**FAKE NEWS DETECTION USING NLP**

**INTRODUCTION**

The problem is to implement fake news detection using NLP. The goal is to enable detection of fake news. This project involves data collection, data pre-processing, feature engineering, clustering algorithms, visualization, and interpretation of results.

In this phase the design to innovation and data flow of customer segmentation is going to be done.

**DATASET**

The data is obtained from https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset

**COLUMNS USED**

From Fake.csv, True.csv data the following columns are used

* title
* text
* subject
* date

**LIBRARIES USED**

The Python 3 environment comes with many helpful analytics libraries installed and several

helpful packages to load.

The essential libraries used in this project are :

1. pandas
2. numpy
3. matplotlib.pyplot
4. seaborn
5. nltk
6. re
7. string
8. sklearn.model\_selection import train\_test\_split
9. sklearn.metrics import classification\_report
10. keras
11. keras.preprocessing import text,sequence
12. keras.models import Sequential
13. keras.layers import Dense,Embedding,LSTM,Dropout
14. warnings
15. os

**TRAIN AND TEST**

Training the dataset by describe(), isnull().sum(), drop(), show(), and by using k-means

algorithm we train the data

Testing the data by importing sklearn.cluster from k-means with ensuring the plot range and axis

labels producing the k value, scattering the data by kmeans.cluster\_centers and producing 3D

plot.

**REST OF THE EXPLANATIONS**

1. **Data Collection:** The process involves gathering sample news data, which includes information about the data provided from the column of the above “CSV” files(fake.csv , true.csv)
2. **Data Preprocessing:** The task involves removing html contents, remove unrelated texts, removal of stop words, lemmatization, truncate data without labels, stemming.
3. **Feature Engineering:** Data preparation and cleaning, handling missing values, and the transformation of categorical features into numerical representations are all part of the task.

**ALGORITHMS USED**

Apply clustering algorithms like K-Means, NLP algorithms, WebCrawlers etc… to detect and analyze the data.

**Visualization**: Visualize the news by using elements like scatter plots, bar

charts, and other techniques.

**Interpretation:** Analyze and interpret the characteristics of each news like number of words etc… to derive actionable insights from the news.



