Fake News Detection

Ankitha Sudarshan[1], **Chaitanya Deepthi Chadalavada**[2]
[1,2]Purdue University, Fort Wayne
{sudaa01,chadc01}@pfw.edu

Abstract

Fake news has become a prevalent issue in today's world, particularly in the digital age where information can be easily disseminated through various social media platforms. In recent years, deep learning models such as Convolutional Neural Networks (CNNs) and BERT (Bidirectional Encoder Representations from Transformers) have shown significant success in this field. In this study, we compare the performance of BERT-based models, traditional CNNs and an ensemble model for the task of fake news detection using a large-scale benchmark dataset. We evaluate the performance of these models using standard classification metrics such as accuracy, precision, recall, and F1-score.

1 Introduction

With the lack of gate-keeping in the digital age, misinformation spread can have serious consequences, ranging from misinformed public opinions to political manipulation and even violence. This has significant implications for our society, as it can manipulate public opinion, mislead people, and even incite violence. Therefore, the identification of fake news has become a crucial issue, and researchers are actively exploring various approaches to tackle this problem. With the advancement of Natural Language Processing (NLP) techniques, there has been growing interest in developing automated methods for identifying fake news.

The goal of identifying fake news through different models will help us to develop automated techniques that can accurately distinguish between real and fake news articles. This can help prevent the spread of misinformation and promote fact-based journalism. By addressing this problem, we can help to safeguard the integrity of news reporting and promote a more informed and educated public.

2 Related Work

Researchers have proposed various techniques for detecting fake news, ranging from rule-based methods to machine learning and deep learning models.

One approach proposed by Shu et al. (2017)(1) is based on data mining techniques that use linguistic features, user reputation, and propagation patterns to identify fake news. Aldwairi and Alwahedi (2018)(2) proposed a machine learning-based approach that uses sentiment analysis, user reputation, and text-based features to detect fake news. Zhou and Zafarani (2020)(3) provided a comprehensive survey of fake news, covering fundamental theories, detection methods, and opportunities. Seo et al. (2021)(4) conducted a comparative study of several BERT-based models for fake news detection. They investigated the impact of different pre-training objectives and fine-tuning strategies on the performance of the models. Finally, Singh et al. (2021)(5) provided a survey of deep learning techniques for fake news detection. They reviewed various deep learning architectures and preprocessing techniques that have been proposed for detecting fake news. These papers demonstrate the importance of developing effective techniques for detecting fake news and highlight the need for continued research in this area.

3 Methodology

We came up with three approaches to this problem statement namely - BERT, traditional CNNs, and an ensemble model which involves traditional machine learning models and evaluates their effectiveness for fake news detection using a benchmark dataset and provide insights into their respective strengths and weaknesses. BERT, a pre-trained transformer model, has shown impressive results in various NLP tasks, including natural language inference and sentiment analysis. In contrast, CNNs are a widely-used neural network architecture for

NLP, particularly for text classification tasks. The ensemble technique is used to give us a reflection of the performance of a combination of several machine learning models.

The BERT-base-uncased, CNN models and the ensemble models are trained on the training set of the FakeNewsNet dataset and evaluated on the test set using standard classification metrics. We fine-tuned the BERT model on the dataset and use it to generate features for classification using a fully connected layer. In the BERT model, we have added 3 linear ReLu layers to the pre-trained bert-base-uncased model to perform a specific downstream task like binary classification, addition of a new classifier layer to the pre-trained BERT model is necessary. In contrast, we use traditional CNN models with multiple convolution layers followed by max-pooling and a fully connected layer for classification. We created a custom CNN model with 6 convolutional layers and 2 dense layers along with dropouts. This model was coupled with GloVe embeddings of dimension 100. The ensemble technique involved a combination of three classifiers namely - Passive Aggressive Classifier,Multinomial Naive Bayes and the Random Forest Classifier. The choice of these machine learning models was a conscious choice made as each of the models have a unique strength and complement each other.

Note that the choice of model architecture and specifics can have an impact on performance.

4 Dataset

The dataset we are using consists of real and fake news articles from Kaggle. It is a large-scale dataset with 20800 articles that contains 2 sub-datasets, each representing: (1) fake news with 10387 articles(0), (2) real news(1) consisting of 10413 articles.

5 Performance Metrics

We evaluate the performance of these models using standard classification metrics such as accuracy, precision, recall, and F1-score. Our experimental results and the findings of this study can contribute to the development of more accurate and reliable fake news detection systems and aid in addressing the spread of misinformation in digital media.

