**Project REPORT**

***0Customer Hotel tweets0Text0Analysis***

**Task-B:-**

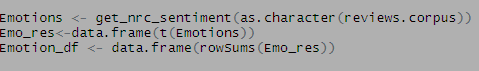
Project: Sentiment analysis and text analysis, additionally alludes as assessment mining, is a sub AI task where we need to figure out which is the general opinion of a given report. Utilizing machine learning procedures and normal language handling we can remove the abstract data of a record and attempt to arrange it as per its extremity, for example, positive, nonpartisan or negative.

It is an extremely helpful investigation since we could decide the general conclusion about the hotel tweets sentiments. In this venture we have to perform machine learning and understand the emotions into "positive" or "negative" notion by structure a model dependent on probabilities.

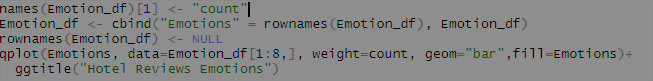
**Ans-1:-**

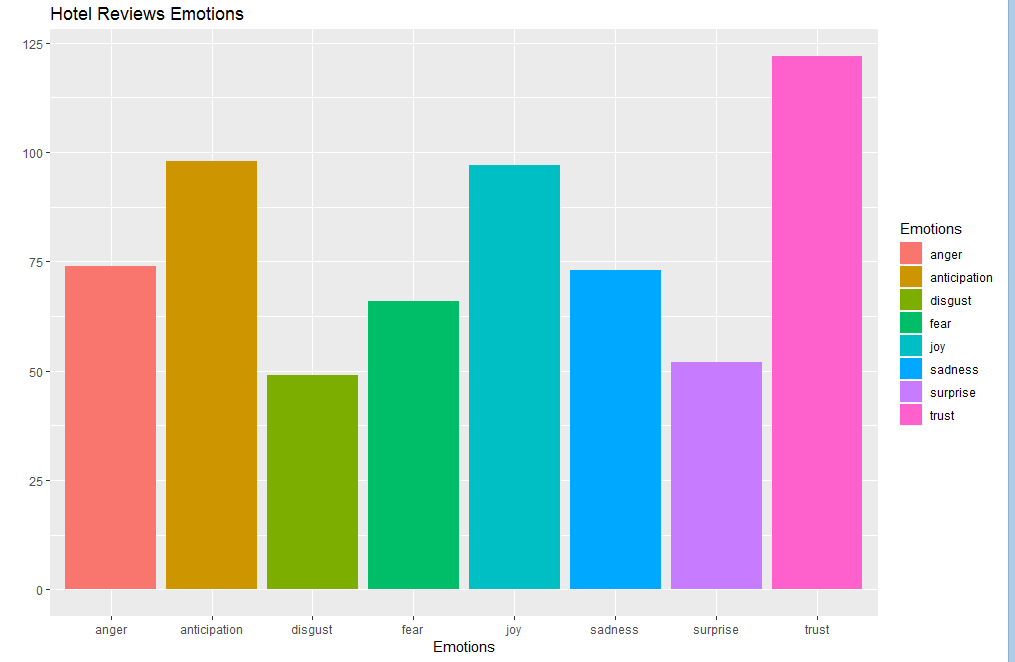
**(1):-**

1. We have Imported the Hotel tweets.csv file by using read\_csv() function on R studio. Then we have installed and loaded the required libraries for this project.
2. Then we have extracted the positive & negative tweet from the given data set and it stored them into two diverse data frame and joint the both data frames (joint the positive & negative tweet into single data-frame for the text analysis).
3. Then we have cleaned the gained data like changed to the lower cases, removed the stop words, strip the white space, removed punctuation, removed unwanted numbers, removed the special characters from the data frame.
4. Now we have applying the get nrc sentiment() function from Syuzhet package to find out all emotions like disgust, fear, joy, anger, anticipation, sadness, surprise and trust from the hotels tweet.



* Then we have plotted the all these emotions gained above in bar chart with the help of ggplot2 library that available in R language.



****

**Fig: Plot for all eight emotions for hotel tweets**

**(2):- Applied Machine Learning on Hotel tweets data:**

* Now gain we have again imported the Hotel tweets.csv data by read\_csv() in R studio. Its not required to again import but we have started the machine learning task in new phase, so we have imported again. Then we installed & loaded the packages required. In this task we have required to use the e1071 library to perform machine learning but we are using RTextTools package as this library also use the equivalent e1071 package which have included the e1071 library (we can see the documentation RTextTools package by R cran website).
* Now again we have extracted the positive & negative tweet from the given data set and store it into two diverse data frame and these both data frame contain the first 200 tweet as training data set & 63 tweet as testing data set.
* Now the combined positive & negative tweet of training data and positive & negative tweet testing tweet to get the train and test data correspondingly. We preprocessed with cleaning the both test & train data like changed to the lower cases, removed the stop word, strip the white spaces, removed the all punctuation, removed unwanted number, removed the all special character.
* After cleaning and prepossessing the train & test tweet data, transformed the hotel tweet data into Document term matrix and then applied the SVM-algorithm (Machine learning algorithm) on the training tweet data.

**In Matrix Changed:**

Tweets.mat=as.matrix(Tweets.spm)

cont=create\_container(Tweets.mat, as.numeric(as.factor(Tweets[,2])),

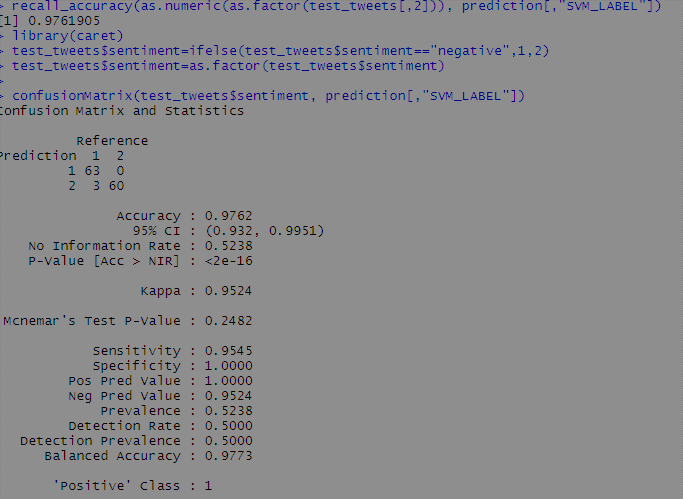
trainSize=1:400, testSize=401:526,virgin=FALSE)

**SVM Machine Learning Code:**

**mod** = train\_models(cont, algorithms=c("SVM"))

Now we have perform the model evaluation to check the model accuracy tested this the SVM model on the test tweet data and figure out the recall accuracy & confusion matrix.

**Code & Output for Model Evaluation:**

****

**Conclusion:**

In this task we have attempted to demonstrate the essential method for mining the hotel tweets positive and negative utilizing KNN algorithm as standard. We could further improve our classifier by attempting to remove more highlights from the hotel tweets, attempting various types of machine learning algorithms, tuning the parameters of the KNN algorithm, or attempting another classifier all together.

**Reference:**

1. Alexander Pak, Patrick Paroubek. 2010, Twitter as a Corpus for Sentiment Analysis and Opinion Mining.
2. Alec Go, Richa Bhayani, Lei Huang. Twitter Sentiment Classification using Distant Supervision.
3. Jin Bai, Jian­Yun Nie. Using Language Models for Text Classification.
4. Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow, Rebecca Passonneau.Sentiment Analysis of Twitter Data.
5. Fuchun Peng. 2003, Augmenting Naive Bayes Classifiers with Statistical Language Models