FAKE NEWS DETECTION USING BERT

Report of Internship

TATA CONSULTANCY SERVICES (TCS)

15-Nov-2021 to 27-May-2022

Submitted by

Mrs. SREETHU P R

Intern Emp ID: 2170603

Final-year Postgraduate student

Centre for Artificial Intelligence

THANGAL KUNJU MUSALIAR COLLEGE OF ENGINEERING KERALA

Under the guidance of

RAJEEV AZHUVATH (TCS Mentor ID: 120914)

&

Prof. SUMOD SUNDAR (TKMCE)

MAY 2022

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 to make my project a successful one. I thank him for steering me to the shore of fulfilment under his protective wings.

With a profound sense of gratitude, I would like to express my heartfelt thanks to my mentor **Rajeev Azhuvath, Tata Consultancy Services (TCS)**, for his expert guidance, constant support and cooperation.

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 immense encouragement. I would like to thank my college guide **Prof. Sumod Sundar**, Assistant Professor, Centre for Artificial Intelligence, TKMCE, for his expert guidance and cooperation.

I also express my thanks to my loving parents, brother 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 is increasing day by day and dissemination of data is rapidly raised.The rise of social networks has accelerated the dissemination of rumours, satires, and false information, increase in the distribution of fake news. Fake news can be broken down into three categories: firstly, the news that is entirely fake, with the purpose to fool the reader by getting him to get confused; secondly are the rumours, which are the information those with ambiguous truth but public acceptance; and the third is those with a witty individual who creates parodies with sarcasm and irony as well as satires. So the identification of such news is real or fake is an important task in the digital life. The fake news may be on different domains such as political domain, entertainment domain, sports domain etc. Spreading of this type of news causes so many problems in our society. One such example is the 2016 presidential election in the United States, where 37 million Facebook users believed and shared fake news created for personal advantage.The false information might be used to damage countries' economies, weaken people's trust in their governments, or promote a specific product to make huge profits.So that early detection of the fake news reduces the impact. The detection of fake news is increasing a serious concern for the news industry and journalists, and tools for detecting fake news have become critical. Various studies regarding the machine learning and deep learning algorithms are found in the literature. All these algorithms are focusing mainly on feature extraction. Therefore, extraction of relevant features is an important task for effective classification task. Generalising a learning model by identifying patterns in a text will help to differentiate fake news from real one. Fake news detection using BERT and LSTM techniques are the most competitive study happened now.Here I do fake news detection using BERT, DistilBERT and some machine learning algorithms such as Naive bayes, decision tree,random forest,SVM and Logistic regression, also do a comparison study with four set of datasets such as Twitter dataset,LIAR dataset,Kaggle dataset and ISOT dataset.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Fake news is information that is incorrect or misleading and is presented as news. Fake news is frequently published with the intent of harming a person's or entity's reputation or profiting from advertising income.When spectacular media accounts were widespread in the 1890s, the phrase was coined.However, the phrase has become increasingly widely used to refer to any type of misleading information, including inadvertent and unconscious mechanisms, as well as by high-profile individuals to refer to any news that is unfavourable to their personal viewpoints.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 that are misunderstood as genuine, as well as items with sensationalist or clickbait headlines that aren't supported by the text. Because of the wide variety of fake news, current scholars have begun to choose the term "information disorder" as a more neutral and informative word.With the rise of social media, particularly the Facebook News Feed, the incidence of fake news has increased, and this disinformation has progressively filtered into the mainstream media. Political polarisation, 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 a BuzzFeed investigation, top fake news stories concerning the 2016 US presidential election garnered more Facebook engagement than top stories from major news outlets. It also has the potential to erode public trust in serious news coverage. Former US President Donald Trump is credited with popularising the word by using it to describe any negative media coverage of himself. Because to 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."

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

- Satire or parody ("no intention to cause harm but has potential to fool")
- False connection ("when headlines, visuals or captions don't support the content")
- Misleading content ("misleading use of information to frame an issue or an individual")
- False context ("when genuine content is shared with false contextual information")

- Impostor content ("when genuine sources are impersonated" with false, made-up sources)
- Manipulated content ("when genuine information or imagery is manipulated to deceive", as with a "doctored" photo)
- Fabricated content ("new content is 100% false, designed to deceive and do harm")

Scientific denialism, defined as the act of generating incorrect or misleading information to unintentionally maintain strong pre-existing ideas, is another possible explanatory category of fake news.
The popularity and usage of online media platforms is increasing day by day and dissemination of data is rapidly raised. Fake news has existed for a long time. Indeed, this problem has been addressed since its emergence, which has resulted in negative consequences in both the technological and political realms.So the identification of fake news spread over these platforms should be dealt with more attention.Facebook has previously taken efforts to counteract the spread of misinformation on its website (in certain countries) by collaborating with third-party fact-checkers to analyse and score the veracity of articles and postings on the social media platform.Artificial intelligence methods help to identify the spread of fake news in these online platforms through deep automated techniques. Deep learning algorithms such as CNN, BERT, LSTM together with other machine learning techniques are used for the effective detection of fake news with good accuracy. These techniques help to explore more features that can be used for analysing fake data from real content.Here do a analysis of fake news detection using BERT.BERT is an acronym for Bidirectional Encoder Representations from Transformers. It is intended to condition both left and right context to pre-train deep bidirectional representations from unlabeled text. 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. The Transformer architecture underpins BERT.BERT has been pre-trained on a large corpus of unlabeled text, which includes the entire Wikipedia (2,500 million words!) and the Book Corpus (800 million words).Half of BERT's success is due to this pre-training stage. This is because as a model is trained on a huge text corpus, it begins to pick up on deeper and more intimate understandings of how the language works. This information serves as a swiss army knife for practically any NLP work.BERT is a model that is "deeply bidirectional." Bidirectional indicates that during the training phase, BERT learns information from both the left and right sides of a token's context. A model's bidirectionality is critical for completely comprehending the meaning of a language.We may fine-tune it by adding a few more output layers to construct cutting-edge models for a number of NLP problems.

Rest of this paper is organized as 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. The experimental results on various classification algorithms are discussed in Section 4.

Chapter 2

Literature Review

In this section, several studies of fake news detection utilizing both machine learning and deep learning techniques are discussed.
Sebastian Kula et. al [1] proposed a fake news detection model derived from the BERT (Bidirectional Encoder Representations from Transformers) architecture.Here uses various types of pre-trained embeddings of the BERT for word embeddings and on the RNN or on the BERT/ RoBERTa pooler for document embeddings. In this paper, training was performed on data classified as true and false. Firstly, models for detecting fake news in article titles were created, followed by the models for detecting fake news in articles' contents. This technique gives n F1 score of 0.9612, this shows the reliability.
Nguyen Manh Duc Tuan et al.[2] proposed a multi-modal fusion with BERT and attention mechanism for Fake News Detection. It is a A new multimodal approach for detecting fake news has been developed. They obtain feature representations from many modalities using neural networks. The attention mechanism is utilised to combine multimodal features, which are then placed in a sigmoid layer for classification. They employ the BERTweet model to extract feature representations from sentences, and a VGG-19 network to extract feature representations from visuals. They suggest a scaled dot-product attention mechanism for both texts and images, as well as a selfattention mechanism for images, because they believe that all components of images are related in nonphotoshopped images. To improve the accuracy of fake news detection, textual and visual representations, as well as three attention outputs, are integrated.
Rohit Kumar Kaliyar et al. [3] proposed a method for Fake news detection in social media with a BERT-based deep learning approach. For better learning, the proposed model combines BERT and three parallel blocks of 1d-CNN with varying kernel-sized convolutional layers and distinct filters. Their model is based on a pre-trained bidirectional transformer encoder word embedding model (BERT).They use BERT as a sentence encoder, which can accurately extract a sentence's context representation for the detection of fake news, a deep neural network with a bidirectional training technique could be the best and most accurate solution. With the powerful capacity to capture semantic and long-distance relationships in phrases, the suggested method increases the performance of fake news identification. The findings of the classification show that FakeBERT produces more accurate results, with an accuracy of 98.90 %.
Wesam Shishah [4] presented Fake News Detection Using BERT Model with Joint Learning. For detecting fake news in articles, a novel BERT with a combined learning-based model is

presented. The proposed method can detect bogus news in both lengthy and short pieces. Rather of presenting sequences utilising the first hidden states of BERT, all hidden states with dynamic range attention mechanisms are used to compute weights. To improve generalisation, RFC and NER task models are combined with BERT via a common parameter layer in collaborative learning. 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.

