Fake News Detection and Evaluation with Confusion Matrix

Karen Marlyn Vinod,
B.E Computer Engineering,
St. Francis Institute of Technology
(affiliated to University of Mumbai)

Period of Internship: 25th August 2025 - 19th September 2025

Report submitted to: IDEAS – Institute of Data Engineering, Analytics and Science Foundation, ISI Kolkata

1. Abstract

The exponential growth of digital content has made it increasingly difficult to distinguish between authentic and fake news. This project aims to address this issue by applying various machine learning techniques to detect fake news. Several models, including Logistic Regression, Random Forest, AdaBoost, XGBoost, and LightGBM, were trained using word embeddings (Word2Vec) and TF-IDF vectorization to represent text data. Accuracy, precision, recall, F1-score, and confusion matrices were used for comparative analysis.The results demonstrate that TF-IDF vectorization, when combined with boosting-based models (XGBoost, LightGBM, and AdaBoost) outperformed traditional classifiers, achieving accuracy levels above 99%. This project demonstrates how machine learning can help reduce misinformation on digital platforms and establishes the foundation for future research in trust-aware AI systems.

2. Introduction

Misinformation poses a significant threat to information integrity, public opinion and social trust[1].With the rapid spread of misinformation through digital media, automated solutions are now crucial for detecting and limiting the spread of fake content.

This project uses Natural Language Processing (NLP) and Machine Learning (ML) techniques to detect fake news. Vectorization was done using Word2Vec and TF-IDF, and performance was assessed using a variety of classifiers and boosting strategies[2][3].

Training received during internship:

- Python Programming Fundamentals – Covered data types, loops, data structures, and object-oriented programming (OOPs).
- Data Analysis – Explored data manipulation and analysis using NumPy and Pandas.
- Introduction to Machine Learning – Conducted hands-on labs on regression and classification techniques.
- Large Language Models (LLMs) – Learned fundamentals and gained practical experience with Ollama.
- Data Science Project Lifecycle – Understood end-to-end workflow: problem definition, data collection & preprocessing, exploratory data analysis, model building & evaluation, and deployment.

3. Project Objective

The main objectives of this project are:

- To preprocess and vectorize news data to facilitate efficient machine learning.
- To train and evaluate multiple classification models, focusing on boosting algorithms.

- To evaluate model performance using various metrics such as accuracy, precision, recall, F1 score, and confusion matrix.
- To analyze the effect of feature extraction methods (Word2Vec vs. TF-IDF) on model performance.
- To demonstrate the advantages of boosting techniques in improving classification accuracy.

4. Methodology

4.1 Dataset Preparation

The project was based on the ISOT Fake News Dataset, which contains two CSV files (True.csv and Fake.csv) with over 12,600 news articles each. Articles from Reuters.com were labeled as true, while those from unreliable outlets (flagged by Politifact and Wikipedia) were labeled as fake.

Steps performed:

- Merging: Fake and true news files were combined into a single dataset.
- Data Cleaning: A custom function (wordopt) was implemented to remove punctuation, stopwords, and apply lemmatization for text normalization.
- Shuffling: Articles were randomized to avoid ordering bias.
- Labeling: Class labels were assigned as Fake = 0 and True = 1.

4.2 Preprocessing

Two different feature engineering approaches were applied:

1. Word2Vec embeddings: to capture the meaning and relationships between words.
2. TF-IDF (Term Frequency - Inverse Document Frequency): to show the importance of a word in relation to all the articles in the dataset.

Finally, the dataset was split into 75% training data and 25% testing data.

4.3 Model Training

- Logistic Regression and Random Forest as baseline models.
- AdaBoost, XGBoost, and LightGBM as advanced boosting models.
- Hyperparameters: tuned where applicable (n_estimators, learning_rate, max_depth).

4.4 Evaluation

- Metrics: Accuracy, Precision, Recall, F1-score.
- Confusion matrices plotted for each classifier.
- Comparative analysis between Word2Vec and TF-IDF results.

4.5 Tools Used

- **Python** (Pandas, NumPy, Matplotlib, Scikit-learn, XGBoost, LightGBM)
- **Google Colab** for development
- **GitHub** for code repository and documentation

5. Data Analysis and Results

This section presents a comprehensive summary of the findings from descriptive analysis and the results of machine learning model evaluation. The analysis was carried out on the ISOT Fake News Dataset containing two CSV files: *True.csv* with over 12,600 legitimate news articles from reuters.com, and *Fake.csv* with more than 12,600 fabricated articles from various unreliable sources. Each record included an article title, full text, publication date, and category label. Articles were collected primarily from 2016–2017, and while the data were cleaned, punctuation inconsistencies and errors in the fake news text were deliberately preserved to retain linguistic authenticity.

