

Fake News Detection: Detailed Project Report

This report outlines the complete approach for building a fake news detection system using natural language processing and machine learning techniques. The primary goal of this project was to accurately classify news articles as either Real or Fake based on their textual content.

a) Data Preprocessing Steps

The dataset was provided as two CSV files for training and testing. Data quality checks were performed to remove rows with missing text or labels. The text was then normalized through several steps: converting all text to lowercase to ensure case-insensitive matching, removing punctuation and numeric digits using regular expressions to focus on words, and mapping the categorical labels to numeric values (Fake=0, Real=1) for model compatibility. Finally, to convert text into meaningful numerical features, a TF-IDF (Term Frequency–Inverse Document Frequency) vectorizer was applied. This technique assigns higher weight to words that are frequent in a document but rare across the corpus, capturing the importance of discriminative terms. The vectorizer was limited to the top 5000 features and configured to ignore English stop words.

b) Modeling Approach and Reasoning

To leverage the strengths of different machine learning algorithms, an ensemble method was adopted. Three diverse classifiers were selected:

- **Logistic Regression:** A strong baseline for text classification tasks, excelling in linearly separable data.
- **Multi-Layer Perceptron (MLP):** A feed-forward neural network capable of capturing non-linear relationships that simple linear models might miss.
- **XGBoost:** An efficient gradient boosting algorithm well-known for handling complex decision boundaries and interactions among features.

These models were combined in a **Voting Classifier** with hard voting, meaning the final prediction is based on the majority vote of the individual models. This approach increases robustness and reduces the risk of overfitting that might occur if only a single model were used.

c) Performance Evaluation

The ensemble model was trained on the preprocessed training data and evaluated on a separate test set. Key performance metrics demonstrate the effectiveness of the approach:

- **Accuracy:** 0.995 – indicating that 99.5% of test samples were correctly classified.
- **Precision:** 0.993 – showing that when the model predicted Real, it was correct 99.3% of the time.
- **Recall:** 0.997 – meaning the model successfully identified 99.7% of actual Real news articles.
- **F1-score:** 0.995 – the harmonic mean of precision and recall, reflecting a balanced performance.

The confusion matrix $\begin{bmatrix} 966 & 7 \\ 3 & 1004 \end{bmatrix}$ reveals very few misclassifications, underscoring the model's reliability.

Conclusion

By carefully preprocessing the text, selecting a mix of complementary models, and employing an ensemble voting strategy, the system achieved near-perfect accuracy in detecting fake news. The combination of TF-IDF features with Logistic Regression, MLP, and XGBoost ensured that both linear patterns and complex non-linear relationships were captured. This methodology demonstrates that integrating multiple algorithms can yield superior performance compared to relying on any single model.