iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** V **Month of publication:** May 2020

DOI: http://doi.org/10.22214/ijraset.2020.5006

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

Sentimental Analysis of Hotel Reviews

Saurabh Pratap Rao[1], Vaishali Singh[2], Shatabdi Singh[3], Vanshika Garg[4], Mr. Nand Kishore Yadav[5]
[1, 2, 3, 4]*Fourth Year ,* [5]*Assistant Professor, Department of Computer Science, RKGIT , Ghaziabad , U.P,*

Abstract: ***The comments and reviews on hotels are one of the most important sources of information in hotel reservation planning. Sentimental analysis is a tool for the classification of reviews provided by reviewers. The necessary information which is important about a hotel can be fetched from hotel reviews. In this project, Data mining technique is used to analyse the positive and negative words from the reviews of all the Hotels on the TripAdvisor website. Mining hotel reviews is desirable to gain knowledge about customer expectations and support effective management of customer relationships.***
Keywords: ***Machine Learning, Data Mining, Web Crawler, Natural Language Processing, Sentiment Analysis***

I. INTRODUCTION

A good hotel image depends upon the wellness of the hotel services. Hotel booking on website has become one of the most important economic use. Sharing on net has become a major tool for conveying ideas about a particular product or services. Now a days, it has been noticed that immediate growth in online group discussions and review of sites where a decisive characteristic of a customer's review is their sentimental or opinion, for example if the review contains words like 'Good', 'Verygood','nice 'then it is likely to be a positive comment. Whereas if the review contains words like 'Bad', 'worse', 'very bad', 'poor' it is likely to be a negative comment.

An unusal way of performing Sentimental Analysis is to detect and classify extreme opinions, which represents the most negative and the most positive opinions about a topic or an individual.

However, Trip advisor's star rating does not convey the definite action of the customer. Most of the ratings are vague , large chuck of reviews fall in the range of 3.5 to 4.5 and very few reviews get below or above. We seek to turn words and reviews into significant measurements. We have developed this model with a supervised sentimental components that are capable of classifying a review as positive or negative with accuracy.

We also determine the polarity of the review that evaluates the review as recommended or not recommended using semantic orientation.

A phrase has a positive semantic orientation when it has good associations (e.g. "Excellent, Awesome") and a negative semantic orientation when it has bad associations (e.g. "Bad"). Next step is to assign the given review to a class, positive or negative, based on the average semantic orientation of the phrases extracted from the review. If the average is positive, the prediction is that the review posted is positive. Otherwise, the prediction is that the item is negative.

Sentimental analysis deals with analysing sentiments, thoughts and the outlook of a user or an author from a given piece of textual data in the form of hotel reviews for Sentimental analysis with opinion mining from customer's perspective. Sentimental analysis use the technique of natural language processing for interpretation of free text and making it analysable and also it help to identify opinions and sentiments online to help them understand what customers think about their products and services.

It helps hotel managers to have a good understanding of customer's needs, to discover areas for further improvement and improve service quality.

II. RELATED WORK

Our model is inspired by the prior work on semantic orientation and unsupervised classification of reviews. In semantic orientation classification of reviews is done in positive and negative. [2]

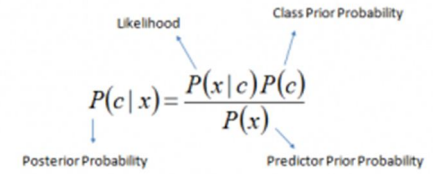
A review is classified on the basis of average semantic orientation of the phrases that are extracted from the reviews by adjectives and adverbs. The prediction is said to be positive or negative depending upon the classification.

A. Different Models for the Project

For modelling data from TripAdvisor after algorithms, we will implement two models:

Naïve-Bayes classifier is a model that analyses the Bayesian probability of words and a support vector machine which is known in the realm of textual analysis.

1) *Naive-Bayes classifier:* In this text classification it is assumed that the attributes are independent of each other in the class.



$$P(c|X) = P(x_1|c)\times P(x_2|c)\times\cdots\times P(x_n|c)\times P(c)$$

Two variants of Naïve-Bayes:

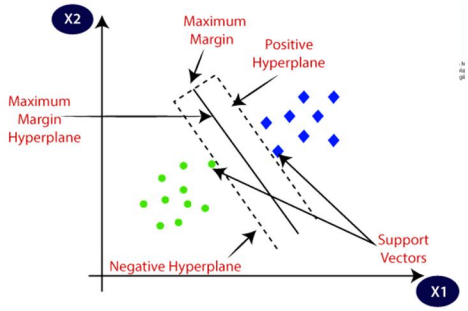
a) Naïve-Bayes multinomial distribution with Laplace smoothing

b) Bernoulli distribution are discussed.

