABBREVIATIONS

API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformers
CNN	Convolutional Neural Network
CSV	Comma Separated Values
DT	Decision Tree
GPT	Generative Pre-trained transformer
GRU	Gated Recurrent Unit
GUI	Graphical User Interface
KNN	k Nearest Neighbours
LR	Logistic Regression
LSTM	Long Short Term Memory
NER	Named Entity Recognition
NLP	Natural Language Processing
PPCA	Probabilstic Principal Component Analysis
RF	Random Forest
RFC	Relational Features Classification
RNN	Recurrent Neural Network
SVM	Support Vector Machine
TSV	Tab Separated Values

Chapter 1

INTRODUCTION

1.1 GENERAL BACKGROUND

Fake news is information that is incorrect or misleading and is presented as news. Fake news is incorrect or misleading information which is presented as news. Fake news is frequently published to harm a person's or entity's reputation or profit from advertising income. When spectacular media accounts were widespread in the 1890s, the phrase was coined [3]. Moreover, high-profile persons increasingly use this term to refer to any misleading information, including unintended and unconscious mechanisms and any news that is unfavourable to their opinions [4]. Furthermore, disinformation is a sneaky sort of propaganda that involves conveying misleading information with a malicious goal and is occasionally created and promoted by hostile foreign entities, particularly during elections. Fake news can include humorous pieces misunderstood as genuine and items with sensationalist or clickbait headlines that are not supported by the text. Because of the wide variety of fake news, current scholars have begun to choose "information disorder" as a more neutral and informative word. With the rise of social media, particularly the Facebook news feed, fake news has increased, and this disinformation has progressively filtered into the mainstream media [5]. Political polarization, post-truth politics, motivated reasoning, confirmation bias, and social media algorithms have all been linked to the propagation of fake news. By competing with actual news, fake news might lessen its impact. According to BuzzFeed data, the top fake news articles concerning the 2016 US presidential election garnered Facebook engagement is higher than that of top stories from major news outlets. It also has the potential to strip public trust in serious news coverage. Former US President Donald Trump is credited with popularizing the word by describing any negative media coverage of him. Because of Trump's misuse of the word, the British government has decided to avoid using it because it is "poorly defined" and "conflates a variety of erroneous information, from legitimate error to foreign meddling." PolitiFact is a fact-checking website that measures the authenticity of claims made by elected officials and others by using its Truth-O-Meter [6]. The objective of Truth-O-Meter is to indicate the statement's relative accuracy. The scale has six levels of trustworthiness, from lowest to highest: TRUE – The statement is correct, and nothing important is missing, MOSTLY TRUE – The statement is correct, but it needs further information or clarification, HALF TRUE – The statement is partially correct, but it omits key details or presents information out of context, MOSTLY FALSE – While the statement contains some truth, it leaves out important details that might create a different impression,

FALSE - The assertion is incorrect, PANTS ON FIRE — The remark is untrue and makes an absurd claim. Like Truth-O-Meter, sometimes Flip-O-Meter were used in PolitiFact. In Flip-O-Meter, there are three categories: No Flip: There is no significant shift in position. The candidate has maintained a high level of consistency; Half Flip: An inconsistency in words or a partial shift of position; Full Flop: A total shift of perspective; a complete flip-flop. But several studies say that Truth-O-meter is better than Flip-O-meter.

According to preliminary studies by Claire Wardle of First Draft News, there are seven forms of fake news [7]:

- Satire or parody (" it has no malicious purpose, it can deceive")
- False connection ("when the content is not supported by the headlines, images, or captions")
- Misleading content ("the misrepresentation of facts in order to frame a situation or a person ")
- False content ("when authentic material is accompanied by misleading contextual data ")
- Impostor content ("when real sources are impersonated by made-up fake source")
- Manipulated content ("When authentic information or imagery is altered to deceive," as in a "doctored" photograph)
- Fabricated content ("The new content is completely misleading and intended to deceive and damage people.")

Scientific denialism, defined as the act of generating incorrect or misleading information to maintain strong pre-existing ideas unintentionally, is another possible explanatory category of fake news. Online media platforms' popularity and usage are increasing daily, and data dissemination is rapidly rising. For a long time, there has been fake news. Indeed, this problem has been addressed since its inception, with disastrous effects in both the technological and political worlds. So the identification of fake news spread over these platforms should be addressed more. Facebook has previously made efforts to counteract the spread of misinformation on its website (in certain countries) by collaborating with third-party fact-checkers to analyze and score the veracity of articles and postings on the social media platform [8]. Artificial intelligence methods help identify the spread of fake news on these online platforms through deep automated techniques. Deep learning algorithms such as CNN, BERT, and LSTM, together with other machine learning techniques, effectively detect fake news with reasonable accuracy. These techniques help explore more features that can be used to analyze fake data from the actual content. The Transformer architecture underpins BERT. BERT is an acronym for Bidirectional Encoder Representations from Transformers. It is intended to condition both left and proper context to pre-train deep bidirectional representations from the unlabeled text [9]. As a result, using just one additional output layer, the pre-trained BERT model may be fine-tuned to generate state-of-the-art models for a wide range of Natural language processing tasks.

BERT was pre-trained on a massive collection of unlabeled text, including the entire Wikipedia (2,500 million words) and the Book Corpus (800 million words) [10]. Half of

BERT's success is due to this pre-training stage. BERT is a model which "deeply bidirectional". Bidirectional indicates that BERT learns information from both the left and right sides of a token's context during the training phase. A model's bidirectionality is critical for thoroughly understanding the meaning of a language. We may fine-tune it by adding a few more output layers to construct cutting-edge models for many NLP problems.

The rest of this work is organized into the following sections: Section 2 reviews the different machine learning and deep learning techniques for fake news detection. Section 3 includes dataset description and detailed methodology. Section 4 describes the data preprocessing step. The classification algorithms are explained in section 5. Section 6 includes the results and discussion and section 7 explains the conclusion and future work.

1.2 OBJECTIVES

BERT and DistilBERT are two pre-trained transformer models. They are used in various NLP applications. The objective of proposed model is to design a multi-domain fake news detection system using BERT and DistilBERT. The transformer models are applied in dataset that contains news from political domain, entertainment domain, sports domain, medical domain, etc.Finally the performance of proposed model is compared with machine learning models such as Random forest, Logistic regression, Naive Bayes, SVM and Decision Tree. The tasks carried out in the experimentation of the proposed model are summarized as follows.

- Pre-processed ISOT, LIAR and Kaggle datasets.
- Performed fake news detection in multi-domain using BERT and DistilBERT.
- Compare the outputs of BERT and DistilBERT with machine learning algorithms like Random forest, Logistic regression, Naive Bayes, SVM and Decision Tree.

Chapter 2

RELATED WORKS

In this section, several studies of fake news detection utilizing both machine learning and deep learning techniques are discussed. Kaliyar et al. proposed a method for Fake news identification in social media with a BERT-based deep learning approach [12]. The proposed model combines BERT and three parallel blocks of 1d-CNN with varying kernel-sized convolutional layers and distinct filters for better learning. Their model is based on a pre-trained bidirectional transformer encoder word embedding model (BERT). They use BERT as a sentence encoder to accurately extract a sentence's context representation to detect fake news. With its powerful capacity to capture semantic and long-distance relationships in phrases, their work increased the performance of fake news identification. The classification findings show that FakeBERT gives an accuracy of 98.90%. Shishah presented Fake News Detection Using BERT Model with Joint Learning [13]. An unconventional BERT with a combined learning-based model is presented to detect fake news in articles. The proposed method can detect fake news in both lengthy and short pieces. Rather than presenting sequences to utilize the initial hidden states of BERT, all hidden states with dynamic range attention mechanisms are used to compute weights. Relational features classification (RFC) and named entity recognition (NER) task models are combined with BERT via a common parameter layer in collaborative learning to improve generalization. A novel framework called SPR-encoder is used in the suggested strategy to change the dynamic attention range of k layers in the BERT model for constructing the task's context vector and exploiting prior information in the given pre-trained model. Two mask matrices are used to extract the required feature presentation of the RC layer for creating the RFC model. Mehta et al. proposed a transformer-based architecture for fake news classification [14].They discussed and addressed the various aspects of transfer learning in the suggested model and presented an architecture to classify fake news. Approaches that focus on text classification utilize contextual word embeddings because the context of the events is critical in determining the news's legitimacy. This is accomplished by language models such as ELMo and BERT, which have increased performance in various NLP tasks.BERT is the first ELMo-based language representation that is deeply bidirectional and unsupervised. Tuan et al. proposed a multimodal fusion with BERT and an attention mechanism for fake news detection [15]. A new multimodal approach for detecting fake news has been developed. They obtain feature representations from many modalities using neural networks. The attention mechanism combines multimodal features, placed in a sigmoid layer for classification. They employed the BERTweet model to extract feature representations from sentences and a VGG-19 net-

work to extract feature representations from visuals. They suggested a scaled dot-product attention mechanism for both texts and images and a self-attention mechanism for images because they believe that all components of images are related in non-photoshopped images. To improve the accuracy of fake news detection, textual and visual representations and three attention outputs are integrated. J Briskilal et al. proposed an ensemble model for classifying idioms and literal texts using BERT and RoBERTa [16]. An idiom is a phrase whose true meaning differs from the one delivered. Rule-based generalization is utilized in idiom recognition and context-based classification to classify idioms and literal phrases. Crowdsourcing has lately been used to detect idiomatic language sentiment annotations. This approach was used to identify 5000 often recurring idioms in total. Several approaches to classifying idioms and literals have been proposed, but none of them has used ensemble pre-trained models like BERT and RoBERTa. This work aimed to use an ensemble method to categorize idiomatic and literal sentences accurately. Trueman et al. proposed an attention-based C-BiLSTM for fake news detection [17]. Deep learning approaches have the advantage of automatically recognizing features. These methods determine the meaning of a word while considering its context. Attention mechanisms [18], in particular, have emerged as one of the most powerful strategies in natural language processing. They are generally utilized in conjunction with recurrent neural networks to anticipate the most important information in a succession of inputs. This work tackles the topic of detecting fake news in a multi-class context. Improve the accuracy of fake news detection by combining attention processes with convolutional bidirectional recurrent neural networks in future. Umar et al. proposed Fake News Stance Detection Using Deep Learning Architecture such as CNN-LSTM [19]. They suggested a technique that automatically classifies news stories as agree, disagree, unrelated, or discussed based on their position labels. The level of agreement between the headline and the body given to headlines is used to classify them. The proposed model is based on observations of how to discover the relevancy of articles by looking for keywords in headlines. Some of the headline keywords can be used to identify crucial sentences in the text of the article. Kumar et al. proposed a model for Fake news detection using deep learning models [20]. Their investigation carefully selected seven sentiment categorization models, including versions of the convolutional neural network (CNN) and long short-term memory (LSTM) architectures. CNN models are frequently used for image classification and detection and text categorization. Because of their limited information retention power and disappearing and ballooning gradient concerns, simple RNNs were not used in their scenario.As a result, they used LSTMs and their variation, bidirectional LSTMs, to filter out these difficulties using RNNs. MaxPooling, the most used pooling approach, is employed in this network for pooling.It has done by applying a max filter to the initial representation's (usually) nonoverlapping subregions. Furthermore, instead of using the rectified linear unit (ReLU) activation function to map the results, they used the Leaky ReLU activation function because negative values become zero when using the ReLU activation function. It immediately reduced the model's accuracy and ability to fit or train from the data properly. They have used ensembling to put their combinations together. The method of ensembling in various networks has proven to be quite effective in improving a network's performance.

The advantages, limitations and possibilities work of five recent related works are summarized in table 2.1.

Title	Technique used	Advantages	Disadvantages/Future work
Implementation of the BERT-derived architectures to tackle disinformation challenges [1] - 2021	BERT, RoBERTa, RNN	This model is solid and reliable, ready to use in real-time fake news detection systems.	In future retrain the model effectively and can be used in various domains
Multimodal Fusion with BERT and Attention Mechanism for Fake News Detection [2] - 2021	BERTweet model and VGG-19 network.	Scale dot product attention mechanism to capture the relationship between text features and visual features.	Very ambiguous when using the picture to express the content of the tweet.
FakeBERT: Fake news detection in social media with a BERT-based deep learning approach [3] - 2021	Single-layer CNNs with BERT	Faster training of model and lower cross-entropy loss. It is convenient to handle large-scale structure as well as unstructured text. It effectively addresses ambiguity.	Not detect the occurrence of fake news for multi-label datasets.
Attention-based C-BiLSTM for fake news detection [7] - 2021	C-BiLSTM	Captures local, global, and temporal meaning of the sentence using C-BiLSTM. Attention mechanism helps to memorize long input sequence.	Small Dataset used.
A transformer-based architecture for fake news classifcation [5] -2021	BERT	Explored both binary as well as multi-label classifcation	In future hyper parameter tuning of the BERT and subsequent layers on the model.

