Fake News Classification Using Machine Learning

Student Name: Bharath

Semester: 5th Semester B.Tech

Internship Project Title: News Article Classification (Fake/Real)

1. Introduction

Fake news spreads quickly online and can dangerously impact public perception by distributing misinformation. This project aims to build an effective machine learning model to classify news articles as either fake or real using natural language processing techniques.

2. Abstract

A labeled dataset of fake and real news articles was used. Text was cleaned (punctuation removed, lowercased, stopwords removed) with NLTK, converted to numbers via TF-IDF vectorization, then classified using a Naive Bayes model. A Streamlit web app was built to allow users to paste news text for instant predictions.

3. Tools and Technologies Used

- Programming Language: Python

- Development Platform: Jupyter Notebook

- Libraries: pandas, numpy, NLTK, scikit-learn, joblib, streamlit

4. Methodology and Implementation Steps

1. Data Collection: Collected Fake.csv and True.csv from Kaggle

2. Data Labeling & Merging: Labeled fake articles as 0 and real as 1. Merged and shuffled both datasets

3. Text Preprocessing: Used NLTK to clean text (remove punctuation, stopwords, lowercase)

4. Feature Extraction: Applied TF-IDF vectorization

5. Model Training: Trained a Naive Bayes classifier

6. Model Evaluation: Evaluated using accuracy, F1-score, precision, and recall

7. Model Saving: Saved model and vectorizer using joblib

8. Web App Development: Built a Streamlit app to classify input news in real time

Fake News Classification Using Machine Learning

5. Results

- The model showed high accuracy in classifying both fake and real articles

- The Streamlit app gives quick and clear predictions to users

6. Conclusion

This project helped me understand how to clean text, train machine learning models, save them, and build web apps. It improved my skills in NLP, model deployment, and applying machine learning to real-world problems.

7. Future Enhancements

- Try advanced models like Logistic Regression, Random Forest, or BERT

- Expand the dataset with articles from multiple sources

- Add a feature to input URLs and auto-extract news content

- Host the app online using Streamlit Cloud