ABSTRACT

This study explores methods for identifying and mitigating the spread of misinformation regarding COVID-19 through the implementation of advanced machine learning techniques. Misinformation during the pandemic has posed significant challenges to public health efforts, highlighting the urgent need for effective fake news detection systems. The research begins with meticulous data collection from reliable sources to build a comprehensive dataset, followed by rigorous preprocessing to enhance data quality. Various predictive models, including XGBoost and Multinomial Naïve Bayes Classifier are employed to classify news as either fake or genuine. The models are fine-tuned through hyperparameter optimization to maximize their performance.

Evaluation metrics such as accuracy, precision, and recall are utilized to assess the effectiveness of each model. Results indicate that the chosen models successfully identify fake COVID-19 news with high precision, demonstrating the potential for these techniques to combat misinformation. The findings also reveal insights into patterns of misinformation dissemination, which could inform strategies for public awareness campaigns and policymaking.

By contributing to the growing field of fake news detection, this research underscores the importance of leveraging technology to address societal challenges. Future studies could expand on these results by incorporating larger datasets or exploring additional models. The outcomes of this study have practical implications for enhancing public trust in information sources during critical times, such as a global health crisis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: INTRODUCTION

1.1. Research background

Social media, being one of the primary sources of news consumption in the current digital era, has grown into a leading mode of information dissemination. Its fast-paced, bite-sized content fosters high emotional engagement but often discourages critical thinking. As a result, users are more susceptible to misinformation and fake news, which can influence public opinion, political discourse, and decision-making processes. The widespread dissemination of false information has become a major concern, especially since fake news is increasingly mimics credible sources, making it difficult to distinguish between accurate and misleading content. Due to the rapid pace at which false information is spread on social media, it has become necessary to come up with an effective detection process to neutralize its negative effects.

1.2. Problem statements

The occurrence of fake news is a big challenge because it has the power to sway public opinion and create confusion, particularly in times of crisis. Fake news on sensitive topics such as health and politics has been known to destroy institutions and spread harmful narratives. In the case of the COVID-19 pandemic, for example, there was a burst of fake news that created public misinformation and distrust in health authorities. Social media sites, as a matter of their algorithm-driven content, have provoked the spread of fake news with unprecedented swiftness, underscoring the need for sophisticated detection and verification tools at an accelerated rate. Existing models of detection are plagued by an inability to properly differentiate between actual and fake news, prompting sophisticated work in more precise algorithms for effective identification of fake news.

1.3. Objectives of research

This research aims to:

- Evaluate the effectiveness of different algorithms in detecting sophisticated fake news during COVID-19.
- Compare the accuracy, efficiency, and robustness of various detection models.
- Identify the most suitable algorithm for reliable and scalable fake news detection.

1.4. Significant of research

The findings of this research will contribute to the ongoing efforts to combat misinformation by identifying the most effective fake news detection methods. This study is particularly relevant for social media platforms, policymakers, and fact-checking organizations, as it provides insights into algorithmic approaches that can enhance the accuracy of news verification. By improving fake news detection, this research aims to promote more reliable information dissemination and reduce the harmful effects of misinformation on society

CHAPTER II: LITERATURE REVIEW

In this paper, use deep learning models (Vanilla RNN, GRU, and LSTM) to detect fake news on social media. They test these models on the LIAR dataset, which contains short, labeled statements. GRU performs better than RNN and LSTM but is still beaten by CNNs. The study notes that LSTM is slow, while GRU is a faster and more efficient alternative.[1]

According to this paper [2], it introduces CSI, a model that detects fake news by analyzing text, user interactions, and source credibility. It has three parts: Capture (tracks user engagement), Score (rates user trustworthiness), and Integrate (combines data for better accuracy). Unlike older methods that focus on just one factor, CSI improves detection by using all three. Tests on real-world data show it outperforms other models and can be used for more than just fake news detection.

