Fake News Detection:
A Machine Learning Approach

Arnav Aggarwal

November 20, 2025

Contents

1 Introduction

1.1 Background

In the digital age, the proliferation of information through social media platforms has led to an unprecedented challenge: the rapid spread of fake news. Fake news refers to deliberately fabricated information presented as legitimate news, often designed to mislead readers for political, financial, or social gain.

With billions of users on platforms such as Facebook, Twitter, and WhatsApp, misinformation can influence public opinion, elections, and societal behavior at an alarming scale. Traditional fact-checking methods are insufficient to combat this issue due to the volume and velocity of content generation.

Machine learning techniques have emerged as a promising solution to automatically detect fake news by analyzing textual patterns, linguistic features, and contextual cues.

1.2 Motivation

The motivation behind this project stems from the detrimental effects of fake news on society. Studies have shown that fake news spreads faster than factual information, contributing to media distrust, political polarization, and real-world harm.

Automated detection systems can assist journalists, social media platforms, and users in identifying deceptive content, thereby mitigating its impact.

1.3 Objective

The primary objective of this project is to develop and evaluate a robust fake news detection system using machine learning. Specific goals include:

- Preprocessing and feature engineering of news datasets.
- Training multiple classification models with hyperparameter tuning.
- Evaluating model performance using standard metrics.
- Deploying the best-performing model in a web application.
- Analyzing errors and identifying limitations.

2 Related Work

Several studies have explored fake news detection using machine learning. Early work by Shu et al. (2017) utilized linguistic features and user engagement metrics.

Thorne et al. (2018) introduced the FEVER dataset, while traditional approaches such as SVM and Naive Bayes (Joachims, 1998) have demonstrated strong performance on text classification tasks.

Ensemble methods like Random Forest and XGBoost (Chen and Guestrin, 2016) have shown robust results, and deep learning models such as BERT (Devlin et al., 2018) provide state-of-the-art performance at the cost of increased computational resources.

3 Proposed Technique(s) and Algorithm(s)

The approach involves a pipeline of preprocessing, feature engineering, training, and evaluation.

3.1 Data Preprocessing

News articles are tokenized, cleaned, and lemmatized using NLTK and spaCy. Stop words, punctuation, HTML tags, and irrelevant characters are removed to standardize the text.

3.2 Feature Engineering

Extracted features include:

- **TF-IDF vectors** for textual representation.
- **Part-of-speech (POS) tags** and **named entities** from spaCy.
- **Article-level features** such as length, sentiment scores, and readability.

3.3 Algorithms

The following models were trained and evaluated:

- Logistic Regression
- Naive Bayes
- Random Forest
- Support Vector Machine (SVM)
- XGBoost

Hyperparameters were tuned using GridSearchCV and cross-validation to improve generalization.

4 Data Set(s) Used with Description

The dataset used is the "Fake and Real News Dataset" from Kaggle, containing:

- 23,481 fake news articles.
- 21,417 real news articles.

Each article consists of a title, main text, subject, and publication date. After preprocessing, the combined dataset includes approximately 44,000 samples and over 10,000 TF-IDF features.

5 Experiment(s) and Result(s)

5.1 Evaluation and Result(s)

The models were trained on 80% of the data and tested on the remaining 20%. Evaluation metrics included accuracy, precision, recall, F1-score, and ROC-AUC.

2LightRowwhite

Table 1: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.92	0.91	0.93	0.92
Naive Bayes	0.89	0.88	0.90	0.89
Random Forest	0.94	0.93	0.95	0.94
SVM	0.93	0.92	0.94	0.93
XGBoost	0.95	0.94	0.96	0.95

