

Proof of Work: Fake News Detection Project

May 15, 2025

1 Introduction

This project focuses on developing a machine learning model to detect fake news, addressing the growing concern of misinformation in an era of exponential data generation. The objective is to create a binary classification model capable of accurately distinguishing between real and fake news articles, leveraging Python and machine learning techniques.

2 Dataset

The dataset comprises two directories: one containing true news articles and the other containing fake news articles, totaling 44,898 instances. True news is labeled as 0, and fake news as 1. The distribution is relatively balanced, with 21,417 true news instances and 23,481 fake news instances, ensuring a robust foundation for model training.

3 Data Preprocessing

The data preparation process involved several key steps:

- Loaded the dataset using the Pandas library.
- Retained only the essential 'text' and 'label' columns, discarding irrelevant data.
- Concatenated the true and fake datasets into a single dataframe and shuffled it to eliminate order-based biases.
- Applied TF-IDF Vectorization to transform the text data into numerical features, configured with a maximum of 50,000 features and an n-gram range of (1,2). Due to memory constraints, only the first 35,000 samples were used.

4 Model Selection

The Multinomial Naive Bayes algorithm was selected for this task. This supervised learning algorithm is particularly effective for high-dimensional text data and discrete features like word counts, making it an ideal choice for classifying news articles based on their textual content.

5 Training and Evaluation

The dataset was split into training and testing sets. The model was trained on the vectorized training data using the Multinomial Naive Bayes classifier. Performance was assessed on the test set using the following metrics:

- Precision
- Recall
- F1-score
- Accuracy

6 Results

The model achieved an accuracy of 95% on the test dataset. The classification report highlights the following metrics:

- **Real News (0):** Precision = 0.96, Recall = 0.94, F1-score = 0.95
- **Fake News (1):** Precision = 0.95, Recall = 0.96, F1-score = 0.95

The training accuracy was approximately 95.95%, indicating good generalization without significant overfitting.

7 Conclusion

This project successfully demonstrates the application of machine learning to detect fake news, achieving a high accuracy of 95% using the Multinomial Naive Bayes algorithm. The combination of TF-IDF Vectorization and a well-suited classifier proved effective for this binary classification task. Future work could explore alternative algorithms, advanced feature engineering techniques, or larger datasets to further enhance performance.