For this paper, it detects fake news in Indonesian articles using the XGBoost algorithm. Their dataset contains 500 news articles (half real, half fake), with 80% used for training. After cleaning the text, they achieve 92% accuracy by fine-tuning the model. The study suggests using larger datasets and improving text preprocessing for better results.[3]

With this paper, it uses machine learning to detect fake news on Facebook. They apply Naïve Bayes, Decision Tree, Random Forest, and Logistic Regression to a labeled dataset from GitHub. Random Forest achieves the highest accuracy (92.4%), followed by Logistic Regression (90.7%). The authors find that using n-grams and larger datasets improves accuracy and suggest applying these methods to other platforms.[4]

This paper [5] presents CSI (Capture, Score, Integrate), a model that detects fake news by analyzing text content, user responses, and source credibility. It consists of three modules: Capture, which uses LSTMs to track user engagement patterns; Score, which assigns a "suspiciousness score" to users; and Integrate, which combines these insights for classification. Unlike traditional methods that focus on one aspect, CSI merges all three for improved accuracy. Experiments on real-world datasets show CSI outperforms existing models. The study highlights its flexibility and ability to classify articles while identifying suspicious users.

This study [6] combines BERT (for understanding text) with LightGBM (for classification) to detect fake news in article titles and full texts. The model cleans data by removing symbols, links, and stopwords before converting text into embeddings. Tested on multiple datasets (including ISOT and LIAR), it outperforms traditional machine learning and deep learning models. The authors highlight the importance of fine-tuning BERT and suggest hybrid models for better accuracy.

CHAPTER III: METHODOLOGY

2.1. Data Collection

For this study, the data being used in this training and testing is all **secondary data** that have been collected for a competition using data source from social platform such as **X**, **Facebook**, **Instagram**, and etc. The data set consists of **8,000** articles, with each article labeled as either 'fake' or 'real'. It was collected from the period of COVID-19 Ethical considerations were considered by using publicly available datasets and ensuring no personal data was collected.

2.2. Data Preprocessing

The data was preprocessed:

- Removing special characters
- Convert to lowercase
- Check for missing values
- Remove extra whitespaces and line break from text
- Normalizing multiple punctuation into single punctuation
- Check Emoji usage
- Classify <URL> link

2.3. Prediction Models

We extracted text features using **BERT** and **TF-IDF** as tokenization methods to compare both of the tokenization methods while using two classifications algorithms of **Naïve Bayes classifier** and gradient boosting techniques by using the model by **XGBoost**.

2.3.1. XGBoost

With XGBoost, we use multiple parameters to change its behavior and also its usage of the resource such as:

- Tree_method: The tree construction algorithm used in XGBoost
- Device: can be set to desired device to trained on.
- Max_bin: increasing this number improve the splits
- N_jobs: change the number of cores to be used for computation
- Eval_metric: evaluation metric for validation data

- Early_stopping_rounds: with this, the model can stop training if there is no progress in n-amount of rounds

2.3.2. Naïve Bayes Classifier

For Naïve Bayes, we use its Multinomial one in order to train and test the data set which includes its parameters such as:

- Fit_prior: with this, you can change it for the model to learn class prior probabilities or a uniform prior.

2.4. Hyperparameter Tuning

In order to tune the hyperparameter for each model, we use GridSearchCV library in order to find the best hyperparameter for each model within the input we put in for it to evaluate if it is the best hyperparameter to be used.

2.4.1. XGBoost Classifier

Hyperparameter tuning for XGBoost:

- N_estimators
- Max_depth
- Learning_rate
- Subsample
- Colsample_bytree

2.4.2. Multinomial Naïve Bayes Classifier

Hyperparameter tuning for Naïve Bayes Classifier includes:

- Alpha: is a parameter for controlling the level of smoothness of applied to probabilities. Adding alpha ensures word that does not appear get a small amount of non-zero probability

2.5. Evaluation Metrics

For assessing the performance of each model, the evaluation metric we are going to be used for each model is:

Accuracy: determine the accuracy of the model's predictions

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

Precision: evaluate the quality of positive prediction of the model

$$Precision = \frac{TP}{TP + FP}$$

Recall: measure the model's ability to find all positive instances

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: evaluate the model performance through the mean of precision and recall

$$F1 - Score = \frac{2(Precision * Recall)}{Precision + Recall}$$

Confusion Matrix: evaluate the count-based performance of classification model by comparing predicted value and actual value providing a detailed breakdown of correct and incorrect classifications

ROC – AUC Score: Evaluate the model threshold-based performance by finding the **area under the curve (AUC)** in **Receiver operating characteristic curve (ROC)** to distinguish between positive classes and negative classes.

$$True\ Positive\ Rate\ (TPR) = \ Recall\ = \frac{TP}{TP + FN}$$

$$False\ Positive\ Rate\ (FPR) = \frac{FP}{FP + TN}$$

CHAPTER III: RESULTS AND DISCUSSION

3.1. Exploratory Data Analysis

In this analysis, we explore a dataset containing information about COVID-19 news articles. The dataset has 8000 records and 3 columns, with features such as tweet label, label indicating whether the article is fake or real. Our aim is to identify any patterns in the dataset between feature and feature.