5.1 Descriptive Analysis

The dataset was first explored to understand the distribution of news articles across categories. Figure **1** presents a pie chart illustrating the proportions of subjects in the news dataset. The dataset demonstrates a balanced representation of true and fake articles, ensuring fairness in model training.

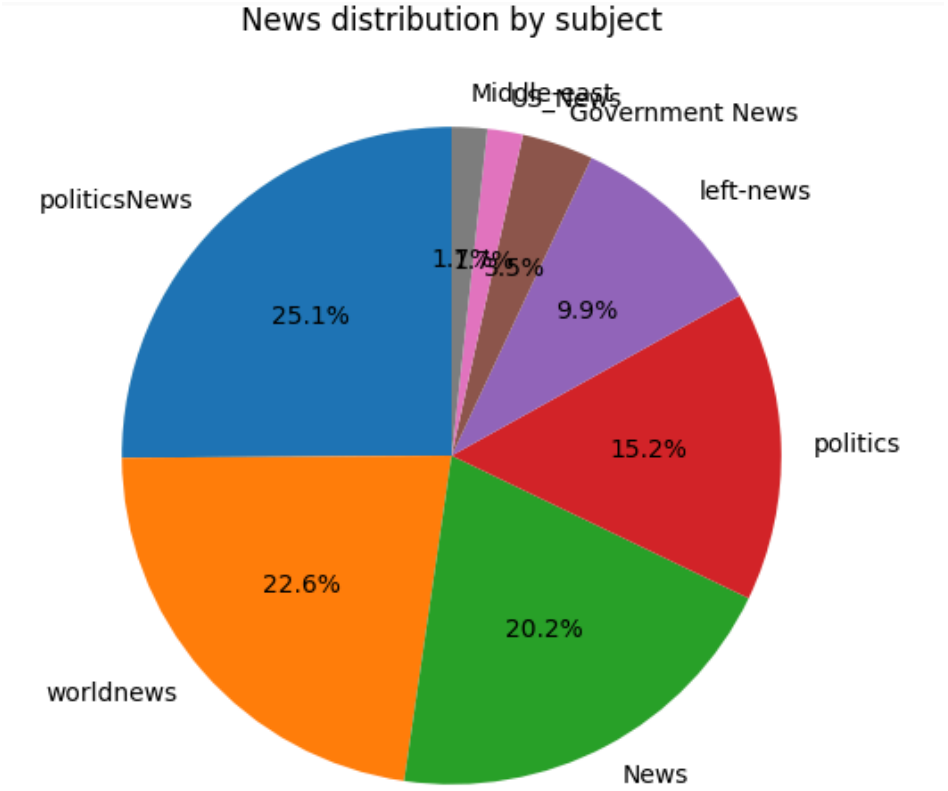

Figure 1 – Distribution of news subjects across the dataset

5.2 Machine Learning Model Evaluation

Multiple machine learning algorithms were trained and evaluated. Model performance was evaluated using four metrics: Accuracy, Precision, Recall, and F1 Score. Confusion matrices were used for visualisation. Experiments were conducted with two types of text representations: Bag-of-Words (raw text features) and TF-IDF (weighted word importance).

5.2.1 Model Performance using Word2Vec

Model	Accuracy	Precision	Recall	F1 Score
Logistic Regression	0.9381	0.9457	0.9353	0.9405
Random Forest	0.9383	0.9370	0.9455	0.9412
AdaBoost	0.9079	0.9159	0.9072	0.9115
XGBoost	0.9293	0.9368	0.9273	0.9320
LightGBM	0.9147	0.9255	0.9101	0.9177

5.2.2 Model Performance using TF-IDF Vectorization

Model	Accuracy	Precision	Recall	F1 Score
AdaBoost (TF-IDF)	0.9945	0.9973	0.9922	0.9947
XGBoost (TF-IDF)	0.9950	0.9976	0.9928	0.9952
LightGBM (TF-IDF)	0.9951	0.9976	0.9930	0.9953

