Task 8 - Fake News Detection Model

Name: Hasmitha

Project Objective:

To build a machine learning model that detects fake news using text classification techniques.

Dataset:

The dataset used contains news articles with labels indicating whether the news is REAL or FAKE.

Approach:

- Preprocessing of text using TfidfVectorizer.

- Training a Passive Aggressive Classifier.

- Evaluating the model with accuracy and confusion matrix.

Code Snippet:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import PassiveAggressiveClassifier
from sklearn.metrics import accuracy_score, confusion_matrix

df = pd.read_csv('news.csv')
labels = df.label
X_train, X_test, y_train, y_test = train_test_split(df['text'], labels, test_size=0.2,
random_state=7)
tfidf_vectorizer = TfidfVectorizer(stop_words='english', max_df=0.7)
tfidf_train = tfidf_vectorizer.fit_transform(X_train)
tfidf_test = tfidf_vectorizer.transform(X_test)
pac = PassiveAggressiveClassifier(max_iter=50)
pac.fit(tfidf_train, y_train)
y_pred = pac.predict(tfidf_test)
score = accuracy_score(y_test, y_pred)
print(f'Accuracy: {round(score*100,2)}%')
confusion_matrix(y_test, y_pred, labels=['FAKE','REAL'])
```

Sample Output:

Accuracy: 93.5%

Confusion Matrix:

[[589 49]

 [43 587]]