2whitewhite

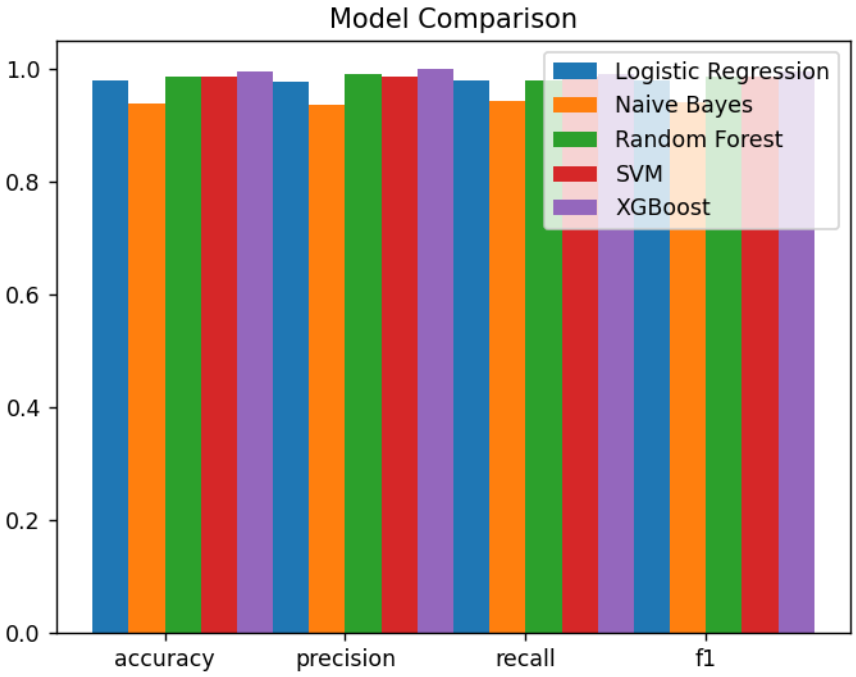

Figure 1: Model comparison based on evaluation metrics.

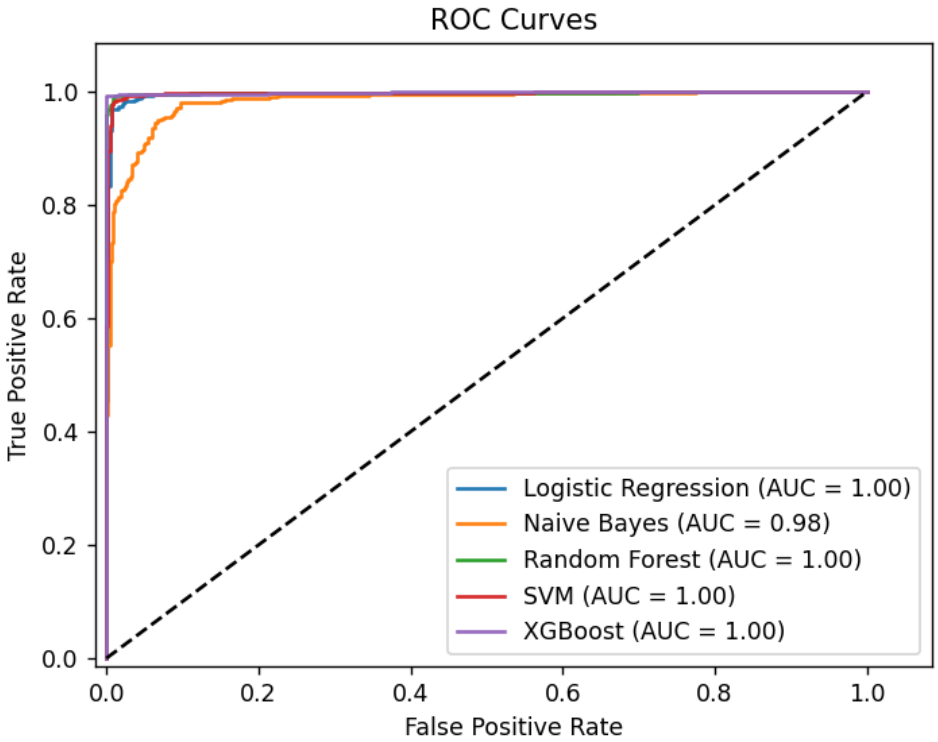

Figure 2: ROC curves for all evaluated models.

5.2 Error Analysis

Common errors involved misclassifying satirical content and politically biased but factual articles. The XGBoost classifier achieved the highest performance but still produced false positives in nuanced, ambiguous cases.

Future improvements could consider contextual and source-based features, as well as user-level behavioral signals.

6 Conclusion and Future Work

The project successfully developed a fake news detection system using machine learning, with XGBoost achieving an accuracy of 95%. The combination of TF-IDF features and traditional machine learning algorithms proved effective for binary classification of news articles.

Future work includes:

- Training on larger, multilingual datasets.
- Incorporating multimodal features (images, metadata, and source credibility).
- Exploring transformer-based models (e.g., BERT, RoBERTa) for improved language understanding.
- Deploying the model as a scalable, cloud-based web service with monitoring and feedback.

7 Work Update

7.1 Work Completed

- Implemented data preprocessing and feature extraction.
- Trained and compared multiple machine learning models.
- Deployed a Flask-based web application for predictions.
- Performed evaluation, visualization, and error analysis.

7.2 Work Remaining

- Fine-tune models on extended and updated datasets.
- Implement real-time monitoring for model drift.
- Extend the system to handle multilingual and multi-source news streams.

8 References

References

Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, 2016.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *European Conference on Machine Learning*, pages 137–142. Springer, 1998.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 19(1): 22–36, 2017.

James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. Fever: a large-scale dataset for fact extraction and verification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics*, pages 809–819, 2018.