Fake News Detection Using NLP

Phase-1 Documentation Submission
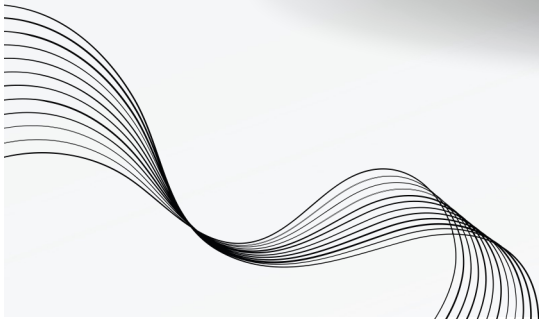

Presented By: Team Member
Muhesh S

CONTENT

INTRODUCTION:

- This project aims to develop a machine learning model that can distinguish between genuine and fake news articles.

- The model will analyze the titles and text of news articles and classify them as either genuine or fake.
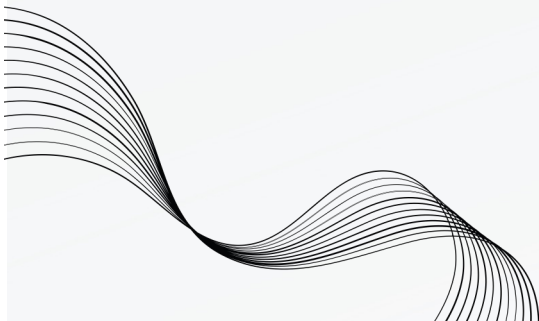
Data Description:

- The dataset used in this project is obtained from Kaggle.

https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

- It contains news articles along with their labels indicating whether they are genuine or fake.

Modules

01   **Data Collection Module**

02   **Data Preprocessing Module**

03   **Feature Extraction Module**

04   **Model Building Module**

05   **Documentation Module**

Data Collection Module:

- This module involves collecting the dataset from Kaggle. The dataset contains news articles along with their labels (genuine or fake).
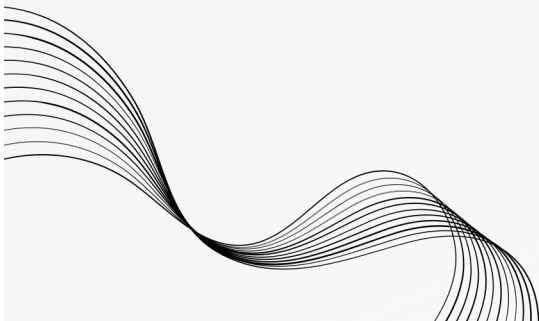
Data Preprocessing Module:

- This module involves cleaning and preprocessing the textual data to prepare it for analysis.

- This includes tasks such as tokenization, stop word removal, stemming, etc.
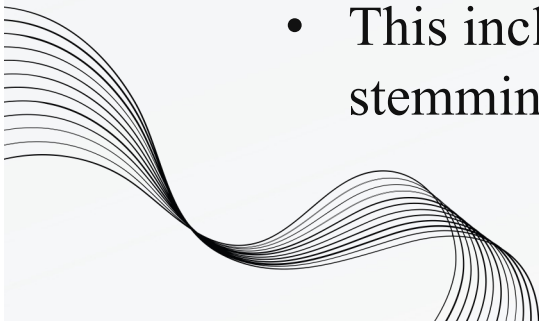
Feature Extraction Module:

- This module involves converting the preprocessed text into numerical features that can be used by a machine learning algorithm.

- Techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings can be used for this purpose.

Model Building Module:

- This module involves selecting a suitable machine learning algorithm for the task (e.g., Logistic Regression, Random Forest, or Neural Networks) and training it on the preprocessed data.

```
Epoch 1/5
102/102 [==============================] - 204s 1s/step - loss: 0.1709 - Accurac
y: 0.7617 - Precision: 0.7751 - Recall: 0.6818 - val_loss: 0.1052 - val_Accuracy:
0.8667 - val_Precision: 0.7959 - val_Recall: 0.9512
Epoch 2/5
102/102 [==============================] - 151s 1s/step - loss: 0.0343 - Accurac
y: 0.9753 - Precision: 0.9758 - Recall: 0.9706 - val_loss: 0.0192 - val_Accuracy:
0.9667 - val_Precision: 0.9524 - val_Recall: 0.9756
Epoch 3/5
102/102 [==============================] - 152s 1s/step - loss: 0.0079 - Accurac
y: 0.9988 - Precision: 1.0000 - Recall: 0.9973 - val_loss: 0.0048 - val_Accuracy:
1.0000 - val_Precision: 1.0000 - val_Recall: 1.0000
Epoch 4/5
102/102 [==============================] - 109s 1s/step - loss: 0.0052 - Accurac
y: 1.0000 - Precision: 1.0000 - Recall: 1.0000 - val_loss: 0.0087 - val_Accuracy:
0.9778 - val_Precision: 0.9535 - val_Recall: 1.0000
Epoch 5/5
102/102 [==============================] - 108s 1s/step - loss: 0.0043 - Accurac
y: 1.0000 - Precision: 1.0000 - Recall: 1.0000 - val_loss: 0.0075 - val_Accuracy:
0.9778 - val_Precision: 0.9535 - val_Recall: 1.0000
```
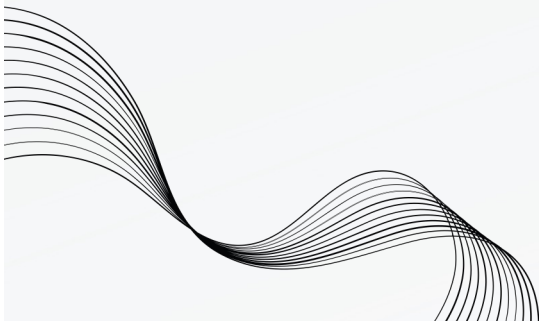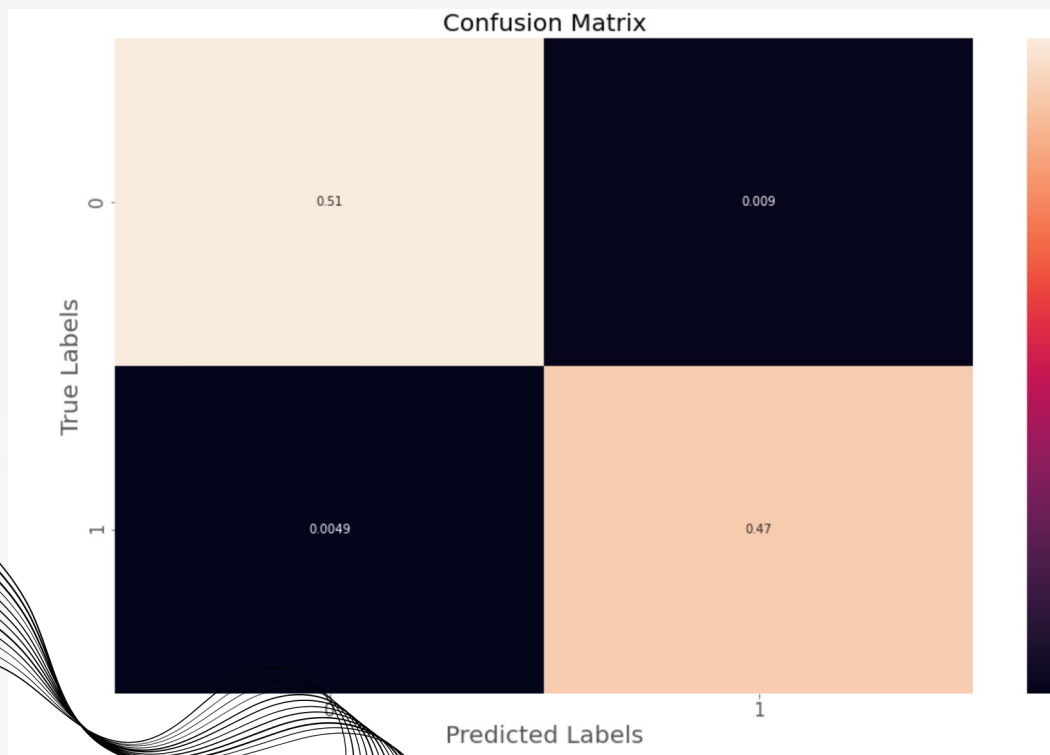
Model Training History
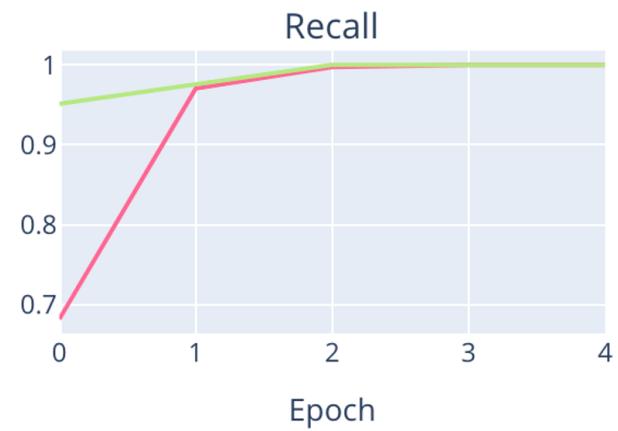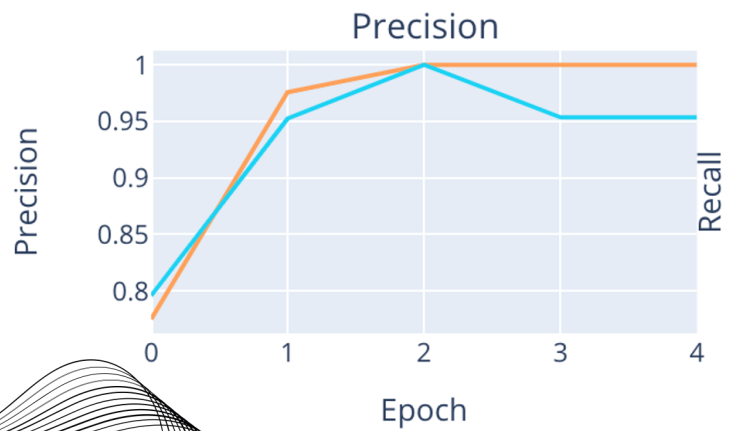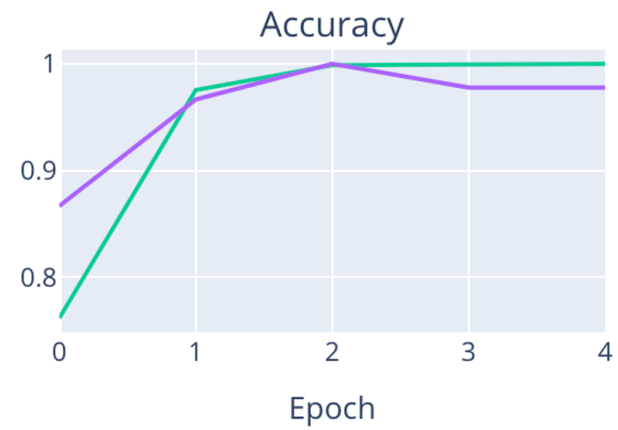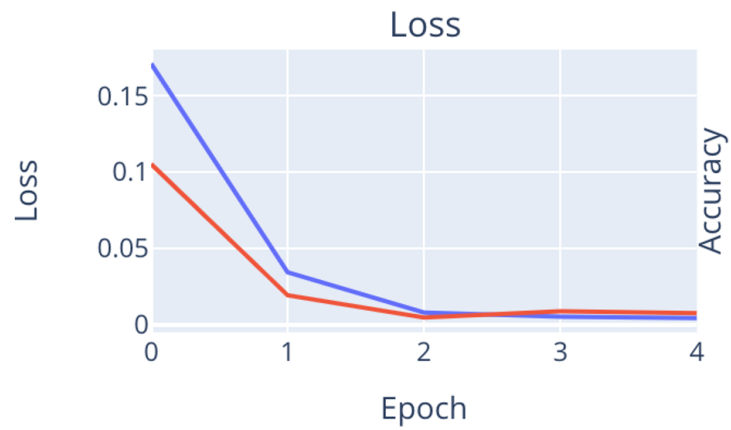
Evaluation Module:

- This module involves evaluating the performance of the trained model using metrics like accuracy, precision, recall, F1-score, and ROC-AUC.

- **Accuracy on testing set:** 0.9860801781737194

- **Precision on testing set:** 0.9812413154238073

- **Recall on testing set:** 0.9897220275636534