Text Mining and Sentimental Analysis for Online Reviews

In the age of internet, reviews have occupied a more prominent place while making any kind of decision. People usually check out for public reviews and research experience from peer groups before making any kind of decisions. These decisions can range from as small as which movie to go for over a weekend to which college/university to choose for their degree. The reviews on the internet can display a wide range of emotions and it usually is a laborious task to go through them before we decide. We are planning to provide a solution to this problem using R while Extracting sentiment out of review using Natural Language Processing.

To this project, we are using the reviews data from Kaggle on Hilton Hawaiian Village Beach Resort located in Honolulu. Data has 13701 rows and 2 columns (review and review date). We will be using R Packages and functionalities to perform text mining and sentimental analysis on this data. Goal of this project is to perform text mining, analyze the words/ emotions contributing to the sentiment of the review and assign a sentiment to each review.

We are planning to use the following R packages to achieve different functionalities w.r.t the project.

a. dplyr for Data manipulation
b. ggplot2 for Data Visualization
c. tidyr for Data Cleansing
d. reshape2 for data transformation
e. tidytext for analysis and visualization of text
f. purrr for working with functions and vectors
g. scales: to override the ggplot2 defaults.
h. Lubridate: for working with dates and times.
i. widyr: for co-occurrence and co-relations among words
j. snowballC for word stemming
k. wordcloud for analyze and visualize keywords

We plan on utilizing R's data manipulation, transformation and cleansing capabilities combined with text analysis techniques like stemming, lemmatization, stop words, bigrams, and trigrams to achieve at a sentiment for a review. In the process we want to make sure that we preserve the emotion of the review by only removing the stop words from the review. Furthermore, we are planning to visualize the importance and association between different words using graphs, charts, plots, and frequency counts.

Source: https://www.kaggle.com/datasets/shrinidhisaravanan/hilton-hawaiian-village

Data: https://www.kaggle.com/code/kerneler/starter-hilton-hawaiian-village-048bdbd4-9/data