- Distribution of multiple punctuation count

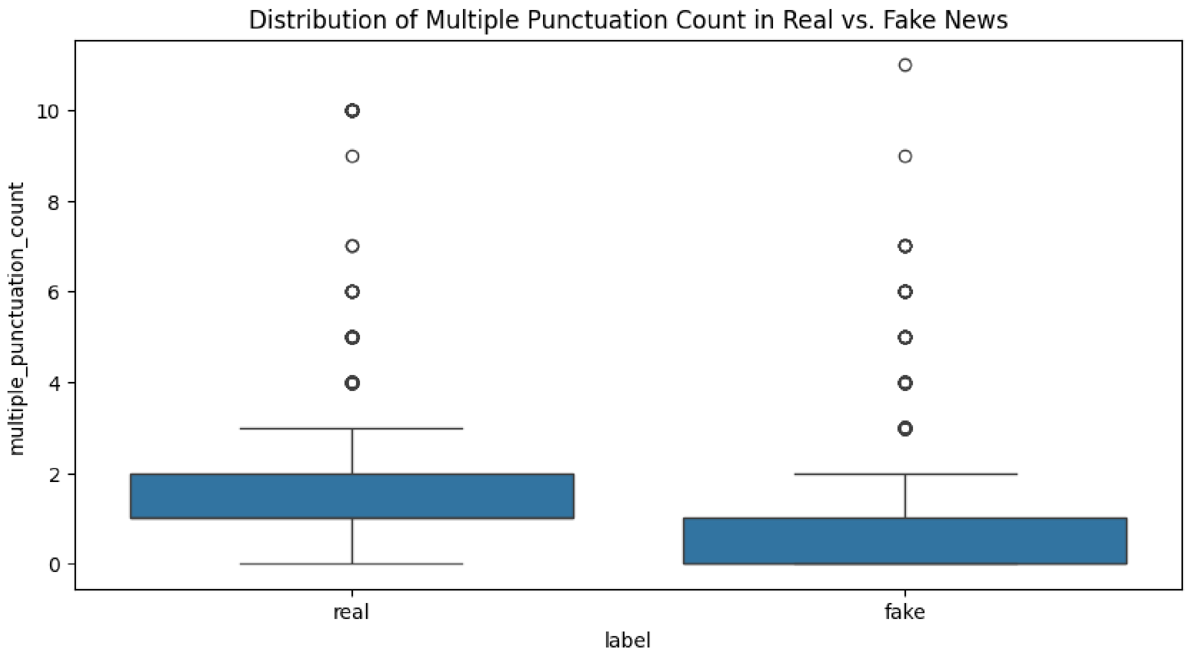

Figure 1: Distribution of Multiple Punctuation Count

For the box plot above, it describes how depending on the amount of punctuation being used in a tweet, it drives the emotional appeal or the urgency of the news toward people. Moreover, it shows that real news either have a low or high punctuation count whereas the fake news consistently have their punctuation count in the middle.

- WordCloud

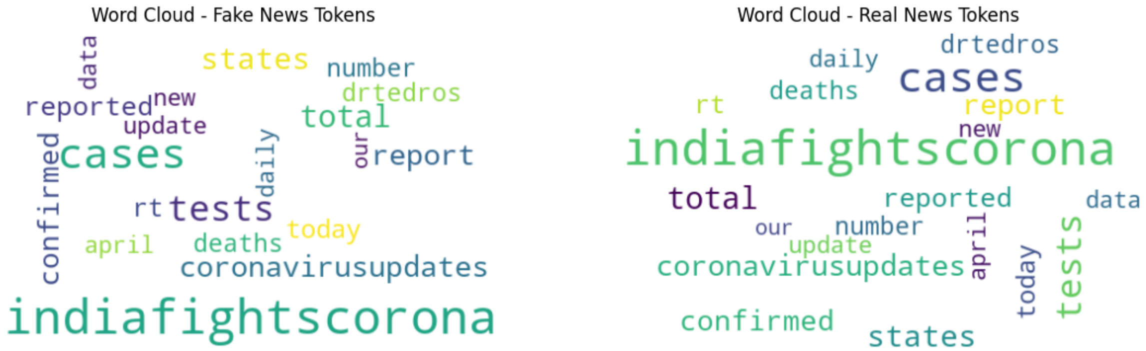

Figure 2: Word Cloud

Based on this word cloud, we see that majority of real news use precise and direct word to describe their news with no abstraction whereas in the fake news, we see that they use many abstract word to describe the news. Although, both sides have very similar pattern of word uses, but the nuances is in the way that they use the word to describe it accurately or abstractly.