Divyam Mehta et al. [5] proposed a transformer-based architecture for fake news classifcation.They discuss and address the many aspects connected to transfer learning in the suggested model, as well as present an architecture to classify bogus news. Approaches that focus on text classification utilising 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 a range of NLP tasks. BERT is the first deeply bidirectional, unsupervised language representation and is based on ELMo.

J Briskilal et al. [6] proposed an ensemble model for classifying idioms and literal texts using BERT and RoBERTa . An idiom is a phrase whose true meaning differs from the one delivered. Rule-based generalisation is utilised 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 have used ensemble pre-trained models like BERT and RoBERTa. The goal of this work is to use an ensemble method to accurately categorise idiomatic and literal sentences.

Tina Esther Trueman et al. [7] proposed an attention-based C-BiLSTM for fake news detection. Deep learning approaches have the advantage of automatically recognising features. These methods determine the meaning of a word while taking into account its context. Attention mechanisms, in particular, have emerged as one of the most potent strategies in natural language processing. They're generally utilised in conjunction with recurrent neural networks to anticipate the most important information in a succession of inputs. This work tackles the topic of detecting bogus news in a multi-class context. Improves the accuracy of fake news detection by combining attention processes with convolutional bidirectional recurrent neural networks. .

Muhammad Umar et al. [8] proposed Fake News Stance Detection Using Deep Learning Architecture such as CNN-LSTM. They suggested a technique that automatically classifies news stories as agree, disagree, unrelated, or discuss 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.

Sachin Kumar et al. [9] proposed a model for Fake news detection using deep learning models. As part of their investigation, they carefully selected 7 models for sentiment categorization, which include versions of the convolutional neural network (CNN) and long short-term.memory (LSTM) architectures. CNN models are frequently used for image classification and detection, as well as 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 its variation, bidirectional LSTMs, to filter out these difficulties using RNNs. MaxPooling, the most used pooling approach, is employed in this network for pooling.It's 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, use the Leaky ReLU activation function because when using the ReLU activation function, negative values becoming zero immediately reduced the accuracy of the model as well as its ability to fit or train from the data properly. They use 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 and limitations/future 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	TFaster training of model and lower cross-entropy loss. It is useful to handle large-scale structure as well as unstructured text. It effectively addresses ambiguity.	Not detect the instances 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

Chapter 3

Methodology

3.1 Proposed Architecture

The proposed system is used to classify the news as fake or real using different machine learning techniques, DistilBERT and BERT . The entire system can be divided into six: Data pre-processing,Word cloud formation, finding of most frequent word, selection of algorithm, training and testing data and finally classification and performance analysis.

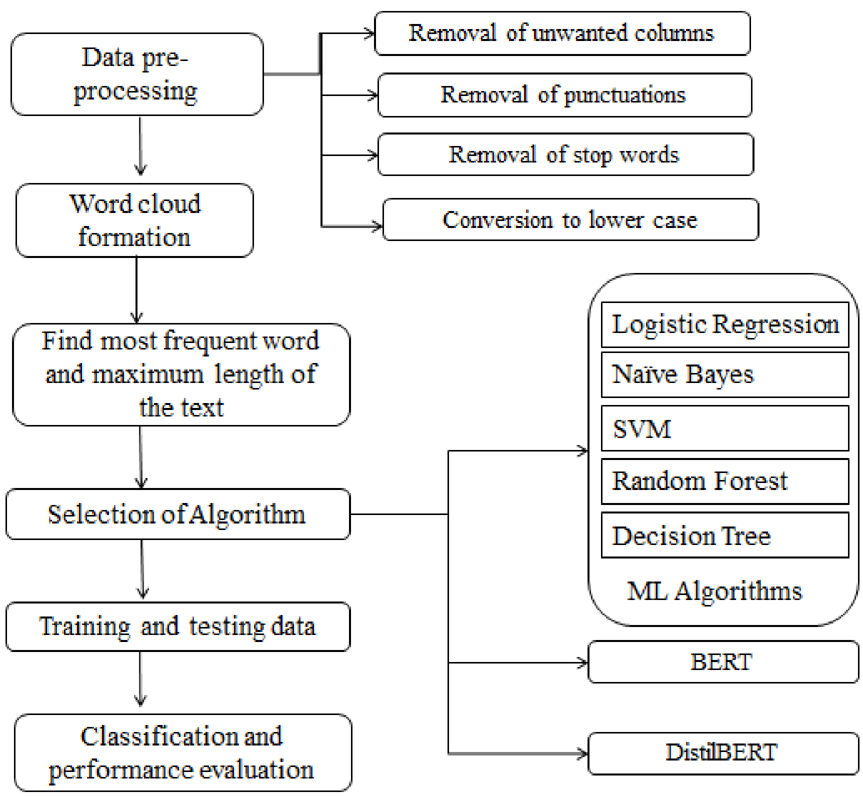

Figure 3.1: System Architecture

The dataset contains text data and they are may not be in same format. So that data preprocessing place a main role in the proposed methadology. Removal of unwanted columns,removal of punctuation, removal of stop words and conversion to lower case are the pre-processing steps done in the input dataset.Here I takes four sets of dataset from that three dataset needs preprocessing.After these process all the text data becomes in same format. So the analysis becomes more easier. Then the next step is word cloud formation.Word Clouds are visual displays of text data – simple text analysis. Word Clouds display the most prominent or frequent words in a body of text (such as a State of the Union Address). Typically, a Word Cloud will ignore the most common words in the language ("a", "an", "the" etc).The remaining words are displayed in a "cloud" with the font size of the word (and-or the coloring of the characters in the word) depicting the relative frequency of occurrence of each target word in the source material.Word cloud helps to find the most frequent word but also plot a bar-graph for finding the most frequent word. Model creation is the backbone of the proposed model. The main focus is on creating BERT model then LSTM, some machine learning models such as Naïve Bayes, Logistic regression, Decision tree, Random forest and SVM. Then train and test the three set of dataset. Then do classification using these models and performance evaluation takes place.

3.2 Frameworks 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 NVDIA 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 pretrained 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. NumPy can be used to perform a wide variety of mathematical operations on arrays. It adds powerful data structures to Python that guarantee efficient calculations with arrays and matrices and it supplies an enormous library of high-level mathematical functions that operate on these arrays and matrices.. pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.NLTK is a toolkit build for working with NLP in Python. It provides us 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.

ISOT dataset

The dataset contains two types of articles fake and real News. This dataset was collected from realworld sources; the truthful articles were obtained by crawling articles from Reuters.com (News website). As for the fake news articles, they were collected from different sources. The fake news articles were collected from unreliable websites that were flagged by Politifact (a fact-checking organization in the USA) and Wikipedia. The dataset contains different types of articles on different topics, however, the majority of articles focus on political and World news topics. The dataset consists of two CSV files. The first file named “True.csv” contains more than 12,600 articles from reuter.com. The second file named “Fake.csv” contains more than 12,600 articles from different fake news outlet resources. Each article contains the following information: article title, text, type and the date the article was published on. To match the fake news data collected for kaggle.com, we focused mostly on collecting articles from 2016 to 2017. The data collected were cleaned and processed, however, the punctuations and mistakes that existed in the fake news were kept in the text.

