Elevate Labs
AI & ML Internship Project : News Article Classification (Fake/Real)

Danny Noby Joseph
dannynobyj@gmail.com

Introduction
In the digital age, misinformation spreads rapidly across social media and online platforms. Identifying fake news is critical to maintain public trust and prevent the spread of false narratives. This project aims to develop a machine learning model that classifies news articles as Fake or Real using Natural Language Processing techniques.

Abstract
This project leverages a labeled dataset of news articles to build a binary text classification system. The raw text is cleaned using NLP techniques such as stopword removal and stemming (via NLTK), and vectorized using TF-IDF. A Multinomial Naive Bayes classifier is trained on this processed data. A user-friendly web application is developed using Streamlit to allow real-time predictions with minimal input. The app also provides an explanation by highlighting the top words influencing each prediction.

Tools Used
Python – Programming language
Pandas – Data handling
NLTK – Text cleaning (stopwords, stemming)
Scikit-learn – ML models and vectorization
Streamlit – Web interface

Steps Involved
Labeled dataset (Fake_Real_News_Data.csv) sourced from Kaggle
Data Preprocessing (1-Real news, 0- Fake news)
Text converted to lowercase, numbers/punctuation removed
Stopwords removed and stemming applied using NLTK
Feature Extraction
TF-IDF vectorization used to convert text to numeric form
Model Training
Trained with Multinomial Naive Bayes and optionally Logistic Regression
Model Evaluation
Evaluated using accuracy, F1-score, and confusion matrix
Streamlit app built to input text, make predictions, and explain results
(Instead of Jupyter Notebook, a regular .py is used to satisfy the streamlit app compatibility)

Conclusion
This project successfully demonstrates how machine learning and NLP can be applied to combat misinformation. The web interface makes the tool accessible and interpretable to users.

Images

```
[nltk_data] Downloading package stopwords to /Users/admin/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
                                                text  label                                       cleaned_text
0  Donald Trump endorsed an unabashedly noninterv...      1  donald trump endors unabashedli noninterventio...
1  While some Justice Department investigations a...      1  justic depart investig adversari new model col...
2  Trump Proudly Declares: Most Of The People I'v...      0  trump proudli declar peopl ive insult deserv a...
3  Donald Trump spent a day in January 2014 hobno...      1  donald trump spent day januari hobnob politici...
4  Notable names include Ray Washburne (Commerce)...      1  notabl name includ ray washburn commerc dallas...
Accuracy: 0.8816101026045777
Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.87      0.88       639
           1       0.87      0.89      0.88       628

    accuracy                           0.88      1267
   macro avg       0.88      0.88      0.88      1267
weighted avg       0.88      0.88      0.88      1267
```