International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

FAKE DETECT: A DEEP LEARNING ENSEMBLE MODEL FOR FAKE NEWS DETECTION

[1] A.Bhagya Sree, [2] Kadthala Rasagna,[3] Vasalamarri Likitha, [4] Chintalapudi Sai Keerthi,

[5] S.Shireesha

[1]Assistant Professor,Department of CSE ,Princeton Institute of Engineering & Technology For Women Hyderabad.

[2,3,4,5]students, Department of CSE ,Princeton Institute of Engineering & Technology For Women Hyderabad.

ABSTRACT

Pervasive usage and the development of social media networks have provided the platform for the fake news to spread fast among people. Fake news often misleads people and creates wrong society perceptions. The spread of low-quality news in social media has negatively affected individuals and society. In this study, we proposed an ensemble-based deep learning model to classify news as fake or real using LIAR dataset. Due to the nature of the dataset attributes, two deep learning models were used. For the textual attribute "statement," Bi-LSTM-GRU-dense deep learning model was used, while for the remaining attributes, dense deep learning model was used. Experimental results showed that the proposed study achieved an accuracy of 0.898, recall of 0.916, precision of 0.913, and F-score of 0.914, respectively, using only statement attribute. Moreover, the outcome of the proposed models is remarkable when compared with that of the previous studies for fake news detection using LIAR dataset.

I.INTRODUCTION

Progression and advancement of the hand-held devices and high-speed Internet have exponentially increased the number of digital media users. According to digital global report 2020, the number of users for digital media reached 4.75 billion, and the social media users reached 301 million in 2020 [1]. This digitalization converts the world into the global village. Due to this advancement, individuals are just one click away from the information worldwide. Despite several advantages, this transformation has raised some challenges. Fake news is one of the challenges faced by the digital community nowadays. Fake news is pervasive propaganda that spreads misinformation online, using social media like Facebook, twitter, and Snapchat to manipulate public perceptions. Social media can have two sides for news consumption, i.e., can be utilized to update the community about the latest news and, on the other hand, can be a source of spreading false news. However, social media is a low cost, quick access, and fast distribution of news and information and to know what is happening worldwide. Moreover, due to its simplicity and lack of control on the Internet, it allows "fake news" to be widespread. Fake news

has become a focal point of discussion in the media over the past three years due to its impact on the 2016 US Presidential election [2]. Reports showed that human's capability for detecting deception without special assistance is only 54% [3]. Therefore, there is a need for an automated way to classify fake and real news accurately. Some studies have been conducted but still there is a need for further attention and exploration. The proposed study attempts to eliminate the spread of rumors and fake news and helps people to identify the news source as trustworthy or not by automatically classifying the news. The organization of this paper is as follows. Section 2 includes a review of previous studies. Section 3 explains the proposed methodology, which contains the "LIAR" dataset description, preprocessing, and classification models used. Section 5 includes experimental setup results and discussion. Finally, Section 6 contains the conclusion of this paper.

II.LITERATURE SURVEY

One of the earlier studies on fake news detection and automatic fact-checking with more than a thousand samples was done by [4] using LIAR dataset. The dataset contains 12.8 K human-labeled short statements from POLITIFACT.COM. The statements were labeled in six different categories, such as pants fire, false, barely true, half true, mostly true, and true. The study used several classifiers such as logistic regression (LR), support vector machine (SVM), a bidirectional long short-term memory (Bi-LSTM) networks model), and a convolutional neural network (CNN) model. For LR and SVM, the study used the LIBSHORTTEXT toolkit and showed significant performance on short text classification problems. The study compared several techniques using text features only and achieved an accuracy of 0.204 and 0.208 on the validation and test sets. Due to overfitting, the Bi-LSTMs did not show good performance. However, the CNN outperformed all models, resulting in an accuracy of 0.270 on the holdout data splitting. Similarly, another study compared three datasets such as LIAR datasets, fake or real news dataset [5], and the dataset generated by collecting fake news and real news from Internet [6]. The study made a comparison among various conventional machine learning models such as SVM, LR, decision tree (DT), AdaBoost (AB), Naive Bayes (NB), and K nearest neighbor (KNN), respectively, using lexical, sentiment, unigram, and bigram techniques with term frequency and inverse document frequency (TF-IDF). Furthermore, several CNN models such as NN, CNN, LSTM, Bi-LSTM, hierarchical attention network (HAN), convolutional HAN, and character level C-LSTM were also used with Glove embedding and character embedding to train the model. They found that the performance of the LSTM model highly depends upon the size of the dataset. The result showed that NB, with n-gram (bigram TF-IDF), features produced the best outcome of approximately 0.94 accuracy with the combined corpus dataset. Conversely, the study by [4] indicated that the CNN model outperformed the LIAR dataset. However, the study by [6] showed that the CNN model is the second-best for all the datasets. The NB model showed the best performance for

