***FAKE NEWS DETECTION USING NLP***

***TEAM MEMBER***

***731721106013 – POONGOTHAI.A***

***PHASE-4 PROJECT SUBMISSION***

***PROJECT: Fake News Detection***



***PROBLEM STATEMENT:***

***The problem is to develop a fake news detection model using a Kaggle dataset. The goal is to distinguish between genuine and fake news articles based on their titles and text. This project involves using natural language processing (NLP) techniques to preprocess the text data, building a machine learning model for classification, and evaluating the model's performance.***

***INTRODUCTION:***

* ***Fake news detection is a critical task in today's information age. It involves the identification and verification of misleading or fabricated information presented as factual news.***
* ***With the proliferation of social media and digital platforms, the spread of fake news has become a significant concern, impacting public discourse and decision-making.***
* ***In this context, advanced technologies, such as natural language processing and machine learning, play a vital role in developing tools and algorithms to detect and combat fake news effectively.***
* ***This is a challenging and evolving field that aims to safeguard the integrity of information and promote responsible journalism in the digital era***

***Phase-4 Work:***

**Feature Selection:**

**Feature selection is a critical step in the process of building a fake news detection model. It involves choosing the most relevant and informative features from the text data to be used as input for the machine learning model. In this context, features are typically derived from the textual content of news articles**.

**Model Training:**

**Model training involves the process of teaching the machine learning algorithm to distinguish between genuine and fake news based on the selected features. Different algorithms and approaches can be considered for this task.**

**Evaluation:**

**Evaluating the fake news detection model is essential to assess its performance and reliability in real-world applications. Several metrics can be used to evaluate the model's effectiveness:**

***DATA SOURCE:***

**Employing robust methodologies to assess the credibility and authenticity of information providers to enhance the accuracy of news verification processes.**

**DATA INFORMATIONS**:

[**https://colab.research.google.com/drive/1Y-vkaF9aA82-\_AiOlaDTutNsF4qWFOOF#scrollTo=66zEsLXY-IxY&line=1&uniqifier=1**](https://colab.research.google.com/drive/1Y-vkaF9aA82-_AiOlaDTutNsF4qWFOOF#scrollTo=66zEsLXY-IxY&line=1&uniqifier=1)

**SOURCE CODE:**

***import pandas as pd***

***not\_fake = pd.read\_csv("../input/fake-and-real-news-dataset/True.csv")***

***fake = pd.read\_csv("../input/fake-and-real-news-dataset/Fake.csv")***

***X = not\_fake["title"].tolist() + fake["title"].tolist()***

***y = [0] \* len(not\_fake) + [1] \* len(fake)***

***from sklearn.model\_selection import train\_test\_split***

***X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2)***

***!pip install simple\_nlp\_library***

***from simple\_nlp\_library import preprocessing, embeddings***

***stop\_words = preprocessing.stop\_words()***

***vectors = embeddings.vectors()***

***X\_train\_vec = [embeddings.tokens\_vector(vectors, preprocessing.semantic\_tokens(stop\_words, x)) for x in X\_train]***

***X\_test\_vec = [embeddings.tokens\_vector(vectors, preprocessing.semantic\_tokens(stop\_words, x)) for x in X\_test]***

