

Instruction Set

You should follow following instructions in order to run our code on your machine.

● Softwares to be used :

- R
- RStudio
- Java (should be there in the pc for a particular library we have used)
- MS-Excel

● Common Instructions:

- Open our code in RStudio or any other text editor for R.
- Make sure that you have installed all the libraries in your system which are being used in the code.
 - If haven't done yet, use **install.packages('package_name')** command.
- The working directory should be properly set for reading the csv files.
 - In order to set the correct working directory path you can traverse to the working directory by clicking on the **three dots symbol** present in the right bottom sub-window and then clicking **more** and selecting **set as working directory** from the drop down menu.
 - Than you will see the path to your working directory in the **Console**.

● Code-wise Instructions :

- **Dictionary-Based-Final.R**

- **INPUT :** A CSV file consisting of a Reviews stored in a column named '**Reviews**'.
- **RUNNING INSTRUCTIONS :**
 - In **line number 12** add the path to your input csv file.
 - In **line number 14** add the name of your input csv file.
- **OUTPUT :**
 - Output will be a CSV file named '**polarity_assigned.csv**'. You can change the name of output file if you want by going to **line number 67** and changing the name.
 - This file will have two columns one consisting of the Reviews and the other the Polarity assigned using the lexicon based approach.
 - You can find the file in the working directory set by you in line number 12.

○ Machine-Learning-Based.R

- **INPUT :** A CSV file consisting of **training dataset** that is it should consist of two columns necessarily. One is '**Reviews**' and other is '**Polarity**'.
- **RUNNING INSTRUCTIONS :**
 - In **line number 21** add the path to your input csv file.
 - In **line number 25** add the name of your input csv file.
 - Select upto **line number 50** and hit '**ctrl + Enter**'.
 - Below that you can find separate 4 line codes for different algorithms. So you can run whatever you want. It is properly commented that which four line is for which part.

- **OUTPUT:** Output will be two tables in the console. One consisting of training set predictions and actual values and other same for the test set.

○ Multinomial_Logistic_Regression_Overall_Score.R

■ INPUT :

- A CSV file consisting of **training dataset**.
- A CSV file consisting of a Reviews stored in a column named '**Reviews**'.

■ RUNNING INSTRUCTIONS :

- In **line number 22** add the path to your input CSV file in the path_1 variable.
- In **line number 27** add the name of your training data CSV file.
- In **line number 60** add the path to your raw csv file whose polarities you have to predict in the path_2 variable.
- In **line number 65** add the name of your raw csv file.

■ OUTPUT :

- Company classified as bad, average or good with respect to employees reviews printed on the console.
- Overall Score of the company out of 5 printed on the console.
- A CSV file named 'new_data.csv' consisting of individual review's polarity. You can change the name of output file if you want by going to **line number 105** and changing the name.
- This file will have two columns one consisting of the Reviews and the other the Polarity assigned using the lexicon based approach.
- You can find the file in the working directory set by you in line number 12.

○ Random-Forest_Overall_Score.R

■ INPUT :

- A CSV file consisting of **training dataset**.
- A CSV file consisting of a Reviews stored in a column named '**Reviews**'.

■ RUNNING INSTRUCTIONS :

- In **line number 22** add the path to your input CSV file in the path_1 variable.
- In **line number 27** add the name of your training data CSV file.
- In **line number 58** add the path to your raw csv file whose polarities you have to predict in the path_2 variable.
- In **line number 66** add the name of your raw csv file.

■ OUTPUT :

- Company classified as bad, average or good with respect to employees reviews printed on the console.
- Overall Score of the company out of 5 printed on the console.
- A CSV file named 'new_data.csv' consisting of individual review's polarity. You can change the name of output file if you want by going to **line number 111** and changing the name.
- This file will have two columns one consisting of the Reviews and the other the Polarity assigned using the lexicon based approach.
- You can find the file in the working directory set by you.

○ Noun_Extraction.R

- **INPUT :** A CSV file consisting of a Reviews stored in a column named '**Reviews**'.
- **RUNNING INSTRUCTIONS :**
 - In **line number 16** add the path to your input csv file..
 - In **line number 18** add the name of your input csv file.
- **OUTPUT :**
 - Output will be the list of important nouns from the reviews in a list named **imp_nouns**.

● **Some General Information :**

- The code for Random Forest algorithm may take some time to run so don't lose patience.
- Also the code for Noun Extraction will also take some time to learn.(It may take more than 5 minutes too.)
- All the codes are properly commented. So you can easily understand the use of each and every command.
- In case of any conceptual information please see the report submitted.
- You can find all the codes in your PC and also in the google drive we have shared with you.