**BIG DATA ANALYSIS AND APPLICATIONS**

A)How to import required libraries and perform DataCleaning and preprocessing like removing unnecessary Columns or Data,Removing Twitter Handles( @user ), Removing punctuation, numbers, special characters, Removing stop words, Tokenizing, and Stemming, TFIDF vectors, POS tagging, checking for missing values ,train/test split of data.

B)Imported the required libraries for performing sentimental analysis tasks on the given data set. I have used libraries like import pandas as pd import numpy as np import re import string import keras import nltk from nltk import sent\_tokenize from nltk import word\_tokenize #from numpy import arrays from keras.models import Sequential from keras.layers.core import Activation,Dropout,Dense from sklearn.model\_selection import train\_test\_split from keras.preprocessing.text import Tokenizer #plotting libraries from matplotlib import pyplot as plt import seaborn as sns import nltk import nltk nltk.download('stopwords') from nltk.corpus import stopwords #Global parameters STOPWORDS = set(stopwords.words('english')) nltk.download("popular")

FOR DATACLEANING AND PREPROCESSING I HAVE USED SOME FUNCTIONS LIKE:

#changes in lower case tweet.lower() #regular expression is used to remove urls punctuations and all #removes user references and #from tweet tweet=re.sub(r'@\w+|#','',tweet) #used to remove urls tweet=re.sub(r"http\S+|www\S+|https\S+",'',tweet,flags=re.MULTILINE) #used to remove punctuations tweet=tweet.translate(str.maketrans('','',string.punctuation)) #used to remove multisapces tweet=re.sub(r'\s+',' ',tweet) #used to remove single characters tweet=re.sub(r'\s+[a-zA-Z]\s+',' ',tweet) #converting in list tweet=list(tweet.split(" ")) #to remove stop words filtered\_words=[w for w in tweet if not w in STOPWORDS] #printing only filtered words #removing numeric from given data sets alpha\_words=[w for w in filtered\_words if w.isalpha()] return " ".join(alpha\_words) #POS TAGGING #Tokenizing words : tokenized\_words = word\_tokenize(sentence)

#Generating output for TF\_IDF : X = vectorizer.fit\_transform(sentences).toarray()

#TRAIN/TEST SPLIT DATA history=model.fit(X\_train,Y\_train,batch\_size=batch\_size,epochs=epochs,verbose=1,validation\_split=0.1) score=model.evaluate(X\_test,Y\_test,batch\_size=batch\_size,verbose=1)

DATA VISUALIZATION AND ANALYSIS FOR CRITICAL STEPS:

plt.plot(history.history['accuracy']) plt.plot(history.history['val\_accuracy']) plt.title('model accuracy') plt.ylabel('accuracy') plt.xlabel('epoch') plt.legend(['train','validation'],loc='upper left') plt.show()

plt.plot(history.history['loss']) plt.plot(history.history['val\_loss']) plt.title('model loss') plt.ylabel('loss') plt.xlabel('epoch') plt.legend(['train','validation'],loc='upper left') plt.show()

C)IN TRAIN/TEST SPLIT OF DATA I have faced some problems while importing libraries and labeling

D)SCREENSHOTS:

![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%206.46.24%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%206.46.35%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%206.46.48%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.04.50%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.00%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.00%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.07%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.13%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.19%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.25%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.31%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.38%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.43%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.49%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.05.55%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.06.01%20PM.png>)  
![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/output%20screenshots/Screenshot%202021-09-13%20at%207.06.05%20PM.png>)

F)VIDEO LINK

![](<https://github.com/saidurga-kanuganti/icp-03/blob/main/Mon%20Sep%2013%202021%2021_46_41.webm>)

G) I have learned that how to manipulate data from the given data set