- Top Token

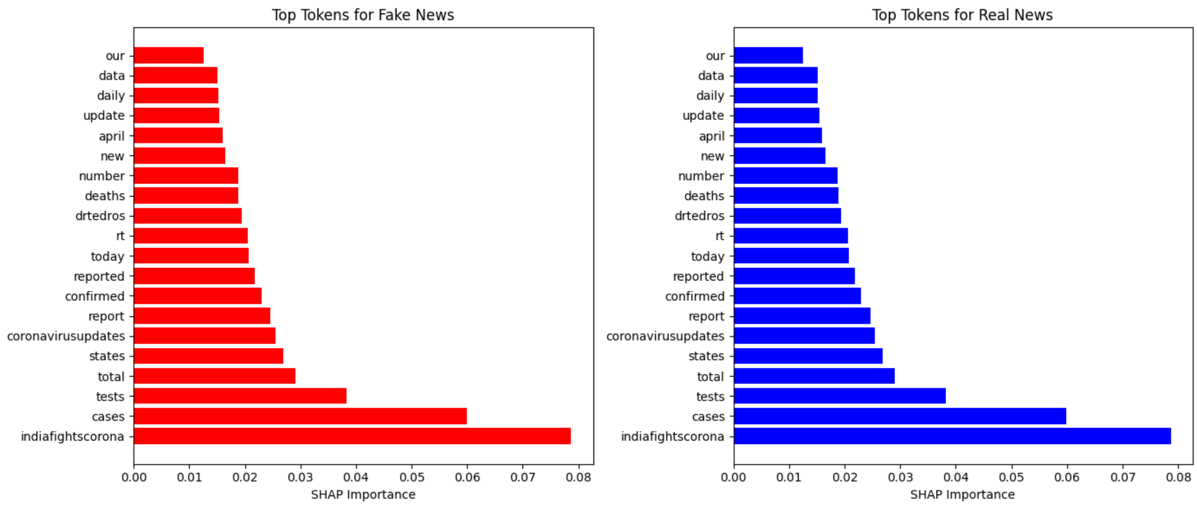

Figure 3: Top Token

We see that in this graph the overlapping of the word being used by both sides are getting harder to notice especially in the top usage of the word.

- Emoji Presence

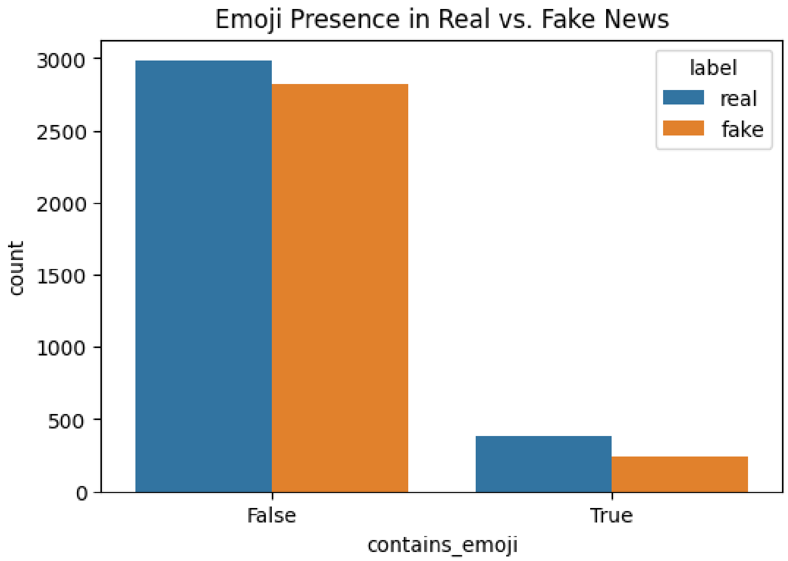

Figure 4: Emoji Presence

For this figure, we can see that especially false news really like the usage of emoji in their tweet to boost urgency of the news to the reader to give the public a panic. Meanwhile, truthful news does not use emoji that much as not using it gives off professionalism and trust.