Table 2.1: Review of deep learning techniques for fake news detection

Chapter 3

METHODOLOGY

3.1 PROPOSED MODEL

The proposed model is used to classify the news as fake or real using BERT and DistilBERT in multi-domain scenario. Finally the accuracy is compared with the accuracy values of different machine learning techniques such as Logistic Regression, Naive Bayes, SVM,Random Forest and Decision Tree. Figure 3.1 shows the framework of the proposed model and the entire experimentation setup is described by fig 3.2.

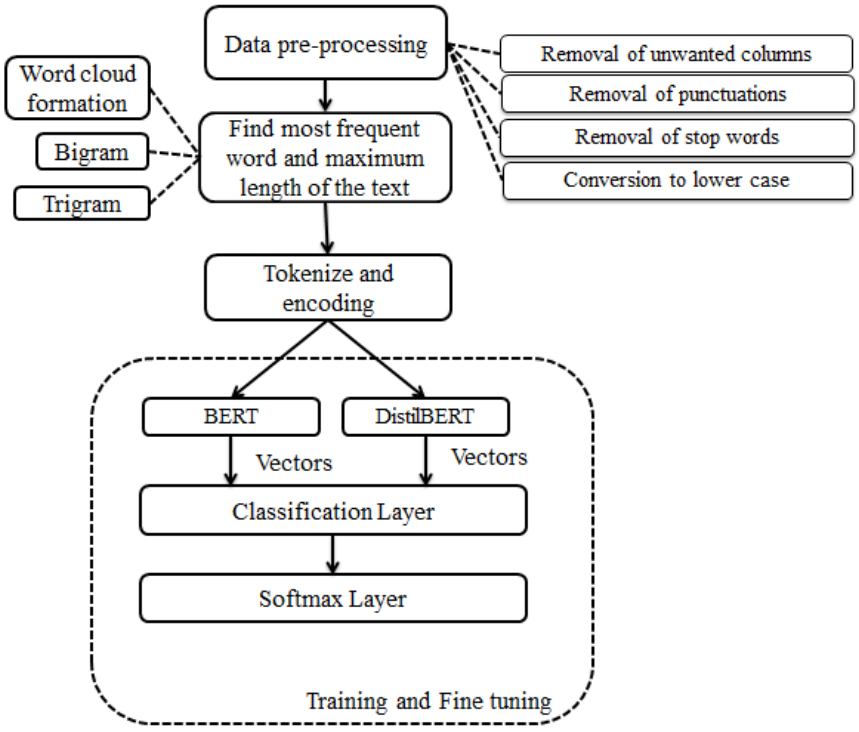

Figure 3.1: Framework of proposed model

The dataset contains text data, which may not be in the same format. So, data preprocessing plays a prominent role in the proposed model. In the preprocessing steps in the

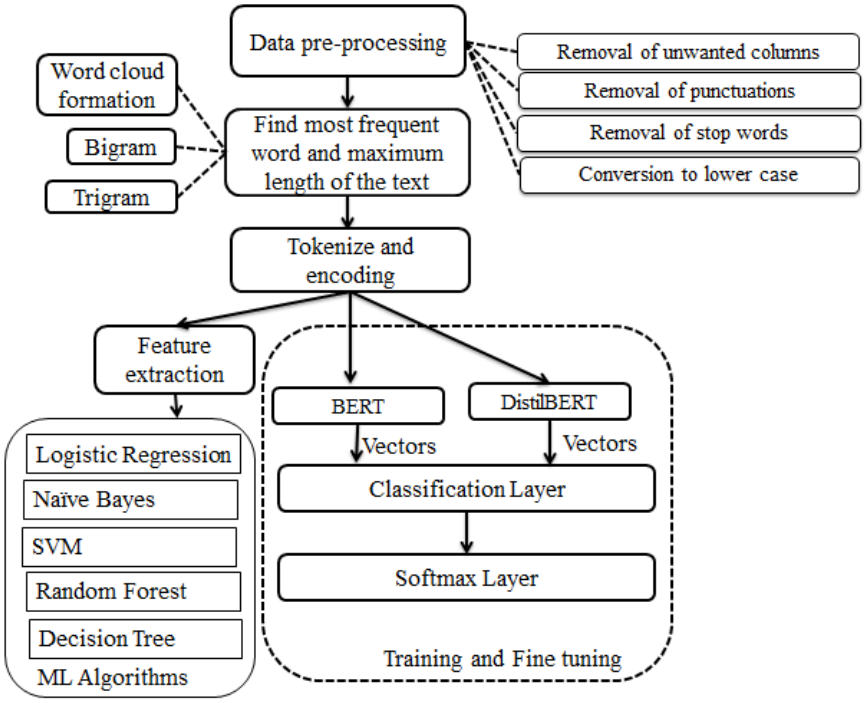

Figure 3.2: Overview of experimentations performed

input dataset, unwanted columns, punctuation, and stop words are removed, and all upper case letters are converted into lower case. After this process, all the text data becomes in the same format. So the analysis becomes more accessible. Then the next step is word cloud formation. Simple text analysis is represented by word clouds and visual representations of text data. Word clouds show the most important or frequently used words in a passage of text (such as a State of the Union Address). The most popular words in the language are usually ignored in a Word Cloud ("a", "an", "the", etc.). The remaining words are shown in a "cloud," with the font size (and colouration of the characters in the word) representing each target word's relative frequency of occurrence in the source material. Word cloud helps find the most frequent word and plot a bar graph for finding the most frequent word. Model creation is the backbone of the proposed model. The main focus is to design and experiment models based on BERT, DistilBERT and ML algorithms such as Naıve Bayes, Logistic regression, Decision tree, Random forest and SVM. Then four sets of the dataset are used for training and testing. Then classification using these models and performance evaluation were carried out.

3.2 FRAMEWORK AND DATASETS

3.2.1 Frameworks Used

The framework used in the project is Keras with Tensorflow, PyTorch as background in the Google Colab and Power edge server with NVIDIA TESLA V100 GPU. The library files used are HuggingfaceTransformer, NumPy, Pandas, Matplotlib, NLTK, etc. The Hugging Face transformers package is an immensely popular Python library providing pre-trained models

such as BERT that are extraordinarily useful for a variety of natural language processing (NLP) tasks. It previously supported only PyTorch, but as of late 2019, TensorFlow 2 is supported as well. Numerous mathematical operations can be carried out on arrays with NumPy. It provides a vast library of high-level mathematical functions that work on these arrays and matrices, as well as strong data structures that ensure efficient calculations with arrays and matrices in Python. Pandas is a data analysis and manipulation software package created for the Python programming language. It includes specific data structures and procedures for working with time series and mathematical tables. Matplotlib is a plotting library for Python programming language and NumPy is the numerical mathematics extension. For integrating charts into programmes utilizing all-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK, it offers an object-oriented API. NLTK is a toolkit built for working with NLP in Python. It provides us with various text processing libraries with a lot of test datasets. A variety of tasks can be performed using NLTK, such as tokenizing, parse tree visualization, etc.

3.2.2 Datasets

There are four set of datasets used for the study: ISOT dataset, LIAR dataset, Kaggle dataset and Twitter dataset. Only Twitter dataset is a pre-processed dataset.

Twitter dataset

Twitter dataset is a pre-processed dataset.It contains two set of CSV files : shorttextpreprocessedtrain for training and shorttextpreprocessedtest for testing.Training set consist of 21390 Fake news and 3946 Real news.Testing set consist of 5379 Fake news and 987 Realnews. It is a class imbalance dataset.

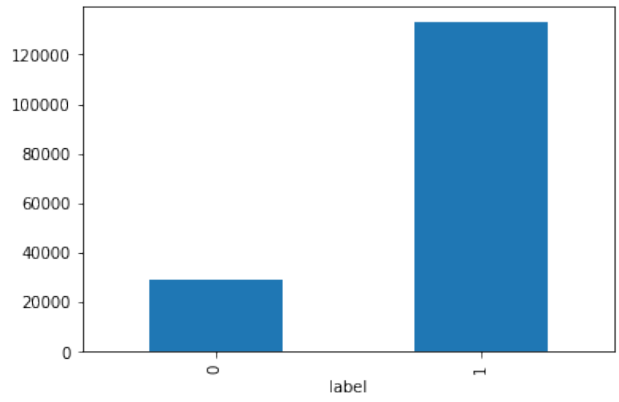

Figure 3.3: True-Fake classes on training dataset

ISOT dataset

The dataset includes both fake news and legitimate content. This dataset was compiled from reliable sources; the accurate articles were culled from the news website Reuters.com. The phoney news pieces were gathered from a variety of sources. The false news reports were gathered from shady websites that Wikipedia and the American fact-checking group

Politifact identified as unreliable. The collection includes several articles kinds on various subjects. The majority of articles, however, concentrate on political and international news issues. There are two CSV files in the dataset. More than 12,600 items from reuter.com in the first file, "True.csv." More than 12,600 stories are included in the second file, "Fake.csv," which comes from various sources used by fake news outlets. Each article includes the following details: the article's title, text, format, and publication date.

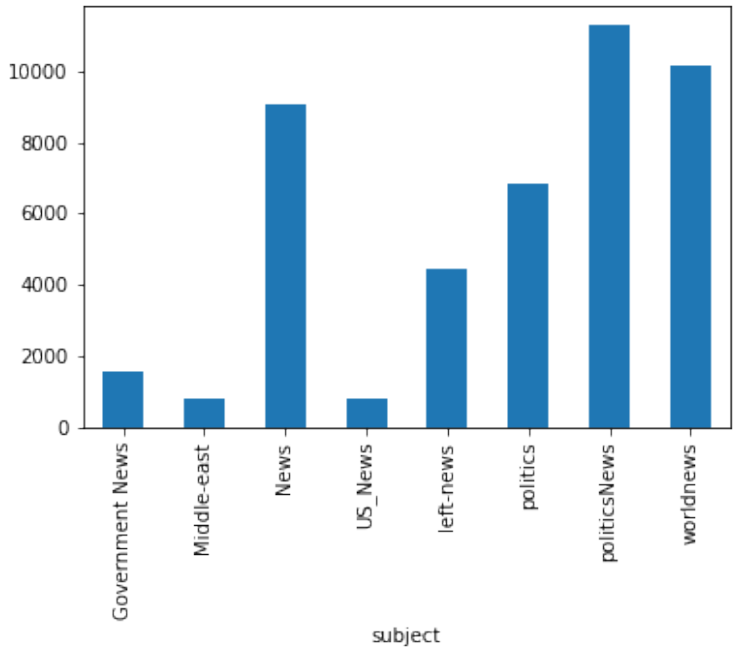

Figure 3.4: Categories in ISOT dataset

LIAR dataset

LIAR is a publicly available dataset for fake news detection. 12.8K short statements that were hand labelled over the course of ten years in a variety of circumstances were gathered from POLITIFACT.COM, which offers a thorough analysis report and links to the relevant source papers for each case. This dataset can also be used to verify study findings.

It includes three sets of TSV files: test, valid, and train (10240 data) (10240). The main purpose of a tab-separated values (TSV) file, a text format, is to store data in a table structure with one line of text for each row in the table. Values for the field are separated by tab characters in the record. The semantics of table columns might be revealed through header rows. TSV files work well as a data interchange format between applications that employ spreadsheets or structured tables. These tab-separated value fields may include text, numbers, statistics, or other types of information. Although data fields stored in CSV files are separated by commas rather than by tabular spaces, the TSV file format is widely supported and extremely comparable to CSV file types. Both fall within the category of delimiter-separated values.

