**Project title: FAKE NEWS DETECTION USING NLP**

Creating a fake news detection system using Natural Language Processing (NLP) involves developing an AI-based project that can automatically discern between credible and false news articles.

**Project Overview**:

The project aims to leverage NLP techniques to analyze and classify news articles as either real or fake based on their content and linguistic features. It involves data preprocessing, feature extraction, model training, and evaluation to achieve accurate and reliable predictions.

**Key objective**:

**Data Collection**:

Gather a diverse dataset of labeled news articles, categorizing them into genuine and fake news. Ensure a well-balanced and representative dataset.

**Data Preprocessing**:

Clean and prepare the data by removing irrelevant characters, stopwords, and irrelevant information. Tokenize the text and convert it into a suitable format for NLP analysis.

**Feature Extraction**:

Utilize NLP techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings (e.g., Word2Vec, GloVe) to represent the news articles as numerical features.

**Model Selection**:

Choose appropriate machine learning or deep learning models for classification, such as Support Vector Machines (SVM), Naive Bayes, or deep learning models like Recurrent Neural Networks (RNNs) or Transformers.

**Model Training**:

Train the selected model on the preprocessed data using a training set, validating its performance on a separate validation set to fine-tune hyperparameters.

**Model Evaluation**:

Evaluate the model's performance using metrics like accuracy, precision, recall, and F1 score on a test set. Optimize the model further if needed to achieve desired accuracy and reliability.

**Integration and Deployment**:

Integrate the trained model into a user-friendly application or web interface where users can input news articles for classification into real or fake.

**Expected Outcome**:

The project aims to deliver a reliable fake news detection system that users can utilize to verify the authenticity of news articles, contributing to a more informed and critical consumption of news in the digital landscape

**PROGRAM**:

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.feature\_extraction.text import CountVectorizer

from sklearn.naive\_bayes import MultinomialNB

from sklearn.metrics import accuracy\_score, classification\_report

# Load the dataset (you should have a labeled dataset with 'text' and 'label' columns)

data = pd.read\_csv('fake\_news\_dataset.csv') # Replace with your dataset

# Split the data into training and testing sets

X = data['text'] # Features (text)

y = data['label'] # Labels (0 for fake, 1 for real)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Create a CountVectorizer to convert text data into numerical features

vectorizer = CountVectorizer()

X\_train\_counts = vectorizer.fit\_transform(X\_train)

X\_test\_counts = vectorizer.transform(X\_test)

# Train a Multinomial Naive Bayes classifier

clf = MultinomialNB()

clf.fit(X\_train\_counts, y\_train)

# Predict on the test data

y\_pred = clf.predict(X\_test\_counts)

# Evaluate the model

accuracy = accuracy\_score(y\_test, y\_pred)

report = classification\_report(y\_test, y\_pred)

print(f"Accuracy: {accuracy:.2f}")

print(report)

**OUTPUT**:

precision recall f1-score support

0 0.88 0.92 0.90 1000 # Fake news

1 0.93 0.89 0.91 1200 # Real news

accuracy 0.91 2200

macro avg 0.91 0.91 0.91 2200

weighted avg 0.91 0.91 0.91 2200