Domain:Aritificial intelligence

# Project name: fake news Detection using NLP

# Problem definition:

# Fake news detection using natural language processing (NLP).

# NLP involves leveraging machine learning techniques to analyse text and identify patterns indicative of misinformation or deception

# verify if news is current old news can be Misleading if presented as recent.

# Design and thinking :

* This project involves 6 steps. They are Data source, Data preprocessing,Feature Extraction, Model selection, Model Training, Evaluation

1.Data source

* Gather a diverse dataset of new article or social media posts categorizing them as either real or fake .
* You’ll need labeled data to train and evaluate the NLP model
* But in our case we are already provided by datasets for KAGGLE platform .
* Dataset link: <https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset>with NLP.
* The dataset contain a list of articles considered as “ fake “news.
* They contain 4 columns and 17903 unique values. The dataset in the columns based on
* 1.Tittle
* 2.text
* 3.subject
* 4.date

2.Data pre-processing

* Clean and pre-process the text data by removing stop words, punctuation and special characters
* In this part include 3 things

1.Text cleaning

✓ Remove HTML tags ,if any from the text.

✓ convert text to lowercase to ensure uniformity

✓ Handle or remove any non standard characters,puncutation numbers.

2.Tokenization

✓Split the text into individual words tokens.

✓This allows the model to processes the text at the word level.

3.Stop word removal

✓ Remove common stop words (eg., “the”,”and”,”in”)that do not carry’s

Significant

3.Feature Extraction

* Transform the convert text into numerical features that the common model machine learning model.
* It methods include TF-IDF(Term frequency -inverse Document frequency)

1.Tf-IDF : it measures the importance of a word within document relatives to its importance in the entire corpus.

* .It balances the frequency of a word in a document with its rarity in the corpus making suitable to its identifying signifying terms

2. Word embedding : word. Embedding like word2vec ,glove and fast test capturing

Semantic relationship between word

* These embedding provide dense vector representation of vector representation of words and you can average or concatenate them to represent documents .

4.Model selection

* In this part we are choosing a suitable model for Random forest for fake news detection.
* Model selection involves 4 steps

✓. It is a group of decision tree from a subset of randomly selected a training dataset .

✓. It combines a multiple tree to form a forest hence the random forest to give high accuracy

To the proposed the model

✓.This type of random forest is ensemble learning.

5.Model training

* Here, we first create an instance of the random forest model ,with the hyper parameter .
* We then fit this to our training data we pass both the targets variables ,so the model can learn
* At this point we have a trained Random forest model ,but we need to find out whether it is making fake news prediction

7 Steps involved in this model training

✓. Import the Skirt-learn python libraries

✓.import the data

✓.Read/clean/adjust the data

✓.Train the classifier model of the training data

✓.create the Random forest model object

✓. Fit the model

✓.predict

6.Evaluation

* After a training a data evaluate the final random forest model and on the dataset to obtained an unbiased estimate to its performance
* Use appropriate evaluation such as acuracy,precision, recall ,f1-score, and Roc-AUC

Conclusion:

✓.These are all the steps are involved in fake news detection using NLP.

✓.To tackle the increasing false information on the internet the machine learning model distinguish as Input such as real or fake news.

✓.These are the things that we are understood in the project ,By this way we are going to done our project