3.2. Results of Prediction Model

3.2.1. Confusion Matrix

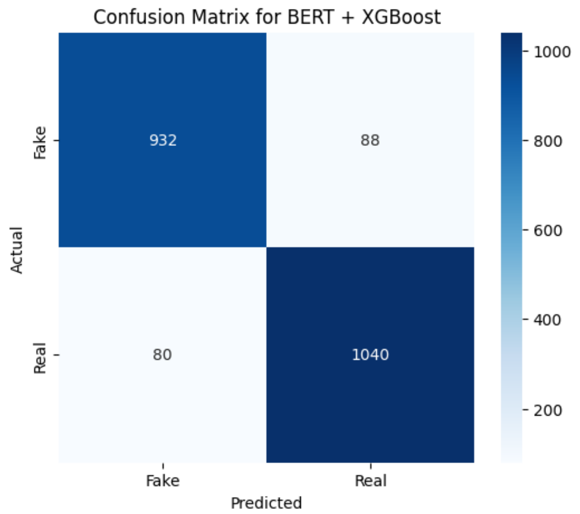

Figure 5: BERT + XGBoost Confusion Matrix

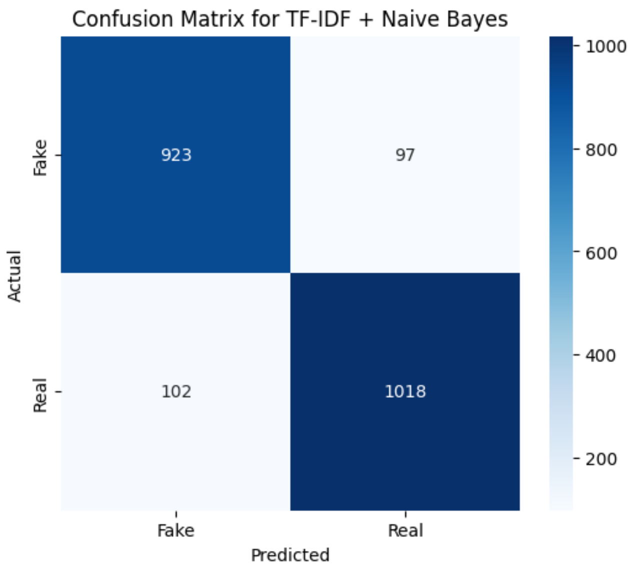

Figure 6: TF-IDF + Naive Bayes Confusion Matrix

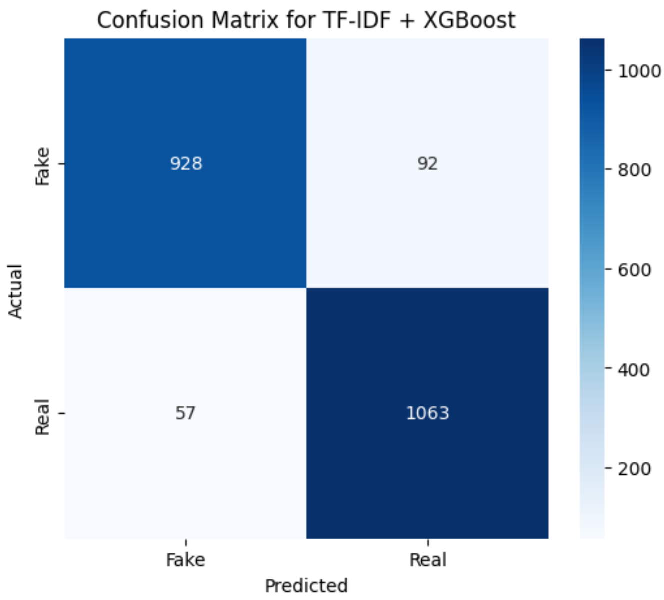

Figure 7: TF-IDF + XGBoost Confusion Matrix

Confusion matrices visually display the classification results of different models by showing:

- **True Positives (TP):** Correctly predicted real samples
- **False Positives (FP):** Incorrectly predicted fake samples
- **True Negatives (TN):** Correctly predicted fake samples
- **False Negatives (FN):** Incorrectly predicted real samples

Each confusion matrix corresponds to a different model:

- **Figure 5: BERT + XGBoost Confusion Matrix**
  - Predicted vs. actual labels are displayed, with a strong diagonal presence, indicating good model performance.
- **Figure 6: TF-IDF + Naïve Bayes Confusion Matrix**
  - The matrix shows a slightly higher error rate compared to BERT + XGBoost.
- **Figure 7: TF-IDF + XGBoost Confusion Matrix**
  - Appears to have fewer false negatives, suggesting better generalization.