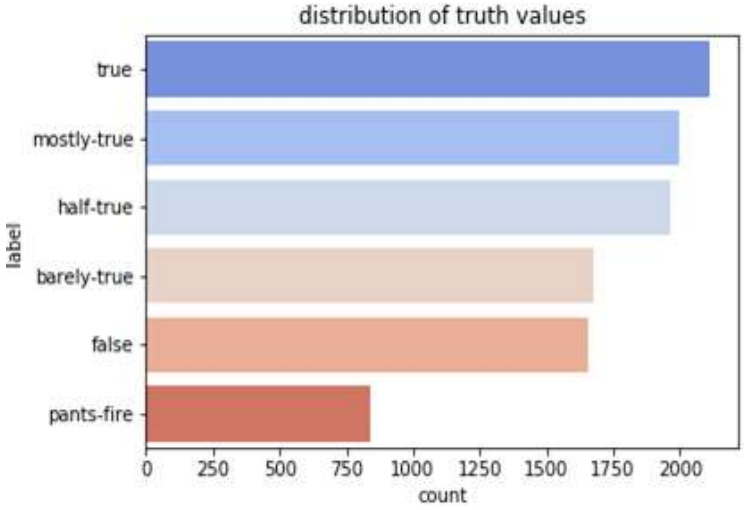

Figure 3.5: Label distribution on LIAR Dataset

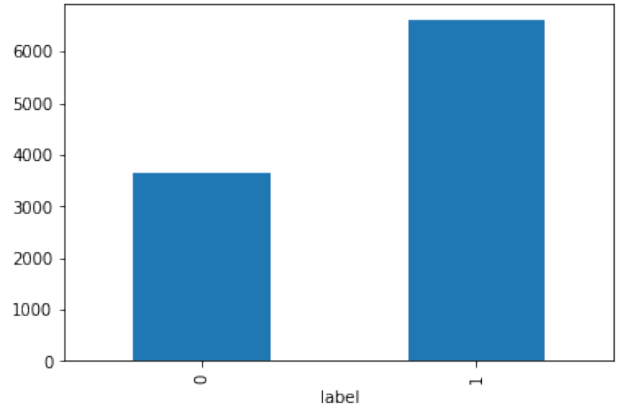

Figure 3.6: True-Fake classes on training dataset

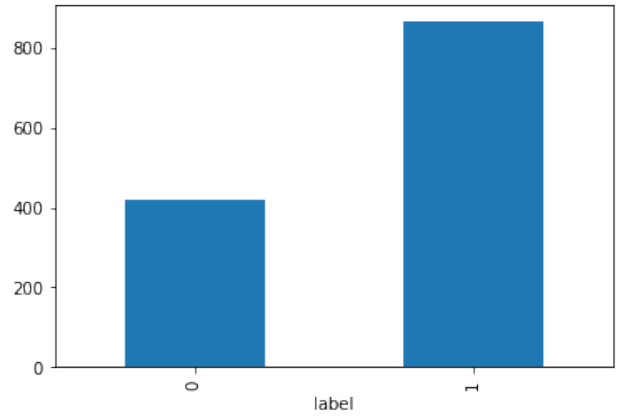

Figure 3.7: True-Fake classes on validation dataset

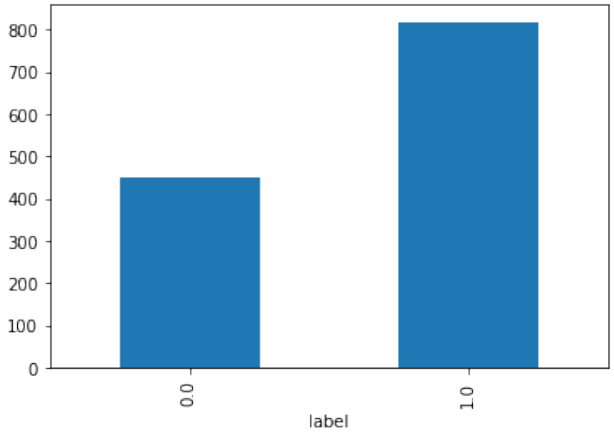

Figure 3.8: True-Fake classes on testing dataset

Kaggle Dataset

Contains two sets of CSV files: Train(20776) and Test(5201) Training and testing dataset have the attributes; id: unique id for a news article title: the title of a news article author: author of the news article text: the text of the article; could be incomplete label: a label that marks the article as potentially unreliable

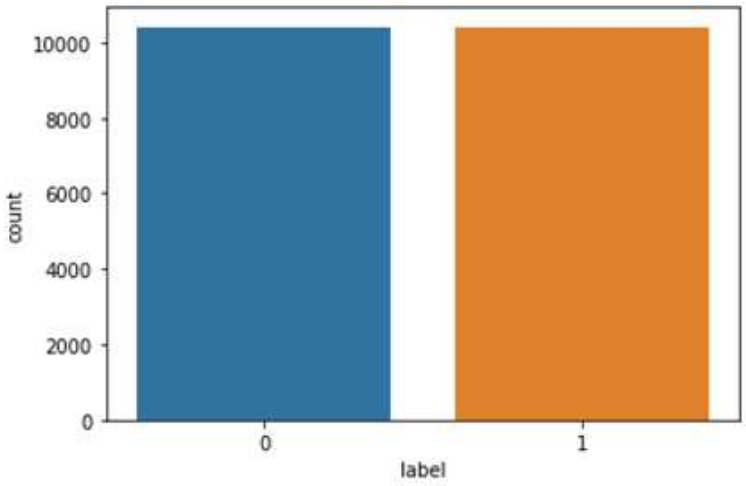

Figure 3.9: True-Fake classes on Kaggle dataset

Chapter 4

DATA PREPROCESSING AND WORD CLOUD FORMATION

4.1 DATA PREPROCESSING

There are four set of datasets are used for analysis.ISOT, LIAR and Kaggle dataset needs preprocessing.

4.1.1 Pre-processing on ISOT dataset

The ISOT dataset also consist of so many columns but only we want the text column and the label corresponding to the text. So that removing unwanted columns are the first preprocessing step. Then remove punctuational marks,stop words and also convert all the upper case letters into lower case.

Out[6]:

	title	text	subject	date	target
0	Republican ex-Treasury chief Paulson slams Tru...	WASHINGTON (Reuters) - Henry Paulson, a Republ...	politicsNews	June 25, 2016	0
1	SHOCK POLL In MUST WIN State Of FLORIDA: Hispa...	Apparently the Black Lives Matter terror group...	left-news	Jul 11, 2016	1
2	MEDALS OF VALOR: President Trump Honored Agent...	It s great to have a president who appreciates...	politics	Jul 27, 2017	1
3	Newsweek Just Made Their BEST Cover Ever And ...	Newsweek has never been a publication to shy a...	News	November 9, 2017	1
4	Trump says he believes Cuba responsible for at...	WASHINGTON (Reuters) - President Donald Trump ...	politicsNews	October 16, 2017	0

Figure 4.1: ISOT Dataset before pre-processing

Out[17]:

	title	text	target
0	republican extreasury chief paulson slam trump...	washington reuters henry paulson republican u ...	0
1	shock poll must win state florida hispanic tur...	apparently black life matter terror group mana...	1
2	medal valor president trump honored agent offi...	great president appreciates special agent poli...	1
3	newsweek made best cover ever people freaking	newsweek never publication shy away controvers...	1
4	trump say belief cuba responsible attack hurt ...	washington reuters president donald trump said...	0

Figure 4.2: ISOT Dataset after pre-processing

4.1.2 Preprocessing on LIAR dataset

The LIAR dataset consist of so many columns but only we want the text column and the label corresponding to the text. So that removing unwanted columns are the first preprocessing step. Then remove punctuational marks,stop words and also convert all the upper case letters into lower case.The news were classified into 6 categories: True, half-true, Mostly-true, Barely-true, False, Pants-fire. But here we do binary classification, so map true and Mostly true into True (0) category and others into fake(1). Also give name to the columns like text and label.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	12134.json	barely-true	We have less Americans working now than in the...	economy,jobs	vicky-hartzler	U.S. Representative	Missouri	republican	1	0	1	0	0	an interview with ABC17 News
1	238.json	pants-fire	When Obama was sworn into office, he DID NOT u...	obama-birth-certificate,religion	chain-email	NaN	NaN	none	11	43	8	5	105	NaN
2	7891.json	false	Says Having organizations parading as being so...	campaign-finance,congress,taxes	earl-blumenauer	U.S. representative	Oregon	democrat	0	1	1	1	0	a U.S. Ways and Means hearing
3	8169.json	half-true	Says nearly half of Oregons children are poor.	poverty	jim-francesconi	Member of the State Board of Higher Education	Oregon	none	0	1	1	1	0	an opinion article
4	929.json	half-true	On attacks by Republicans that various program...	economy,stimulus	barack-obama	President	Illinois	democrat	70	71	160	163	9	interview with CBS News

Figure 4.3: LIAR Dataset before pre-processing

	text	label
0	we less americans working 70s	1
1	when obama sworn office did not use holy bible...	1
2	says having organizations parading social welf...	1
3	says nearly half oregons children poor	1
4	on attacks republicans various programs econom...	1

Figure 4.4: LIAR Dataset after pre-processing

4.1.3 Pre-processing on Kaggle dataset

The Kaggle dataset also consist of so many columns but only we want the text column and the label corresponding to the text. So that removing unwanted columns are the first preprocessing step. Then remove punctuational marks,stop words and also convert all the upper case letters into lower case.

	id	title	author	text	label
0	0	House Dem Aide: We Didn't Even See Comey's Let...	Darrell Lucus	House Dem Aide: We Didn't Even See Comey's Let...	1
1	1	FLYNN: Hillary Clinton, Big Woman on Campus - ...	Daniel J. Flynn	Ever get the feeling your life circles the rou...	0
2	2	Why the Truth Might Get You Fired	Consortiumnews.com	Why the Truth Might Get You Fired October 29, ...	1
3	3	15 Civilians Killed In Single US Airstrike Hav...	Jessica Purkiss	Videos 15 Civilians Killed In Single US Airstr...	1
4	4	Iranian woman jailed for fictional unpublished...	Howard Portnoy	Print \nAn Iranian woman has been sentenced to...	1

Figure 4.5: Kaggle Dataset before pre-processing

	title	text	label
0	house dem aide didnt even see comeys letter ja...	house dem aide didnt even see comeys letter ja...	1
1	flynn hillary clinton big woman campus breitbart	ever get feeling life circle roundabout rather...	0
2	truth might get fired	truth might get fired october 29 2016 tension ...	1
3	15 civilian killed single u airstrike identified	video 15 civilian killed single u airstrike id...	1
4	iranian woman jailed fictional unpublished sto...	print iranian woman sentenced six year prison ...	1

Figure 4.6: Kaggle Dataset after pre-processing

4.2 WORD CLOUD FORMATION

Word clouds or tag clouds are graphical representations of word frequency that give greater prominence to words that appear more frequently in a source text. The larger the word in the visual the more common the word was in the document(s).

