FAKE NEWS DETECTION USING MACHINE LEARNING

Submitted by: Jyothi Neelam

Submission Date: 20-06-2025

This project aims to detect fake news using a Logistic Regression classifier.

The model is trained on a dataset of real and fake news and uses TF-IDF vectorization

to convert text data into numerical format. This helps in identifying patterns in language

that distinguish fake news from real news.

```
import pandas as pd
import numpy as np
import re
import string

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, classification_report

df = pd.read_csv('fake_or_real_news.csv')
df['text'] = df['text'].apply(lambda text: re.sub(r'[%s]' % re.escape(string.punctuation), '', text.lower()))
X = df['text']
y = df['label'].apply(lambda x: 1 if x == 'REAL' else 0)

vectorizer = TfidfVectorizer(stop_words='english', max_df=0.7)
X_vec = vectorizer.fit_transform(X)
X_train, X_test, y_train, y_test = train_test_split(X_vec, y, test_size=0.2, random_state=42)

model = LogisticRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

print("Accuracy:", accuracy_score(y_test, y_pred))
```