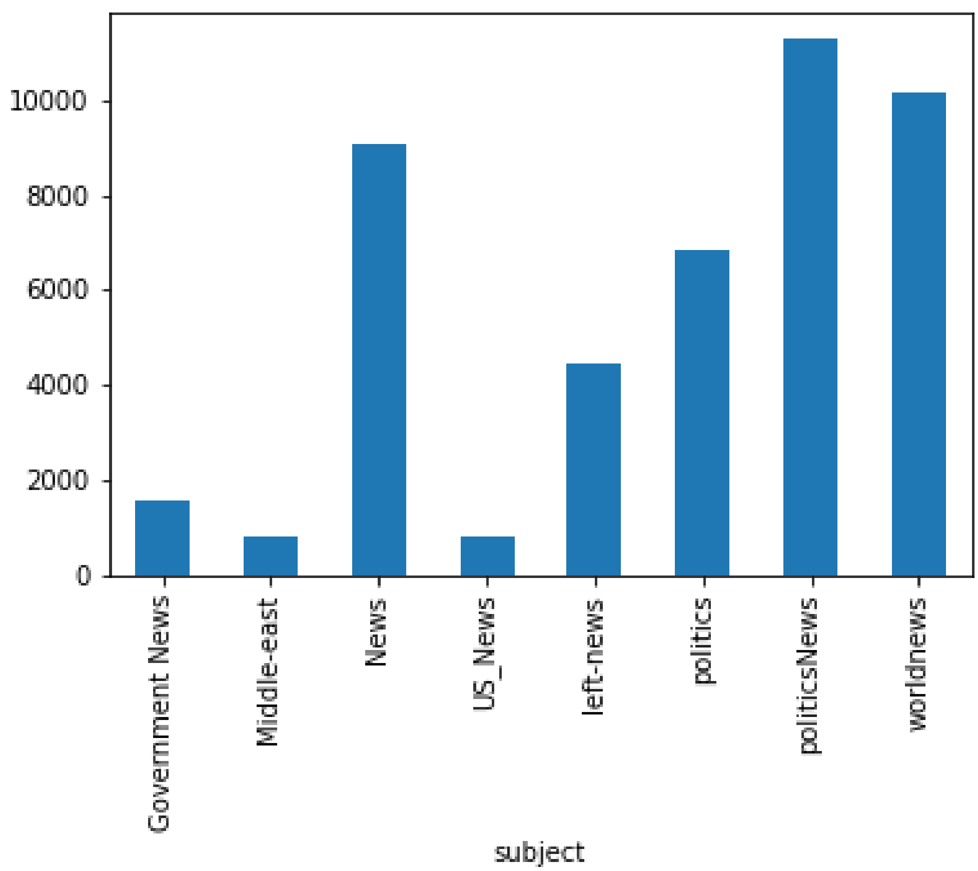

Figure 3.2: Categories in ISOT dataset

LIAR dataset

LIAR is a publicly available dataset for fake news detection. A decade-long of 12.8K manually labeled short statements were collected in various contexts from POLITIFACT.COM, which provides detailed analysis report and links to source documents for each case. This dataset can be used for fact-checking research as well. It contains 3 set of TSV files: train(10240 data), valid (1284) and test (10240). A tab-separated values (TSV) file is a text format whose primary function is to store data in a table structure where each record in the table is recorded as one line of the text file. The field's values in the record are separated by tab characters. Header rows may provide information about the semantics of table columns. TSV files function well as a data exchange format between programs that use structured tables or spreadsheets. These tab-separated value fields may contain a variety of data including text, mathematical, statistical, or scientific data. The TSV file format is widely supported and is very similar to CSV file formats, though data fields stored in CSV files are separated by commas rather than tabular spaces. Both are a type of delimiter-separated value format.

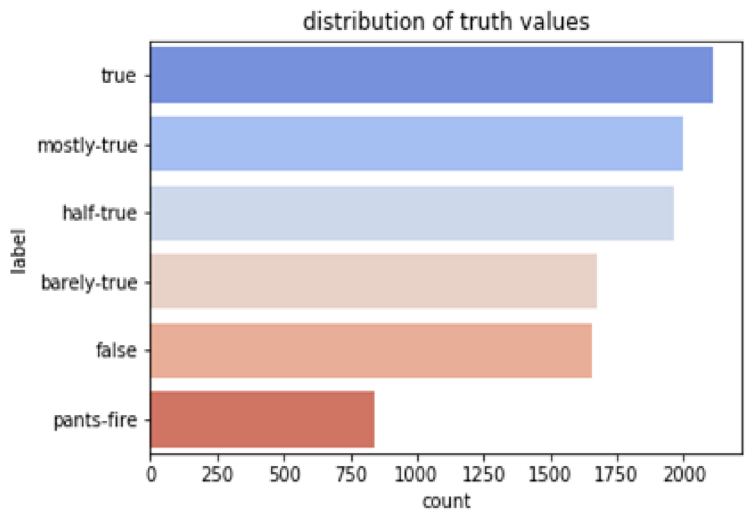

Figure 3.3: Label distribution on LIAR Dataset

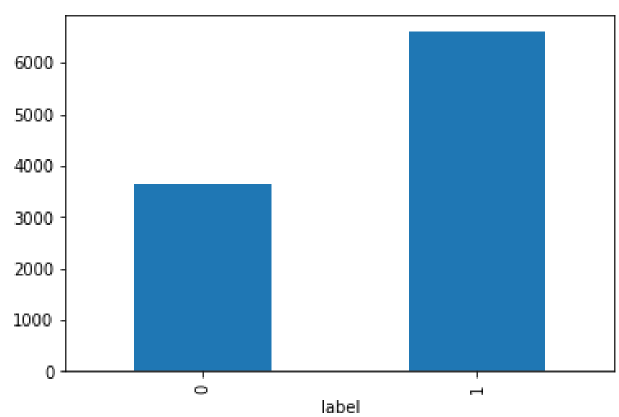

Figure 3.4: True-Fake classes on training dataset

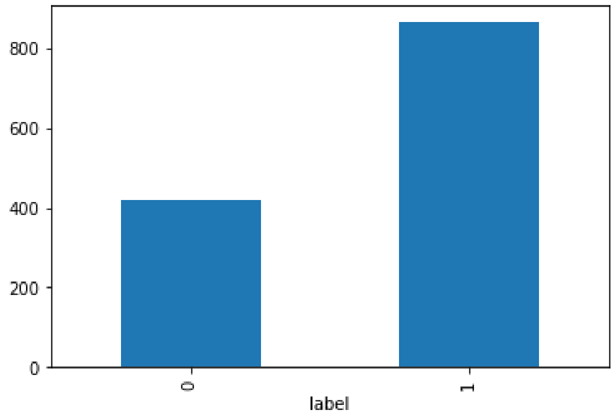

Figure 3.5: True-Fake classes on validation dataset

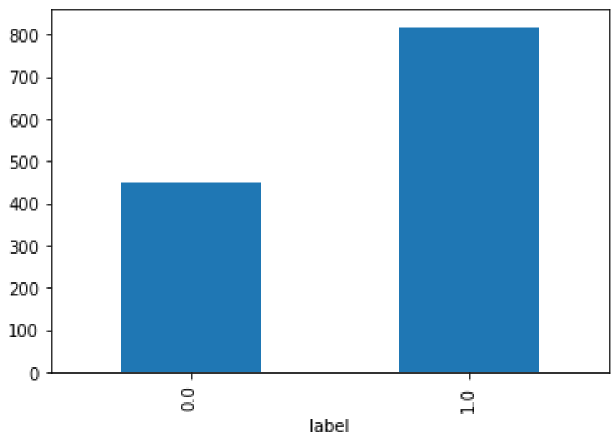

Figure 3.6: True-Fake classes on testing dataset

Twitter dataset

Twitter dataset is a pre-processed dataset.It contains two set of CSV files : shorttextpre-processedtrain 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.