Word cloud on Twitter dataset

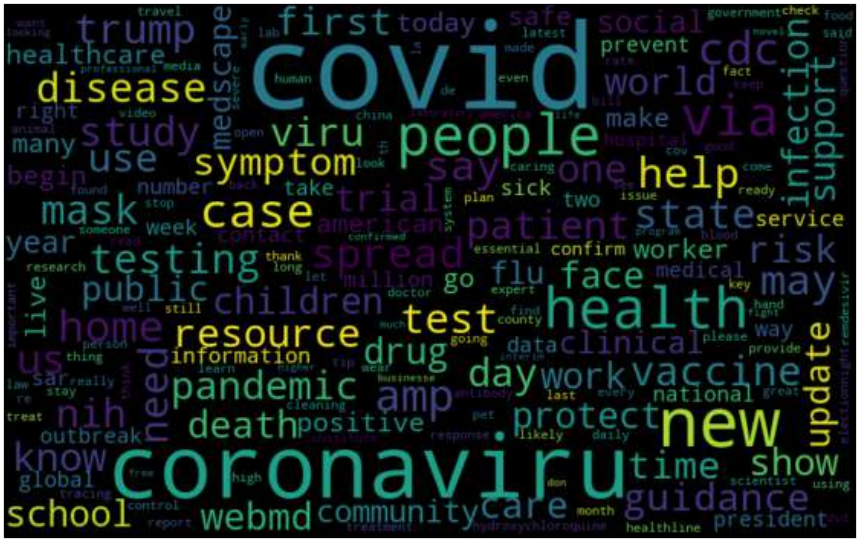

Figure 4.7: Word cloud on real news of Twitter dataset

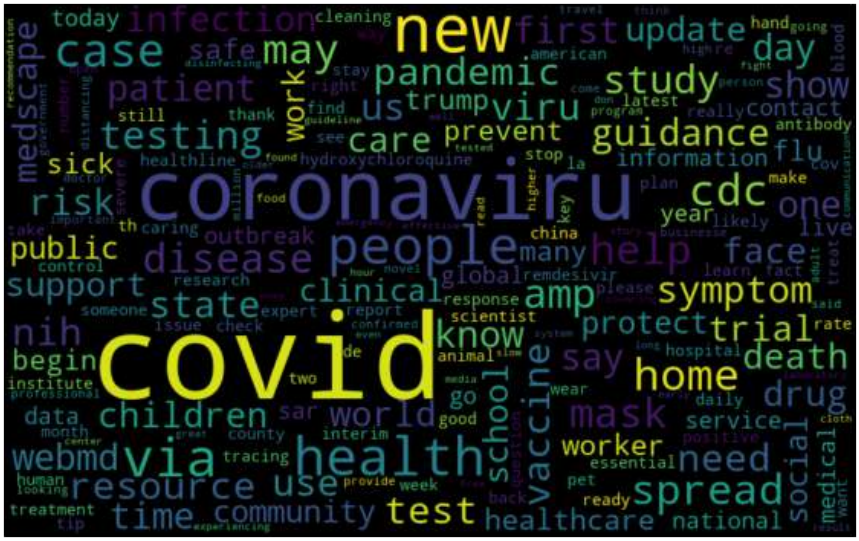

Figure 4.8: Word cloud on fake news of Twitter dataset

Word cloud on ISOT dataset

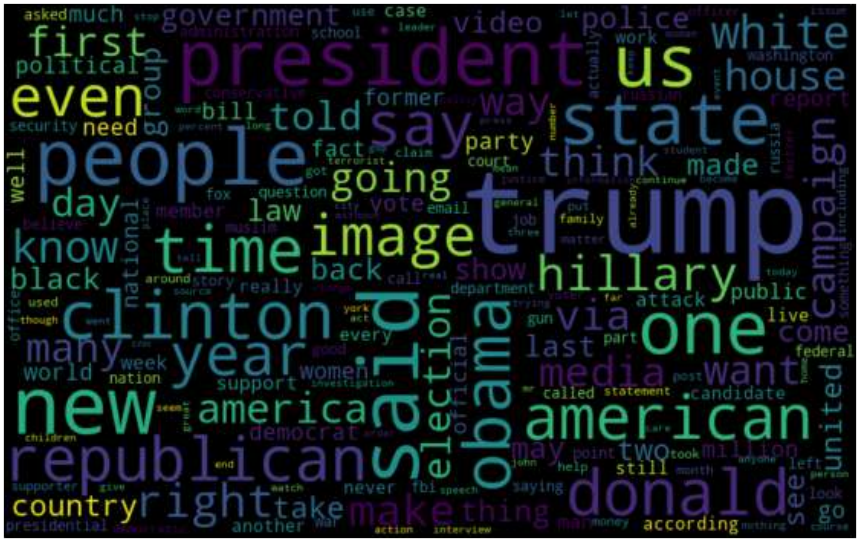

Figure 4.9: Word cloud on real news of ISOT dataset

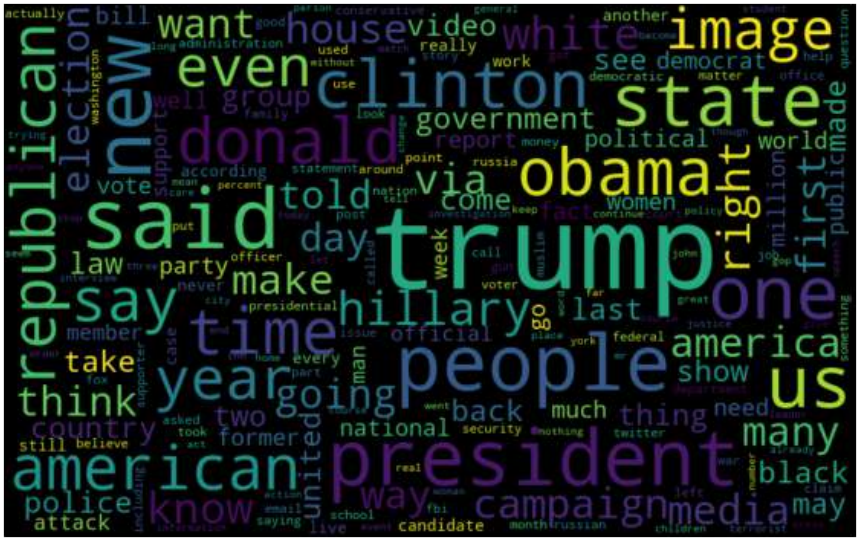

Figure 4.10: Word cloud on fake news of ISOT dataset

Word cloud on LIAR dataset

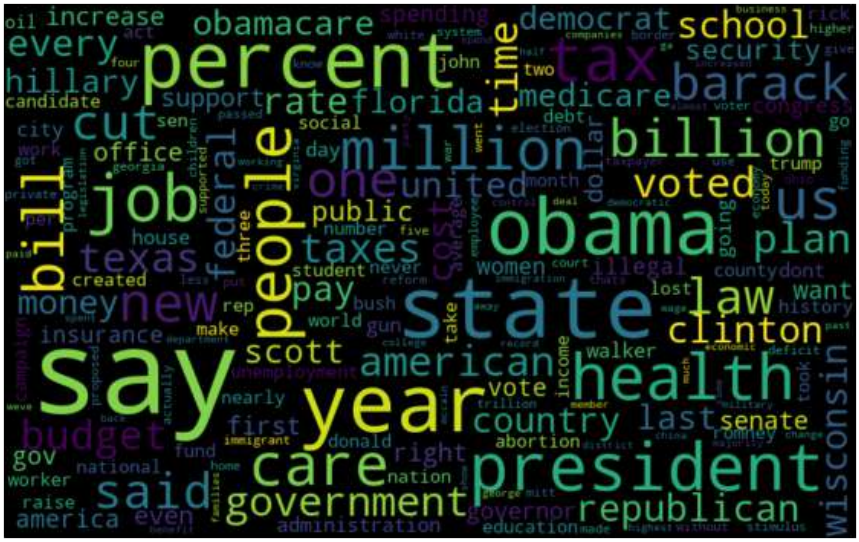

Figure 4.11: Word cloud on real news of LIAR dataset

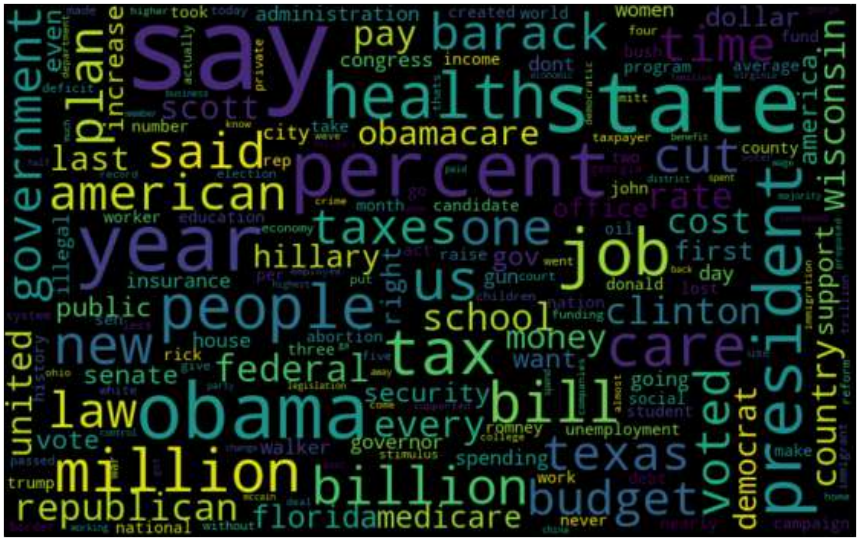

Figure 4.12: Word cloud on fake news of LIAR dataset

Word cloud on Kaggle dataset

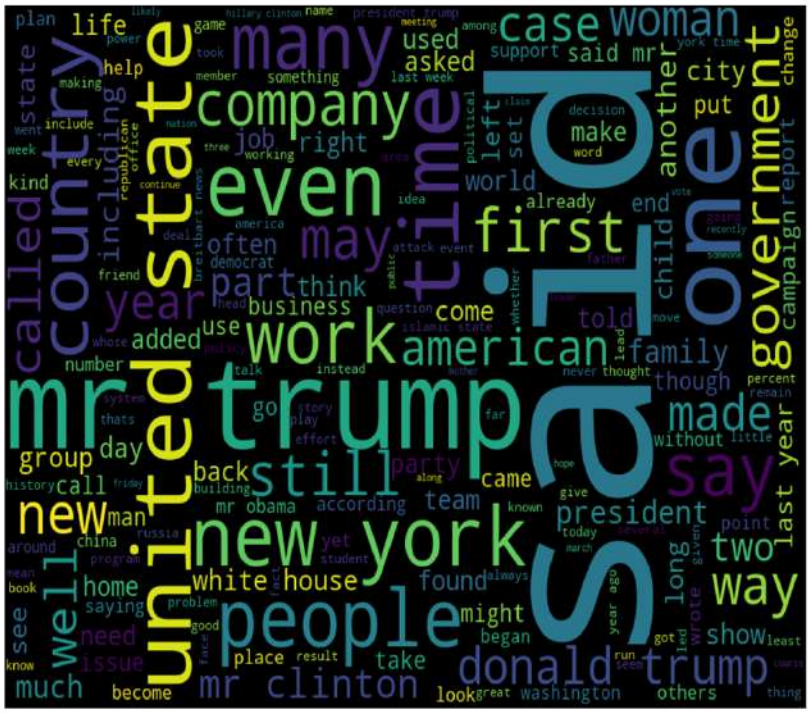

Figure 4.13: Word cloud on real news of Kaggle dataset

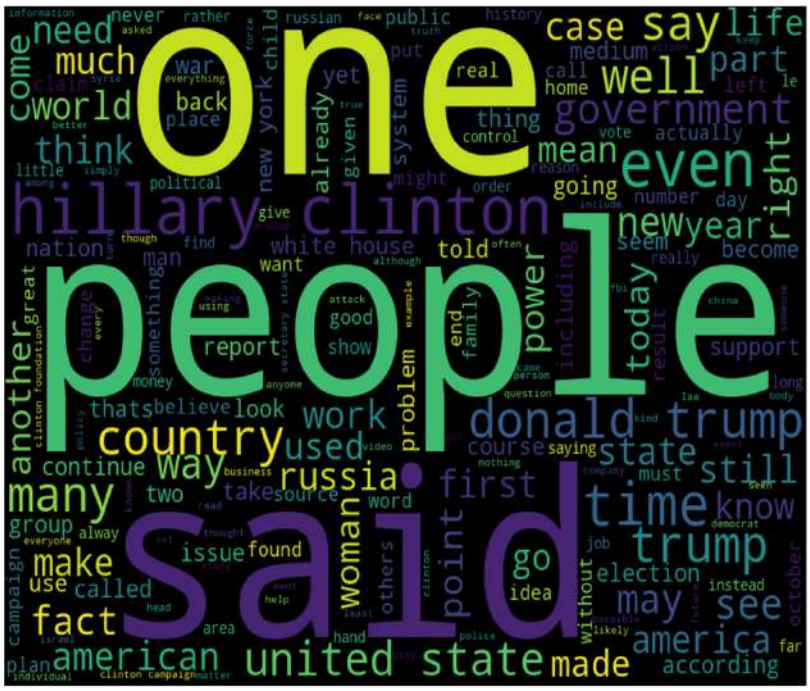

Figure 4.14: Word cloud on fake news of Kaggle dataset

Chapter 5

CLASSIFICATION ALGORITHMS

5.1 MACHINE LEARNING ALGORITHMS

Naïve Bayes Classifier

Naive Bayes classifier is one of the most basic fastest classification algorithm used in machine learning and commonly used for text classification.The Naive Bayes algorithm is basically on Bayes theorem and included in unsupervised learning model [30].It is a probabilistic classifier and it predicts based on the probability of the object. The each pair of features being classified by Naive Bayes algorithm is independet of others. This algorithm needs a large set of training dataset for text classification.

Logistic regression

Logistic regression included in the family of supervised learning approach [31].The logistic function is also known as sigmoid function. Logistic regression is used for the prediction of categorical dependent variable from a set of independent variables. So that the result is a categorical or discrete value. That is it can be a Yes or No, 0 or 1, True or False and so on. But it does not gives the exact value like 0 or 1, it gives a probabilistic range between 0 and 1. Logistic regression varies from linear regression only based on its usage. Instead of fitting a regression layer it uses a "S" shaped logistic function for predicting two maximum values (0 or 1). The key feature of logistic regression is the ability of giving probabilities and categorise new data using both continuous and discrete datasets. Logistic regression has the ability to categorise the observations based on many forms of data and rapidly identifies the most efficient variable for classification.

Decision trees

Decision tree is a rule-based algorithm for classification and regression problems [30]. It is a type of supervised learning model, that continuously separate data based on a parameter. Decision tree contains decision node and leaves. Decision node used to separate the data and leaves represents the final outcome or decisions of the tree. It uses the values in each feature to split the dataset to a point where all data points with same class are grouped. Decision tree has verity of applications so it can be used as a general-purpose predictive modeling tool. Based on some conditions decision tree determines multiple ways to segment

a data. The goal of decision tree is to develop a target variable by learning simple decision rules from data attributes.

