Project Report

on

Fake News Detection

Neeraj Singh, 11707114, KM031, Roll no 40

1. Collection of Database

I took the dataset from DataFlair.com. The dataset consists of all the districts in India and the number of each type of crime committed. This dataset is of the year 2014. The URL for the dataset is:

<https://github.com/15hu/Criminal-Behaviour-Analysis-and-Segmentation/blob/master/crime_data.csv>

1. Implementation Code with Comment Lines

# -\*- coding: utf-8 -\*-

"""

1.Name- Arun Yadav

Reg.id-11711181

2.Name-Neeraj Singh

Reg.id-11707098

"""

"""

Spyder Editor

This is a temporary script file.

"""

# Dataset Downloaded from Kaggle.com

import pandas as pd

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

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.linear\_model import PassiveAggressiveClassifier

from sklearn.svm import LinearSVC

from sklearn.ensemble import RandomForestClassifier

from sklearn.tree import DecisionTreeClassifier

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import accuracy\_score

df=pd.read\_csv('C:\\Users\\LENOVO\\Downloads\\news.csv')

#Get shape and head

df.shape

df.head()

#Get the labels

labels=df.label

labels.head()

#Split the dataset

x\_train,x\_test,y\_train,y\_test=train\_test\_split(df['text'], labels, test\_size=0.2, random\_state=7)

#Initialize a TfidfVectorizer.For the stopping word

tfidf\_vectorizer=TfidfVectorizer(stop\_words='english', max\_df=0.7)

Classifier=[PassiveAggressiveClassifier,DecisionTreeClassifier,RandomForestClassifier,LinearSVC,LogisticRegression]

Name=['PassiveAggressiveClassifier','DecisionTreeClassifier','RandomForestClassifier','LinearSVC','LogisticRegression']

# Fit and transform train set, transform test set

tfidf\_train=tfidf\_vectorizer.fit\_transform(x\_train)

tfidf\_test=tfidf\_vectorizer.transform(x\_test)

for i in range(0,5):

if(i==0):

p=Classifier[i](max\_iter=50);

else:

p=Classifier[i]();

p.fit(tfidf\_train,y\_train)

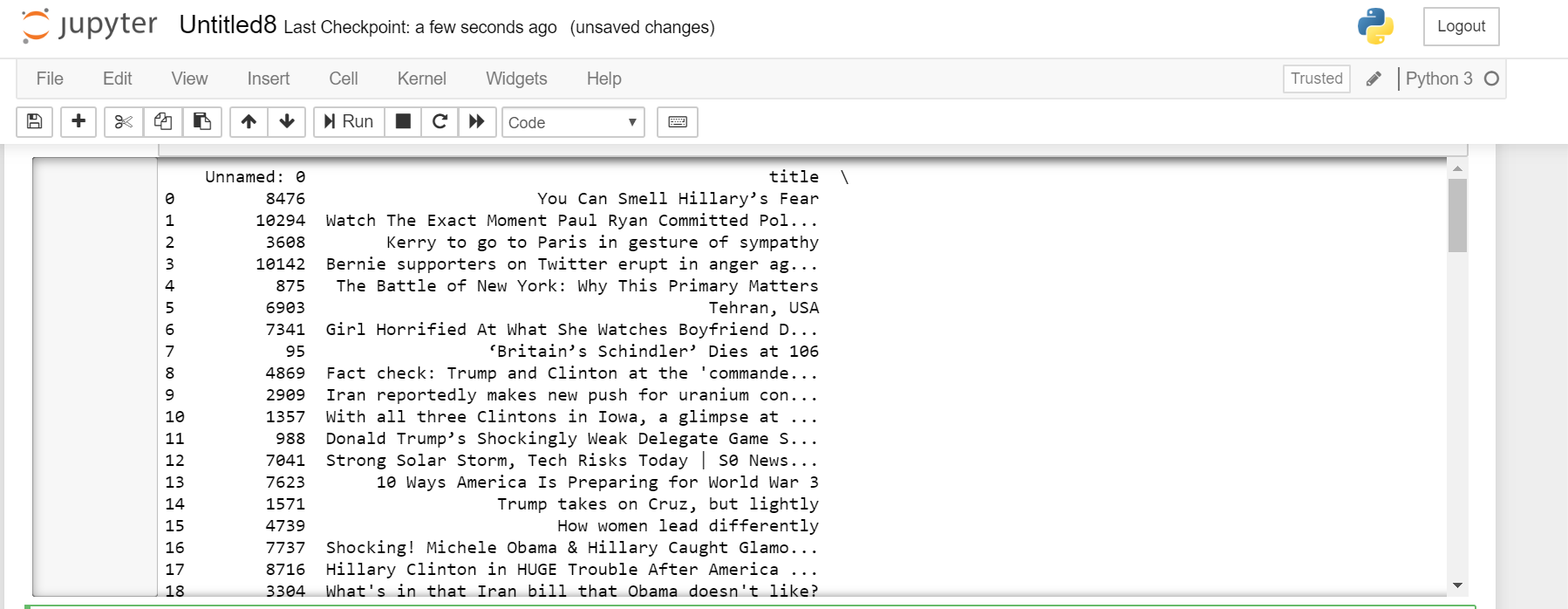
#DataFlair - Predict on the test set and calculate accuracy

y\_pred=p.predict(tfidf\_test)

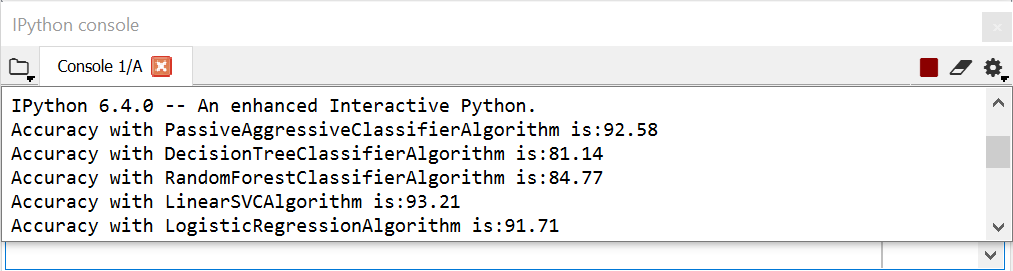
score=accuracy\_score(y\_test,y\_pred)

print('Accuracy with '+Name[i]+'Algorithm is:'+str(round(score\*100,2)))

3.Output



4.Output



4.Github Repository

The link of my project repository is:

<https://github.com/Arun1717/int247-machine-learning-project-2020-kem031-rollno_40_53/blob/master/Fake_News_Detection.py>