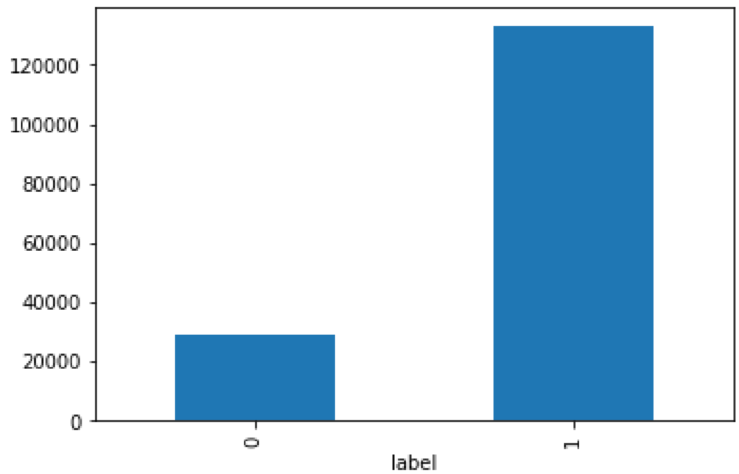

Figure 3.7: True-Fake classes on training dataset

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

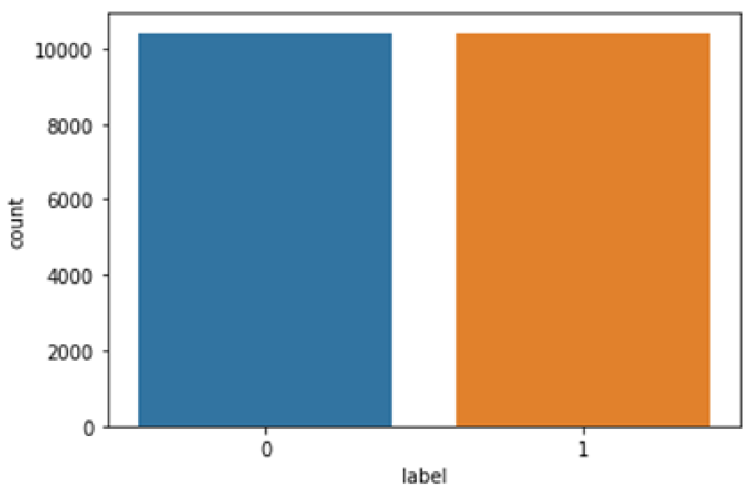

Figure 3.8: True-Fake classes on Kaggle dataset

Chapter 4

Data preprocessing and word cloud formation

4.1 Data Preprocessing

There are three set of datasets are used for analysis.LIAR and ISOT dataset needs pre-processing.

4.1.1 Pre-processing 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 4.1: 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 4.2: LIAR Dataset after pre-processing

4.1.2 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 4.3: 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 4.4: ISOT 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 LIAR dataset

