Project Title: Fake News Detection Using Machine Learning

Problem Statement:

The proliferation of misinformation and fake news poses a significant challenge in today's digital age. The goal of this project is to develop a machine learning model capable of identifying fake news articles and distinguishing them from genuine news sources.

Project Overview:

Fake news detection involves analysing textual content to identify misleading or deceptive information. By leveraging natural language processing (NLP) techniques and machine learning algorithms, I aim to build a robust model that can effectively classify news articles as either real or fake based on their content and context.

I am going to to perform exploratory data analysis to understand the distribution and characteristics of real and fake news articles. analyse the frequency of words, topics, and sentiments in different types of news articles. explore correlations between news features and identify potential patterns or biases.

For this I need to extract relevant features from the textual content of news articles, such as word frequency, n-grams, and sentiment scores, Preprocess text data by removing stopwords, punctuation, and special characters, and performing stemming or lemmatization.

Data Collection: For this project, i will gather a dataset of news articles labelled as either real or fake from reliable sources such as reputable news agencies, fact-checking organizations, and datasets available online. The dataset will include features such as article text, title, publication date, author information, and source credibility, dataset will be submitted in the next submission

Data Preprocessing:

Preprocess text data by removing stopwords, punctuation, and special characters.

Perform stemming or lemmatization to standardize word forms.

Use techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings to represent textual features numerically.

Type of Problem:

Fake news detection is a classification problem where I aim to predict whether a news article is real, or fake based on its content.

I will use supervised learning techniques and labelled data to train our classification model. If labelled data is not available, I may explore semi-supervised or unsupervised learning approaches.

Conclusion:

Fake news detection is a challenging yet critical task in today's information-driven society. By developing a robust machine learning model, I aim to empower users to identify trustworthy news sources and combat the spread of misinformation effectively. Through diligent data collection, preprocessing, model training, and evaluation,