Project Report: News Article Classification (Fake/Real)

---

1. Introduction

In the digital era, fake news has become a major concern due to its rapid spread and harmful consequences. With the rise of social media and open publishing platforms, misinformation can influence public opinion, manipulate behavior, and even cause real-world damage. To counter this, our project aims to build a machine learning-based solution that classifies news articles as real or fake using Natural Language Processing (NLP) techniques.

2. Abstract

This project leverages NLP and machine learning to detect and classify fake news articles. We used a publicly available dataset with labeled news samples and applied text preprocessing, TF-IDF vectorization, and Logistic Regression for classification. The project also includes a dynamic and interactive Streamlit web interface allowing users to input or upload news data, receive predictions, and understand the reasoning behind the classification through keyword explanation.

3. Tools Used

- **Programming Language**: Python 3.x
- **Libraries**: NLTK, Scikit-learn, Pandas, NumPy, Joblib
- **Interface**: Streamlit
- **Model**: Logistic Regression
- **Vectorizer**: TF-IDF

4. Steps Involved in Building the Project

1. **Dataset Collection**: A labeled dataset of real and fake news was sourced from Kaggle.
2. **Text Preprocessing**: Cleaned text by removing punctuation, lowercasing, tokenizing, removing stopwords, and lemmatizing using NLTK.
3. **Feature Extraction**: Converted textual data into numerical form using TF-IDF vectorization.
4. **Model Training**: Trained a Logistic Regression model to classify news articles.
5. **Model Evaluation**: Evaluated model accuracy, precision, recall, and F1-score.
6. **Web GUI Development**: Created a modern GUI using Streamlit, supporting:
    - Single news article prediction
    - CSV upload for bulk classification
    - Result explanation (keywords and weights)
    - Downloadable report generation

5. Conclusion

The News Article Classification (Fake/Real) project successfully demonstrates the use of machine learning and NLP for real-world misinformation detection. With an interactive GUI and explanation-based predictions, the tool provides transparency and usability. This project can be further enhanced by integrating deep learning models, multilingual support, and real-time news scraping to strengthen its practical applications.