3.2.2. Performance Table

Table 1: Model Performance Table

NLP + Model	Accuracy	True Label (1)			False Label (0)		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
TF-IDF + XGBoost	93%	92%	95%	93%	94%	91%	93%
TF-IDF + Naïve Bayes	91%	91%	91%	91%	90%	90%	90%
Bert + XGBoost	92%	92%	93%	93%	92%	91%	92%

Key Observations:

- **TF-IDF + XGBoost** achieves the highest accuracy **(93%).**
- **BERT + XGBoost** has strong recall **(92%-93%).**
- **Naïve Bayes** shows slightly lower performance **(91% accuracy).**

3.2.3. ROC-AUC Score

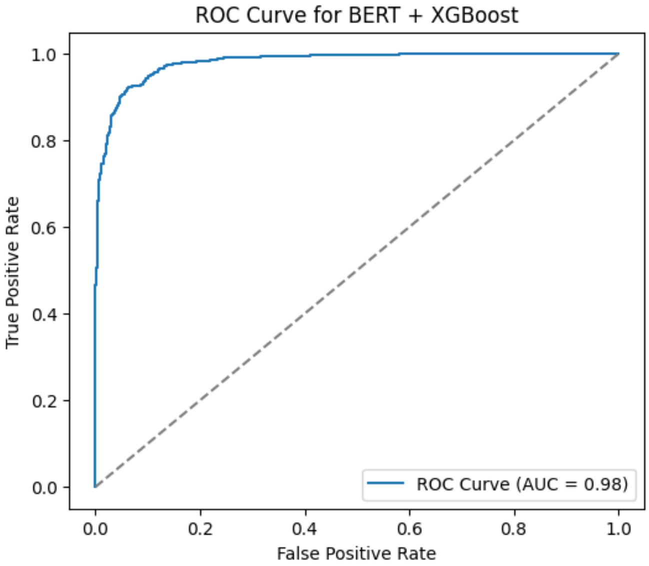

Figure 8: BERT + XGBoost ROC Curve

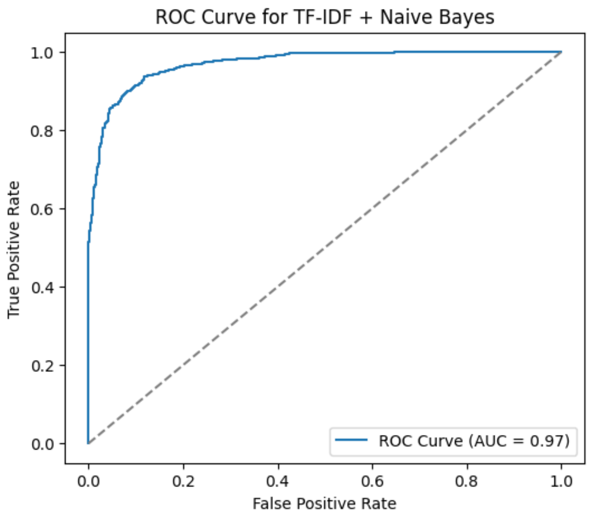

Figure 9: TF-IDF + Naive Bayes ROC Curve

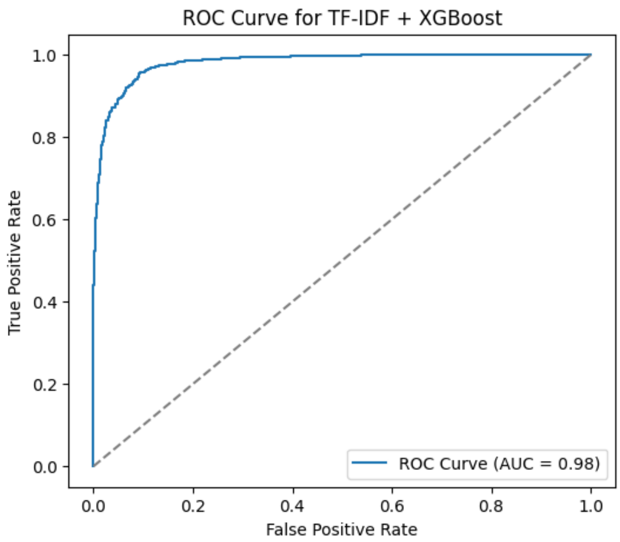

Figure 10: TF-IDF + XGBoost ROC Curve

The ROC curves in the image are part of a study analyzing how well different machine learning models can detect fake COVID-19 news. These curves show how accurately the models classify news as real or fake, based on their ability to balance true positives (correctly identifying fake news) and false positives (mistakenly labeling real news as fake).

- With BERT + XGBoost and TF-IDF + XGBoost both achieved impressive AUC score of 0.98, meaning they are highly effective in distinguish fake news.
- TF-IDF + Naïve Bayes performed quite well with an AUC score of 0.97.

CHAPTER IV: CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This study evaluated various machine learning models for detecting fake COVID-19 news using different text preprocessing, tokenization, and classification techniques. Our results indicate that TF-IDF + XGBoost achieved the highest accuracy (93%), making it the most reliable model for fake news detection. BERT + XGBoost demonstrated strong recall (92%-93%), ensuring it captures more fake news instances, while Naïve Bayes had slightly lower performance (91% accuracy).

The ROC-AUC scores further confirm the effectiveness of BERT + XGBoost and TF-IDF + XGBoost, both achieving an AUC of 0.98, indicating their strong ability to distinguish between real and fake news. TF-IDF + Naïve Bayes performed well but slightly lower, with an AUC of 0.97.

Additionally, analysis of word usage, punctuation, and emoji presence highlighted patterns in fake news content, such as the use of abstract wording and urgency-driven punctuation and emojis. These insights reinforce the importance of robust detection systems to combat misinformation.