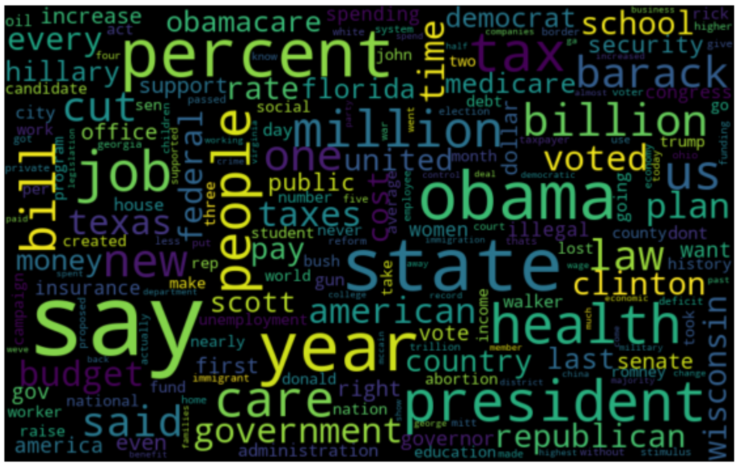

Figure 4.7: Word cloud on real news of LIAR dataset

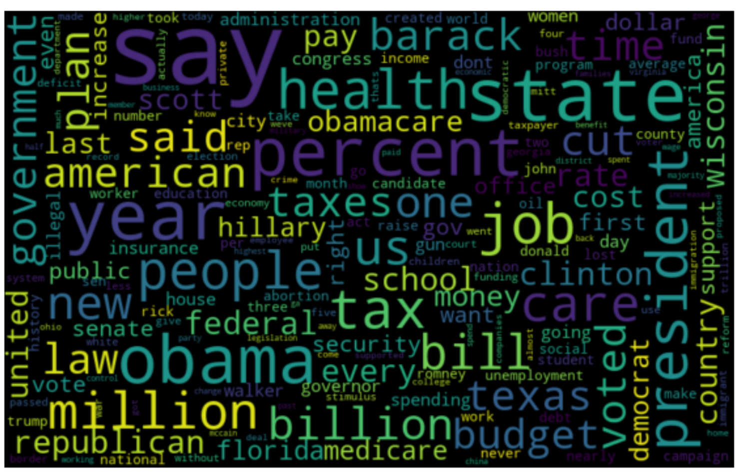

Figure 4.8: Word cloud on fake news of LIAR dataset

Word cloud on ISOT dataset

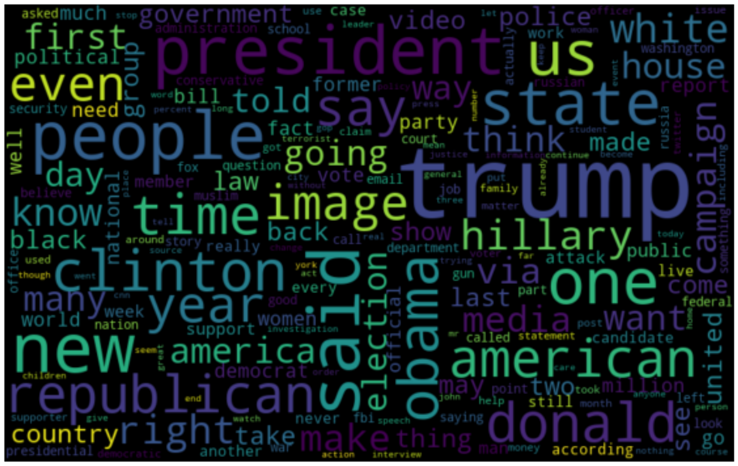

Figure 4.9: Word cloud on real news of ISOT dataset

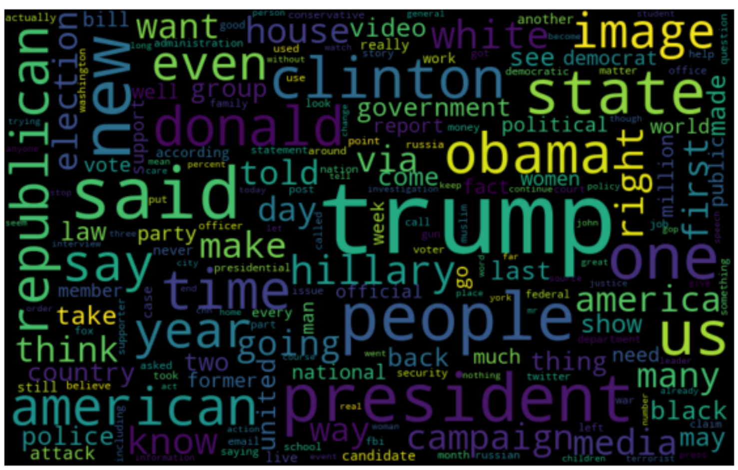

Figure 4.10: Word cloud on fake news of ISOT dataset

Word cloud on Twitter dataset

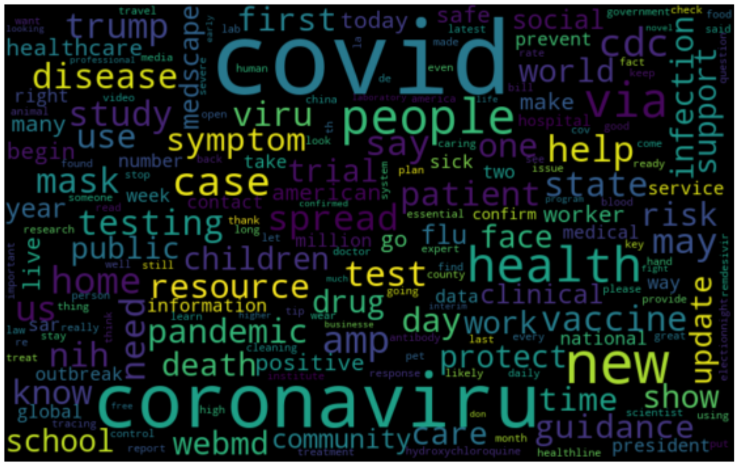

Figure 4.11: Word cloud on real news of Twitter dataset

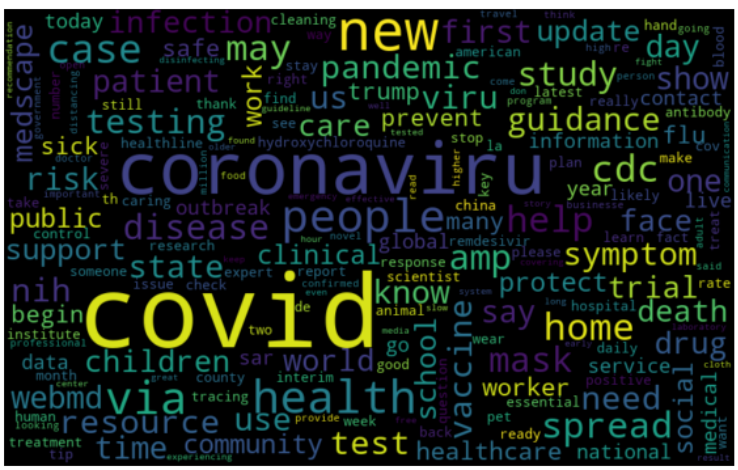

Figure 4.12: Word cloud on fake news of Twitter dataset

Word cloud on Kaggle dataset

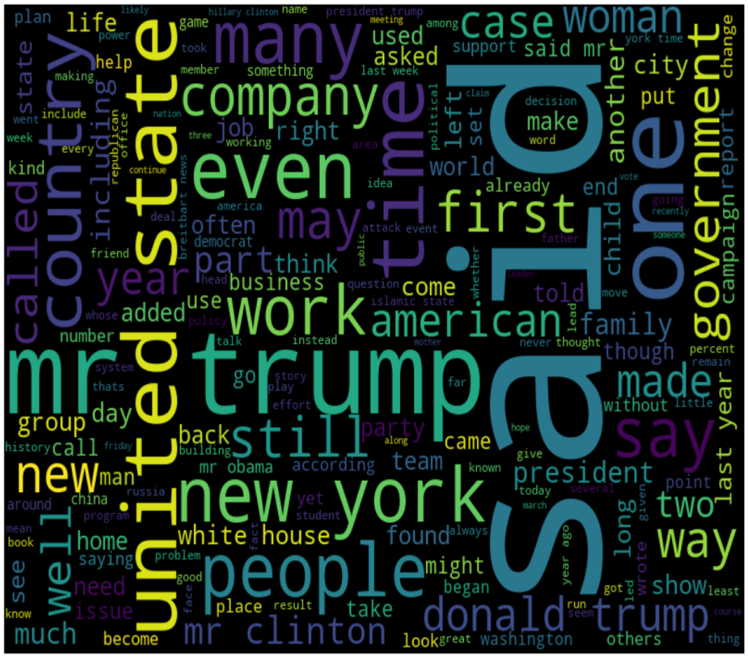

Figure 4.13: Word cloud on real news of Kaggle dataset

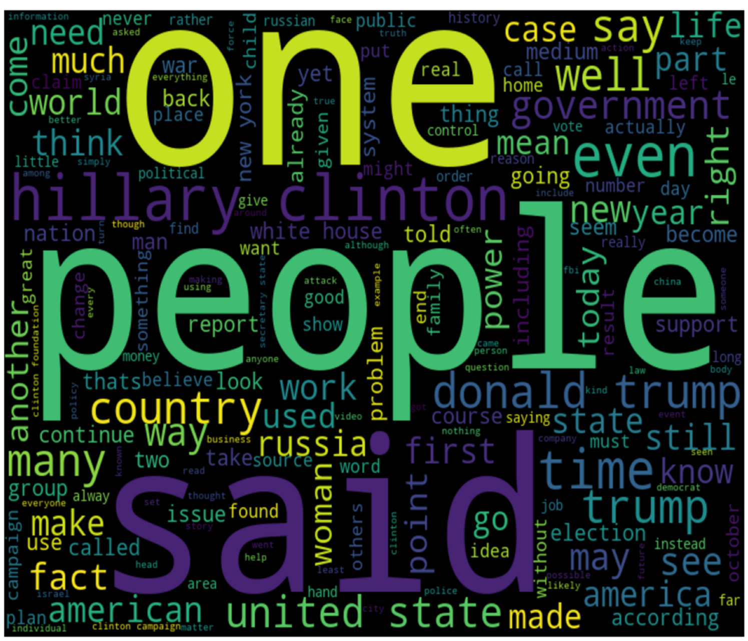

Figure 4.14: Word cloud on fake news of Kaggle dataset

Chapter 5

Results and Discussions

In this work, I tried to classify the fake news and real news with BERT, DistilBERT and five different ML algorithms such as Naive bayes classifier, Logistic regression, Decision tree, Random forest and SVM.The experiments done with four different dataset.

5.1 Classification using BERT

The classification using BERT is done with four different datasets. The design steps involved in the BERT classification process are:

- Import Libraries : This is the primary step in every deep learning and machine learning process. Here import the libraries such as Transformer, numpy, pandas, matplotlib, NLTK, Word cloud, etc.
- Load Dataset: Load the dataset (Any of them from LIAR Dataset, ISOT Dataset, Kaggle Dataset and Twitter dataset)
- Data preparation : In this step removes unwanted columns.
- Clean the entire dataset by Remove stop words, punctuation, and finally convert all uppercase letters into lowercase.
- Word cloud formation : Create word clod for real and fake data separately.
- Find most frequent words in the dataset- using Bigram, Trigram or simply by a bar graph.
- Split the dataset into train and test set
- Tokenize the sentences
- Convert tokenize dataset into torch dataset
- Call the fine-tuned BERT model for text classification from pre-trained model: Here BertForSequenceClassification s used.
- Train the model and evaluate on test dataset.
- Save the fine-tuned model and tokenizer.

- Get prediction by adding a softmax layer for obtaining the output probabilities.