6 Experiments - The CNN Approach

We performed an experiment with the CNN model by running it on 3 and 5 epochs. We did so as it would give us a fair idea of the model's behavior and the impact of epoch number on the results.It made way for us to the process of hyper-parameter tuning. The model run for 3 epochs gave us an accuracy of 63% on the test data and gave an F1-score of 55% for the Fake news articles and 67 for the real articles during prediction. The model run for 5 epochs gave us a slightly better accuracy of 65% on the test data. However we observed that this model was over-fitting the data. This means that the model, instead of learning the data , began to memorize the data. This would not give us good results as the model would not be able to generalize well.

The following graphs Figure 1 and Figure 4 illustrate the model's accuracy and loss of Fake and real news on 3 and 5 epochs respectively.

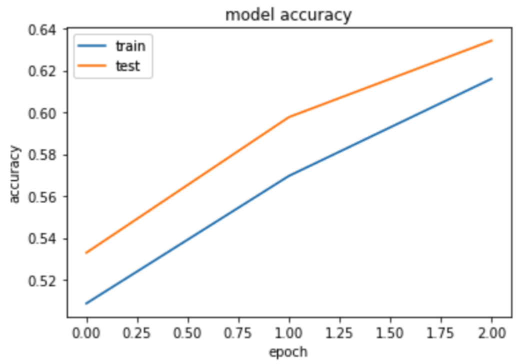

Figure 1: Model Accuracy Graph - 3 epochs

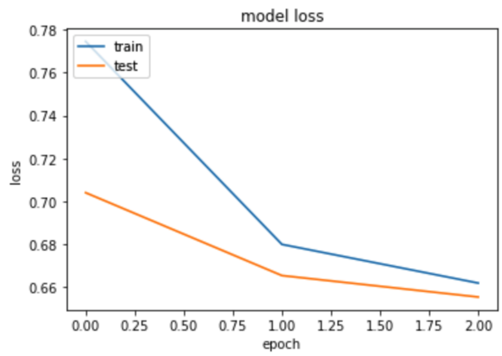

Figure 2: Model Loss Graph - 3 epochs

7 Analysis Done - The CNN Approach

The results indicate that the model trained on 3 epochs performs better at identifying non-fake articles than fake articles, as shown by the higher

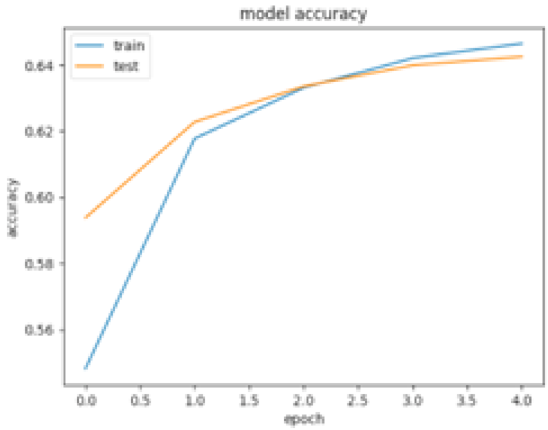

Figure 3: Model Accuracy Graph - 5 epochs

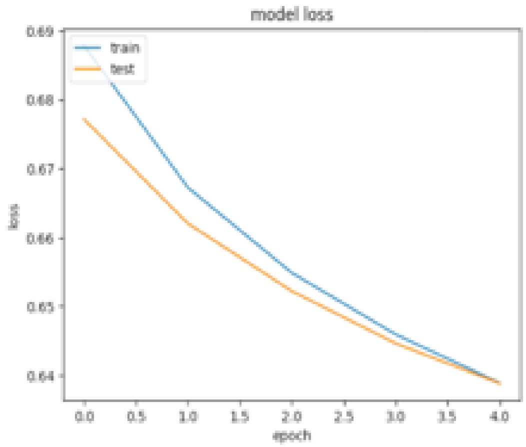

Figure 4: Model Loss Graph - 5 epochs

precision for non-fake articles and the higher recall for fake articles. The addition of GloVe embeddings definitely bolstered the model's prediction capability however, the overall performance of the model is mediocre, with an accuracy of 0.63, indicating that there is room for improvement with the model run on 3 epochs. The precision of 0.77 for real articles and The recall of 0.60 for non-fake articles.

On the other hand, the precision of 0.47 for fake articles and the recall of 0.67 for fake articles.