***from sklearn.neural\_network import MLPClassifier***

***clf = MLPClassifier(hidden\_layer\_sizes=(25), early\_stopping=True)***

***clf.fit(X\_train\_vec, y\_train)***

***import matplotlib.pyplot as plt***

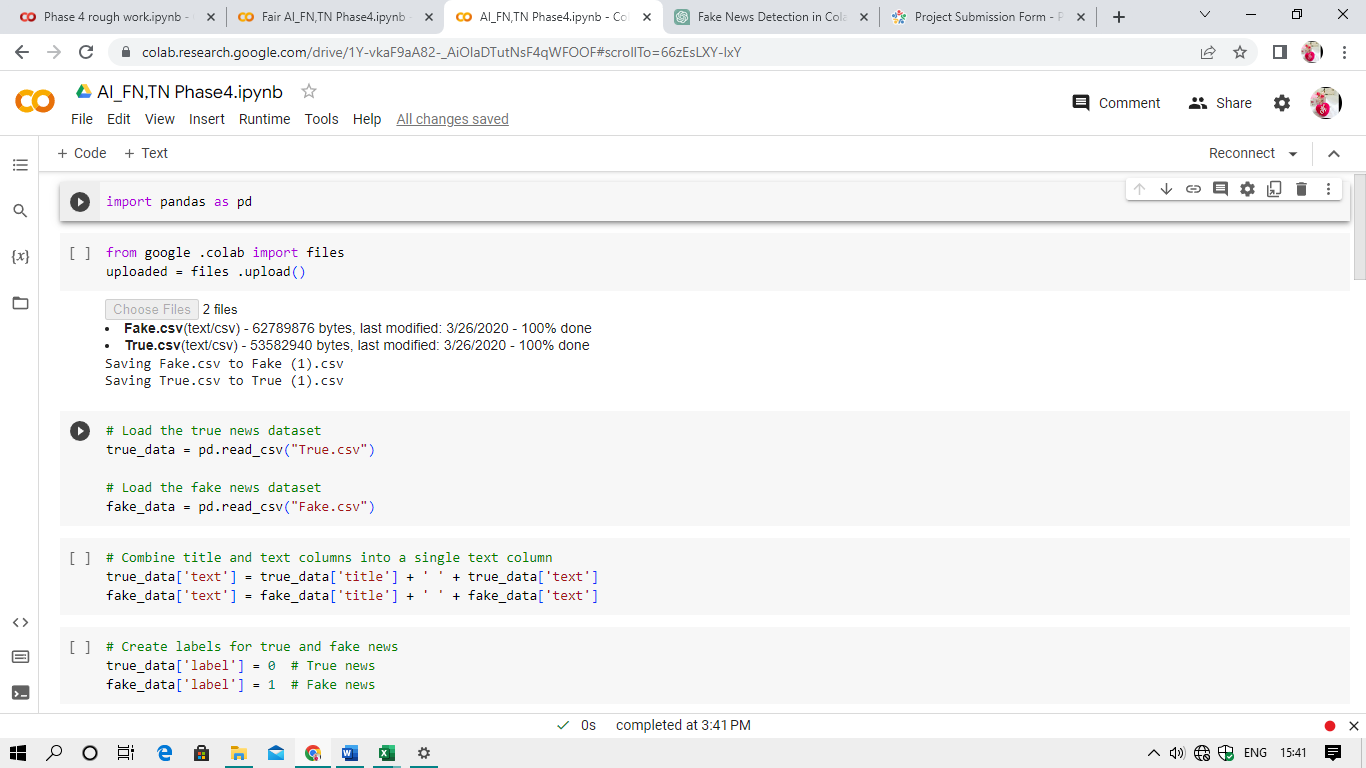
***plt.plot(clf.validation\_scores\_)***

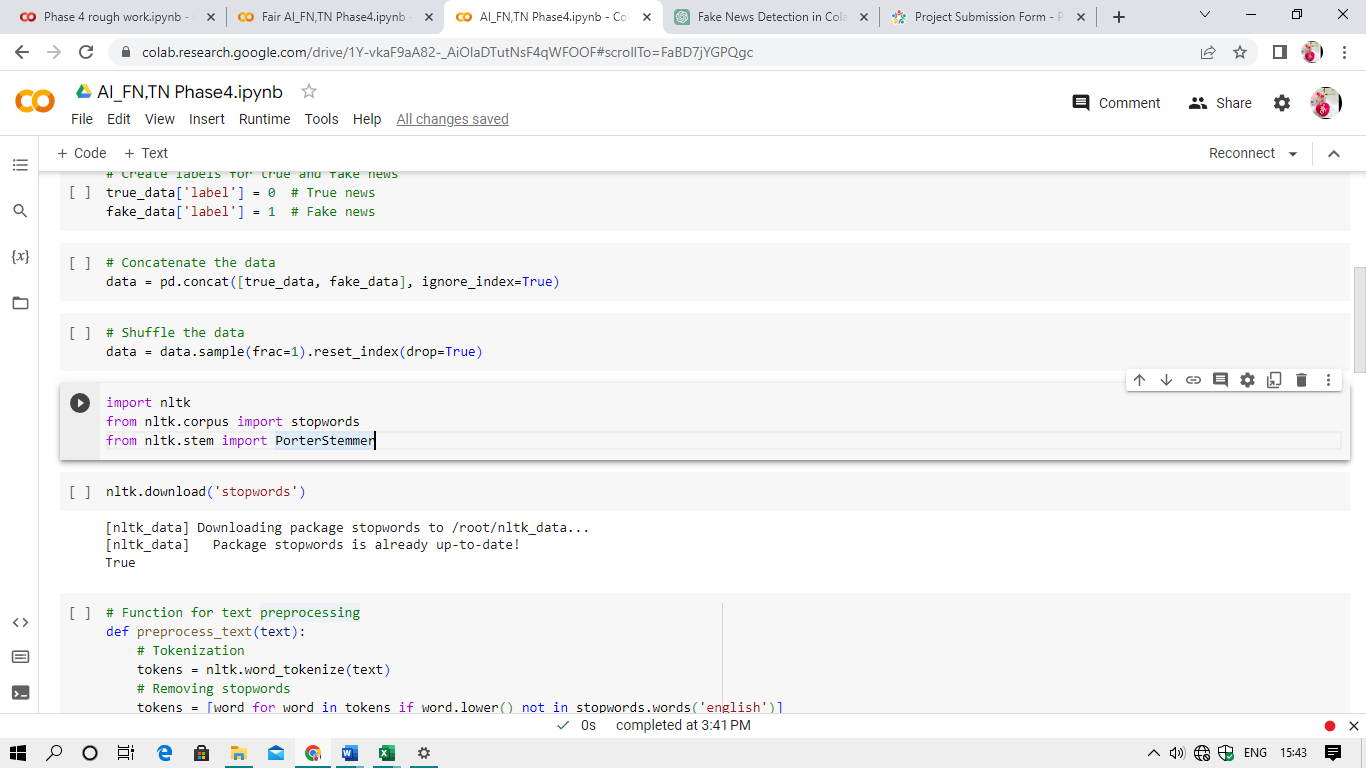
***from sklearn.metrics import accuracy\_score***

