Fake news detection using BERT

Abstract:

The spread of fake news on online platforms is one of the main problems today. As a result, determining whether such news is true or false is essential in modern technological life. Fake news can exist in various domains, including politics, entertainment, sports, Etc. This type of news causes so many problems in our society. So the early detection of fake news reduces the impact. Various studies regarding machine learning and deep learning algorithms are found in the literature. All these algorithms focus mainly on feature extraction. Therefore, the extraction of relevant features is essential for a practical classification task. The ability to generalise a learning model by identifying patterns in a text will aid in distinguishing between fake and real news.

Introduction:

The rise of social networks has accelerated the dissemination of rumours, satires, and false information, increasing the distribution of fake news. False news is divided into three categories: first, actual fake news, which is intended to deceive the reader; second, rumours, which are information with unclear veracity but widespread acceptance; and third, parodies and satires generated by a funny individual. So, identifying such news as real or fake is essential in digital life. One 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 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. Various studies regarding machine learning and deep learning algorithms are found in the literature. The ability to generalise a learning model by identifying patterns in a text will aid in distinguishing between fake and real news. Fake news detection using BERT and LSTM techniques is now the most competitive study. I detect fake news using BERT, LSTM, and machine learning algorithms such as Naive Bayes, decision trees, random forest, SVM, and logistic regression. I also do a comparison study with three datasets: the Twitter dataset, LIAR dataset, Kaggle dataset, and ISOT dataset.

Literature Survey:

Rohit Kumar Kaliyar et al. [1] proposed a method for Fake news detection in social media with a BERT-based deep learning approach. 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, which can accurately extract a sentence's context representation to detect 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 detection. The classification findings 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. A novel BERT with a combined learning-based model is presented for detecting fake news in articles. The proposed method can detect fake news in both lengthy and short pieces. Rather than presenting sequences utilising the first hidden states of BERT, all hidden states with dynamic range attention mechanisms are used to compute weights.

RFC and NER task models are combined with BERT via a standard parameter layer in collaborative learning to improve generalisation. 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 classification. They discuss the many aspects connected to transfer learning in the suggested model and present architecture to classify fake news. Approaches that focus on text classification utilise

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 unsupervised, deeply bidirectional language representation based on ELMo.

Dataset collected:

1. **ISOT dataset:** There are two articles: fake and true news. The true articles were retrieved via crawling articles from Reuters.com; the dataset was compiled from real-world sources (the news website). The fake news items were gathered from untrustworthy websites identified by Politifact (a fact-checking group based in the United States) and Wikipedia. Most of them are about politics and world news."True.csv" holds more than 12,600 Reuters.com stories. "False.csv" has almost 12,600 stories culled from various fake news sources. The following information is included in each article: article title, text, type, and publication date.
2. **LIAR dataset:** LIAR is a publicly accessible dataset for detecting fake news. A tab-separated values (TSV) file is a text format that stores data in a table structure by recording each entry in the table as one line of the text file.
3. It includes three TSV files: train (10240 data), valid (1284), and test (1284).
4. Twitter dataset: The Twitter dataset contains 161743 data related to the Covid-19 pandemic. The disadvantage of this dataset is class imbalance.
5. **Kaggle dataset:** The Kaggle dataset's train.csv (20776) and test.csv (5201) CSV files Both datasets have the following attributes: title: the title of a news story, id: the unique id for a news article, author: the news article's author, the article's text; it may be incomplete, a label indicating that an article is potentially untrustworthy (1: untrustworthy, 0: reliable)

Design steps:

The proposed system is used to classify news as fake or real using different machine learning techniques, BERT and DistilBERT. The entire system can be divided into six parts: Data preprocessing, word cloud formation, word frequency analysis, model creation, training and testing data, and finally, classification and performance analysis are all part of the process. The dataset contains text data, which may not be in the same format. So, data preprocessing plays a prominent role in the proposed methodology. The preprocessing steps on the input dataset are removing unwanted columns, removing punctuation, removing stop words, and converting to lower case. Here I will take four sets of datasets. All the text data becomes the same format after this process. So the analysis becomes more accessible. Then the next step is word cloud formation. Word Clouds are visual displays of text data, a 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 colouring of the characters in the word) depicting the relative frequency of occurrence of each target word in the source material. A word cloud helps find the most frequent word and plots a bar graph for finding the most frequent word. Model creation is the backbone of the proposed model. The main focus is on creating the BERT model, DistilBERT, and some machine learning models such as Naive Bayes, Logistic regression, Decision Tree, Random Forest, and SVM. Then train and test the four sets of datasets. Then

classification takes place using these models, and performance evaluation take place.

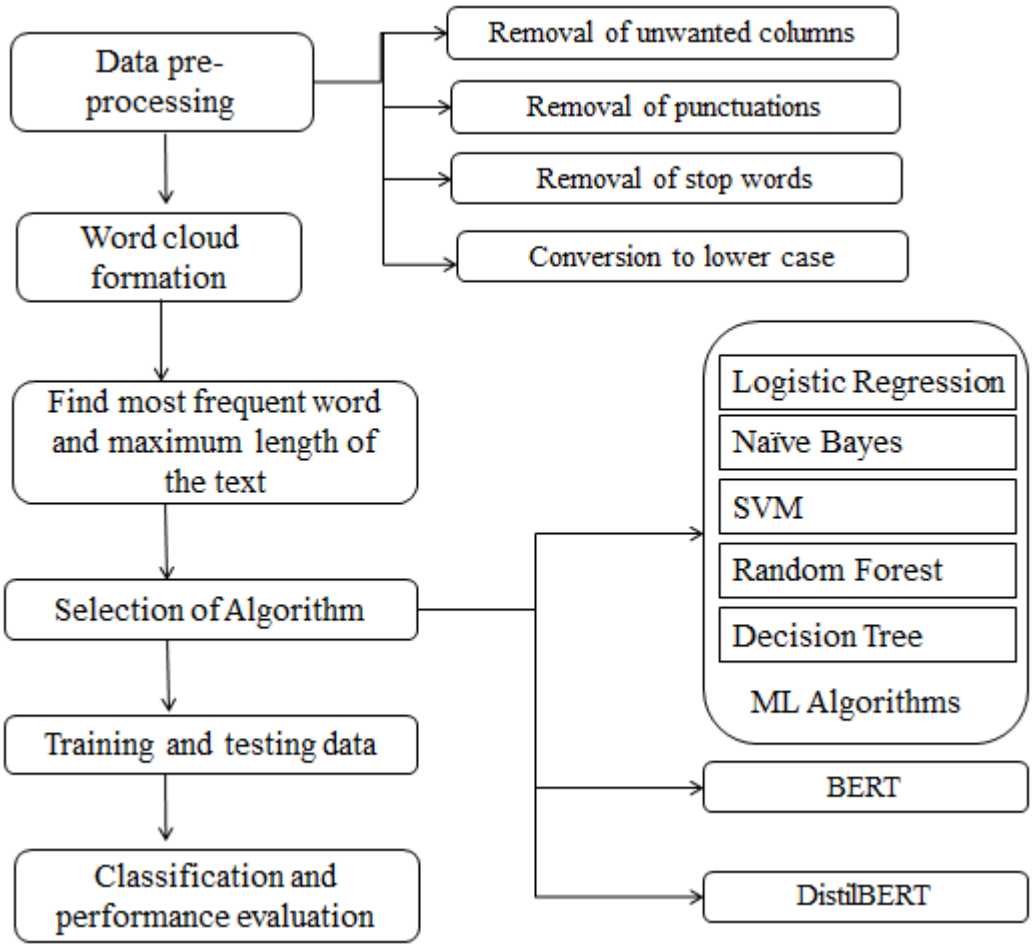

Fig 1: Architecture of Proposed model

Frameworks and Libraries:

- Language: Python
- Framework: - Keras with Tensorflow, PyTorch as background in the Google Colab and Power edge server with NVIDIA TESLA V100 GPU
- Library: - HuggingfaceTransformer, NumPy, Pandas, Matplotlib, nltk

Hands-on details:

- Performed classification using SVM, Naïve Bayes, Logistic regression, Decision Tree, Random Forest on LIAR, Twitter, Kaggle, and ISOT.
- Performed classification using BERT on the Twitter dataset, LIAR dataset, ISOT dataset and Kaggle dataset.
- Performed classification using DistilBERT on the Twitter, LIAR, and ISOT datasets.

Result:

(i) Word cloud formed on ISOT Dataset

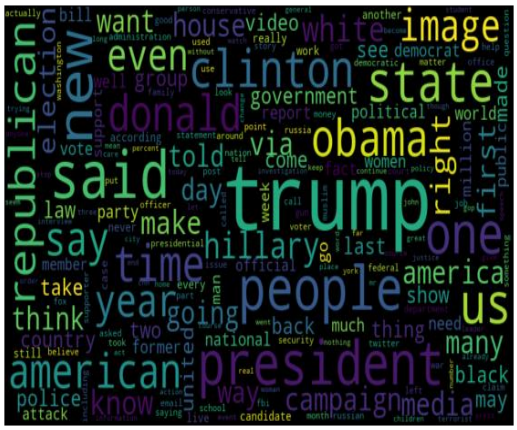

Fig 2: Word cloud fake news dataset

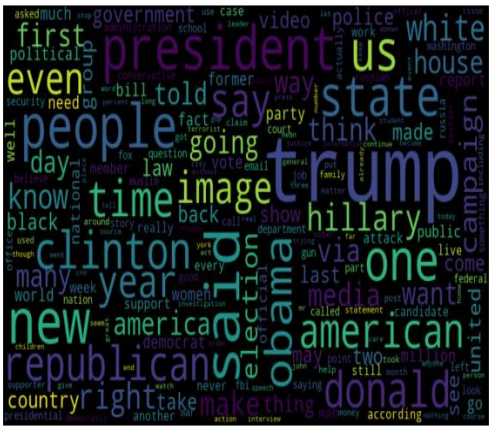

Fig 3: Word cloud real news dataset

(ii) Word cloud formed on LIAR Dataset

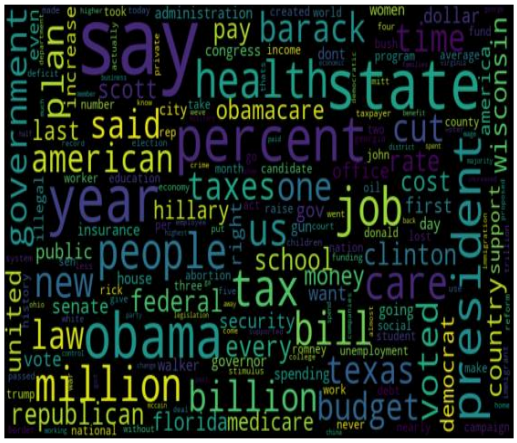

Fig 4: Word cloud on fake news dataset

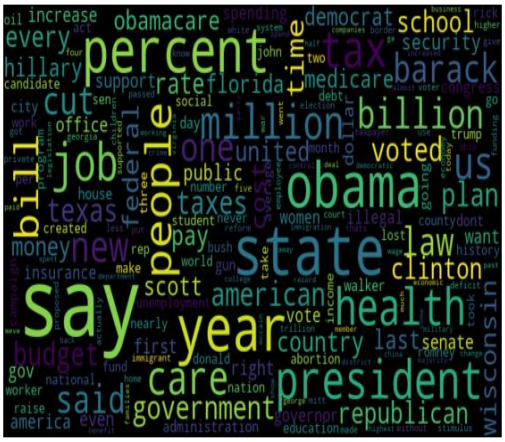

