Fake News Detection: Model Report

1. Objective:

The objective of this project is to classify news articles as either "Fake" or "True" using machine learning models.

After benchmarking several models, we selected the Random Forest model based on its performance.

2. Data Overview:

- Total Records: 9,393 (4,697 Fake, 4,696 True)
- Dataset Source: Combined from "Fake.csv" and "True.csv"

3. Preprocessing Steps:

- Merged and labeled datasets.
- Removed duplicates and null values.
- Text preprocessing included tokenization, removal of stopwords, lemmatization, and cleaning.

4. Models Benchmarked:

- Random Forest
- XGBoost

5. Evaluation Metrics:

- Precision, Recall, F1-Score, Accuracy (on test dataset).

6. Model Selection:

The Random Forest model achieved the best results:

- Accuracy: 99.91%
- Precision (macro average): 99.91%
- Recall (macro average): 99.91%
- F1-Score (macro average): 99.91%

7. Conclusion:

The Random Forest model outperformed other models in terms of accuracy and robustness, making it the ideal choice for deployment.