International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 01 | Jan 2021 www.irjet.net p-ISSN: 2395-0072

SENTIMENT ANALYSIS OF HOTEL REVIEWS

PRASHANT BHATI[1], PARTH PARASHAR[2], PRIYANSH SINGH[3], RUPAM KUMAR[4],Dr. SUMA SWAMY[5]

[1]Dept. of Computer Science and Engineering, Sir MVIT College, Bangalore

---

ABSTRACT - Travel planning relies on user reviews and comments of the hotels on the web. Thus, knowing this data is incredibly important for hotel management. We tend to introduce a platform that collect comments from the user and creates classified and structured overviews of such comments and facilitates access to its data.

KEY WORDS: Opinion Mining, Hotel Reviews, Pre-processing, Feature Extraction.

1. INTRODUCTION

Now a days, a business concern or a service based firm needs feedback from its customers. Increasing of business are requiring for additional variety of services and products. So, the organization should hassle of concerning the reviews, and ratings given by its user to extend the business, as they are required for additional variety of services and products. The service-consumers will mention their feelings and reviews on online-portals. By performing the opinion mining and sentiment analysis on these details we will predict the rating of that organization. One recommender system is required for generating the ratings in precise and correct manner. For a hotel business, reviews regarding numerous aspects like Cleanliness, Maintenance, Behavior, Food, Hospitality, Room neatness, Response from the staff of Hotel, etc. plays a significant role for recommender system. The Customer's feeling with respect to a hotel depends upon the facilities they got from that hotel such as price, location, cleanliness, and facilities of the hotel, services provided by the hotel like laundry, complimentary breakfast, free wi-fi, bar/lounge, babysitting rooms etc. The sentiments can be expressed in the form of excellent, good, average, poor, terrible etc. Basically, the customers want to convey their sentiment with these rating and review.

LITERATURE SURVEY

Lumi lee from china (2016) enquired and analyzed 72 research paper related to the tourism and hospitality that was published in research journals between January 2008 and December 2015. He analyzed and reviewed the topic-related to characteristic of tourism and hospitality online reviews in different marketing segments and used heuristic systematic model (HSM) to divide and sum up the features that affect consumer's belief in previous HTOR studies. They believed that their suggested ideas will help in the identification of research topic in extended HTORs literature and pointing out possible direction for future studies.

Piang nd bm. (2017) offers a foundation for understanding the operational challenges and recognize several research path for social media analytics in hospitality and tourism area. They comparatively examined information quality related to online reviews on entire hotel population in Manhattan, New York using three major online review platforms which are TripAdvisor, Expedia and yelp through text analytics. The authors state that there exists an enormous inconsistency in the representation of the hotel industry on these platforms. In addition, online reviews differ greatly in terms of their sentiment, semantic features, linguistic characteristic, rating and relationship between the features. A study made an observed analysis to discover the relationship between customer feeling and online customer rating for hotel (Geetha *et al.*, 2017). The researchers considered using the lexicon of positive and negative words used by studies in (Hu and Liu, 2004).

As for the classification algorithm, the researchers utilize **D.kumar** classification technique to use the lexicon of word in order to match words from the document against the lexicon. Then, the classifier assigned the probability of the words being positive or negative. The result of the study finds that there is consistency between customer rating and actual customer's feelings towards the hotel.

k. Pin Lon and Bhang (2016) enforced Russian sentiment mining system towards tourism dataset as the evaluation sample. The proposed model was then compare with support vector machine and logistic regression models to compared their performance. As for the result, the projected sentiment model shell the 2 different models with accuracy of 81.3% and F-measure at 81.8%.

Chang, Ku, and Chen (2017) state that the real value of social media information is hardly acknowledge because of overloaded information and existing literature in analyzing hotel reviews rarely provides deciding information and promoting insight to enhance business services. The authors projected Associates in Nursing integrated framework consisting of many methods to realize process insight into hotel reviews and rating. The result unconcealed that the projected approach shell baseline algorithm rule with high exactness and recall price that are unit 0.95 and 0.96.

Divyashree, L, and Majumdar (2017) used data mining and sentiment analysis techniques to analyses the polarity of the words from the reviews of all hotel on the TripAdvisor website. The studies use the data from Bengaluru Bangalore District Karnataka Hotels dataset containing 6 attributes and

109 instances. The researcher used Partition around medoids (PAM) Algorithm to cluster the extracted data and utilizing J48 algorithm for the data classification which is then applied to the clustered dataset. As for the result, 86% of the words in the data was classified as positive words while 14% was classified as negative words.

Sentiment analysis was employed in order to decompose the reviews into totally different dimensions to live the service quality and performance of the hotel based on SERVPERF model. As the result, different dimensions of user reviews have considerably different effect in forming user evaluation and driving content generation.

Gao, Hao, and Fu (2015) have made an effort to discover and explore a set of three web services which are known as Alchemy, Text2Data and Semantron. The authors then compare the sentiment analysis capabilities of these three services using reviews gain from TripAdvisor. They discover that these three web services have high accuracies since most of the reviews are classifies as positive.

