Module 8: Text Mining

Assignment

Module 8: Assignment

Perform the following activities in R

Task 1: Text acquisition and wrangling

→ Load RestReviews.tsv data into R (use read.delim2 function).

Download Link

→ Convert it into a corpus or Vcorpous (also called volatile corpus, which is nothing but a corpus which resides in RAM, and is distroyed when R object containing it is destroyed.)

→ Perform the following text cleaning techniques

o   Remove punctuations

o   Remove numbers

o   Convert it into lower case

o   Remove stopwords

o   Strip white spaces

o   Also convert the words to stem values

→ Create a DocumentTermMatrix using it

Task 2: Build a classifier which can classify reviews into good or bad(liked or not)

→ Create a data frame using the DTM and the 'liked' column of our data

→ Use Naïve Bayes classification algorithm