Step	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
200	0.146000	0.004917	0.999220	0.999253	0.998721	0.999787
400	0.014800	0.006016	0.999109	0.999147	0.998295	1.000000
600	0.009300	0.003091	0.999332	0.999360	0.998721	1.000000
800	0.003900	0.000694	0.999777	0.999787	0.999573	1.000000
1000	0.002100	0.001011	0.999777	0.999787	0.999573	1.000000
1200	0.000100	0.000736	0.999889	0.999893	0.999787	1.000000
1400	0.005600	0.000840	0.999777	0.999787	0.999573	1.000000
1600	0.004200	0.000091	1.000000	1.000000	1.000000	1.000000

Figure 5.1: Output of BERT classifier on ISOT Dataset during Training

```
***** Running Evaluation *****
  Num examples = 8980
  Batch size = 20
/home/administrator/anaconda3/lib/python3.9/site-packages/
to gather along dimension 0, but all input tensors were sc
  warnings.warn('Was asked to gather along dimension 0, bu
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 5.2: Output of BERT classifier on ISOT Dataset during Validation

Step	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
200	0.010400	0.029362	0.996172	0.995628	0.997497	0.993766
400	0.039100	0.118688	0.973476	0.968840	0.999337	0.940150
600	0.067200	0.020631	0.996445	0.995944	0.996877	0.995012
800	0.021100	0.015175	0.996992	0.996581	0.993800	0.999377
1000	0.006400	0.014516	0.998086	0.997819	0.997508	0.998130
1200	0.029000	0.009180	0.998086	0.997819	0.997508	0.998130
1400	0.014100	0.005111	0.998906	0.998755	0.997512	1.000000

Figure 5.3: Output of BERT classifier on Kaggle Dataset during Training

```
{'epoch': 1.0,
 'eval_accuracy': 0.9989062072737216,
 'eval_f1': 0.9987546699875467,
 'eval_loss': 0.005110885016620159,
 '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 5.4: Output of BERT classifier on Kaggle Dataset during Validation

Step	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
200	0.165000	0.566074	0.870904	0.906618	0.932110	0.882484
400	0.307200	0.239885	0.897515	0.926883	0.939344	0.914749
600	0.259000	0.212299	0.909824	0.936550	0.935920	0.937181
800	0.249300	0.200490	0.911169	0.938490	0.923199	0.954297
1000	0.233900	0.229428	0.912695	0.939261	0.928206	0.950583
...	...	...	...	...	...	...
6600	0.144100	0.136385	0.946624	0.962404	0.962755	0.962054
6800	0.142900	0.138613	0.947684	0.963349	0.958539	0.968208
7000	0.138800	0.141027	0.948046	0.963365	0.964791	0.961945
7200	0.138400	0.137223	0.948072	0.963426	0.963742	0.963110
7400	0.150100	0.132515	0.948589	0.963967	0.959584	0.968390
7600	0.133100	0.133747	0.948124	0.963421	0.964828	0.962017

Figure 5.5: Output of BERT classifier on Twitter Dataset during Training

```
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 5.6: Output of BERT classifier on Twitter Dataset during Validation

Step	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
200	0.672900	0.679516	0.644531	0.771787	0.656183	0.936834
400	0.665300	0.635513	0.640625	0.739745	0.690885	0.796043
600	0.656800	0.661908	0.641602	0.781678	0.641602	1.000000
800	0.654200	0.652114	0.641602	0.781678	0.641602	1.000000

Figure 5.7: Output of BERT classifier on LIAR Dataset during Training

```
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 5.8: Output of BERT classifier on LIAR Dataset during Validation

5.2 Classification using DistilBERT

The classification using DistilBERT is done with Three different datasets. The design steps involved in the DistilBERT classification process are:

- Import Libraries : Like Bert classification process first import the libraries such as Transformer, numpy, pandas, matplotlib, NLTK, Word cloud, etc.
- Load Dataset: Load the dataset (Any of them from LIAR Dataset, ISOT Dataset and Twitter dataset)
- Data preparation : In this step removes unwanted columns.
- Clean the entire dataset by Remove stop words, punctuation, and finally convert all uppercase letters into lowercase.
- Word cloud formation : Create word clod for real and fake data separately.
- Find most frequent words in the dataset- using Bigram, Trigram or simply by a bar graph.
- Split the dataset into train and test set.

- Encode with DistilBERT.
- Find the maximum length of the sentance in the dataset.
- Tokenize the sentences.
- Convert labels and encodings to Tensorflow dataset.
- Fine tune with native Tensorflow: Here TFDistilBertForSequenceClassification is used for classification
- Train the model and evaluate on test dataset.
- Get prediction by adding a softmax layer for obtaining the output probabilities.

```
Epoch 1/10
1965/1965 [==============================] - 625s 314ms/step - loss: 0.0196 - accuracy: 0.9938
Epoch 2/10
1965/1965 [==============================] - 617s 314ms/step - loss: 0.0038 - accuracy: 0.9991
Epoch 3/10
1965/1965 [==============================] - 618s 314ms/step - loss: 0.0011 - accuracy: 0.9997
Epoch 4/10
 155/1965 [=>............................] - ETA: 9:29 - loss: 1.3566e-04 - accuracy: 1.0000
```

Figure 5.9: Output of DistilBERT classifier on ISOT Dataset during Training

```
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 5.10: Output of DistilBERT classifier on ISOT Dataset during Validation

```
Downloading: 100%                                   256M/256M [00:07<00:00, 36.3MB/s]
All model checkpoint layers were used when initializing TFDistilBertForSequenceClassification.

All the layers of TFDistilBertForSequenceClassification were initialized from the model checkpoin
If your task is similar to the task the model of the checkpoint was trained on, you can already u
Epoch 1/3
6066/6066 [==============================] - 2571s 421ms/step - loss: 0.1220 - accuracy: 0.9529
Epoch 2/3
6066/6066 [==============================] - 2557s 422ms/step - loss: 0.0666 - accuracy: 0.9748
Epoch 3/3
6066/6066 [==============================] - 2556s 421ms/step - loss: 0.0423 - accuracy: 0.9845
<keras.callbacks.History at 0x7feb9e2067d0>
```

Figure 5.11: Output of DistilBERT classifier on Twitter Dataset during Training

```
[ ] #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 5.12: Output of DistilBERT classifier on Twitter Dataset during Validation

```
All model checkpoint layers were used when initializing TFDistilBertForSequenceClassification.

All the layers of TFDistilBertForSequenceClassification were initialized from the model checkpoint
If your task is similar to the task the model of the checkpoint was trained on, you can already us
Epoch 1/10
1726/1726 [==============================] - 313s 175ms/step - loss: 0.6538 - accuracy: 0.6411
Epoch 2/10
1726/1726 [==============================] - 302s 175ms/step - loss: 0.6205 - accuracy: 0.6589
Epoch 3/10
1726/1726 [==============================] - 302s 175ms/step - loss: 0.5426 - accuracy: 0.7386
Epoch 4/10
1726/1726 [==============================] - 302s 175ms/step - loss: 0.3891 - accuracy: 0.8330
Epoch 5/10
1726/1726 [==============================] - 302s 175ms/step - loss: 0.2744 - accuracy: 0.8988
Epoch 6/10
1726/1726 [==============================] - 302s 175ms/step - loss: 0.2042 - accuracy: 0.9253
Epoch 7/10
1726/1726 [==============================] - 302s 175ms/step - loss: 0.1734 - accuracy: 0.9341
Epoch 8/10
1726/1726 [==============================] - 302s 175ms/step - loss: 0.1549 - accuracy: 0.9390
Epoch 9/10
1726/1726 [==============================] - 302s 175ms/step - loss: 0.1504 - accuracy: 0.9405
Epoch 10/10
1726/1726 [==============================] - 301s 175ms/step - loss: 0.1224 - accuracy: 0.9523
<keras.callbacks.History at 0x7ff520a7f610>
```