2) *SVM:* SVM is the support vector machine which uses linear kernel. The best models used for prediction of textual information is provided by SVM. Document vectors are sparse and it is very well explained with SVM.

III. SEMANTIC ORIENTATION

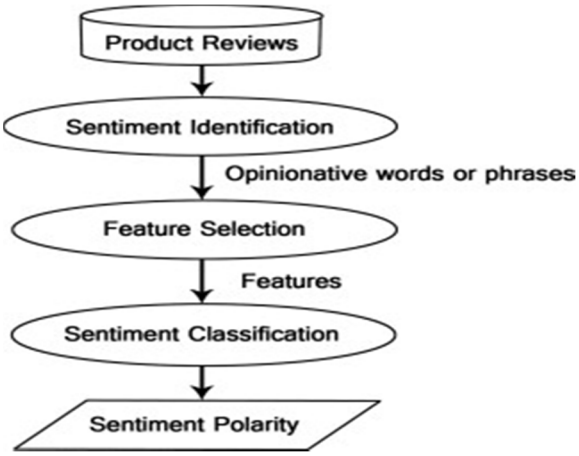
Classification of positive and negative terms in sets. The part-of-speech tagger is predicated in the review. Mutual dependence between two instances of random variables is measured by Pointwise Mutual Information (PMI). Here positive results show high correlation, zero result shows no dependence and negative result shows that the correlation is opposite.



$$L(w) = \sum_{i=1} \underbrace{max(0, 1 - y_i[w^T x_i + b])}_{\text{Loss function}} + \underbrace{\lambda ||w||_2^2}_{\text{regularization}}$$

IV. KEYWORD BASED METHODOLOGY

Analysis of sentences on the basis of keywords is important. The differentiation of the words on the basis of their polarity is recorded and classified in their respective types.



ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue V May 2020- Available at www.ijraset.com

V. MACHINE LEARNING

Machine learning is an AI which makes the system able to improve & learn from past records without itself being programmed explicitly. Processes are similar as data mining as well as predictive modelling.
Extraction of raw data and turning it into useful knowledge is a part of a machine learning.

VI. METHODOLOGY

In this project, the website www.tripadvisor.in is used, it uses Jsoup. It contains a total of six attributes & 109 instances. Attributes comprises of name of the hotel, its reviews, ratings, address, hotel link and the names of all the hotels.
Clustering and classification is the main method used behind this project. It analyses and predictions are made about the hotel type.
Algorithm used for clustering data is Partition around Medoids[PAM].

J48 algorithm for classification is used.
It is happened as:

A. *Extraction*

Connection with trip advisor website is created using Jsoup. Screen scraping takes place. The extracted data is in CSV file format.

B. *Clustering*

Splitting of the data elements into groups of similar kind. CLRANS, PAM or k-means are some partitioning clustering methods.

C. *Classification*

Sorting data is done in this using J48 algorithm which is based on decision tree.

D. *Natural language processing*

It is used for interaction between computers and human languages. Analysis of text and predictive analysis is handled.
NLP identifies positive and negative and is useful as translators for language translation.

VII. PERFORMANCE METRICES

The technologies used are-

A. *Python*

It is an interpreted high-level language. A dynamic type system is featured by python. It is –

1) Object oriented
2) Imperative
3) Has large standard libraries

Two libraries are used in this analysis-

a) Panda
b) NLTK (Natural language Toolkit)

The function applied in this is sentiment intensity analyser () from the nltj. sentiment vader package.
Using apply () and *lambda,* results are transformed and stored as "reviews".
Sentiment score is calculated for every review. There are negative, positive, neutral and compound scores for every review. If compound score is less than 0.2, review turns out to be negative.
If it ranges between -0.2 and 0.2, then the review is said to be neutral.

