Fake News Detection

Name: Parth Azad ID: E23CSEU1032

Name: Dhruv ID: E23CSEU1031

Guide: Dr. Purushottam Kumar

Introduction:

Fake news has become an alarming issue in the digital world. It misleads people, promotes false narratives, and sometimes leads to serious social and political consequences. This project aims to tackle this growing problem using Machine Learning techniques to automatically detect and classify fake news.

Objective:

The main objective is to build a machine learning model that can distinguish between real and fake news. The project involves preprocessing news articles, converting them into numerical representations (vectors), and training machine learning models.

Tools and Libraries Used:

- Pandas: For data manipulation.
- Matplotlib & Seaborn: For visualization.

- NLTK: For text preprocessing.

- Scikit-learn: For machine learning algorithms and evaluation.
- WordCloud: For visualizing frequent words.

Dataset:

The dataset consists of news articles labeled as real or fake. It includes columns such as title, subject, date, text, and class, where 1 indicates real news and 0 indicates fake news.

Methodology:

1. Data Preprocessing: Cleaning the text, removing stopwords, and tokenizing.

2. Data Visualization: Creating word clouds for both real and fake news.

3. Feature Extraction: Converting text into vectors using TF-IDF.

4. Model Training: Logistic Regression and Decision Tree classifiers are used for training

5. Model Evaluation:

Accuracy scores and confusion matrix are used to evaluate the models.

Sample Code Snippets:

Data Preprocessing:

```python
def
  preprocess_text(text_data):
  preprocessed_text = []

  for sentence in text_data:

    sentence = re.sub(r'[^\w\s]', '', sentence)

        preprocessed_text.append(' '.join(token.lower() for token in sentence.split() if token not
in stopwords.words('english')))

    return
preprocessed_text
```
TF-IDF Vectorization:
```python
from sklearn.feature_extraction.text import TfidfVectorizer

vectorization = TfidfVectorizer()

x_train = vectorization.fit_transform(x_train)
x_test = vectorization.transform(x_test)
```

Conclusion:

The project successfully built a machine learning model capable of detecting fake news with high accuracy. The Decision Tree classifier provided 99.5% accuracy on the test data. This system can help combat the spread of misinformation by automatically identifying fake news articles.