Figure 5.13: Output of DistilBERT classifier on LIAR Dataset during Training

```
[ ] #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 5.14: Output of DistilBERT classifier on LIAR Dataset during Validation

5.3 Performance analysis

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 and ML algorithm is done through the accuracy measurement only.

5.3.1 Performance analysis table of BERT

Training

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

Table 5.1: Performance analysis of BERT Model during Training

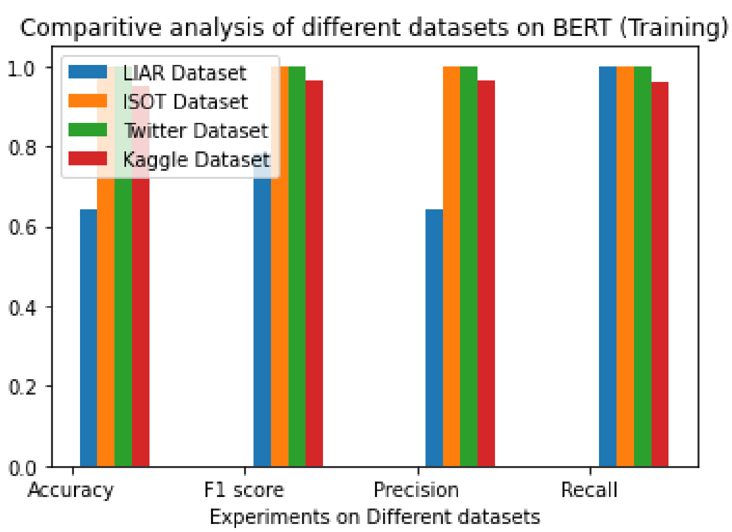

Figure 5.15: Comparison of BERT algorithms on all Datasets during training

Validation

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

Table 5.2: Performance analysis of BERT Model during Validation

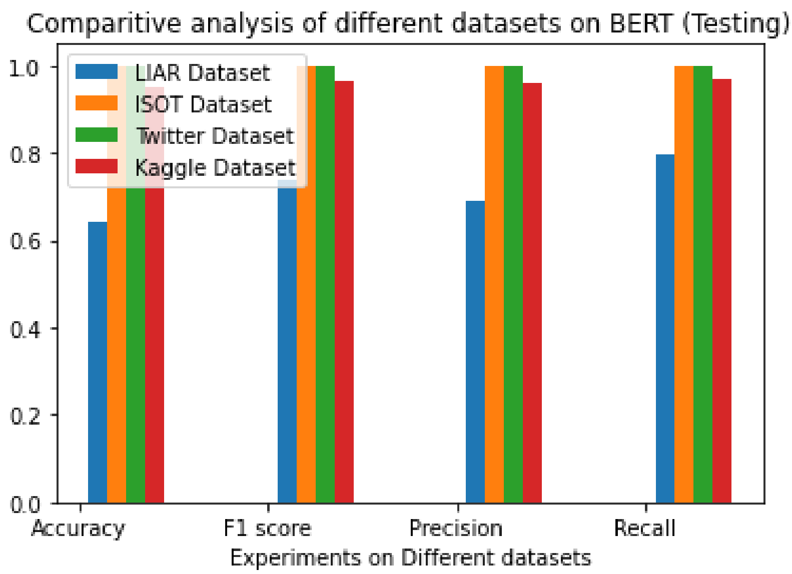

Figure 5.16: Comparison of BERT algorithms on all Datasets during validation

In the case of BERT evaluates the accuracy,f1-score, precision and recall, 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 much during validation.

5.3.2 Performance analysis table of DistilBERT

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

Table 5.3: Performance analysis of DistilBERT Model during Training and validation process

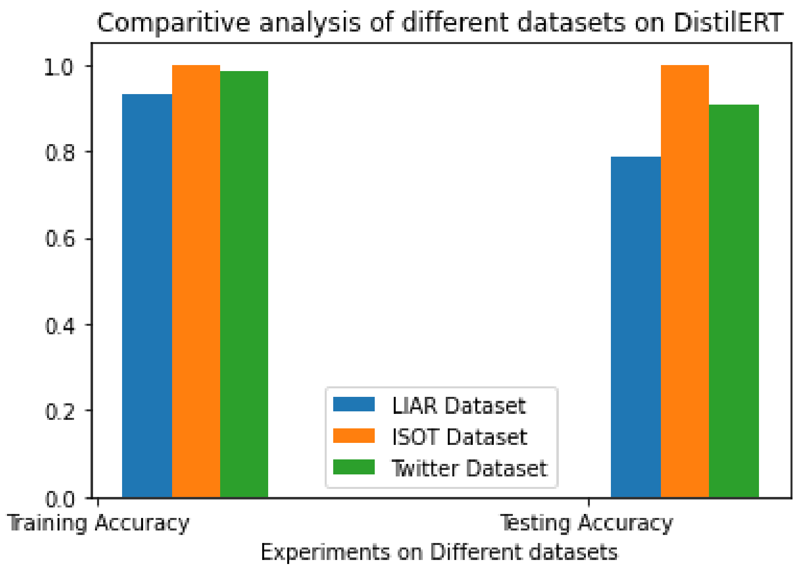

Figure 5.17: Accuracy of DistilBERT algorithms on all Datasets

In the case of DistilBERT only evaluates the accuracy and the experimental results given by the classification using DistilBERT model gives a better performance on ISOT and twitter dataset but the performance of LIAR dataset is much during validation as same as BERT Model.

5.4 Classification using ML Algorithms

5.4.1 Experiment on ISOT dataset

The experiment conducted on ISOT dataset with different machine learning algorithms shows better accuracy.The accuracy obtained are: Naïve Bayes: 94.65%,Logistic Regression:98.73%,Decision Tree:99.57%,Random Forest:99.21 %,SVM: 99.55%.The below figure shows the confusion matrix of SVM, that provides maximum accuracy for ISOT dataset.

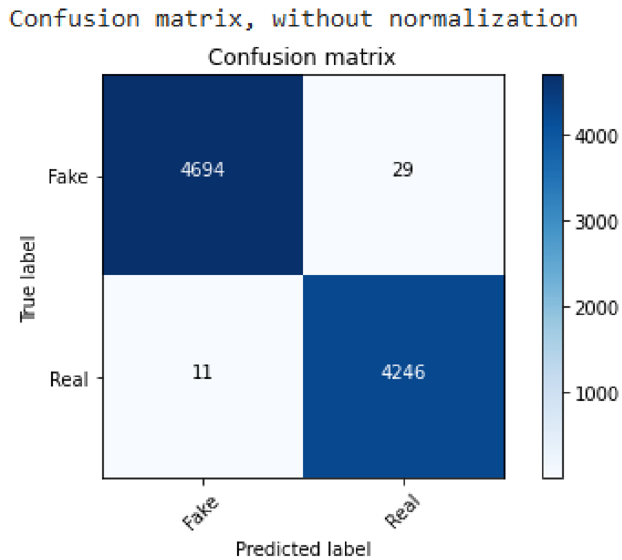

Figure 5.18: Confusion matrix of SVM on ISOT dataset

5.4.2 Experiment on LIAR dataset

The experiment conducted on LIAR dataset with different machine learning algorithms shows lower accuracy.The accuracy obtained are: Naïve Bayes: 62.83%,Logistic Regression:58.09%,Decision Tree:61.48%,Random Forest:54.22 %,SVM: 58.25%.The below figure shows the confusion matrix of Naive Bayes algorithm, that provides maximum accuracy for LIAR dataset.

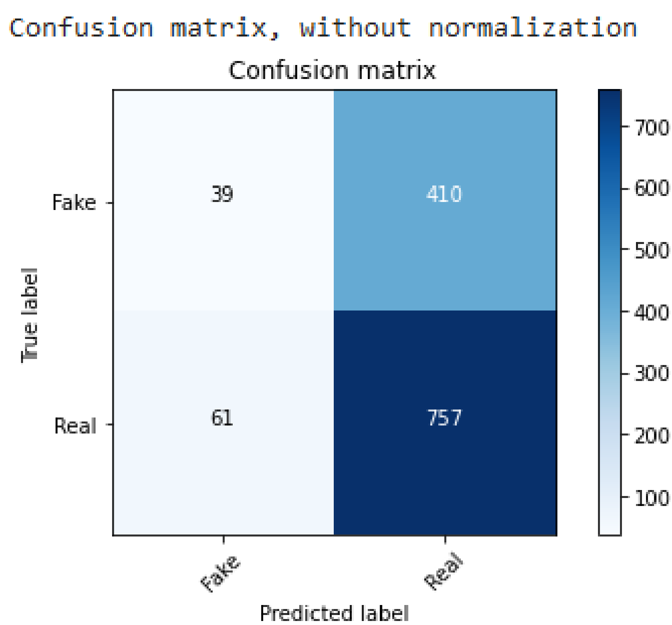

Figure 5.19: Confusion matrix of Naive Bayes on LIAR dataset

5.4.3 Experiment on Twitter dataset

The experiment conducted on Twitter dataset with different machine learning algorithms shows good accuracy.The accuracy obtained are: Naïve Bayes: 90.51%,Logistic Regression:93.3%,Decision Tree:87.41%,Random Forest:93.75 %,SVM: 94.03%.The below figure shows the confusion matrix of SVM, that provides maximum accuracy for Twitter dataset.