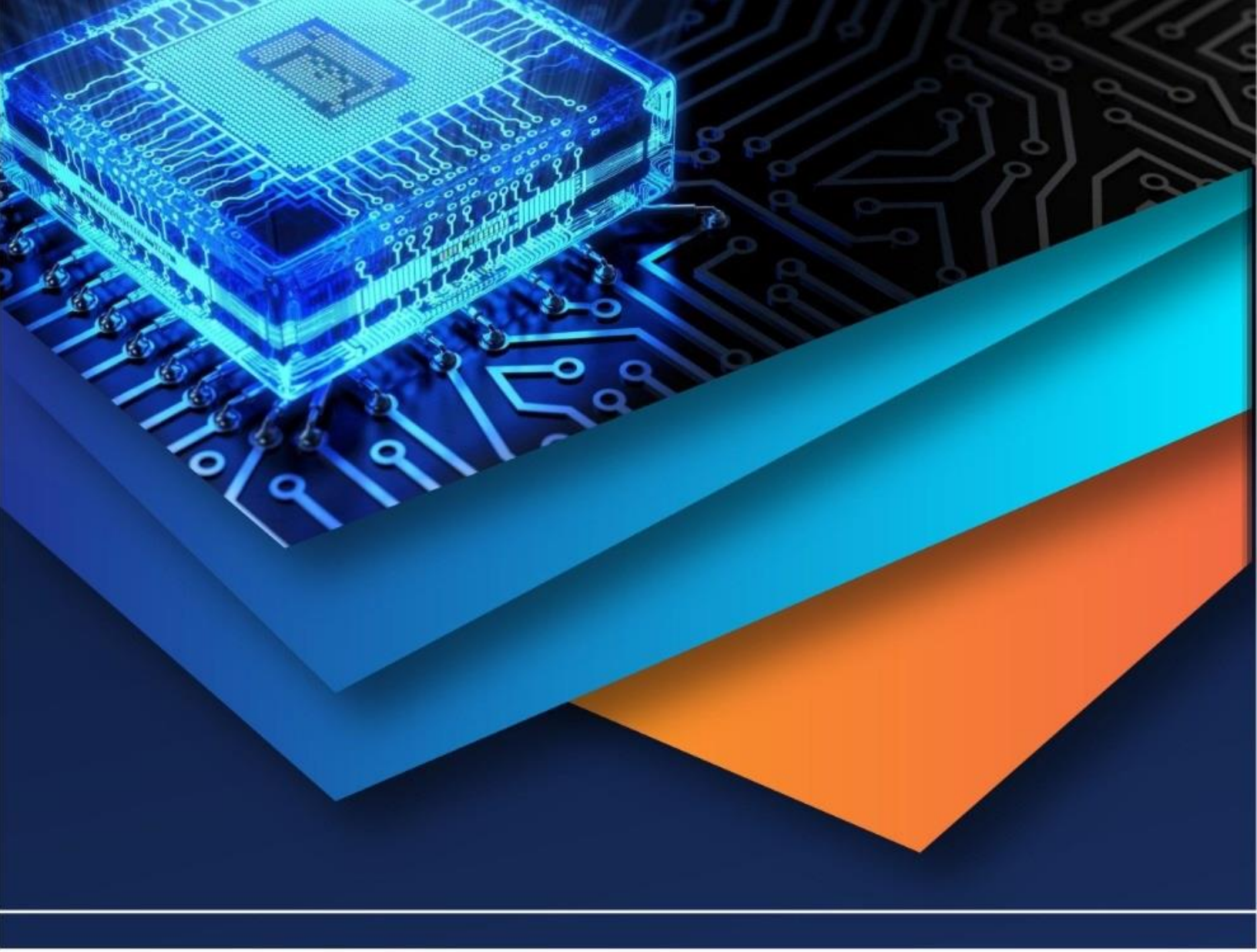
B. Mongo DB/Firebase(Database)
It stores data in JSON type of documents. Hence the documents are flexible, changeable, variable.

VIII. DISCUSSION

In this paper, we have discussed the better performance of Naive-Bayes model over SVM in our dataset [4]. For future experiments, inclusion of sentimental lexicons and non-word tokens are used. It captures semantic similarities in reviews.

REFERENCES

[1] Opinion mining and sentiment analysis by Bo pang and Lillian lee
[2] Thumbs Up Thumbs Down by D.Turney
[3] Learning Word Vectors for Sentiment Analysis by Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts
[4] Trip Advisor Dataset : http://times.cs.uiuc.edu/~wang296/Data



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY