**Hotel review sentiment analysis**

**CE350: Data Warehousing and Data Mining**

Nehal mahida (17CE052) , Chandresh Mendapara (17CE057) , Karan Mundhava (17CE060) , Jaydip Pansuriya (17CE064)

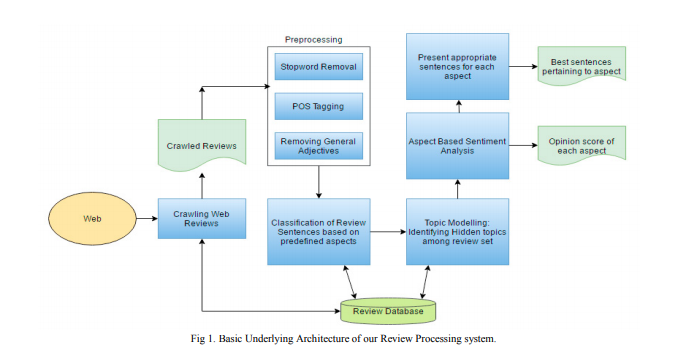
<https://github.com/chandreshmendapara/Sentiment-Analysis-Of-Restaurants-Review>

**ABSTRACT**

This project addresses the problem of sentiment analysis of particular Hotel. In our project we will classify each hotel in three category according to customer review. There are three category negative , positive and nutrule. Now a days there is lots of websites which give the review about hotel loke google map , zomato etc. so we combine all those review and generate sentiment of that data this is our goal. When user looking for any hotel at that time it is difficult for them to find out best hotel for him so it is problem and we are going to solve this problem in this project. We take data in both from written and rate out of 5.Analysing this hotel reviews and give sentiment on it is very important for those who like choose hotel on basis of reviews of other.

**Introduction**

From last few years , people are using lots of application like google map , zomato and other and they give there reviews on it so it will helpful to other as well. Most of the time if people like some hotel they give there honest review and give rating according to it so if we can use that reviews and rating for analysis and give sentiment on it so it will helpful to other also. But here is one proble that is all people are not use same platform for giving there reviews and that why we gate reviews from different platforms. In order to use those all reviews we need to combine all them together and analyse them. First of all in order to combine all reviews together we merge all result into one and then we input it into our model. Our model has trained on naïve based algorithm. After processing those input output will be the sentiment of particular hotel. So any user want to see the reviews of particular hotel then he just need to search that hotel name in our project after that he/she will get sentiment of that particular hotel. In this process we take care about each and every review of user. Because lots of the time there is redundancy in data so cleaning process is very important here. We need to take care about cleaning and removing unnecessary data from data set. After all this process we train model and use it for checking review of particular hotel.

****

**What is sentiment analysis?**

Sentiment Analysis is the process of ‘computationally’ determining whether a piece of writing is positive, negative or neutral. It’s also known as **opinion mining**, deriving the opinion or attitude of a speaker.



**Why sentiment analysis?**

**Business:**In marketing field companies use it to develop their strategies, to understand customers’ feelings towards products or brand, how people respond to particular hotel. So using this companies decide that it is worth to buy the share of that hotel or not.



**Politics:**In political field, it is used to keep track of political view, to detect consistency and inconsistency between statements and actions at the government level. It can be used to predict election results as well!

**Public Actions:**Sentiment analysis also is used to monitor and analyse social phenomena, for the spotting of potentially dangerous situations and determining the general mood of the blogosphere. Public also find that particular hotel or other is good for them or not.

**Prerequisites**

**Basic programming knowledge**

Although Python is highly involved in this mini-project, it is not required to have a deep knowledge in the language, as long as you have basic programming knowledge.

**Installed tools**

For this program, we will need Python to be installed on the computer. We will be using the libraries twitter, nltk, re, csv, time, and json. You are likely to have to install the first two libraries. The rest already come with the Python interpreter. It doesn’t hurt to check that they’re up-to-date though.

**Data set splitting concept**

This is critical to fully understand the process pipeline. You only need to know [the difference between Training and Test data sets](https://medium.com/datadriveninvestor/what-are-training-validation-and-test-data-sets-in-machine-learning-d1dd1ab09bae), and in what context each one is used.

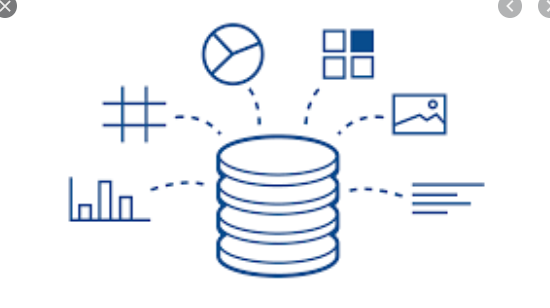
**Basic RESTful API knowledge**

This is not crucial, but it could help. We are using Zomato Api then Yelp Api for getting reviews and rating of particular hotel.

**How to perform Hotel reviews sentiment analysis:**

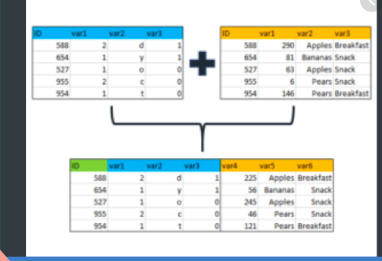
**Data gathering :**

We use different Api for this purpouse and then merge all data into one unit. For example Yelp and zomato Api give such data and we combine all into one unit.



**Merging Data Set into one Unit :**

After receiving data from different sources we need to merge into one set. So in this step we perform merge operation and get all in one unit.



**Removing Special Characters and Numbers:**

In this step we remove all special characters , numbers and symbols which is unwanted. If we can’t do this step then we will not get accurate result and our model was failed. So this is very important step for accuracy purpose.

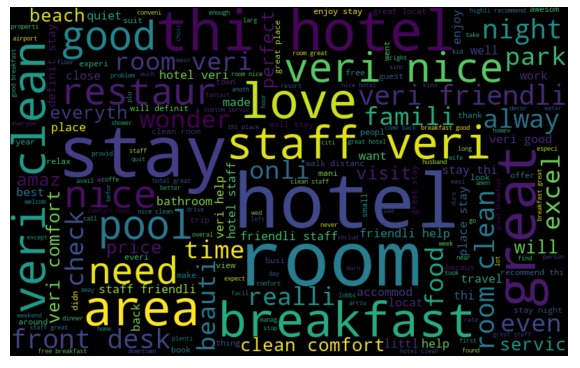
**Tokenization**:

In this step We split whole sentence into small token and store in some collection. After storing all we remove duplicate one and take only necessary words and use it for analysis.

**Join Token :**

We further join Token together and use it for analysis.

**Word Cloud of Positive Reviews**:



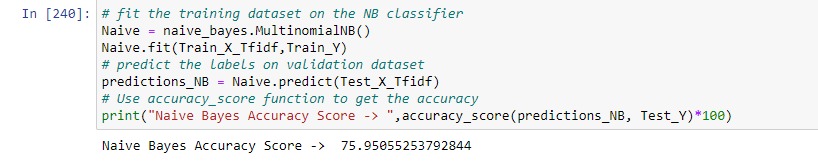
**Word Cloud of Nagative Reviews**:



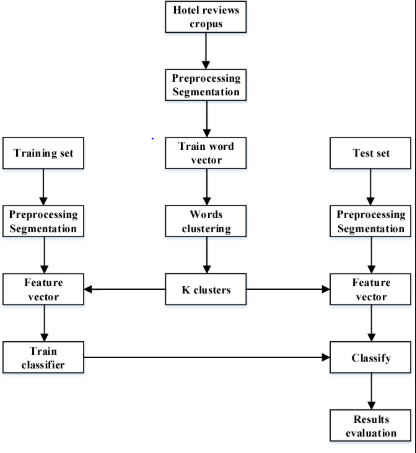
**Naïve Bayes Algorithm for Classification** :

We use Naïve Bayes algorithm train our model.

**Result :**



**Working Process:**



In this project, first we have collected the data and stored them in a file then read the data from the file and word tokenized them using python’s Natural Language Toolkit Library(nltk). There is a list available of positive and negative reviews on internet that will have all the negative or positive words, stored the words in the dataframe and then compared words of my reviews with the respective words in positive and negative dataframes and generated net positive score and negative score for each( I have collected more then 500 reviews) about hotel and then analyzing them and displaying the final result in a pie chart to determine people opinion about particular hotel.

**References:**

<https://www.kaggle.com/jonathanoheix/sentiment-analysis-with-hotel-reviews>

<http://cs229.stanford.edu/proj2014/Vikram%20Elango,%20Govindrajan%20Narayanan,%20Sentiment%20Analysis%20for%20Hotel%20Reviews.pdf>

<https://www.researchgate.net/publication/275955836_Sentiment_Analysis_for_Hotel_Reviews>