***COMSATS UNIVERSITY***

Logo, company name

Description automatically generated

**WAH CAMPUS**

***Submitted By***

|  |  |  |  |
| --- | --- | --- | --- |
| Shayan Babar | M.Arsal | Ali Raza | Saad Rashid |

***Name:***

|  |  |  |  |
| --- | --- | --- | --- |
| Fa19bcs051 | Fa19-BCS-056 | Fa19-BCS-047 | Fa19-BCS-194 |

***Registration No:***

***Class/Section:*** *BSCS/6d*

***Submitted To***

***Teacher Name:*** *Ammara Zamir*

***Date of Submission:*** *14/06/2022*

# Cleaning data

Text

Description automatically generated

Because out data set was small we were able to analyze that the only irregularities it had were specials characters present in it hence using regular expressions we removed than using the above code which is called later in the actual code its self.

# Text blob sentiment analysis and feature extraction

A screenshot of a computer

Description automatically generated with medium confidence

* **getSubjectivity** – finds the subjectivity of the text passed using text blob function extBlob(text).sentiment.subjectivity
* **getPolarity** --finds the Polarity of the text passed using text blob function extBlob(text).sentiment.Polarity
* **features(text)** – ectracts features( positive,negative,neutral,compound) using ntlk function

SentimentIntensityAnalyzer().polarity\_scores(text)

* **len—**feature is founded using string.split()
* **data is maped using dataframe[column].apply(function\_name)**
* **in feature extraction method .map() function is used to map 4 columns in data frame.**

# Label class

Label class is introduced using following code (cols named is” Cleaned\_essay\_Analysis”)

Text

Description automatically generated

Polarity is the output that lies between [-1,1], where -1 refers to negative sentiment and +1 refers to positive sentiment

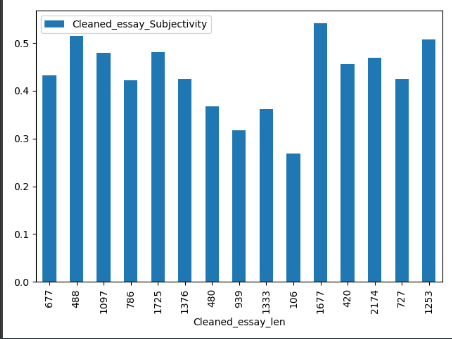
# Mathplotlib

Text

Description automatically generated

A screenshot of a computer

Description automatically generated with medium confidence



*Simple bar chart of subjectivity and length of the essay*

Chart, scatter chart

Description automatically generated

*Scatter plot subjectivity and polarity*

# Seaborn

Text

Description automatically generated

*Same as mathplotlip except the plot time is defined like this sns.plottype(,,,,,)*

Chart, scatter chart

Description automatically generated

# Main code

Text

Description automatically generated

# Csv results

Graphical user interface, text, application

Description automatically generated

Uncleaned scaped data

A picture containing text

Description automatically generated

Cleaned data

Graphical user interface, application, table, Excel

Description automatically generated

After running the code

Graphical user interface, application, table, Excel

Description automatically generated

Values

# Classification using random and accuracy with confusion matrix

Text

Description automatically generated

Before feeding this data into our Machine Learning models I decided to divide our data into features (**X**) and labels (**Y**)

 a function (**forest\_test**) to divide the input data into train and test sets and then train and test a Random Forest Classifier

Text

Description automatically generated

As shown below, training a Random Forest classifier using all the features, led to 100% Accuracy in about 0.9s of training time