Fake News Detection Project Report

Project Title:

Fake vs Real News Detection Using NLP and Machine Learning

Submitted by:

Pasumala Harshitha

harshithapasumala07@gmail.com
+91 9642630995

Date: 21-05-2025

1. Objective

The goal of this project is to build a machine learning model that classifies whether a given news article is **fake** or **real**. With the surge in misinformation online, especially on social media, this system provides a reliable mechanism to flag suspicious news content using NLP techniques.

---

2. Dataset

- **Source**: Kaggle – *Fake and Real News Dataset*
- **Files used**: Fake.csv and True.csv
- **Features**:
    - title – Headline of the news
    - text – Full content of the news article

- subject, date
- Label: 1 for Real, 0 for Fake

---

3. Methodology

3.1 Data Preprocessing

- Combined title and body text for better context.
- Cleaned the text: lowercase conversion, punctuation removal.
- Removed nulls and duplicates.
- Labeled fake news as 0, real news as 1.

3.2 Text Vectorization

- Used **TF-IDF Vectorizer** to transform text into numerical vectors.
- Removed stopwords and filtered high-frequency words (max_df=0.7).

3.3 Model Selection

- Chosen model: **Passive Aggressive Classifier** (highly effective for large sparse data).
- Trained on 80% of the data and tested on the remaining 20%.

3.4 Evaluation

- Metrics used: Accuracy, Confusion Matrix
- Achieved an **accuracy of ~93%**.
- Model performed well in distinguishing fake and real news.

---

4. Results

Metric	Value
Accuracy	~93%
Precision/Recall	Balanced
Confusion Matrix	Low false positives/negatives

---

5. Tools & Libraries

- Python, Jupyter Notebook
- pandas, scikit-learn, matplotlib, nltk

---

7. Conclusion

This project successfully demonstrates a working pipeline for fake news detection using machine learning and NLP. The model performs well, and its deployment through a web app and browser extension makes it practical for real-world use.