Fake news detection using -NLP

Project title	Fake news detection using NLP
Skills taken away from this project	• Data set splitting & Training • Model Evaluation • Model Prediction
Domain	Multimedia
Team Members	S.Sachith J.Devadharshanan M.Mohamed Irfan V.Chandiran

Introduction:

Fake news detection using Natural Language Processing (NLP) is a critical field of research and application aimed at identifying and mitigating the spread of misleading or false information in digital media. With the rapid expansion of social media and online news platforms, the dissemination of misinformation has become a pressing concern. NLP, a sub field of artificial intelligence, plays a pivotal role in addressing this issue by leveraging techniques from linguistics and machine learning to analyze and understand text data.

Objective:

✧ Fake news detection using machine learning is to develop a model or system that can automatically identify and classify news articles or information as either "real" or "fake".

Library Installation:

Import the necessary libraries for this project

```
[1] import numpy as np
    import pandas as pd
    import re
    from nltk.corpus import stopwords
    from nltk.stem.porter import PorterStemmer
    from sklearn.feature_extraction.text import TfidfVectorizer
    from sklearn.model_selection import train_test_split
    from sklearn.linear_model import LogisticRegression
    from sklearn.metrics import accuracy_score
```

Splitting the data set to training & test data:

a) Training the Model: Logistic Regression:

✧ Logistic regression is a statistical method used for analyzing a data set in which there are one or more independent variables that can be used to predict the outcome of a categorical dependent variable..

```
[ ]  model = LogisticRegression()
```

```
▶  model.fit(X_train, Y_train)
```

```
😊  LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
                       intercept_scaling=1, l1_ratio=None, max_iter=100,
                       multi_class='auto', n_jobs=None, penalty='l2',
                       random_state=None, solver='lbfgs', tol=0.0001, verbose=0,
                       warm_start=False)
```

b) Model Evaluation:

✧ Model evaluation is a crucial step in the machine learning pipeline. It involves assessing how well a trained model performs on a data set it has never seen before. The goal is to understand how the model generalizes to new, unseen data, which is essential for its practical application.

```
[ ]  # accuracy score on the training data
     X_train_prediction = model.predict(X_train)
     training_data_accuracy = accuracy_score(X_train_prediction, Y_train)
```

```
▶  print('Accuracy score of the training data : ', training_data_accuracy)
```

```
😊  Accuracy score of the training data :  0.9865985576923076
```

```
[ ]  # accuracy score on the test data
     X_test_prediction = model.predict(X_test)
     test_data_accuracy = accuracy_score(X_test_prediction, Y_test)
```

```
[ ]  print('Accuracy score of the test data : ', test_data_accuracy)

     Accuracy score of the test data :  0.9790865384615385
```

c) Model Prediction System:

✧ A model prediction system refers to the infrastructure and processes that allow a trained machine learning model to make predictions or decisions on new, unseen data. This system typically involves several components working together.

```
▶  X_new = X_test[3]

   prediction = model.predict(X_new)
   print(prediction)

   if (prediction[0]==0):
     print('The news is Real')
   else:
     print('The news is Fake')
```

```
😊  [0]
   The news is Real
```

```
[ ]  print(Y_test[3])

     0
```

Conclusion:

Fake news detection using NLP offers a powerful tool to combat misinformation. By analyzing linguistic patterns, NLP models can discern the authenticity of news articles. However, ongoing refinement and interdisciplinary collaboration are essential for effective implementation in the fight against fake news.