the LIAR dataset with 0.60 accuracy and 0.59 F1-score. For the fake or real news, dataset Char-level C-LSTM showed the best performance with 0.95 accuracy and 0.95 F1-score. LSTM-based models showed the best outcome on the combined corpus dataset, where both Bi-LSTM and C-LSTM produced an accuracy of 0.95 and F1-score of 0.95. Furthermore, another study was performed by Girgis et al. [3] regarding the spread of fake news and used recurrent neural network (RNN) models (Vanilla RNN, Gated Recurrent Unit (GRU)) and long short-term memories (LSTMs) on the LIAR dataset to predict fake news. They compared and analyzed their results with Wangs [4] findings. Although similar results were achieved, GRU (0.217) outperformed the other models. Nevertheless, in comparison with the findings of Wang, they found that CNN is better in terms of speed and outcomes. Similarly, the authors in [7] used the LSTM model on LIAR dataset. They found that adding the speaker profile enhances the performance of the algorithm. The model achieved an accuracy of 0.415.

III.EXISTING SYSTEM

The detection of fake news has become an increasingly important challenge in the era of digital media and social networking. Existing fake news detection systems often employ traditional machine learning algorithms, including Support Vector Machines (SVM), Naive Bayes, and Logistic Regression. These systems primarily rely on feature extraction methods such as TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings to represent news articles and classify them as fake or real. However, these systems face several limitations. One major drawback is that they often struggle to handle the complexity of natural language, such as sarcasm, irony, or the subtle use of language in fake news articles. In many existing systems, there is an over-reliance on handcrafted features, such as word frequency, linguistic cues, and metadata, without considering the broader context or deeper semantics of the articles. This results in reduced accuracy, particularly when the dataset contains a large amount of ambiguous or context-dependent information. Furthermore, while some systems use deep learning models, such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), these models are often used in isolation without leveraging the full potential of combining multiple models or ensemble learning approaches. Ensemble learning techniques, where multiple models are trained and their predictions are aggregated, have proven effective in improving the accuracy and robustness of fake news detection systems.

IV.PROPOSED SYSTEM

The proposed system, Fake Detect, introduces a deep learning ensemble model for fake news detection. The core idea is to combine multiple advanced deep learning models to improve the accuracy and generalizability of fake news classification. The system incorporates a hybrid approach that combines Convolutional Neural Networks (CNN) for capturing local patterns in text, Recurrent Neural Networks (RNN), particularly Long Short-Term Memory (LSTM) networks, for capturing sequential

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

relationships in news articles, and Transformer-based models like BERT (Bidirectional Encoder Representations from Transformers) for understanding context and semantic meaning at a deeper level. The ensemble approach ensures that the system benefits from the strengths of various architectures while mitigating their individual weaknesses. CNNs excel at identifying local word-level patterns and extracting useful features, which is useful for distinguishing fake news articles that often rely on misleading patterns. RNNs, particularly LSTMs, are designed to capture long-range dependencies and sequential relationships in text, which is valuable for understanding the flow and structure of news articles. Meanwhile, Transformer-based models like BERT have the ability to process contextual information bidirectionally, enabling them to capture deeper semantic meanings, even when the language is subtle or ambiguous.

V.SYSTEM ARCHITECTURE

The dataset contains two types of feature such as short textual feature, i.e., statement and other features like speaker job title, subject, and venue. Therefore, the features were initially divided according to the category. For the statement attribute, several NLP techniques like tokenization, lemmatization, and stop word removal were used. However, for the other category of features, different data preprocessing techniques were applied that will be discussed further in the preprocessing section.

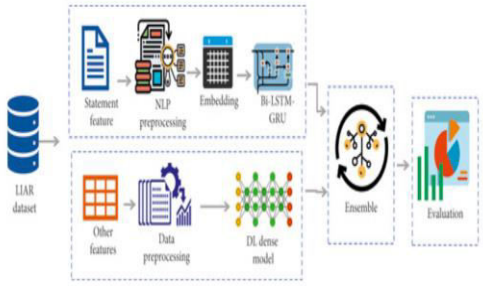
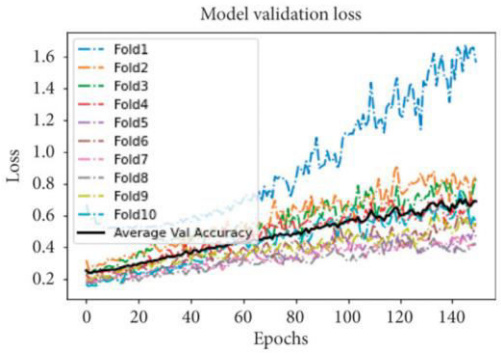
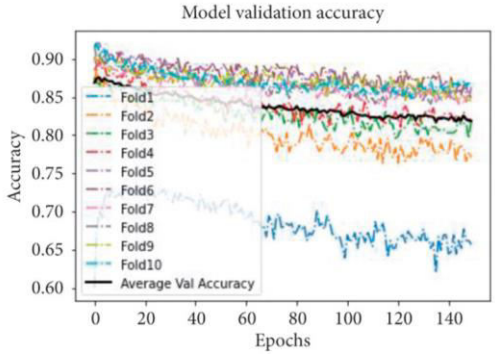
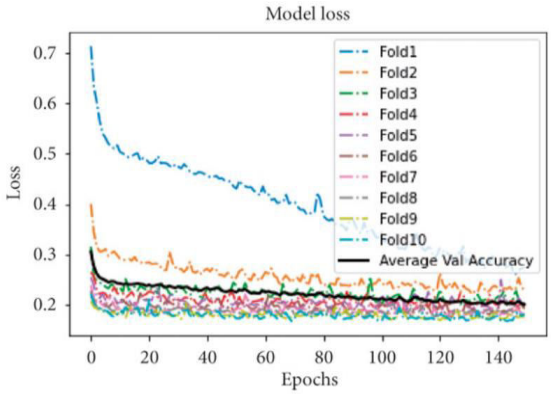


