IBM NAAN MUDHALVAN

ARTIFICIAL INTELLIGENCE-GROUP 3

PROJECT: TEAM MEMBER:

FAKE NEWS DETECTON NAME:SRIRAJ.R

USING NLP REG NO:721921243111

PHASE 3: DEVELOPMENT PART-I



**INTRODUCTION:**

This project aims to develop a fake news detection system using natural language processing techniques to help users differentiate between real and fake news articles.This process involves data collection,text preprocessing and the application of deep learning models for classification.

**ALGORITHM:**

STEP 1: Import the library packages and modules

STEP 2: Load the Datasets

STEP 3: Exploring the Datasets

STEP 4: Split the dataset into training and testing sets

STEP 5: Preprocess the text data

STEP 6:Print the results

**PROGRAM:**

In the Development phase part-1,I’ve started building my project upto data preprocessing.

import pandas as pd

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

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import accuracy\_score, confusion\_matrix, precision\_score, recall\_score

fake\_data = pd.read\_csv('Fake.csv')

fake\_data['label'] = 'FAKE'

true\_data = pd.read\_csv('True.csv')

true\_data['label'] = 'REAL'

data = pd.concat([fake\_data, true\_data], ignore\_index=True)

# Split the dataset into training and testing sets

X = data['title'] + ' ' + data['text'] + ' ' + data['subject'] + ' ' + data['date']

y = data['label']

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

# Preprocess the text data using TF-IDF vectorization

tfidf\_vectorizer = TfidfVectorizer(stop\_words='english')

X\_train\_tfidf = tfidf\_vectorizer.fit\_transform(X\_train)

X\_test\_tfidf = tfidf\_vectorizer.transform(X\_test)

**Libraries and Modules used:**

1.PANDAS :

I import the pandas library first. It is an open source data manipulation and analysis library,widely used for tasks like data cleaning,transformation and analysis.

2.SCIKIT-LEARN :

It is often referrd to as sk-learn which provides a set of tools for various ml tasks including classification,regression,clustering,model selction and more.

3.LOGISTIC REGRESSION:

In this project ,logistic regression method is used to find the find the probability of real and fake news articles.This method is primarily used to predict the probability of an instance belonging to a particular class or category.

4.SK-LEARN MATRICES:

SK-learn matrices includes accuracy score, confusion matrix, precision score, recall score are imported to calculate the accuracy ,prediction and more.

5.TF-IDF VECTORIZATION :

Data preprocessing is done by using TF-IDF vectorization.Term frequency measures the frequency of word in a document.Inverse Document Frequency measures how important a term is across a collection of documents.

**CONCLUSION:**

In conclusion, I have build my project upto data preprocessing and learn about some libraries in python.I’ve also got knowledge in some methods used in NLP.