MAHABUBUL ALAM, TANMIN ALAM ROKTI, NABIL FAIEAZ DIPTA, MD RAKIBUL ISLAM, JUNAN NA BINTE WAHID

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

ABSTRACT

Fake news is false or misleading information that is created to deceive people. It is mostly created to harm a person or organization's reputation. In this era of information technology, the rapid dissemination of news and information through social media platforms such as Twitter, Facebook, Instagram, and so on has become an integral part of our daily lives. We have seen numerous how devastating consequences can be due to the spread of fake news. Rumors can create some unimaginable havoc in real life. There are already some existing works that need more eciency. We want to propose a deep learning-based approach to detect fake news from social media in real time.

RESEARCH OBJECTIVE

The objective of this research is to develop a robust deep learning model capable of accurately detecting fake news in social media posts. By leveraging advanced natural language processing techniques and large-scale datasets, the model aims to identify linguistic and contextual patterns that distinguish false information from truthful content. The project will explore various neural network architectures, including transformers and recurrent neural networks, to optimize performance. Ultimately, the goal is to create a reliable tool that aids in mitigating the spread of misinformation on social media platforms, contributing to a more informed and discerning digital society.

PROBLEM STATEMENT

Fake news, amplified by modern technologies, undermines trust and unity on social media. This study aims to develop a more efficient deep learning model for fake news detection, improving accuracy by incorporating user profiles and focusing on dishonest subjects. The goal is to enhance existing detection methods on social media.
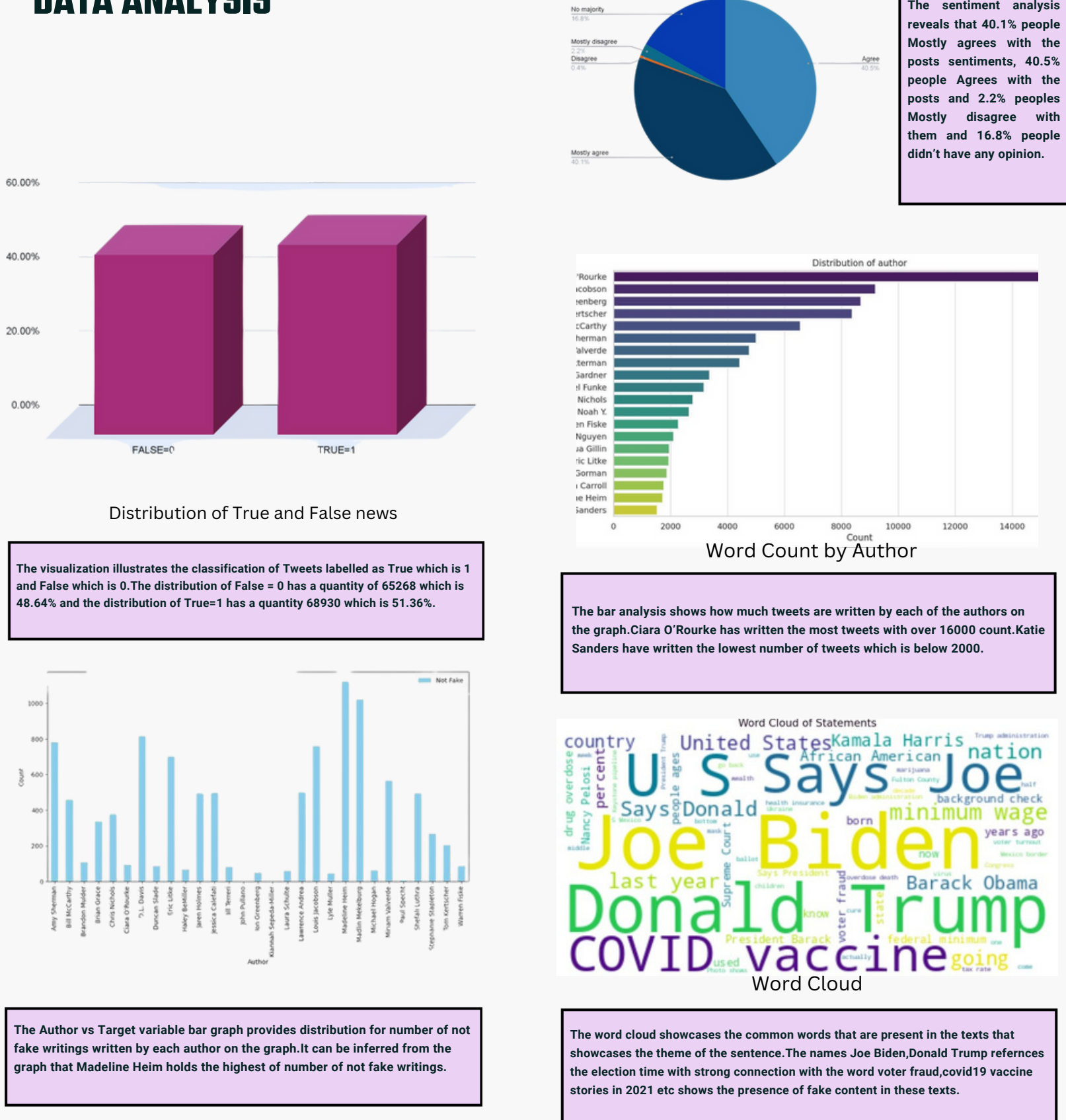
LITERATURE REVIEW