Figure 5.1 Architecture

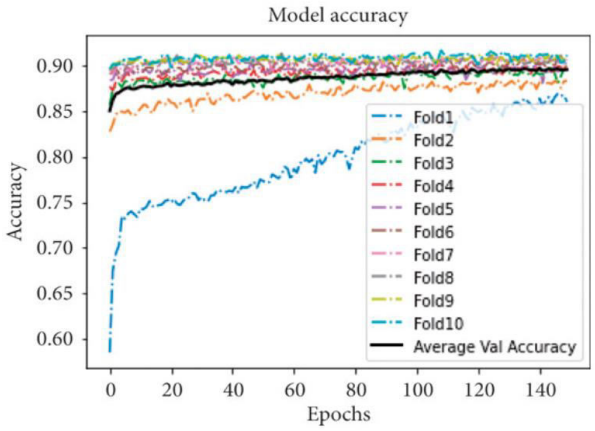
VI.OUTPUT SCREENSHOTS







International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



VII.CONCLUSION

The Fake Detect system, based on an advanced deep learning ensemble model, represents a significant step forward in the detection of fake news. By combining multiple deep learning architectures, including CNNs, LSTMs, and Transformer-based models like BERT, the system is able to process news articles with greater accuracy and handle the complexity of natural language. The use of attention mechanisms further enhances the system's ability to focus on the most relevant parts of an article, making it more effective at detecting subtle patterns in fake news. Furthermore, the integration of meta-information such as source credibility and user engagement adds an extra layer of sophistication, improving the system's overall robustness and adaptability. The Fake Detect model is not only capable of identifying fake news with high accuracy but can also adapt to evolving language patterns and new forms of misinformation. This makes it a valuable tool for media platforms, fact-checkers, and even social media users who wish to verify the authenticity of news content. As fake news continues to proliferate across various digital platforms, systems like Fake Detect will play a crucial role in combating misinformation and helping people make informed decisions.

VIII.FUTURE SCOPE

While the proposed Fake Detect system demonstrates promising results in fake news detection, there are several avenues for further improvement and expansion. One key area for future development is the incorporation of multimodal data, such as images, videos, and social media interactions, into the detection process. Many fake news stories today are accompanied by misleading or doctored images and videos, which can be as deceptive as the textual content. Integrating image and video analysis with text-based analysis would provide a more holistic approach to fake news detection. Another avenue for future work is the adaptation of the system to different languages. While current models are primarily trained on English-language datasets, fake news is a global issue, and the system must be capable of detecting misinformation in multiple languages. Multilingual models, such as mBERT (Multilingual BERT) or XLM-RoBERTa, could be explored to extend the system's reach and effectiveness in diverse linguistic contexts. Additionally, real-time fake news detection is an area where the system could be improved. As news spreads rapidly on social media platforms, it is critical to identify fake news in real-time, often within hours or minutes of publication. This requires the system to be highly efficient and capable of processing large

volumes of data in a timely manner. Optimizing the ensemble model for faster processing and scaling it to handle real-time news feeds would significantly increase its usefulness.

IX.REFERENCES

1. Ruchansky, N., Seo, S., & Liu, Y. (2017). "CSI: A Hybrid Deep Model for Fake News Detection." Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.

2. Patel, S., & Purohit, H. (2020). "Fake News Detection using Deep Learning: A Review." Journal of Artificial Intelligence.

3. Xu, H., & Li, L. (2018). "A Survey on Fake News Detection Techniques." IEEE Access.

4. Wang, W., et al. (2020). "Multi-Level Attention Networks for Fake News Detection." Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics.

5. Yang, X., et al. (2019). "A Survey on Fake News Detection: Techniques and Trends." Future Generation Computer Systems.

6. Khandelwal, M., & Agerri, R. (2021). "Ensemble Learning Models for Fake News Detection." International Journal of Machine Learning and Cybernetics.

7. Lazer, D. M. J., et al. (2018). "The Science of Fake News." Science.

8. Devlin, J., et al. (2018). "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." NAACL-HLT.

9.Jin, Z., et al. (2020). "Fake News Detection on Social Media: A Data Mining Perspective." Proceedings of the 2020 IEEE International Conference on Data Mining.

10. Zhang, X., et al. (2020). "Deep Learning for Fake News Detection on Social Media." IEEE Transactions on Knowledge and Data Engineering.