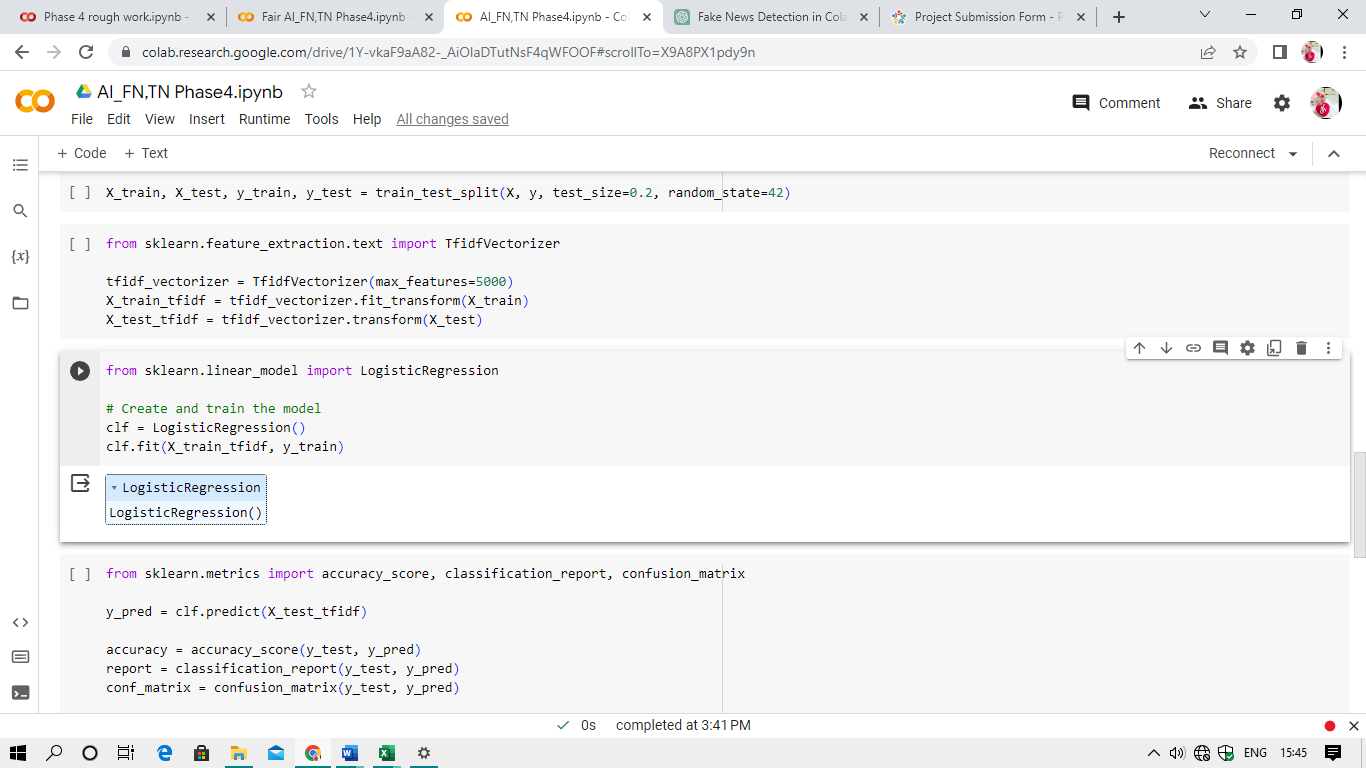
***accuracy\_score(y\_train, clf.predict(X\_train\_vec))***