8 Results and Analysis - The CNN Approach

The methods used to generate fake news are constantly evolving, and new techniques are being developed all the time. This means that CNN models need to be constantly updated and retrained to keep up with the latest trends in fake news generation. Also using an advanced BERT model to identify Fake news articles might help in building more accurate results for a larger scale of news article data. So we are planning to train and test the same dataset with a BERT-based model and try to add more training data to check for possible shortcomings in our analyses.

9 Problems Encountered - The CNN Approach

We observed a slight bias in the training dataset which might have a slight impact on the results. Natural language is often ambiguous and can have multiple interpretations. As we saw, the model is sensitive to the training as we saw a variation when we changed the number of epochs. The model trained on 5 epochs began to over-fit the data.

CNN models in general are very sensitive to hyperparameter tuning and the slightest of variations can lead to variations in the model, which is quite tricky. The call of the number of layers to be added or removed is tricky and significantly impacts the results as too many layers or too less of layers might either end up underfitting or overfitting the data.

10 Experiments - The BERT Approach

The results obtained clearly indicate that the BERT-based uncased model performed better than the CNN model used in the previous step with an accuracy of 80.7%.

The test data we used consists of News articles with complex language which is difficult to classify, but this model classified well with an average accuracy of about 82%. Although this is not a high accuracy value, the model classified real and fake data well where 8 out of 10 times the predictions and the true values matched which meant that the model had a good recall rate. This can also be an indicator that the model is not overfitting on the data.

Overall, from the results, there is still some room for improvement to effectively detect fake news.

11 Results and Analysis - The BERT Approach

As part of the current results and analysis, we created a custom architecture on top of the pre-trained BERT architecture where we defined the 3 linear ReLu layers to customize BERT for our binary classification task. The prediction of this model for lengthy texts of length >=440 words is 62.7% for fake data and 73.3% for real data. The following graphs Figure 3 and Figure 4 illustrate the accuracy and loss of Fake and real news using BERT based model. An interesting observation we made was that the test accuracy decreased from 82% to 80% during the course of the task. Although the accuracy showed such a pattern, BERT still performed better on input data of greater lengths. When we went on another iteration to retrain the model , we observed that the accuracy showed a positive correlation this time and increased from 78% to 80%. Both the times the accuracy came to about 80% .

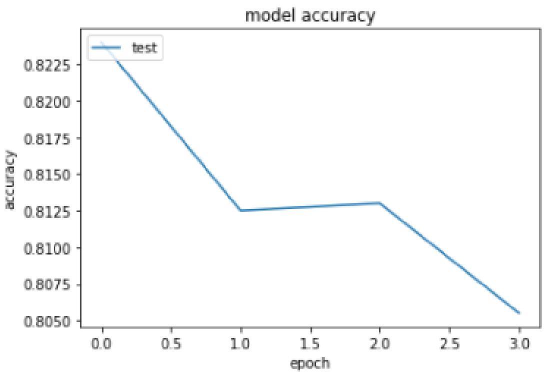

Figure 5: Model Accuracy Graph - Initial Training

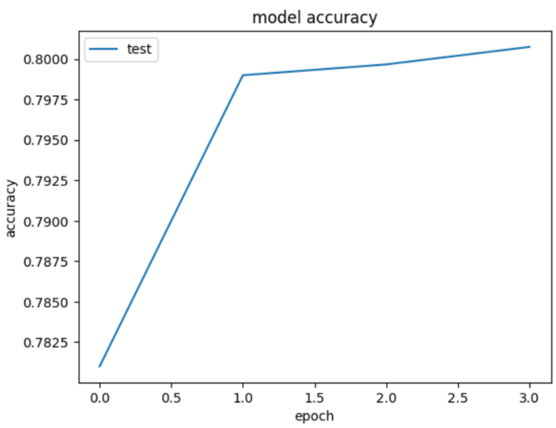

Figure 6: Model Accuracy Graph - Post Retraining

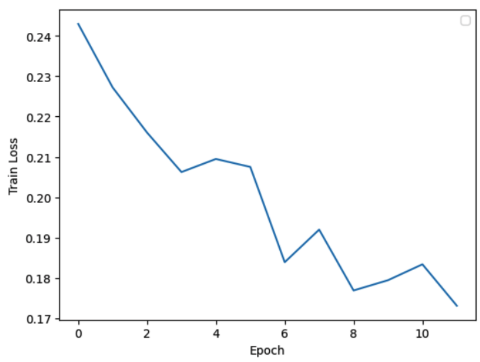

Figure 7: Model Loss Graph

As we continued to use the same dataset as of CNN model to identify fake articles, we progressed to use a BERT-based uncased model this time instead of a CNN model and found an increase in the performance of the model by about 17%. For this model, the loss graph is gradually decreased to reach a minimum which shows that the model is not over fitting the data and the accuracy of the model stood at 80.7%.

