Instructions on how to run our models

We have downloaded our dataset from Kaggle.com, here is the link to download the dataset. https://www.kaggle.com/yelp-dataset/yelp-dataset/data

Initial Preprocessing:

So, for our initial preprocessing we have merged required files from the whole raw dataset. We have considered some attributes and imputed some. Our final file after initial preprocessing is "Yelp.csv".

This are some of the libraries you will require to run Initial_Preprocessing:

1) data.table
2) mice

After installing all the libraries just load them and run the Initial_preprocessing.R file.

Predictive Modelling:

For all our predictive models we have used file "Yelp.csv" which we got after initial preprocessing.

This are some of the libraries you will require to run our predictive models:

3) caret
4) e1071
5) rpart
6) nnet
7) Liblinear
8) randomforest
9) lubridate
10) AppliedPredictiveModeling
11) tree
12) rpart.plot
13) pROC
14) ROCR
15) RColorBrewer
16) party
17) partykit

After installing all the libraries just load them and run the models one by one.

Sentimental Analysis using text mining and Naïve Bayes:

So, for this you will require "NY_IL_review.csv" file which is attached to this packet.

The libraries which are required for this is:

1) tm
2) e1071
3) dplyr
4) caret
5) naivebayes
6) kernlab
7) party
8) FSelector
9) rminer
10) SnowballC
11) RWeka

After installing all the libraries just load them and run the sentimental_analysis_text_mining_naïve_bayes.R file.

Sentimental Analysis using tidytext:

For this you will require "review.csv" file which can be found at the given link. https://www.kaggle.com/yelp-dataset/yelp-dataset/data

The libraries which are required for this is:

1) tm
2) DT
3) SnowballC
4) wordcloud
5) tidytext
6) RColorBrewer
7) dplyr
8) plyr
9) ggplot2
10) textcat

After installing all the libraries just load them and run the sentimental_analysis_tidytext.R file.

Visualization:

For this you will require "yelp_business_attributes.csv" file which can be found at the given link. https://www.kaggle.com/yelp-dataset/yelp-dataset/data

It uses basic plotting libraries which are already present in Rstudio, just run Visualization.R file.