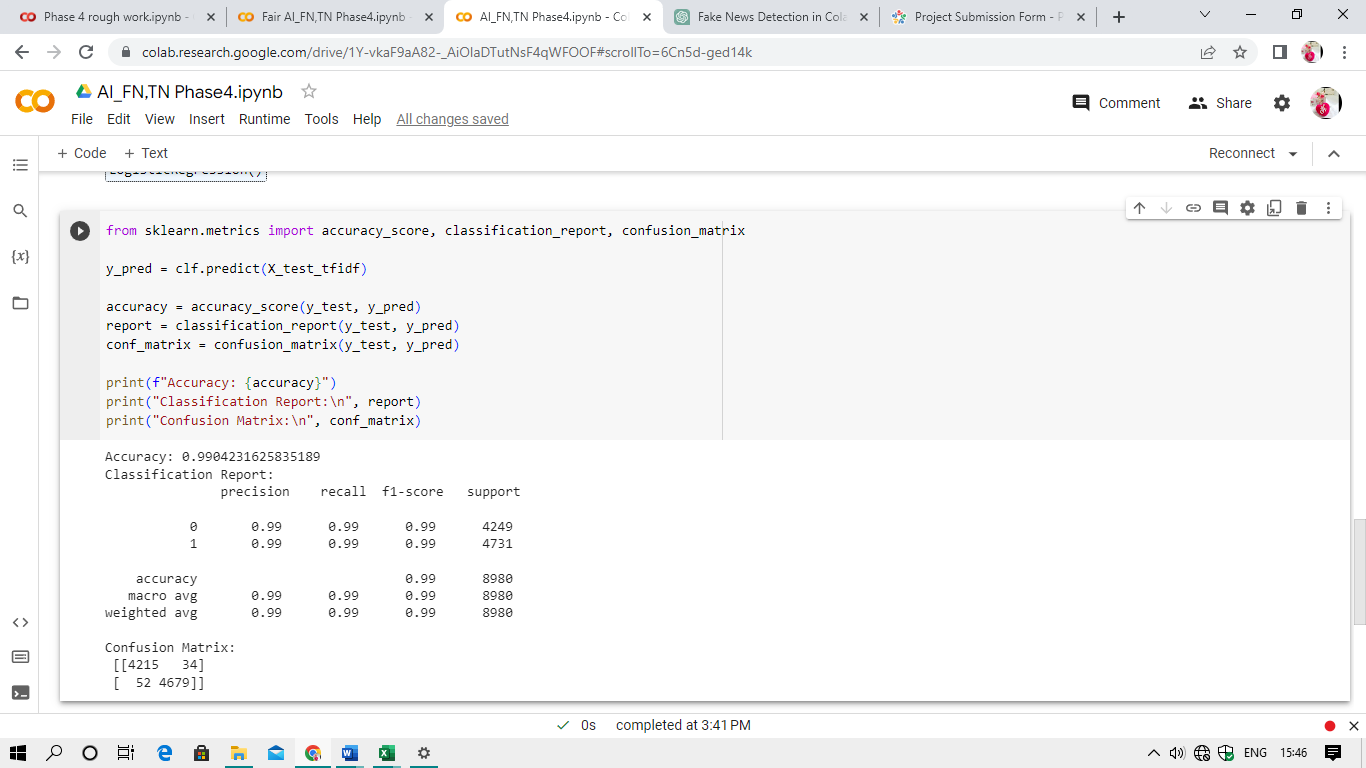
***from sklearn.metrics import accuracy\_score***

***accuracy\_score(y\_train, clf.predict(X\_train\_vec))***









***DATASET LINK:***

[**https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset**](https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset)

***That’s all the informations we gathered…!***