Fig 5: Word cloud real news dataset

(iii) Word cloud formed on Twitter Dataset

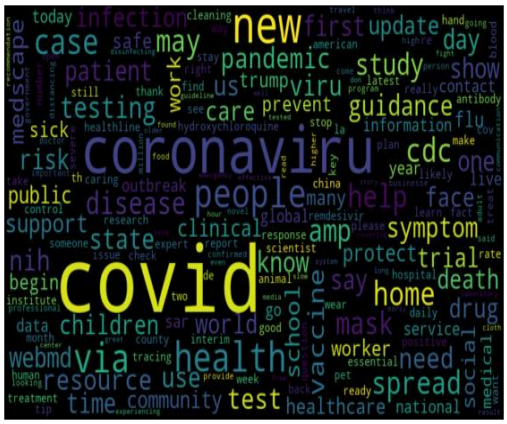

Fig 6: Word cloud on fake news dataset

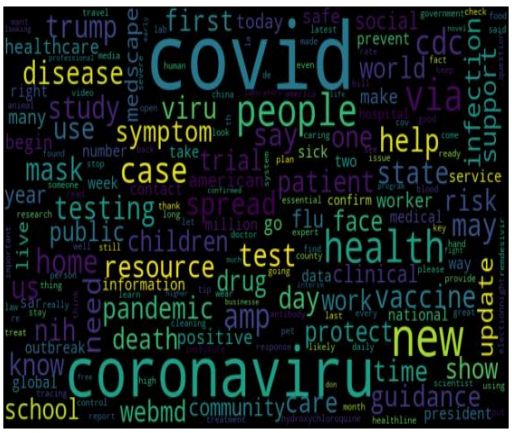

Fig 7: Word cloud real news dataset

(iv) Word cloud formed on Kaggle Dataset

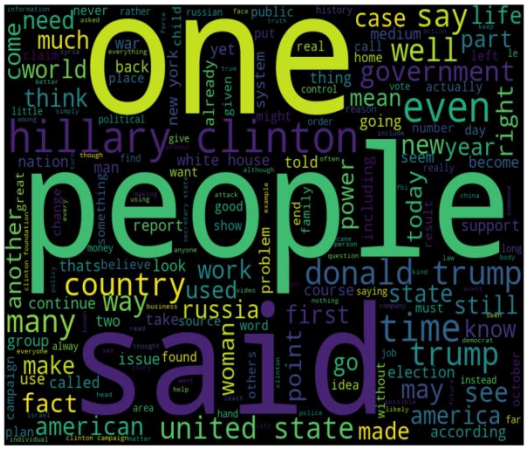

Fig 8: Word cloud on fake news dataset

Fig 9: Word cloud real news dataset

(v) Comparison of ML Algorithms

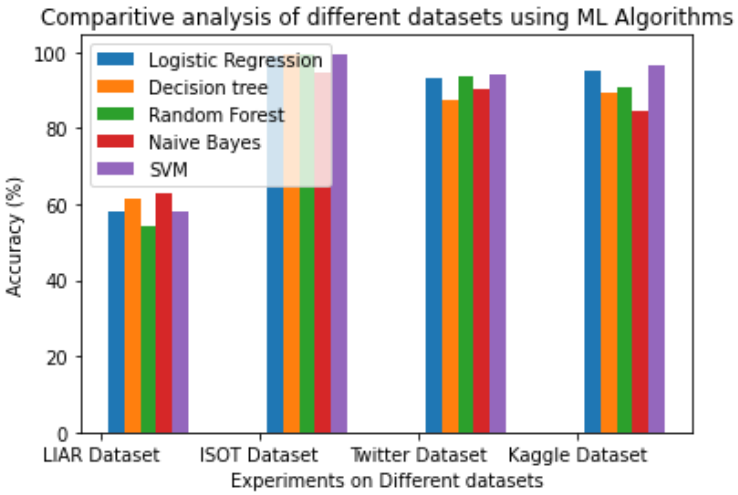

Fig 10: Comparison of ML Algorithms

(vi) Comparison of BERT Algorithms

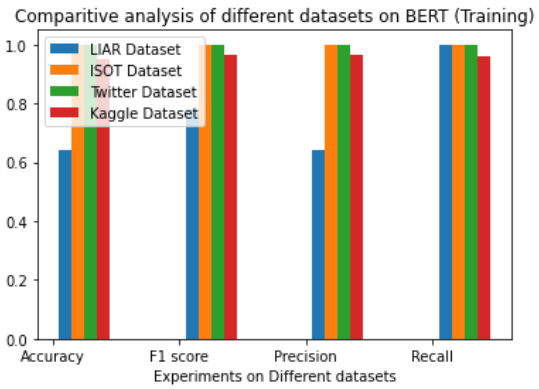

Fig 11:Comparison during training

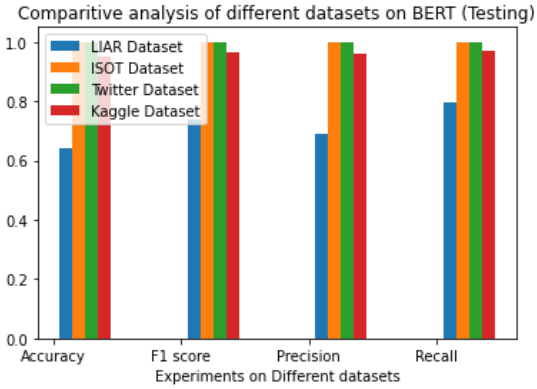

Fig 12: Comparison during testing

(vii) Comparison of DistilBERT Algorithms

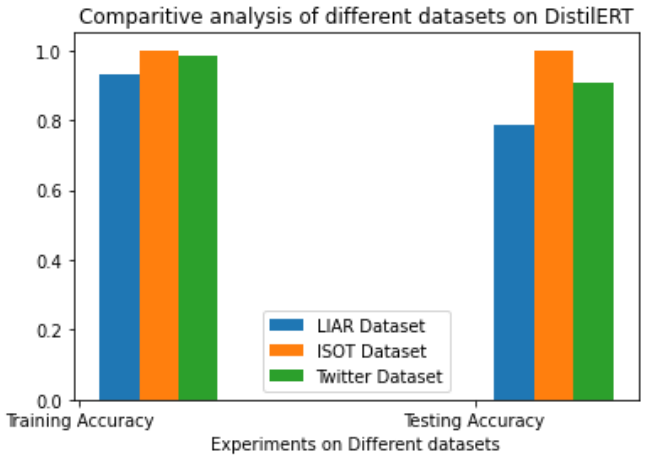

Fig 13: Comparison of DistilBERT Algorithm

(viii) Classification using ML Algorithms

	LIAR Dataset	**ISOT Dataset**	**Twitter Dataset**	**Kaggle Dataset**
Model	Accuracy (in %)			
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 1: Result of Classification using ML Algorithms

(ix) Classification using BERT:

Dataset	**Training**				**Validation**			
	Accuracy	**F1 Score**	**Precision**	**Recall**	**Accuracy**	**F1 Score**	**Precision**	**Recall**
LIAR	0.6416	0.7816	0.6416	1.000	0.6402	0.7397	0.6908	0.7960
ISOT	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Kaggle	0.9989	0.9985	0.9975	1.000	0.998906	0.9987	0.9975	1.000
Twitter	0.94812	0.9634	0.9648	0.9620	0.948593	0.9637	0.959584	0.96839

Table 2: Result of Classification using BERT

(x) Classification using DistilBERT:

Dataset	Training accuracy	Validation accuracy
LIAR	0.9287	0.7868
ISOT	1.00	0.9998
Twitter	0.9845	0.9074

Table 1: Result of Classification using DistilBERT

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 is crucial 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] 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