Random Forest

Random forest is of the well-known supervised machine learning algorithm. It can be used for both classification and regression applications [30]. The performance of random forest is increased by its ensemble learning ability. Ensemble learning method is a type of method that integrate several classifiers to solve a complex problem. Random forest contains a number of decision trees and the decision tree evaluates based on various subsets of the given dataset. Finally takes the average output of all the decision trees and enhances the predicted accuracy of the dataset. Instead of taking the single output of decision trees. the random forest collects the forecasts from each tree and predicts the final output. The accuracy of random forest increased by increasing the number of decision trees in the random forest.

Support Vector Machine (SVM)

Support vector machine also known a SVM is supervised learning used for both classification and regression problems [30]. SVM finds an optimum line or decision boundary. The decision boundary categorises n-dimensional space into classes. So that in future the additional data points are categorised into correct class. The optimal choice boundary is known as hyper plane and SVM helps to determine the extreme points or vectors that used to create the hyper plane. The points create the hyper plane is known as support vectors and the algorithm used to find these support vector is called support vector machine.

5.2 BERT

The primary technological advancement of BERT is the application of Transformer's bidirectional training, a well-liked attention model, to language modelling. As opposed to other attempts, which either examined a text sequence from left to right or combined left-to-right and right-to-left training, this one looks at it from top to bottom. The Transformer encoder reads the entire sequence of words simultaneously, in contrast to directional models, which read the text input sequentially (from right to left or left to right). Although it would be more accurate to describe it as non-directional, it is therefore thought of as bidirectional. This trait enables the model to understand a word's context depending on all of its surroundings (left and right of the word).

There are two variants of BERT:

- BERT Base: 12 layers (transformer blocks), 12 attention heads, and 110 million parameters.Comparable in size to the OpenAI Transformer in order to compare performance.
- BERT Large: 24 layers (transformer blocks), 16 attention heads and, 340 million parameters, a ridiculously huge model.

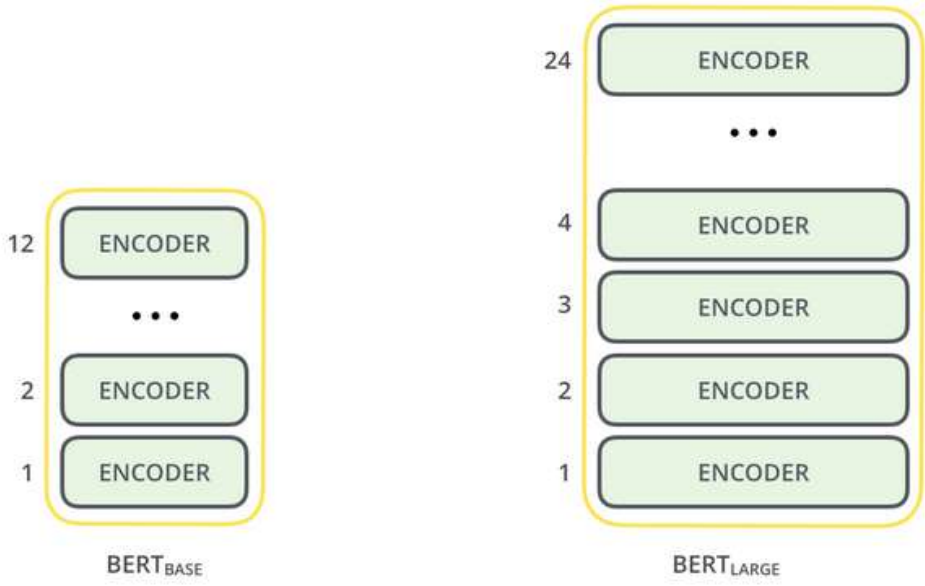

Figure 5.1: Architecture of BERT Base and bert Large model

BERT is basically a trained Transformer Encoder stack. Both BERT model sizes have a large number of encoder layers (which the paper calls Transformer Blocks) – twelve for the Base version, and twenty four for the Large version. These also have larger feedforward-networks (768 and 1024 hidden units respectively), and more attention heads (12 and 16 respectively) than the default configuration in the reference implementation of the Transformer in the initial paper (6 encoder layers, 512 hidden units, and 8 attention heads). The BERT Base architecture has the same model size as OpenAI's GPT for comparison purposes. All of these Transformer layers are Encoder-only blocks.

5.3 DistilBERT

DistilBERT is a small, fast, cheap and light transformer model based on the BERT architecture. It does not have a token-type embedding pooler and retains only half of the layers from Google's BERT. The knowledge distillation is performed during the pre-training phase, reducing the size of a BERT model by 40

```
Model: "tf_distil_bert_for_sequence_classification_1"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 distilbert (TFDistilBertMai  multiple                 66362880
 nLayer)

 pre_classifier (Dense)      multiple                  590592

 classifier (Dense)          multiple                  1538

 dropout_39 (Dropout)        multiple                  0

=================================================================
Total params: 66,955,010
Trainable params: 66,955,010
Non-trainable params: 0
_________________________________________________________________
None
```

Figure 5.2: Architecture of DistilBERT model

Chapter 6

RESULTS AND DISCUSSIONS

In this work, the fake news and real news were classified with five different ML algorithms such as Naive Bayes classifier, Logistic regression, Decision tree, Random forest and SVM, BERT, and DistilBERT. The experiments were done with four different datasets: Twitter, ISOT, LIAR and Kaggle.

6.1 EXPERIMENTATION USING MACHINE LEARNING ALGORITHMS

The confusion matrices of results obtained from various classifiers experimented on several datasets are shown from fig 6.1 to 6.20.

6.1.1 Experiment on Twitter dataset

The experiment conducted on Twitter dataset with different machine learning algorithms shows good accuracy.The accuracies obtained are: Naïve Bayes: 90.51%,Logistic Regression:93.3%,Decision Tree:87.41%,Random Forest:93.75 %,SVM: 94.03%.

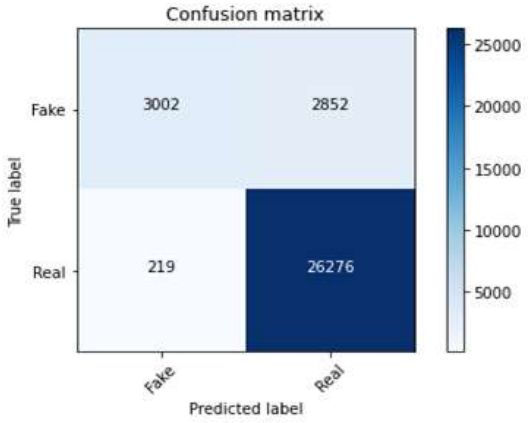

Figure 6.1: Confusion matrix of Naive Bayes

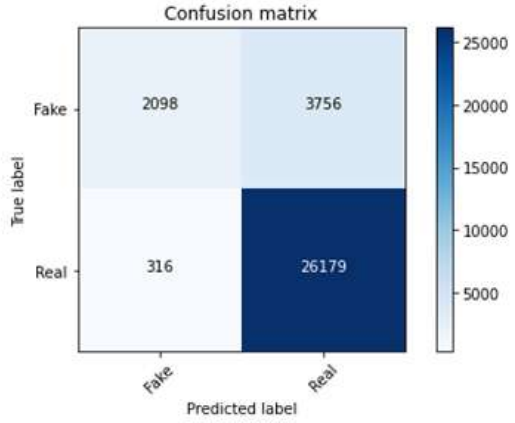

Figure 6.2: Confusion matrix of Decision tree

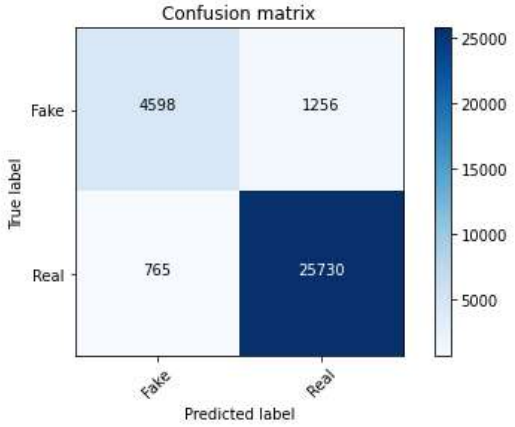

Figure 6.3: Confusion matrix of Random Forest

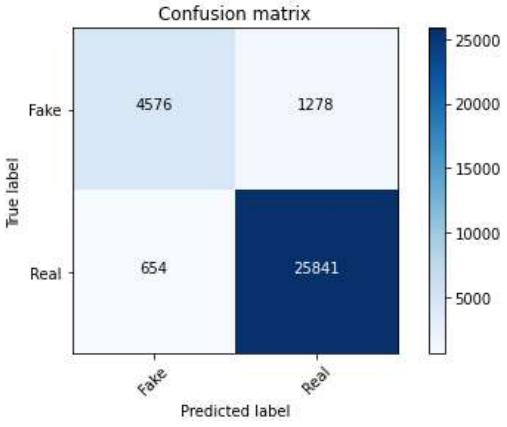

Figure 6.4: Confusion matrix of SVM

6.1.2 Experiment on ISOT dataset

The experiment conducted on ISOT dataset with different machine learning algorithms shows better accuracy.The accuracies obtained are: Naïve Bayes: 94.65%,Logistic Regres-

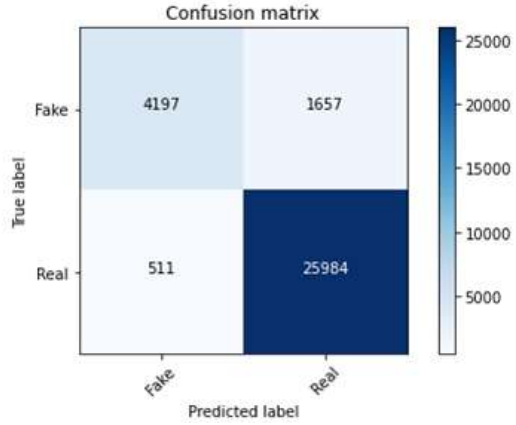

Figure 6.5: Confusion matrix of Logistic regression

sion:98.73%,Decision Tree:99.57%,Random Forest:99.21 %,SVM: 99.55%.

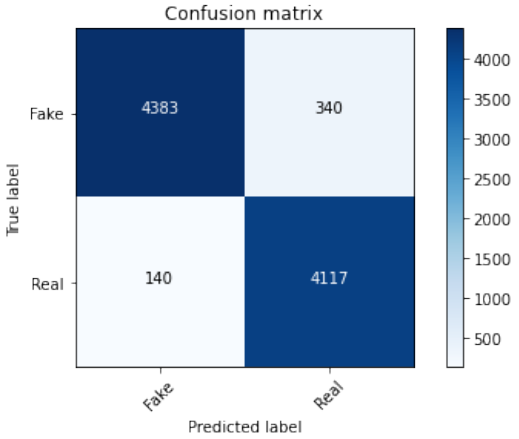

Figure 6.6: Confusion matrix of Naive Bayes

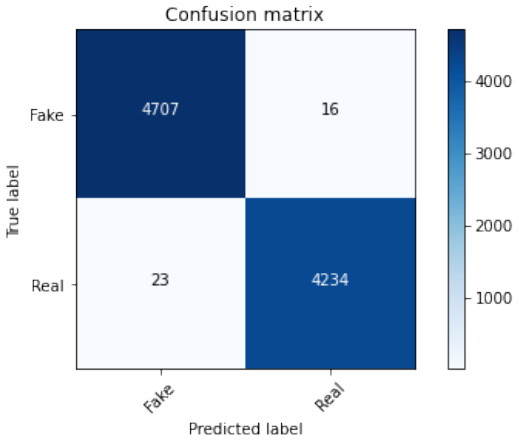

Figure 6.7: Confusion matrix of Decision tree

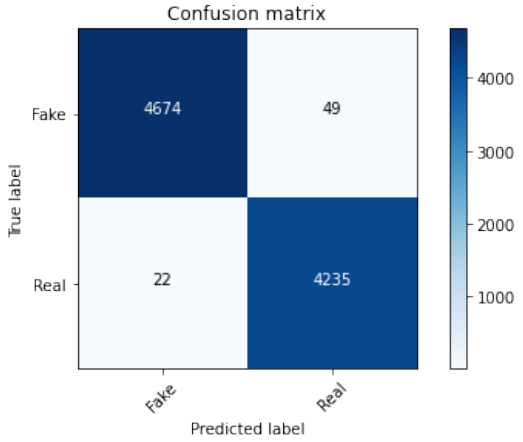

Figure 6.8: Confusion matrix of Random Forest

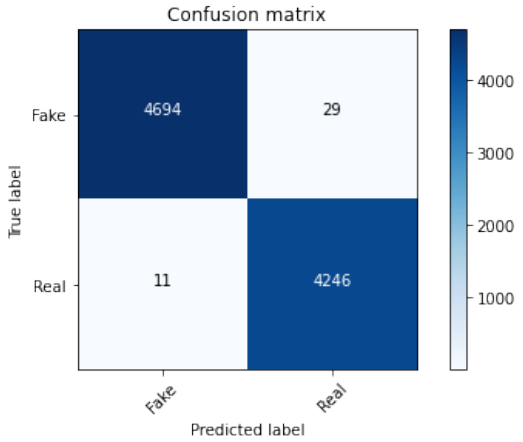

Figure 6.9: Confusion matrix of SVM

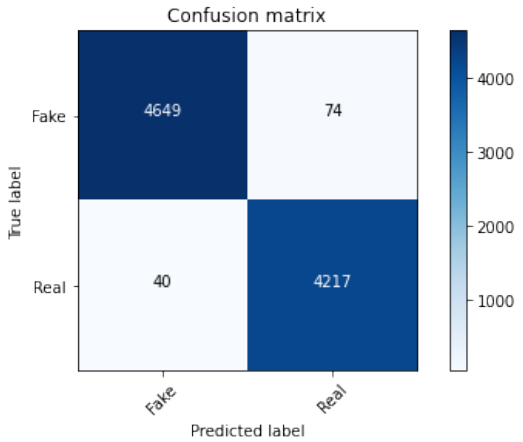

Figure 6.10: Confusion matrix of Logistic regression

6.1.3 Experiment on LIAR dataset

The experiment conducted on LIAR dataset with different machine learning algorithms shows lower accuracy.The accuracies obtained are: Naïve Bayes: 62.83%,Logistic Regression:58.09%,Decision Tree:61.48%,Random Forest:54.22 %,SVM: 58.25%.