Comparsion Table-

AUTHOR NAME	TECHNIQUES	DESCRIPTION
Lumi lee	Random Forest	It has an accuracy of 79.3%
Piang nd bm.	Classification Technique	To examine information quality related to online reviews.
D.kumar	Hotel Reviews using bert model.	Transformer, Pre-trained BERT
k. Pin Lon and Bhang	Optimized hotel review analysis using deep learning.	It has an accuracy of 81.3% and F-measure at 81.8%.
Chang, Ku, and Chen	Hotel reviews Classification using Text Mining(2017)	provides methods to realize process insight into hotel reviews and rating
Divya shree and majumdhar	Machine learning approaches	Analyses the polarity of the word
Gao,Hao and Fu	Alchemy and semantron	Discover web has higher accuracy.

METHODOLOGY

The step-by-step procedure for analyzing public opinion in Twitter is shown below-

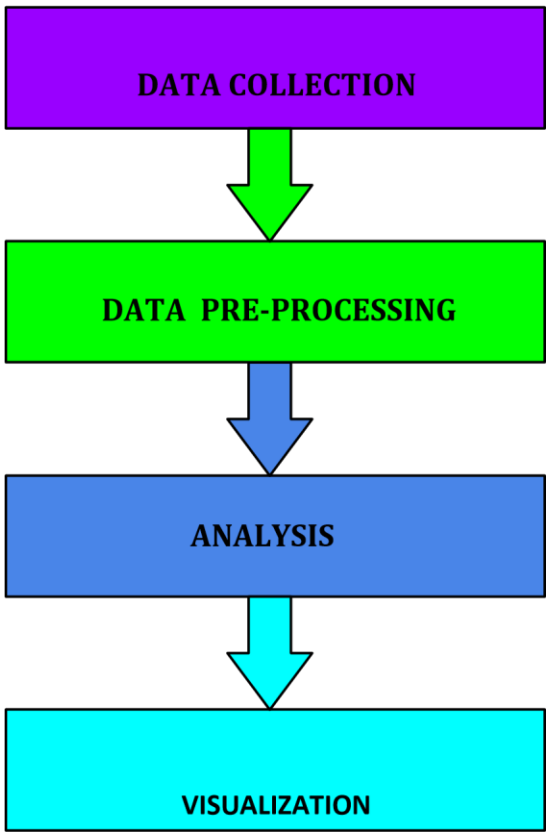

Fig 1. System overview

Data Collection

The process of fetching and measuring information on variables of interest. We are using a dataset which consists of half a million hotel reviews scrapped from Booking.com stored in the form of CSV data.

Data pre-processing

The data collected is messy and full of unnecessary objects which is irrelevant to machine learning classifiers. So, it becomes necessary to first clean the data and remove all redundancies and make it appropriate to feed as input for classifiers. Accuracy of feature extraction also greatly depends on the quality of text data.

Analysis

The main idea behind sentimental analysis is to categorize a review as positive or negative. The most important indicators of sentimental analysis are opinion words which imply whether the review is positive, negative or neutral.

Visualization

The result from analysis is visualized in the form pie charts or bar charts. A structured analysis helps the users to understand complex data more efficiently.

CONCLUSION

In this paper, we came across various steps used to perform opinion mining. After evaluating a number of machine learning approaches, we identified those most suitable to classify public opinions. We observed that it is possible to analyze a large set of reviews using various machine learning approaches in a reliable way wherein we employ a wide word ambiguity, and multipolarity thereby ensuring enhanced accuracy in our future model.

REFERENCES

1. Punita Hani and Mohammed Yasir Arafat: Ranking Hotels using Social Media and sentiment Analysis.

2. Dinkar Prajapati, Markopoulos, Akanksha Chamoun, M. kushbaha "sentiment Analysis of Hotel Reviws in Latin: International journal of advance computer science, pg:-373-383.

3. A.M. El-Halees,"Arabic Text Classification Using Maximum Entropy,"Islam.Univ. J. pg:-157-167,2007.

4. C.J. Hutto and E. Gillbert, "VANDER:A Parsimonius Rule-based Model for sentiment Analysis of Social Media Text," Proc. English int AAAI Conf Weblogs Soc. Media, pp.216-225,2014

range of methodologies to collect, train and classify reviews. After analysing all the techniques, we wish to deploy the best available algorithms on our dataset and compare the accuracy rates produced and select the most promising one thereafter. In doing so, we would also like to mitigate major challenges faced by sentiment analysis such as sarcasm, negations,

5. Jbshoek. Choi & bales Lee, "Hotel reviews using sentiment Analysis using Deep Learning ",pg:-994-1012.

6. Sentiment classification of consumer generated online reviews using topic modelling.

7. H.J. Do and H.J. choi,"Korean twitter emotion classification using automatically built emotion lexicons and fine-grained features,"29th Pacific Asia Conf.Lang.Inf.Comput. Paclic 2015.

8. Sumbal Riaz, Mehvish Fatima, M. Kamran, M.wasif Nisar, "opinion mining on large scale data using sentiment analysis and k-means clustering".

9. Siti Nuradilah Azman et al, "Towards an Enhanced Aspect-based Contradiction Detection Approach for Online Review Content".

10. N. Bekmamedova and G. Shanks, "Social media analytics and business value: A theoretical framework and case study," Proc. Annu. Hawaii Int. Conf. Syst.