

Fake News Detection App - Project Report

1. Overview

The Fake News Detection App is a Natural Language Processing (NLP) project developed to classify news articles as either Fake or Real. It uses TF-IDF vectorization and a Logistic Regression model to predict the authenticity of user-entered headlines or full news articles. The app is deployed using Streamlit and provides real-time predictions, explanation highlights, and basic model performance metrics.

2. Objectives

- Build a machine learning model to identify fake and real news.
- Clean and preprocess text data using NLTK.
- Use TF-IDF for feature extraction.
- Train and evaluate Logistic Regression and Naive Bayes models.
- Develop a Streamlit web interface for user interaction.
- Display prediction results, confidence score, and contributing keywords.
- Provide visualization for performance metrics.

3. Tools & Technologies Used

- Python, Pandas, Numpy, Scikit-learn, NLTK
- Streamlit for UI
- TF-IDF Vectorizer for text representation
- Logistic Regression for classification

4. Project Workflow

1. Data Collection: Fake news dataset collected from Kaggle.
2. Data Preprocessing: Tokenization, stopword removal, and stemming using NLTK.
3. Feature Extraction: Text converted into numerical features using TF-IDF.
4. Model Training: Logistic Regression model trained for classification.
5. Evaluation: Measured accuracy, F1-score, and confusion matrix.
6. App Interface: Streamlit app for user interaction and visualization.

5. Results

■ Prediction Output Example:

Input: "Reports claim that aliens landed in New York City last night, causing panic among residents. Scientists confirm this is false."

Prediction: REAL NEWS (Confidence: 98.61%)

Top Contributing Words: fake, scientists, claim, confirm, report

■ Model Performance Metrics:

Accuracy: 0.40, F1-score: 0.57

6. Conclusion

The Fake News Detection App successfully demonstrates the use of Natural Language Processing and Machine Learning to verify news authenticity. While the model provides accurate real-time predictions, performance can be further improved using a larger and more balanced dataset.

7. Author Details

Prepared By	SHIBIKASRI G
Email	shibiguna22@gmail.com
Phone	7418349930