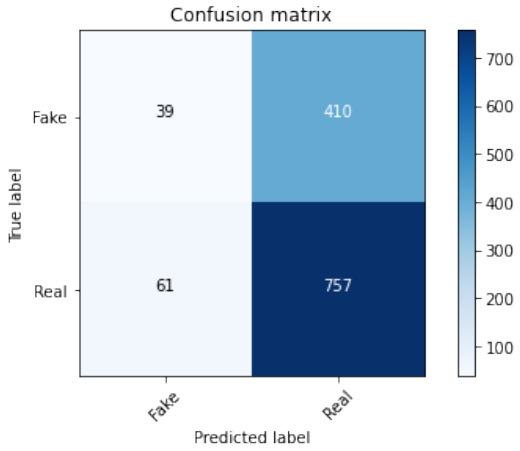

Figure 6.11: Confusion matrix of Naive Bayes

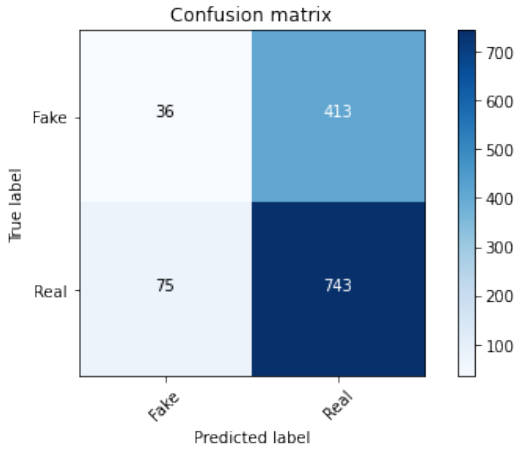

Figure 6.12: Confusion matrix of Decision tree

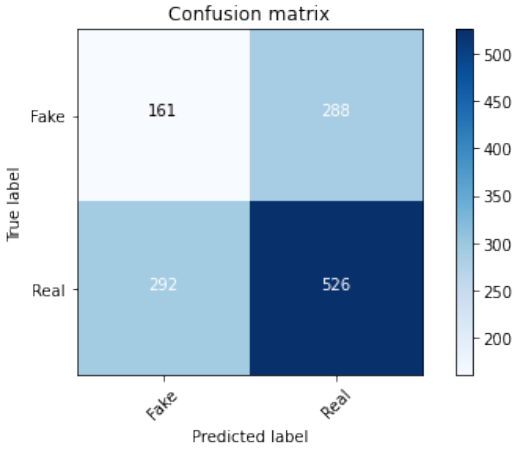

Figure 6.13: Confusion matrix of Random Forest

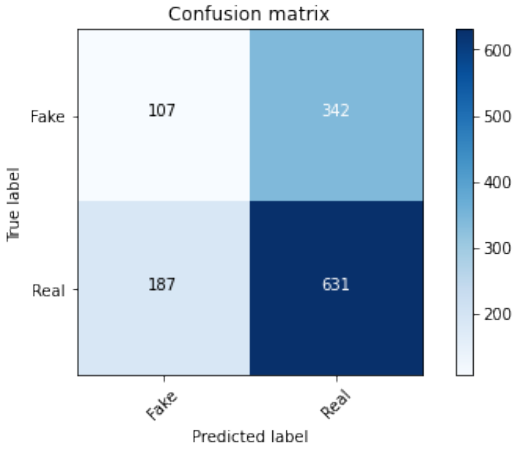

Figure 6.14: Confusion matrix of SVM

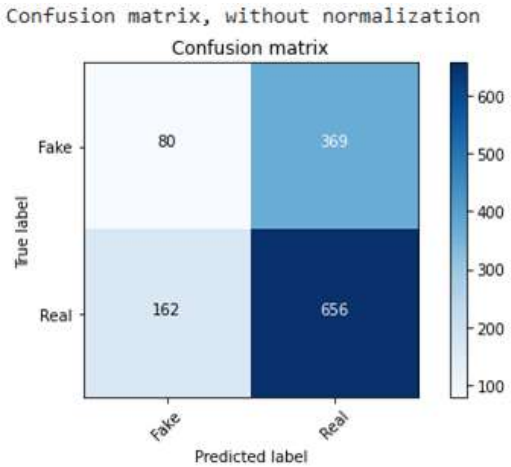

Figure 6.15: Confusion matrix of Logistic regression

6.1.4 Experiment on Kaggle dataset

The experiment conducted on Kaggle dataset with different machine learning algorithms shows lower accuracy.The accuracies obtained are: Naïve Bayes: 84.66%,Logistic Regression:94.87%,Decision Tree:89.36%,Random Forest:90.87%,SVM: 96.58%.

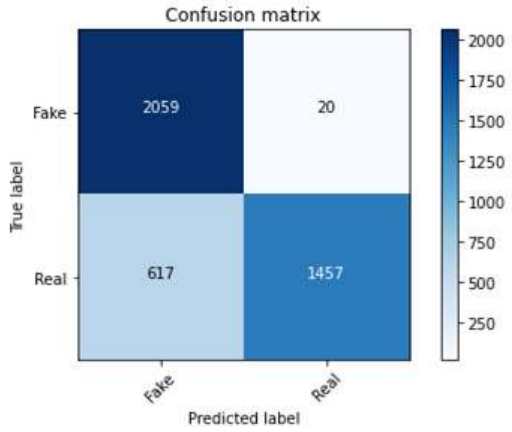

Figure 6.16: Confusion matrix of Naive Bayes

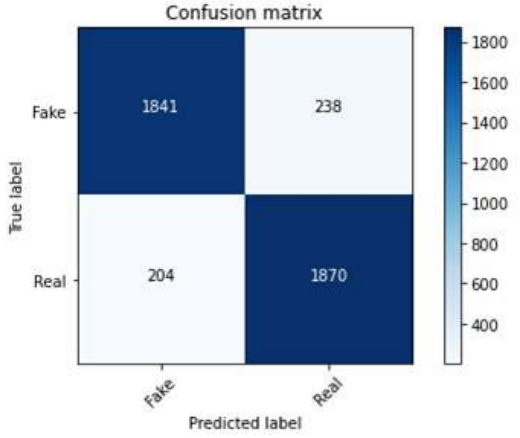

Figure 6.17: Confusion matrix of Decision tree

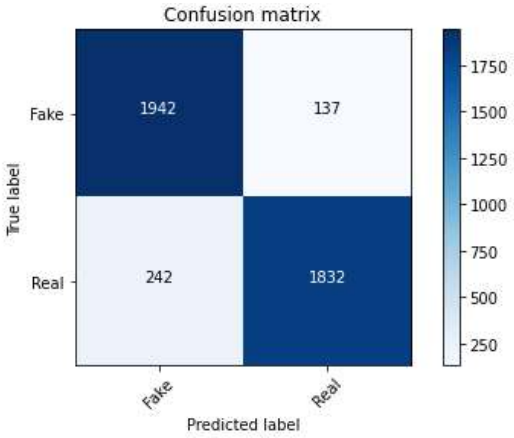

Figure 6.18: Confusion matrix of Random Forest

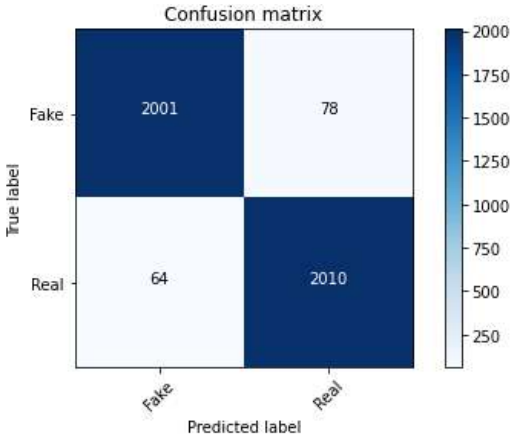

Figure 6.19: Confusion matrix of SVM

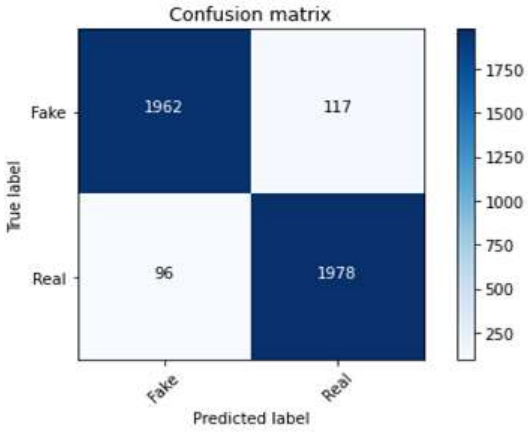

Figure 6.20: Confusion matrix of Logistic regression

6.2 EXPERIMENTATION USING BERT

The classification using BERT is done with four different datasets. First step is the data preprocessing technique. In this step, unwanted columns from the four datasets are removed.

Then the entire dataset cleaned by removing stop words and punctuation and converted all uppercase letters into lowercase. On Third step is Word cloud formation is performed. Word cloud is separately created for real and fake data on all datasets. Then the most frequent words on each dataset is found using Bigram, Trigram or simply by a bar graph technique. On the next step the entire data divided into two: one for training and one for testing. Then tokenize the sentences and converted tokenized dataset into torch dataset. After these processes imported fine-tuned BERT model from the pre-trained model for text classification. Here, BertForSequenceClassifications is used. The model is trained using train dataset and evaluated on the test dataset. Finally the fine-tuned model and tokenizer were saved. The analysis of the BERT model is done based on Precision, Recall, F1 score and accuracy on different datasets. At the same time, performance analysis of DistilBERT and ML algorithm is done through the accuracy measurement only.The output obtained in the experiment are shown from fig 6.21 to 6.28.

```
Out[48]: {'eval_loss': 0.13251514732837677,
          'eval_accuracy': 0.9485893092658202,
          'eval_f1': 0.9639672297542231,
          'eval_precision': 0.9595842956120092,
          'eval_recall': 0.9683903860160233,
          'eval_runtime': 98.2727,
          'eval_samples_per_second': 393.487,
          'eval_steps_per_second': 9.84,
          'epoch': 1.0}
```

Figure 6.21: Output of BERT classifier on Twitter Dataset during Validation

```
Attempted to log scalar metric eval_loss:
9.067665814654902e-05
Attempted to log scalar metric eval_accuracy:
1.0
Attempted to log scalar metric eval_f1:
1.0
Attempted to log scalar metric eval_precision:
1.0
Attempted to log scalar metric eval_recall:
1.0
Attempted to log scalar metric eval_runtime:
56.0933
Attempted to log scalar metric eval_samples_per_second:
160.09
Attempted to log scalar metric eval_steps_per_second:
4.011
Attempted to log scalar metric epoch:
1.0
```

Figure 6.22: Output of BERT classifier on ISOT Dataset during Validation

```
{'epoch': 1.0,
 'eval_accuracy': 0.9989062072737216,
 'eval_f1': 0.9987546699875467,
 'eval_loss': 0.0051108850166201 59,
 'eval_precision': 0.9975124378109452,
 'eval_recall': 1.0,
 'eval_runtime': 241.0123,
 'eval_samples_per_second': 15.174,
 'eval_steps_per_second': 0.759}
```