This literature review covers various aspects of Deep Learning methods, Social media post analysis and fake news exploration. It examines various tweets and fake news patterns and large language models to detect them. Key studies explore detection accuracy associated with large language models, hybrid optimization models. The review discusses information retrieval, data analysis, text processing and fake news analysis. It also addresses gaps and challenges in existing research, emphasizing the need for further exploration validation. Overall, the literature contributes insights into the evolving landscape of Deep learning and fake news detections.
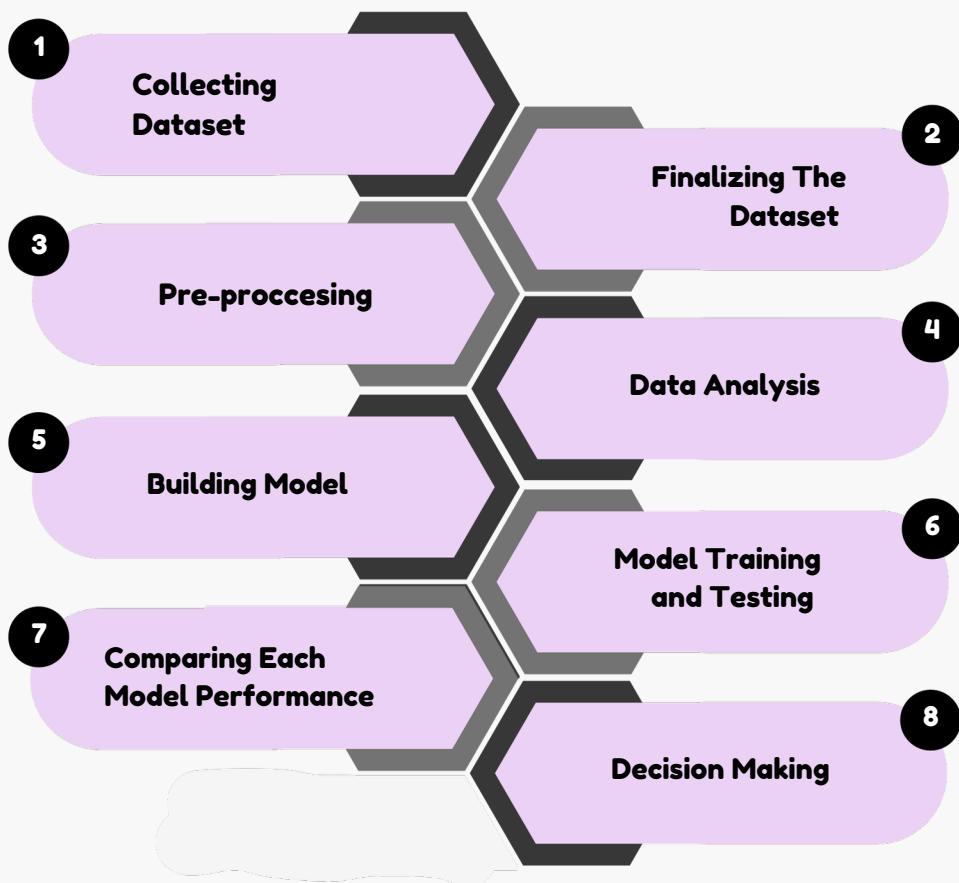
DATA PREPROCESSING