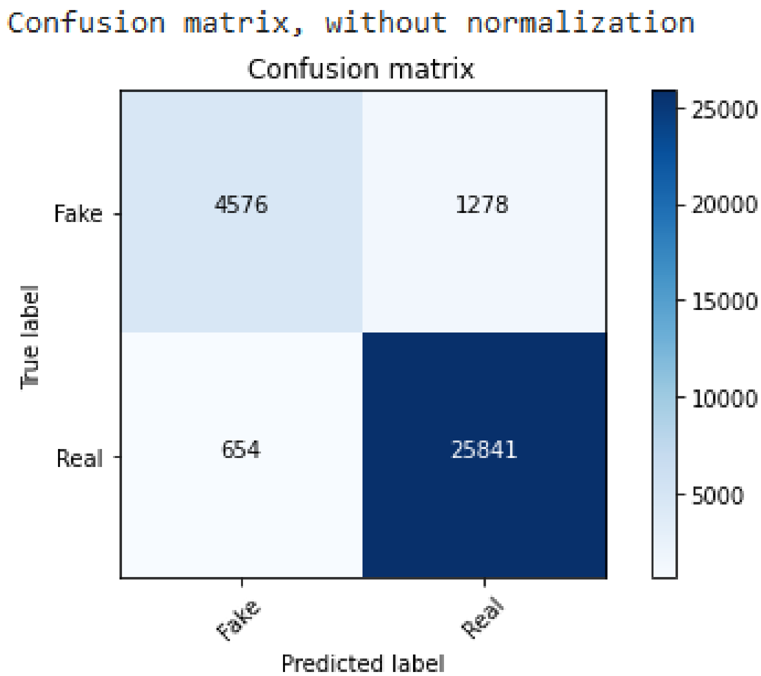

Figure 5.20: Confusion matrix of SVM on Twitter dataset

5.4.4 Experiment on Kaggle dataset

The experiment conducted on Kaggle dataset with different machine learning algorithms shows lower accuracy.The accuracy obtained are: Naïve Bayes: 84.66%,Logistic Regression:94.87%,Decision Tree:89.36%,Random Forest:90.87%,SVM: 96.58%.The below figure shows the confusion matrix of SVM, that provides maximum accuracy for Kaggle dataset.

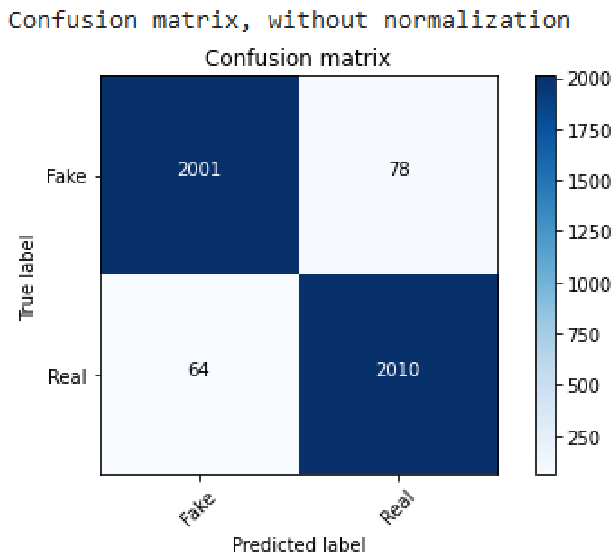

Figure 5.21: Confusion matrix of SVM on Kaggle dataset

Model	**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 5.4: Performance analysis of Machine Learning Algorithms

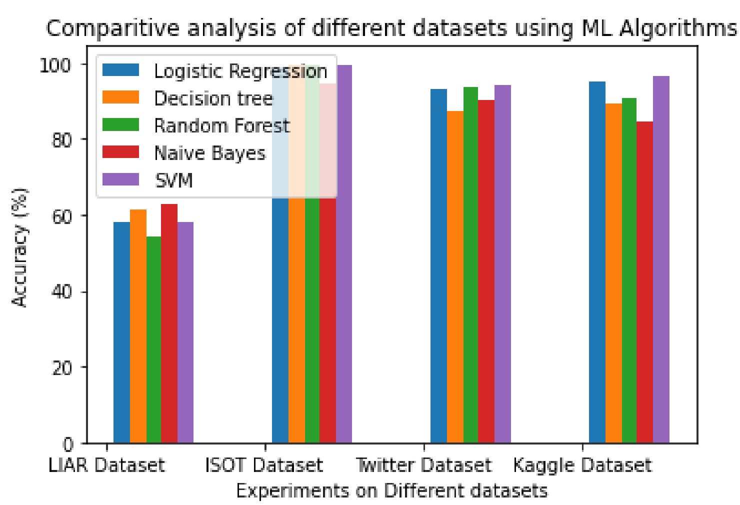

Figure 5.22: Accuracy of ML algorithms on all Datasets

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 in all classifiers.

Chapter 6

Conclusion and Future works

False news detection is essential since many people propagate fake news on social media to deceive the public. It is vital to detect false news to protect individuals or organisations from losing their reputations. On the ISOT, Kaggle, Twitter, and LIAR datasets, I conduct an experimental study of fake news detection using BERT and DistiBERT, 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). According to the experimental investigation, BERT and DistilBERT can be employed as a generalised model for fake news identification. However, the LIAR dataset's performance is substantially poorer than the other three datasets since some of the articles in the LIAR dataset are from the incorrect set of data (PolitiFact's Flip-o-Meter rather than its Truth-o-Meter) but are nonetheless tagged with a truth value. As a result, such data points are useless for training the model. The LIAR dataset is difficult to classify due to a lack of sources or knowledge bases to rely on for verification. Although I focused solely on text analysis in this study, the source plays a crucial role in disseminating bogus news. 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 faces numerous difficulties; one of them is that it depends on the quality of data, which differs among social media platforms. Another is the availability of multilingual and mixed languages.

References

[1] Kula, Sebastian, Rafał Kozik, and Michał Choraś. ”Implementation of the BERT-derived architectures to tackle disinformation challenges.” Neural Computing and Applications (2021): 1-13.

[2] 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).

[3] 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.

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

[5] 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.

[6] 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. v

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

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

[9] 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.