Figure 6.23: Output of BERT classifier on Kaggle Dataset during Validation

```
Out[40]: {'eval_loss': 0.6355127692222595,
          'eval_accuracy': 0.640625,
          'eval_f1': 0.7397454031117398,
          'eval_precision': 0.6908850726552179,
          'eval_recall': 0.7960426179604262,
          'eval_runtime': 5.2521,
          'eval_samples_per_second': 389.937,
          'eval_steps_per_second': 19.611,
          'epoch': 1.0}
```

Figure 6.24: Output of BERT classifier on LIAR Dataset during Validation

6.3 EXPERIMENTATION USING DistilBERT

The classification using DistilBERT is performed with three different datasets such as Twitter, ISOT, LIAR and Kaggle dataset. The primary step is the data preparation process. In this step unwanted columns are removed. Then the entire dataset is cleaned by removing stop words and punctuation and finally all uppercase letters are converted into lowercase. After that word cloud formation performed. Separate word cloud for real and fake data for the three datasets are created. Using Bigram, Trigram or simply by a bar graph the most frequent words in the dataset is found. Then the dataset separated into a train and test set. After that encoded with DistilBERT. Then found the maximum length of text in the dataset. In the tokenization technique, tokenize the sentences and then converted the labels and encodings to Tensorflow dataset. Here TFDistilBertForSequenceClassification model is used. It is fine tuned with native Tensorflow model. Trained and evaluated the model, finally obtained the output probabilities from the softmax layer. The output obtained from the DistilBERT experiments are shown from 6.29 to 6.36.

```
[ ] #Model Evaluation
    model.evaluate(test_dataset.shuffle(len(X_test)).batch(BATCH_SIZE), return_dict=True, batch_size=BATCH_SIZE)

    4044/4044 [==============================] - 790s 195ms/step - loss: 0.0989 - accuracy: 0.9674
    {'accuracy': 0.9674023985862732, 'loss': 0.09888310730457306}
```

Figure 6.25: Output of DistilBERT classifier on Twitter Dataset during Validation

```
In [32]: #Model Evaluation
         model.evaluate(test_dataset.shuffle(len(X_test)).batch(BATCH_SIZE), return_dict=True, batch_size=BATCH_SIZE)

         842/842 [==============================] - 96s 112ms/step - loss: 3.2334e-04 - accuracy: 0.9999

Out[32]: {'loss': 0.0003233425668440759, 'accuracy': 0.9998515248298645}
```

Figure 6.26: Output of DistilBERT classifier on ISOT Dataset during Validation

```
[ ] #Model Evaluation
    model.evaluate(test_dataset.shuffle(len(X_test)).batch(BATCH_SIZE), return_dict=True, batch_size=BATCH_SIZE)

    1151/1151 [==============================] - 100s 86ms/step - loss: 1.6099 - accuracy: 0.6018
    {'accuracy': 0.60178142786026, 'loss': 1.6099258661270142}
```

Figure 6.27: Output of DistilBERT classifier on LIAR Dataset during Validation

```
with tf.device('GPU:1'):
    #Model Evaluation
    model.evaluate(test_dataset.shuffle(len(X_test)).batch(BATCH_SIZE), return_dict=True, batc

390/390 [==============================] - 45s 113ms/step - loss: 0.0643 - accuracy: 0.9905
```

Figure 6.28: Output of DistilBERT classifier on Kaggle Dataset during Validation

6.4 PERFORMANCE ANALYSIS

Performance analysis of ML algorithms on different datasets are obtained by generating confusion matrix and accuracy. The analysis of BERT model is done in the basis of Precision, Recall, F1 score and accuracy on different datasets.While performance analysis of DistilBERT is done through the accuracy measurement only.

6.4.1 Performance analysis table of ML algorithms

Model	Accuracy (%)			
	LIAR Dataset	**ISOT Dataset**	**Twitter Dataset**	**Kaggle Dataset**
Logistic regression	58.09	98.73	93.3	94.87
Decision Tree	61.48	**99.57**	87.41	89.36
Random Forest	54.22	99.21	93.75	90.87
Naïve Bayes	**62.83**	94.65	90.51	84.66
SVM	58.25	99.55	**94.03**	**96.58**

Table 6.1: Performance analysis of Machine Learning Algorithms

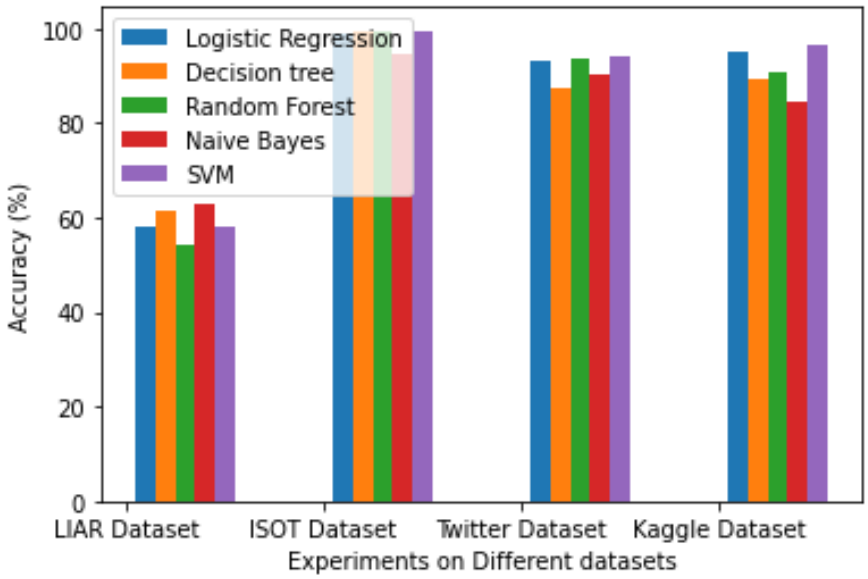

Figure 6.29: Comparative analysis of different datasets using ML Algorithms

Table 6.1 shows the experimental results of machine learning algorithms in different datasets and fig 6.37 shows the graphical representation. From the experiment study it is clear that for ISOT dataset Decision Tree gives the maximum accuracy and Naïve Bayes classifier gives the minimum accuracy.For LIAR dataset Naïve Bayes classifier gives the maximum accuracy and Random forest gives the minimum accuracy.ISOT dataset gives more accurate prediction than LIAR dataset.For Twitter dataset SVM has highest accuracy and Decision tree has lowest accuracy. In the case of Kaggle dataset SVM also have highest

accuracy and Naive Bayes has the lowest accuracy. In overall the LIAR dataset shows very poor accuracy than others datasets in all classifiers.

6.4.2 Performance analysis table of BERT

Training

Dataset	**Accuracy**	**F1-score**	**Precision**	**Recall**
Twitter	**0.94812**	**0.9634**	**0.9648**	**0.9620**
ISOT	**1.000**	**1.000**	**1.000**	**1.000**
LIAR	**0.6416**	**0.7816**	**0.7816**	**1.000**
Kaggle	**0.9989**	**0.9985**	**0.9975**	**1.000**

Table 6.2: Performance analysis of BERT Model during Training

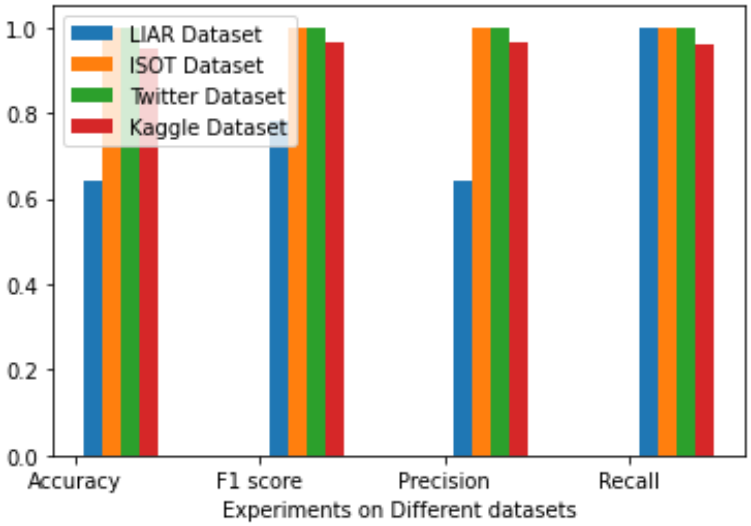

Figure 6.30: Comparative analysis of different datasets on BERT(Training)

Validation

Dataset	**Accuracy**	**F1-score**	**Precision**	**Recall**
Twitter	**0.948593**	**0.9637**	**0.959584**	**0.96839**
ISOT	**1.000**	**1.000**	**1.000**	**1.000**
LIAR	**0.6402**	**0.7397**	**0.6908**	**0.7960**
Kaggle	**0.998906**	**0.9987**	**0.9975**	**1.000**

Table 6.3: Performance analysis of BERT Model during Validation

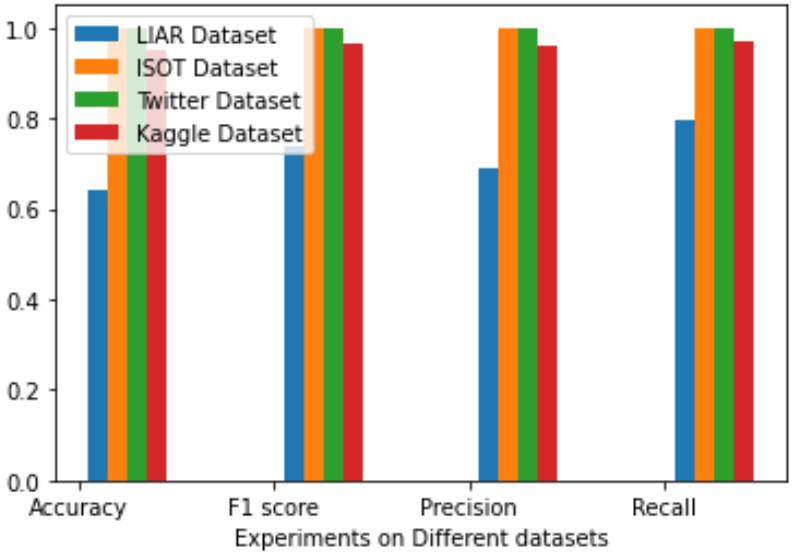

Figure 6.31: Comparative analysis of different datasets on BERT(Validation)

In the case of BERT, evaluates the accuracy,f1-score, precision and recall.Table 6.2 and 6.3 represents the experimental result of BERT model during training and validation stage respectively. Fig 6.38 and 6.39 shows corresponding graphical representations. The experimental results given by the classification using BERT model gives a better performance on ISOT, Kaggle and twitter dataset but the performance of LIAR dataset is lesser during validation.

6.4.3 Performance analysis table of DistilBERT

Dataset	**Training Accuracy**	**Testing Accuracy**
Twitter	**0.9845**	**0.9074**
ISOT	**1.000**	**0.9998**
LIAR	**0.9287**	**0.7868**
Kaggle	**0.9993**	**0.9905**

Table 6.4: Performance analysis of DistilBERT Model during Training and validation process

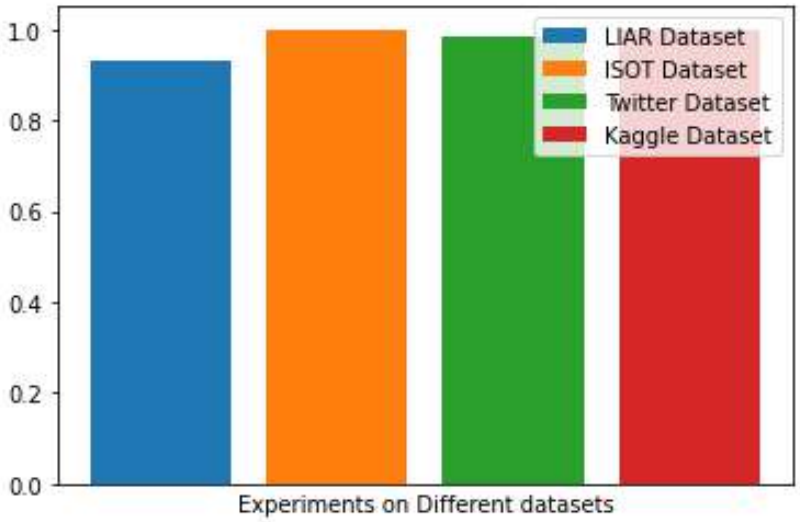

Figure 6.32: Comparative analysis of different datasets on DistilBERT (Training)

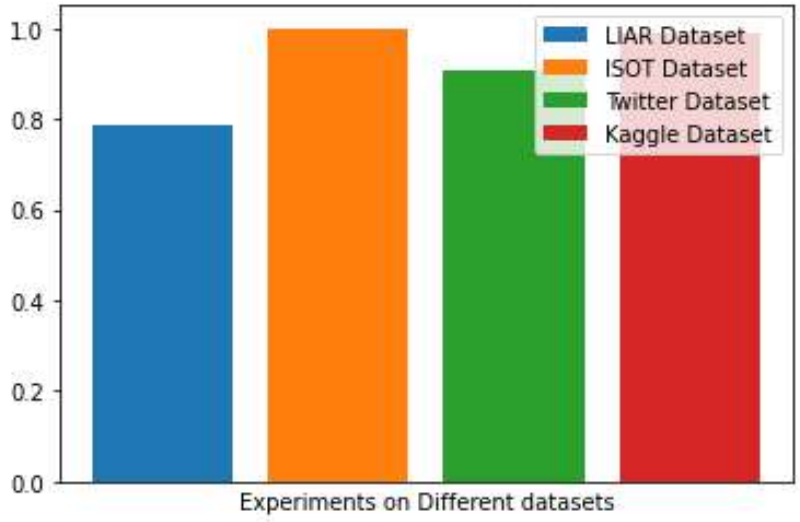

Figure 6.33: Comparative analysis of different datasets on DistilBERT (Validation)

In the case of DistilBERT only evaluates the accuracy.Table 6.4 represents the experimental result of BERT model during training and validation stage respectively. Fig 6.40 and 6.41 shows corresponding graphical representations.The experimental results given by the classification using DistilBERT model gives a better performance on the twitter, ISOT and Kaggle dataset but the performance of LIAR dataset is much lower during validation as same as BERT Model.