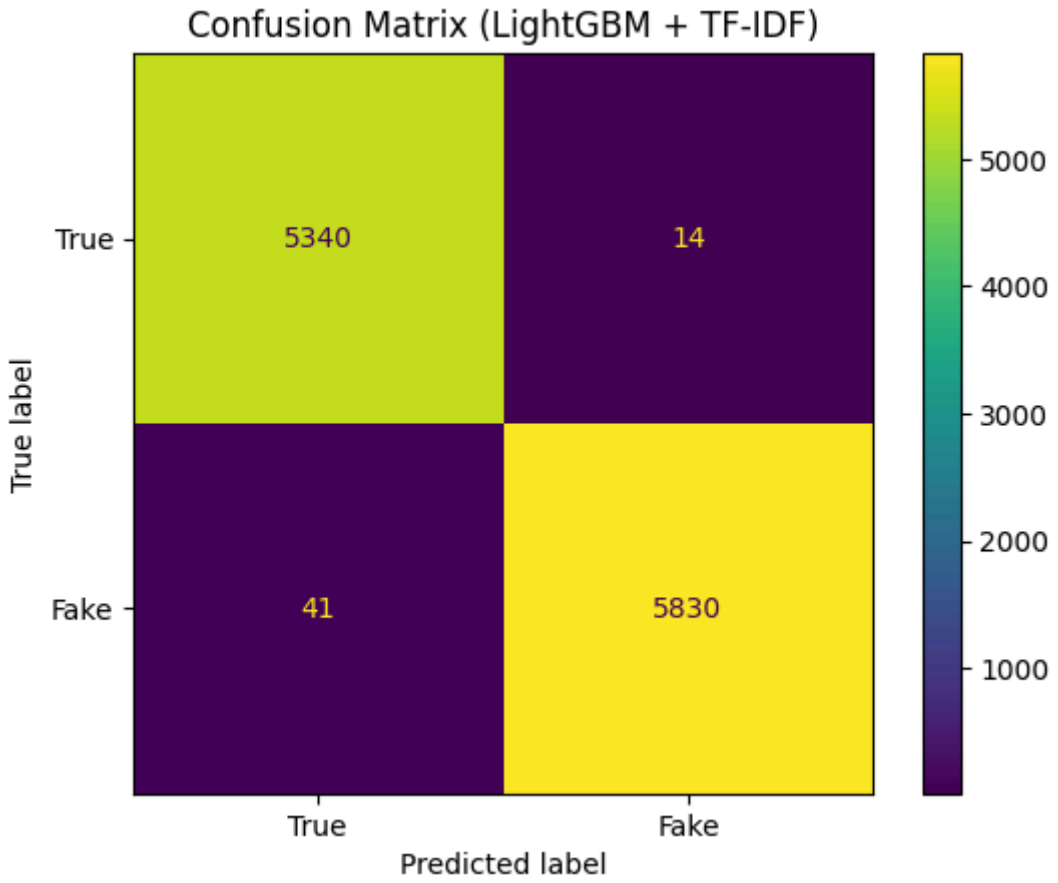

Fig 2 - LightGBM using TF-IDF vectorization

5.3 Comparative Analysis

- Logistic Regression and Random Forest gave strong baseline performance.
- Boosting algorithms (XGBoost, LightGBM, AdaBoost) significantly improved results.
- TF-IDF outperformed Word2Vec for this dataset, leading to higher precision and accuracy.
- Overall, boosting models with TF-IDF achieved above 99% accuracy, with LightGBM having the best accuracy among the models evaluated.

6. Conclusion

This project successfully developed and evaluated a set of machine learning models for fake news detection. The findings clearly demonstrated that learning techniques, particularly boosting algorithms, are highly effective for this task, attaining impressive accuracy and reliability. The choice of vectorization also proved to be a critical factor, with both Word2Vec and TF-IDF yielding strong results.

This foundation could be expanded upon in future research by:

- **Expanding the Dataset:** Incorporating a larger and more diverse dataset, including multilingual sources.
- **Exploring Deep Learning:** Implementing advanced models like LSTMs and Transformers (BERT) to capture more complex information.
- **Real Time Deployment:** Developing a web or mobile application for real-time fake news detection,thus making this technology accessible to a wider audience.

7. APPENDICES

Appendix A – References

[1] X. Zhou and R. Zafarani, "A survey of fake news: Fundamental theories, detection methods, and opportunities," *ACM Computing Surveys*, vol. 53, no. 5, pp. 1–40, Art. no. 109, Sep. 2020, doi: 10.1145/3395046

[2] M. A. B. Al-Tarawneh, O. Al-irr, K. S. Al-Maaitah, H. Kanj, and W. H. F. Aly, "Enhancing fake news detection with word embedding: A machine learning and deep learning approach," *Computers*, vol. 13, no. 9, p. 239, 2024, doi: 10.3390/computers13090239

[3] B. Viha and M. Nirmala, "Tri-Algo guardian ensemble approach for fake news detection in social media," *Journal of Big Data*, vol. 12, no. 1, p. 118, May 2025, doi: 10.1186/s40537-025-01161-2

- Github link for the codes developed: https://github.com/Karen-Vinod-02/Fake-News-Detection-and-Evaluation