Internship Project Report

News Article Classification (Fake or Real)

Submitted by: S Sreelakshmi

Introduction

With the widespread influence of digital media, the spread of fake news has become a critical issue. This project aims to classify news articles as 'Fake' or 'Real' using Machine Learning techniques. It helps build awareness and provides a tool for detecting misinformation.

Abstract

The project uses Natural Language Processing (NLP) and a Logistic Regression model to analyze and classify news articles. We use a labeled dataset consisting of real and fake news to train the model. The TF-IDF vectorizer transforms the text data, and the model is evaluated using precision, recall, and accuracy metrics. The final model can predict whether a news article is real or fake based on the input text.

Tools Used

Python, Pandas, Scikit-learn, NLTK

Steps Involved in Building the Project

1. Dataset Preparation: Combined real and fake news articles into one dataset with labels.

2. Text Preprocessing: Removed stopwords using NLTK for cleaner inputs.

3. Feature Extraction: Used TF-IDF to convert text into numerical format.

4. Model Training: Trained a Logistic Regression model using Scikit-learn.

5. Evaluation: Achieved over 90% accuracy using F1-score and classification report.

6. Prediction: Allowed users to input custom news and check if it's Real or Fake.

Internship Project Report

Conclusion

This project demonstrates how NLP and Machine Learning can effectively detect misinformation in news articles. With simple yet powerful tools, even basic models like Logistic Regression can perform well on text classification tasks.