Model	Accuracy (%)			
	LIAR Dataset	**ISOT Dataset**	**Twitter Dataset**	**Kaggle Dataset**
Logistic regression	58.09	98.73	93.3	94.87
Decision Tree	61.48	99.57	87.41	89.36
Random Forest	54.22	99.21	93.75	90.87
Naïve Bayes	62.83	94.65	90.51	84.66
SVM	58.25	99.55	94.03	96.58
BERT	64.02	**100**	**94.85**	**99.89**
DistilBERT	**78.68**	99.98	90.04	99.05

Table 6.5: Accuracy comparison of all models

Table 6.5 shows the overall analysis of the experiment. From the table, it is clear that BERT and DistilBERT give the maximum accuracy for all datasets. For the LIAR dataset, DistilBERT gives the maximum accuracy of 78.68; for Twitter, ISOT and Kaggle datasets BERT gives the maximum accuracy of 94.85, 100 and 99.89, respectively.

Work	Technique	Accuracy (%)
Jiang et al. [25]	Stacking method (LSTM, CNN, SVM,LR,DT,RF,GRU and KNN)	99.94
Dixit et.al [26]	Flight-based LSTM with PPCA	99
Proposed work	**BERT**	**100**

Table 6.6: Comparison of the proposed technique with the existing technique (ISOT Dataset)

Work	Technique	Accuracy (%)
Saqib et.al [27]	Ensemble Machine Learning algorithms (Decision Tree, Random Forest and Extra tree classifier with feature extraction)	44.15
Rjalaxmi et.al [28]	Optimized LSTM	45.23
Proposed work	**DistilBERT**	**78.68**

Table 6.7: Comparison of the proposed technique with the existing technique (LIAR Dataset)

Work	Technique	Accuracy (%)
Kaliyar et al. [12]	Single-layer CNNs with BERT	98.90
Aman et.al [29]	CNN with Adam optimizer	94.71
Proposed work	**BERT**	**99.89**

Table 6.8: Comparison of the proposed technique with the existing technique (Kaggle Dataset)

Twitter, ISOT, LIAR and Kaggle datasets are considered for the experiment. Table 6.6, 6.7 and 6.8 shows the comparative analysis of the proposed work with the existing work on ISOT, LIAR and Kaggle datasets. In this work, the Twitter dataset is used for an experimental study. From the table, it is clear that the proposed model is an efficient model for fake news classification. Compared with CNN, LSTM and other machine learning models, BERT and DistilBERT are more efficient in handling text data. Because of BERT, the bidirectional transformer model is designed as a masked language model for analyzing various natural language processing tasks.

Chapter 7

CONCLUSION AND FUTURE WORKS

Automated fake news detection is essential since many people propagate false news on social media to deceive the public. It is vital to detect false news to protect individuals or organizations from losing their reputations. The experimentations of fake news detection were conducted on twitter, ISOT, LIAR and Kaggle datasets, mainly using BERT and DistilBERT, as well as a comparative study of different machine learning algorithms such as Naive Bays, Random Forest, Decision Tree, Logistic Regression, and Support Vector Machine (SVM). Observing the results after experimentation, it is found that BERT and DistilBERT can be employed as a model for fake news identification on multi-domain. The pre-training and fine tuning in BERT and DistilBERT improved total performance. The BERT model obtained an accuracy of 94.8%, 100%, 99.89% and 64.02% on the Twitter, ISOT, Kaggle and the LIAR datasets respectively; DistilBERT obtained an accuracy of 90.74%, 99.98%, 99.05% and 78.68% on the Twitter, ISOT, Kaggle and the LIAR datasets respectively. Although the experimentation focused solely on text analysis, the source of the news is crucial in disseminating bogus or not. That is because the likelihood of a fraudulent source making fake news is very high; adding source information in addition to text analysis would improve the proposed model's real-time prediction. We can expand this work to Multimodal analysis (text + photos + voice) in the future because many people prefer to send photographs rather than text. Fake news identification suffers numerous difficulties. The performance of detection depends on the quality of data, that can't be guaranteed from the social media platforms. Also, measuring the fakeness from the text which is available in multilingual and mixed languages is a challenging task.

References

[1] De Oliveira, Nicollas R., Dianne SV Medeiros, and Diogo MF Mattos. "A sensitive stylistic approach to identify fake news on social networking." IEEE Signal Processing Letters 27: 1250-1254(2020).

[2] Ali, Hassan, Muhammad Suleman Khan, Amer AlGhadhban, Meshari Alazmi, Ahmad Alzamil, Khaled AlUtaibi, and Junaid Qadir. "All Your Fake Detector Are Belong to Us: Evaluating Adversarial Robustness of Fake-news Detectors Under Black-Box Settings." (2021).

[3] Ngwainmbi, Emmanuel K. "Fake News Reporting on Social Media Platforms and Implications for Nation-State Building." In Media in the Global Context, pp. 95-123. Palgrave Macmillan, Cham, 2019.

[4] Himma-Kadakas, Marju. "Alternative facts and fake news entering journalistic content production cycle." Cosmopolitan Civil Societies: An Interdisciplinary Journal 9, no. 2 (2017): 25-41.

[5] Pollicino, Oreste, and Elettra Bietti. "Truth and deception across the Atlantic: a roadmap of disinformation in the US and Europe." Italian J. Pub. L. 11 (2019): 43.

[6] Fleming, Jennifer, and Christopher Karadjov. "Focusing on Facts: Media and News Literacy Education in the Age of Misinformation." In Media Literacy in a Disruptive Media Environment, pp. 77-93. Routledge, 2020.

[7] Pourghomi, Pardis, Fadi Safieddine, Wassim Masri, and Milan Dordevic. "How to stop spread of misinformation on social media: Facebook plans vs. right-click authenticate approach." In 2017 International Conference on Engineering MIS (ICEMIS), pp. 1-8. IEEE, 2017.

[8] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

[9] Koshiyama, Adriano, Nick Firoozye, and Philip Treleaven. "Algorithms in future capital markets: a survey on AI, ML and associated algorithms in capital markets." In Proceedings of the First ACM International Conference on AI in Finance, pp. 1-8. 2020.

[10] Kula, Sebastian, Rafa l Kozik, and Micha l Chora´s. "Implementation of the BERT-derived architectures to tackle disinformation challenges." Neural Computing and Applications (2021): 1-13.

[11] Tuan, Nguyen Manh Duc, and Pham Quang Nhat Minh. "Multimodal Fusion with BERT and Attention Mechanism for Fake News Detection." arXiv preprint arXiv:2104.11476 (2021).

[12] Kaliyar, Rohit Kumar, Anurag Goswami, and Pratik Narang. "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach." Multimedia Tools and Applications 80, no. 8 (2021): 11765-11788.

[13] Shishah, Wesam. "Fake News Detection Using BERT Model with Joint Learning." Arabian Journal for Science and Engineering (2021): 1-13.

[14] Mehta, Divyam, Aniket Dwivedi, Arunabha Patra, and M. Anand Kumar. "A transformer- based architecture for fake news classification." Social Network Analysis and Mining 11, no. 1 (2021): 1-12.

[15] Briskilal, J., and C. N. Subalalitha. "An ensemble model for classifying idioms and literal texts using BERT and RoBERTa." Information Processing Management 59, no. 1 (2022): 102756.

[16] Trueman, Tina Esther, Ashok Kumar, P. Narayanasamy, and J. Vidya. "Attention-based C- BiLSTM for fake news detection." Applied Soft Computing 110 (2021): 107600.

[17] Cui, Baiyun, Yingming Li, Ming Chen, and Zhongfei Zhang. "Fine-tune BERT with sparse self-attention mechanism." In Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP), pp. 3548-3553. 2019.

[18] Umer, Muhammad, Zainab Imtiaz, Saleem Ullah, Arif Mehmood, Gyu Sang Choi, and Byung-Won On. "Fake news stance detection using deep learning architecture (CNNLSTM)." IEEE Access 8 (2020): 156695-156706.

[19] Kumar, Sachin, Rohan Asthana, Shashwat Upadhyay, Nidhi Upreti, and Mohammad Akbar. "Fake news detection using deep learning models: A novel approach." Transactions on Emerging Telecommunications Technologies 31, no. 2 (2020): e3767.

[20] Hakak, Saqib, Mamoun Alazab, Suleman Khan, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, and Wazir Zada Khan. "An ensemble machine learning approach through effective feature extraction to classify fake news." Future Generation Computer Systems 117 (2021): 47-58.

[21] Wang, William Yang. "" liar, liar pants on fire": A new benchmark dataset for fake news detection." arXiv preprint arXiv:1705.00648 (2017).

[22] Gundapu, Sunil, and Radhika Mamidi. "Transformer based automatic COVID-19 fake news detection system." arXiv preprint arXiv:2101.00180 (2021).

[23] Singh, Vivek, Rupanjal Dasgupta, Darshan Sonagra, Karthik Raman, and Isha Ghosh. "Automated fake news detection using linguistic analysis and machine learning." In International conference on social computing, behavioral-cultural modeling, prediction and behavior representation in modeling and simulation (SBP-BRiMS), pp. 1-3. 2017.

[24] Heimerl, Florian, Steffen Lohmann, Simon Lange, and Thomas Ertl. "Word cloud explorer: Text analytics based on word clouds." In 2014 47th Hawaii international conference on system sciences, pp. 1833-1842. IEEE, 2014

[25] Jiang, T. A. O., Jian Ping Li, Amin Ul Haq, Abdus Saboor, and Amjad Ali. "A novel stacking approach for accurate detection of fake news." IEEE Access 9 (2021): 22626-22639.

[26] Dixit, Dheeraj Kumar, Amit Bhagat, and Dharmendra Dangi. "Automating fake news detection using PPCA and levy flight-based LSTM." Soft Computing (2022): 1-13.

[27] Hakak, Saqib, Mamoun Alazab, Suleman Khan, Thippa Reddy Gadekallu, Praveen Kumar Reddy Maddikunta, and Wazir Zada Khan. "An ensemble machine learning approach through effective feature extraction to classify fake news." Future Generation Computer Systems 117 (2021): 47-58.

[28] Rajalaxmi, R. R., L. V. Narasimha Prasad, B. Janakiramaiah, C. S. Pavankumar, N. Neelima, and V. E. Sathishkumar. "Optimizing Hyperparameters and Performance Analysis of LSTM Model in Detecting Fake News on Social media." Transactions on Asian and Low-Resource Language Information Processing (2022).

[29] Agarwal, Aman, Mamta Mittal, Akshat Pathak, and Lalit Mohan Goyal. "Fake news detection using a blend of neural networks: An application of deep learning." SN Computer Science 1, no. 3 (2020): 1-9.

[30] Verma, Pawan Kumar, Prateek Agrawal, Ivone Amorim, and Radu Prodan. "WELFake: word embedding over linguistic features for fake news detection." IEEE Transactions on Computational Social Systems 8, no. 4 (2021): 881-893.

[31] Ahmad, Iftikhar, Muhammad Yousaf, Suhail Yousaf, and Muhammad Ovais Ahmad. "Fake news detection using machine learning ensemble methods." Complexity 2020 (2020).