

INTERNSHIP PROJECT – PHASE – 2

TOIPC: News Article Classification (Fake/Real)

1.INTRODUCTION

In today's digital age, news spreads rapidly across the internet, making it easier for misinformation and fake news to influence public opinion. This project aims to detect whether a news article is fake or real using machine learning techniques based on textual content analysis.

2.ABSTRACT

The Fake News Detection project is a binary text classification system built using Python and machine learning. It uses real and fake news articles collected from a Kaggle dataset and applies natural language processing (NLP) to vectorize and classify the text. The system outputs whether a news article is fake or real, with an interactive interface built using Streamlit.

3.TOOLS USED:

- **Language:** Python
- **Libraries:** Pandas, Scikit-learn, Pickle, Streamlit
- **Model:** Logistic Regression
- **Vectorizer:** TF-IDF (Term Frequency–Inverse Document Frequency)
- **Dataset:** Fake and Real News Dataset from Kaggle

4.STEPS INVOLVED IN BUILDING A PROJECT:

1. Data:

Downloaded and combined Fake.csv and True.csv datasets from Kaggle.

2. Preprocessing:

Cleaned the text data, removed null values, and added labels (FAKE = 1, REAL = 0).

3. Vectorization:

Applied TF-IDF vectorization to convert raw text into numerical format.

4. Training:

Trained a Logistic Regression model on 80% of the dataset, validated on the remaining 20%.

5. Saving:

Saved the trained model and vectorizer using Pickle for reuse.

6. Streamlit

Developed a Streamlit web app with a text input box and prediction button to classify new articles.

5.CONCLUSION

The project successfully detects fake news with high accuracy using a simple Logistic Regression model and TF-IDF vectorization. It provides a user-friendly interface for real-time classification, making it useful for quick verification of news content. This can be further enhanced with deep learning or real-time news scraping in future improvements.