Fake news detection using BERT

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. 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 rumors, which are the information 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 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. 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 the early detection of 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 machine learning and deep learning algorithms are found in the literature. All these algorithms are focusing mainly on feature extraction. Therefore, the extraction of relevant features is an important task for the effective classification task. 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. Here I do fake news detection using BERT, LSTM, and some machine learning algorithms such as Naive Bayes, decision tree, random forest, SVM, and Logistic regression, also do a comparison study with three sets of datasets such as the Twitter dataset, LIAR dataset, and ISOT dataset.

Literature review:

Rohit Kumar Kaliyar et al. [1] 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 [2] 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 than presenting sequences utilizing the first hidden states of BERT, all hidden states with dynamic range attention mechanisms are used to compute weights. To improve generalization, 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. [3] 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. Nguyen Manh Duc Tuan et al.[4] 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. J Briskilal et al. [5] 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. [6] 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. [7] 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. [8] 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.

Dataset collected:

1. **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.

2. **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.

3. **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.

Design steps:

The proposed system is used to classify the news as fake or real using different machine learning techniques, LSTM and BERT . The entire system can be divided into six: Data pre-processing,Word cloud formation, finding of most frequent word, model creation, training and testing data and finally classification and performance analysis. 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 three sets of dataset from that two 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 Na¨ıve 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.
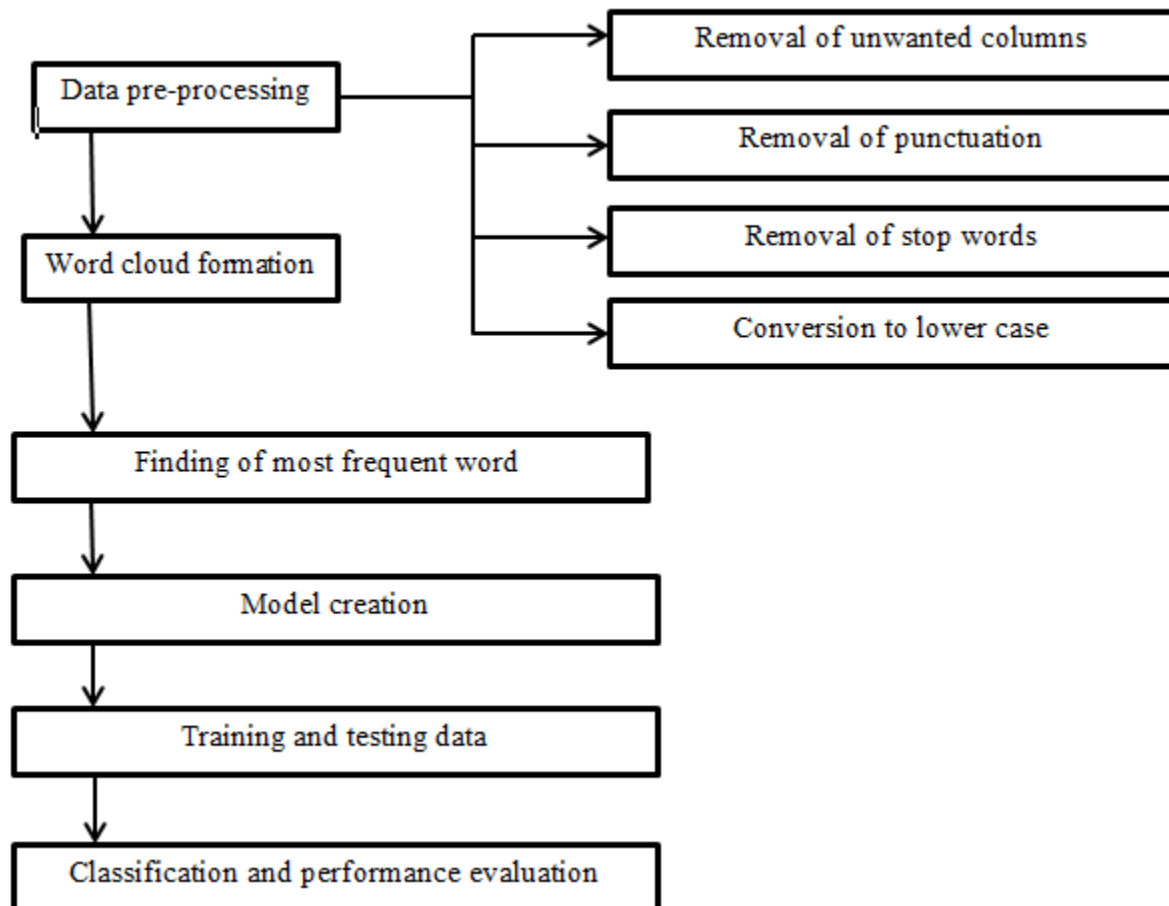
Frameworks and Libraries:

- Language: Python
- Framework: - Keras with Tensorflow, sklearn, pytorch as background in the Google Colab
- Library: - HuggingfaceTransformer, NumPy, Pandas, Matplotlib, nltk

Handson details:

- Performed classification using LSTM, SVM, Naïve Bayes, Logistic regression, Decision Tree, random Forest on LIAR dataset.
- Performed classification using BERT, LSTM, SVM, Naïve Bayes, Logistic regression, Decision Tree, random Forest on ISOT dataset.
- Performed classification using BERT on Twitter dataset.

Result:

(i) Classification using ISOT dataset:
- Step 1: Load training and testing data using pandas.
- Step 2: Perform pre-processing
- Step 3: perform classification with BERT, LSTM, SVM, Naïve Bayes, Logistic regression, Decision Tree and Random Forest.

Model	Accuracy (in %)
LSTM	99.498
Logistic regression	98.73
Decision Tree	99.57
Random Forest	99.21
Naïve Bayes	94.65
SVM	99.55

(ii) Classification using LIAR dataset:
- Step 1: Load training and testing data using pandas.
- Step 2: Perform pre-processing
- Step 3: perform classification with LSTM, SVM, Naïve Bayes, Logistic regression, Decision Tree and Random Forest.

Model	Accuracy (in %)
LSTM	35
Logistic regression	58.09
Decision Tree	61.48
Random Forest	54.22
Naïve Bayes	62.83
SVM	58.25

(iii) Classification using Twitter dataset:
- Step 1: Load training and testing data using pandas.
- Step 2: perform classification with BERT in different epochs with different batch size.

Batch size	Epoch	Accuracy
32	5	0.26
16	5	0.27
5	5	0.31

2	5	0.38
2	10	0.45
32	50	0.27

References:

[1]]. 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.

[2] 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.

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

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

[5] 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.

[6] 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).

Guide Details:

Prof. Sumod Sundar
TKM College of Engineering, Kollam
Email: sumodsundar@tkmce.ac.in
Mob: 8086515716