The preprocessing workflow involves several essential steps to prepare text data for deep learning models. Initially, text is cleaned and normalized by converting it to lowercase, removing punctuation and special characters, and ensuring consistency across entries. Tokenization is then applied to split the text into individual words, with the removal of irrelevant entries to refine the dataset's quality. Next, text vectorization is performed to map words to unique integer IDs. To ensure uniformity in input size, sequences are padded to a fixed length, such as 100 for LSTM and 250 for DNN models. Finally, the dataset is split into training and testing/validation sets, with the training set typically comprising 20% of the data and the remaining 80% allocated for evaluation. This division allows for robust model training and assessment on unseen data, ensuring the model's generalization ability.
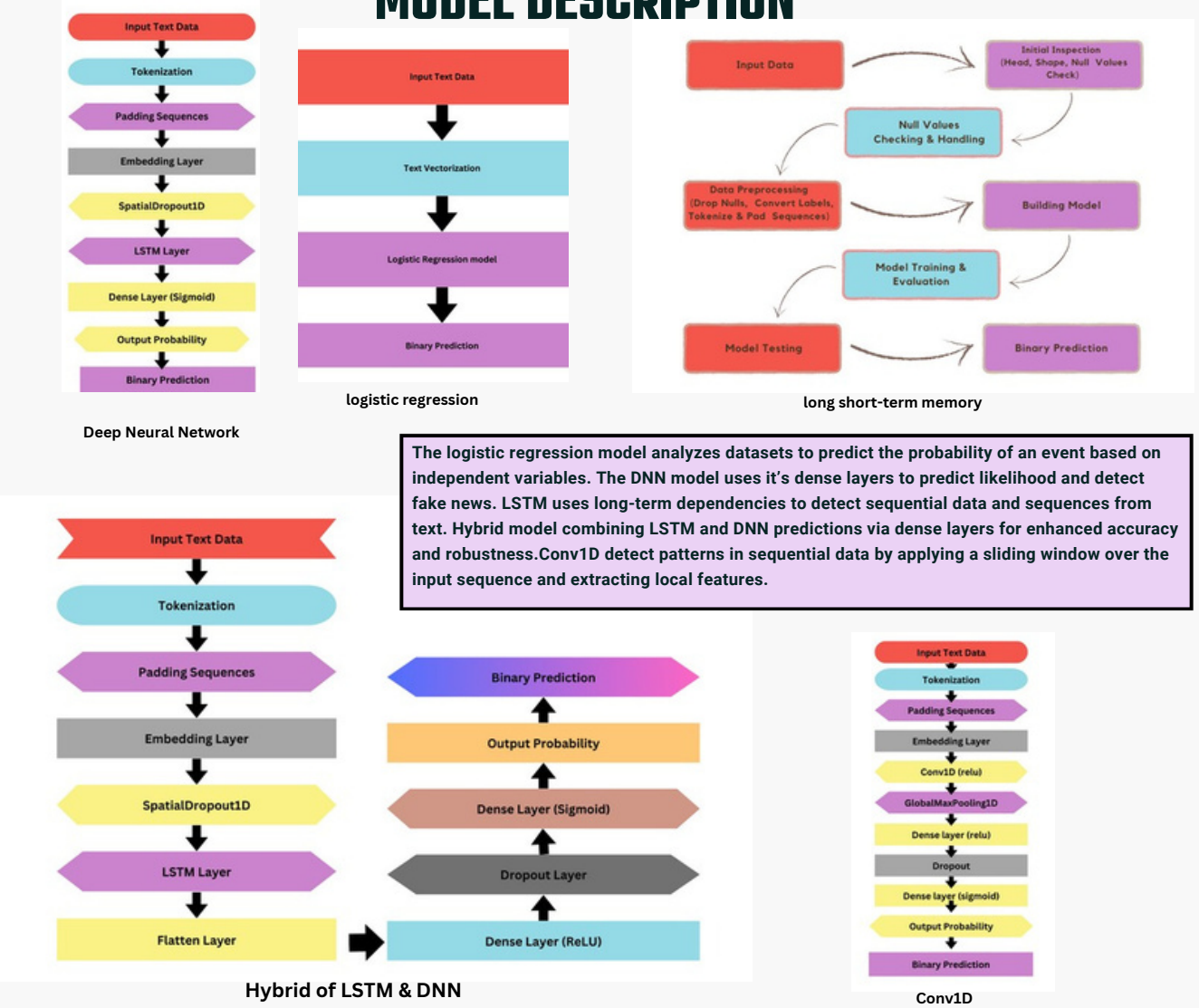
DATA ANALYSIS

