Fake news detection using NLP

Project title	**Fake news detection using NLP**
Skills taken away from this project	• **Python scripting** • **Data Preprocessing** • **Machine learning and NLP**
Domain	**Multimedia**

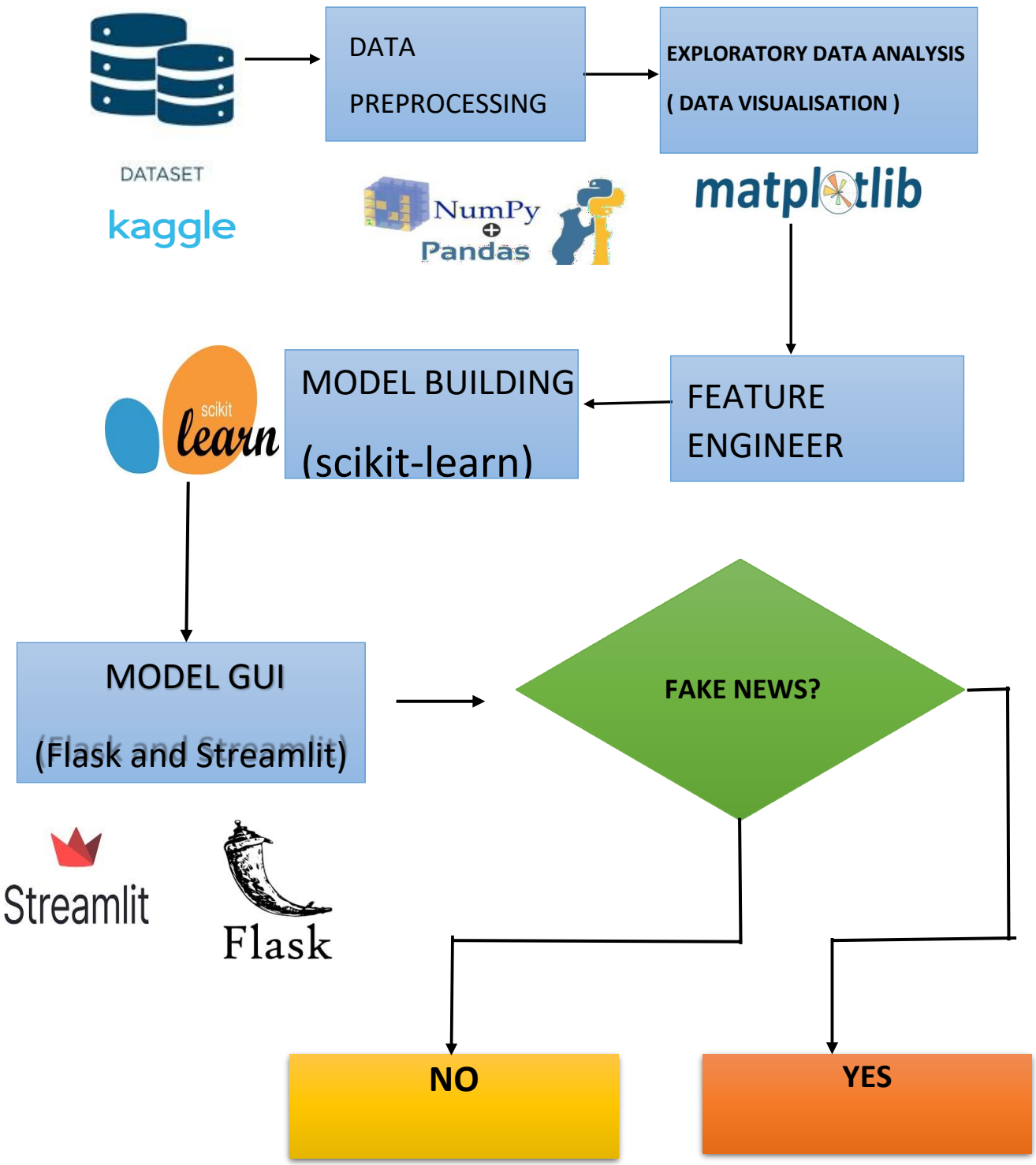
Problem statement:

"Fake News Detection using NLP" is to develop an effective and accurate system that can automatically identify and classify misleading or fabricated information from genuine news articles and textual content. This involves leveraging Natural Language Processing techniques to analyze linguistic features, syntactic structures, and contextual cues in order to make informed assessments about the veracity of the provided information. The goal is to create a reliable tool that aids in combating the proliferation of misinformation, thereby fostering a more informed and trustworthy information ecosystem for users.

Abstract:

With the rapid proliferation of digital media and social networking platforms, the spread of fake news has become a significant challenge in today's information landscape. Detecting and mitigating the impact of fake news is crucial for maintaining the integrity of public discourse and preserving the credibility of reliable sources. This research presents a comprehensive approach for fake news detection employing Natural Language Processing (NLP) techniques.The proposed methodology leverages NLP to extract and analyze linguistic features from textual content, including lexical, syntactic, and semantic attributes. These features are used to train and validate machine learning models, enabling the system to distinguish between genuine and fabricated news articles. Additionally, the model is equipped with a robust preprocessing pipeline that addresses issues such as noise, bias, and imbalance in the dataset.

Design Techniques:



Dataset Link :

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

Approaches

Machine learning like

1.Data understanding

Data understanding is a critical phase in the data analysis process that involves acquiring, exploring, and comprehending the dataset. It helps establish a solid foundation for subsequent analysis and modeling.

2.Data preprocessing

Data preprocessing is a crucial step in preparing raw data for analysis or machine learning tasks. It involves various techniques to clean, transform, and organize data, making it suitable for further processing. Common steps include handling missing values, scaling features, encoding categorical variables, and splitting data into training and testing sets.

3.EDA (Exploratory data analysis)

EDA stands for Exploratory Data Analysis. It's an approach to analyzing datasets to summarize

their main characteristics, often with visual methods. EDA helps in understanding the structure, patterns, and relationships in the data. It involves techniques like summary statistics, visualizations, and data cleaning.

4. Feature Engineering

Feature engineering refers to manipulation — addition, deletion, combination, mutation — of your data set to improve machine learning model training, leading to better performance and greater accuracy.

5. Model building and evaluation

Model evaluation is the process that uses some metrics which help us to analyze the performance of the model. As we all know that model development is a multi-step process and a check should be kept on how well the model generalizes future predictions.

6. Model GUI

A graphical user interface (GUI) is a type of user interface through which users interact with electronic devices via visual indicator representations.

7. Using NLP for analysing the news article

This project will involve investigating and experimenting with different NLP approaches to extract information from news articles, and then classify them

Learning outcomes

1. Developing profusion in phyton programing language and it's data analysis libraries (panda,numpy, matplotlib,seaborn,scikid-learn,NLP libraries and GUI libraries)

2. gaining experience in data preprocessing techniques such as handling missing data outlayer and data normalisation

3. Understanding and visulation the data using EDA technique

4. Learning and applying machine learning techniques such as regression and classification

5. Buliding and optimising ML model using appropriate evaluation Matrix

6. Experience in feature Engineering techniques to create new information representation of the data

7 developing web application using GEY

8. Understanding and the challenges and best practices in the multimedia and how Machine can solve them.