Overall, TF-IDF + XGBoost is the most effective model, balancing accuracy, precision, and recall, making it a strong candidate for scalable and reliable fake news detection systems.

4.1. Recommendation

For future work, we aim to expand our dataset by incorporating multimodal data, including images, videos, and additional metadata, to enhance the contextual understanding of news articles. This integration could improve the model's ability to detect fake news by analyzing visual and textual correlations. Furthermore, we plan to develop our approach as a browser extension to test its effectiveness in real-time article evaluation, providing users with immediate feedback on news credibility. Additionally, we will explore extending our methodology to other domains beyond COVID-19-related news, enabling a more comprehensive fake news detection system applicable to various topics and misinformation trends.

REFERENCES

[1] S. Girgis, E. Amer, and M. Gadallah, “Deep Learning Algorithms for Detecting Fake News in Online Text,” in *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, Cairo, Egypt: IEEE, Dec. 2018, pp. 93–97. doi: 10.1109/ICCES.2018.8639198.

[2] A. Nistor and E. Zadobrischi, “The Influence of Fake News on Social Media: Analysis and Verification of Web Content during the COVID-19 Pandemic by Advanced Machine Learning Methods and Natural Language Processing,” *Sustainability*, vol. 14, no. 17, p. 10466, Aug. 2022, doi: 10.3390/su141710466.

[3] J. P. Haumahu, S. D. H. Permana, and Y. Yaddarabullah, “Fake news classification for Indonesian news using Extreme Gradient Boosting (XGBoost),” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 1098, no. 5, p. 052081, Mar. 2021, doi: 10.1088/1757-899X/1098/5/052081.

[4] A. Jain and A. Kasbe, “Fake News Detection,” in *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, Bhopal: IEEE, Feb. 2018, pp. 1–5. doi: 10.1109/SCEECS.2018.8546944.

[5] N. Ruchansky, S. Seo, and Y. Liu, “CSI: A Hybrid Deep Model for Fake News Detection,” in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, Singapore Singapore: ACM, Nov. 2017, pp. 797–806. doi: 10.1145/3132847.3132877.

[6] E. Essa, K. Omar, and A. Alqahtani, “Fake news detection based on a hybrid BERT and LightGBM models,” *Complex Intell. Syst.*, vol. 9, no. 6, pp. 6581–6592, Dec. 2023, doi: 10.1007/s40747-023-01098-0.