Distribution of True and False news

The visualization illustrates the classification of Tweets labelled as True which is 1 and False which is 0.The distribution of False = 0 has a quantity of 65268 which is 48.64% and the distribution of True=1 has a quantity 68930 which is 51.36%.

The Author vs Target variable bar graph provides distribution for number of not fake writings written by each author on the graph.It can be inferred from the graph that Madeline Heim holds the highest of number of not fake writings.

The sentiment analysis reveals that 40.1% people Mostly agrees with the posts sentiments, 40.5% people Agrees with the posts and 2.2% peoples Mostly disagree with them and 16.8% people didn't have any opinion.

Word Count by Author

The bar analysis shows how much tweets are written by each of the authors on the graph.Ciara O'Rourke has written the most tweets with over 16000 count.Katie Sanders have written the lowest number of tweets which is below 2000.

Word Cloud of Statements

Word Cloud

The word cloud showcases the common words that are present in the texts that showcases the theme of the sentence.The names Joe Biden,Donald Trump refernces the election time with strong connection with the word voter fraud,covid19 vaccine stories in 2021 etc shows the presence of fake content in these texts.

MODEL DESCRIPTION

Deep Neural Network

logistic regression

long short-term memory

The logistic regression model analyzes datasets to predict the probability of an event based on independent variables. The DNN model uses it's dense layers to predict likelihood and detect fake news. LSTM uses long-term dependencies to detect sequential data and sequences from text. Hybrid model combining LSTM and DNN predictions via dense layers for enhanced accuracy and robustness.Conv1D detect patterns in sequential data by applying a sliding window over the input sequence and extracting local features.

Hybrid of LSTM & DNN

Conv1D

Performance Metric Value For All Model:

MODEL	AVG VALIDATION ACCURACY	AVG ACCURACY	AVG PRECISION	AVG RE-CALL	TIMING
Logistic Regression	0.79924	0.99608	0.80060	0.81302	Less than 5 Min
DNN	0.9983	0.99977	0.99978	0.99978	More than 1 Hours
LSTM+DNN	0.99966	0.99825	0.99949	0.99985	More than 1 Hours
Conv1D	0.9989	0.9998	0.9998	0.9998	Less than 1 Hours
LSTM	0.99974	0.9982	0.99854	0.9979	Less than 30 Min

WORK FLOW DIAGRAM

1. Collecting Dataset
2. Finalizing The Dataset
3. Pre-procceing
4. Data Analysis
5. Building Model
6. Model Training and Testing
7. Comparing Each Model Performance
8. Decision Making

CONCLUSION

This thesis presents detection of fake news utilizing deep learning, with a literature review emphasizing the gaps fake news detection and showing the ability to detect them using different deep learning methods. Also creating a vast area to research further to develop more efficient and optimized models.

REFERENCES

1. Cao, J., Sheng, Q., Qi, P., Zhong, L., Wang, Y., & Zhang, X. (2019). False news detection on social media. arXiv (Cornell University). https://doi.org/10.48550/arxiv.1908.10818

2. Ruchansky, N., Seo, S., & Liu, Y. (2017). CSI. ACM Digital Library. https://doi.org/10.1145/3132847.3132877

3. Deligiannis, N., Huu, T. D., Nguyen, D. M., & Luo, X. (2018). Deep learning for geolocating social media users and detecting fake news. In NATO Workshop.