12 Problems Encountered - The BERT Approach

We initially chose to use chunks of data - one with a size of 5000 records and the other with the full dataset of 20800 records to ease the computation and to compare the working of the model with less and more data. The latter gave us better results as the model trained on more data, which naturally understands better and would have seen more varied kinds of data and hence used this for further analysis and experimentation. We did fall short of computational resources and did resort to using the cloud to obtain our required results.

13 Results - The Ensemble Approach

The ensemble model consists of three individual models namely - Passive Aggressive Classifier(PAC), Multinomial Naive Bayes(MNB) and Random Forest(RF) Classifier.This model interestingly gave a high accuracy of 90.1%. It is a little surprising as, this is an ensemble of traditional machine learning models as opposed to our other two models which were more advanced in terms of architecture.

14 Analysis - The Ensemble Approach

The PAC model appears to have the highest accuracy of the three individual models, as it has a high number of true positives and true negatives and relatively low numbers of false positives and false negatives. The MNB model has a very high number of true positives, but a relatively high number of false positives, which indicates that it may be more prone to classifying instances as positive when they are actually negative. The RF model has a high number of true negatives, but a relatively high number of false positives, which suggests that it may be more prone to classifying instances as negative when they are actually positive. This suggests that the ensemble model is able to effectively combine the strengths of the individual models to achieve higher performance in accurately classifying instances.

15 Overall Results

Model	Accuracy
CNN - 3 epochs	0.63
CNN - 5 epochs	0.65
BERT Model	0.8
Ensemble model	0.9

Figure 8: Model evaluation

16 Problems Encountered - The Ensemble Approach

The approach of using an ensemble model to combine the predictions of multiple weak and strong learners does have potential problems. Overfitting may occur if the individual models are trained too much on the training data, while correlated predictions may result in a less effective combination of the models' strengths. Imbalanced data and poor feature selection can also lead to biased predictions and poor performance. Inconsistencies in labeling can create noisy or misleading predictions, and a limited diversity of models may hinder the ability of the ensemble model to achieve higher performance. To mitigate these issues, appropriate methods such as regularization, diverse model selection, addressing data imbalance, careful feature selection, and improving labeling consistency should be used.

17 Future Work

The dataset used in the paper is a bit older version from Kaggle and it can be made modified according to the latest news article data to see how the models perform. Currently, most fake news detection models focus on analyzing text-based data. However, fake news can also be spread through images, videos, and other forms of media. There is a scope to explore ways to incorporate multi-modal data sources into fake news detection models to improve their accuracy. Future research could focus on developing fake news detection models that are more robust to adversarial attacks. One could also explore ways to develop context-aware fake news detection models that can take into account factors such as the source of the content, the language used, and the social and political climate in which it is shared.

18 Conclusion

In this paper, we presented a novel approach to detect fake news using multiple models, including a fine-tuned BERT model, a CNN model, and an ensemble model. Our results revealed that the fine-tuned BERT model achieved a remarkable accuracy of 80%, while the CNN model had an accuracy of around 63-65%. However, the ensemble model significantly outperformed both individual models, achieving an accuracy of 90%. We performed a comprehensive analysis of the individual models' performance, allowing us to gain a better understanding of their strengths and weaknesses. Our findings highlight the potential of combining different models for the fake news detection task, and suggest that fine-tuning pre-trained language models like BERT can be an effective approach. Future research could focus on improving the accuracy of the CNN model or exploring other novel methods to further improve the performance of the ensemble model.

References

[1] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD explorations newsletter*, vol. 19, no. 1, pp. 22–36, 2017.

[2] M. Aldwairi and A. Alwahedi, "Detecting fake news in social media networks," vol. 141, 2018, pp. 215–222.

[3] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Computing Surveys (CSUR)*, vol. 53, no. 5, pp. 1–40, 2020.

[4] J. Seo, S. Jung, S. Lee, and K. Park, "Bert for fake news detection: A comparative study," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2021, pp. 2479–2482.

[5] H. Singh, H. Gupta, P. Singh, and P. Bhattacharya, "Fake news detection using deep learning: A survey," *arXiv preprint arXiv:2104.08436*, 2021.

Github Link: https://github.com/deepthi-912/NLP_FakeNews