NAME : SWATHI SUNDARESWARAN IYER

📰 News Article Classification (Fake/Real)

Internship Project Report

🔹 Introduction

In the current age of digital media, the widespread dissemination of fake news poses a significant challenge to information integrity. This project aims to develop a machine learning model that classifies news articles as **fake** or **real** based on their content. Using natural language processing (NLP) techniques, the system is trained on a labeled dataset and made accessible via a web-based Streamlit app.

🔹 Abstract

This project involves building a binary classifier using Python and machine learning to distinguish between fake and real news articles. The dataset consists of labeled news headlines. After preprocessing the text using NLTK, the data is vectorized with TF-IDF and classified using a **Naive Bayes** algorithm. The trained model is deployed using **Streamlit**, allowing users to test the app in real-time. The system achieves over **90% accuracy**, providing a robust tool to identify misleading information.

🔹 Tools & Libraries Used

- Python 3.x
- NLTK (Natural Language Toolkit)
- Pandas
- Scikit-learn
- Joblib (for model serialization)
- Streamlit (for web deployment)

🔹 Steps Involved in Building the Project

1. **Data Collection**
    - Used the *Fake and Real News Dataset* from Kaggle

- Combined Fake.csv and True.csv, added labels

2. **Text Preprocessing (NLP)**
   - Converted to lowercase
   - Removed punctuation and stopwords using NLTK
   - Cleaned and tokenized text
3. **Feature Extraction**
   - Applied **TF-IDF Vectorization** to transform text into numeric vectors
4. **Model Training**
   - Used **Multinomial Naive Bayes** classifier
   - Trained on 80% of data, tested on 20%
5. **Evaluation**
   - Achieved ~92% accuracy
   - Used classification_report() and confusion_matrix()
6. **Model Deployment**
   - Saved model and vectorizer using joblib
   - Built a user-friendly Streamlit app (app.py)
   - Deployed on Streamlit Cloud for public use

🔹 Conclusion

The Fake News Detector is a lightweight, effective, and accessible web tool that accurately classifies news articles. With real-time performance and a simple UI, it demonstrates practical applications of NLP and machine learning. This project not only enhances understanding of text classification but also shows how to deploy models for real-world usage.

🔹 GitHub Repository

📂 [https://github.com/Swathi9137/Fake-News-Detector-app.git]

🔹 Live App Demo

🌐 [https://fake-news-detector-app-swathiiyer2004.streamlit.app/]
