Text Summarization and Sentiment Analysis of Online Hotel Reviews

Prepared by Group 15:

Prashik Waghmare(11432504),

Rushitha Chalamasetty(11593804),

Shanmukha Sai Sirivadan Guntuboina(11521076),

Vinaykumar Rajeshwar Gadewar(11598536)

Github Project Link:

https://github.com/ShanmukhaGuntuboina/NLP-Project-15.git

I. Motivation

Online reviews influence the traveling experience of consumers, reflecting a potent sort of word-of-mouth referred as electronic word-of-mouth. 95% of consumers, according to TrustYou.com, read online hotel reviews before making hotel reservations. The impact of internet reviews on customers and the hotel sector overall has also been supported by studies. But the main problem is as follows: Yelp is one of the most used sites for rating and categorizing reviews. But these platforms continue to demand that users read a ton of reviews to create own judgments about the amenities worth seeing.

Because of this excessive information, early abandonment of research into online reviews is possible. Furthermore, the quality of reviews varies widely. Certain reviews include unreliable or biased information, while other reviews offer unbiased assessments and useful information. As a result, customers must go through a lot of evaluations and allocate enough mental effort to distinguish between unbiased, high-quality reviews and unreliable or biased ones.

Such a thorough investigation can be time and energy consuming for a consumer, so it would be advantageous for them to have a quick method of processing the large number of internet reviews. To overcome this information overload and lack of efficient resources for review analysis, a summary must be produced from reviews online. Although there are many ways to summarize hotel reviews, there are two significant drawbacks that should be taken into account. Most contemporary travel websites permit readers may vote on how useful internet reviews are. A good review is one that wins sufficient votes, however this quantity alone might be biased.

Votes, for instance, can be strongly impacted by factors like when a review was published. Reviews that have been newly published are unlikely to get more votes overall than earlier reviews since the collection of votes also depends on time about usefulness. Newer reviews, nevertheless, offer latest hotel information, evaluations, perhaps providing more pertinent and helpful information to consumers. Because the variables influencing the usefulness of online reviews been thoroughly examined, making a prediction model is feasible.

The second issue is creating a summary that fully covers all important hotel-related topics. For instance, some beneficial reviews address topics like excellent sleep and helpful room service, but not other factors, such as vicinity, traffic, and other hotel amenities. There have been proposed summarization methods that might be helpful in creating a summary from each review, but they are unable to combine specific details into all-encompassing comprehensive covering. Therefore, a successful text summarizing technique has to carefully choose positive hotel reviews to build a thorough and complete understanding of hotel experiences.

We plan to work on existing summarizers and solve the aforementioned problems that might be a hindrance to the consumers.

II. Objectives

By the end of the project we will achieve the best way of summarization and sentiment analysis of online reviews possible by eliminating the drawbacks in the already existing methods of summarizations.

As described earlier understanding the numerous number of hotel reviews is not a simple job and is notachievable until and unless we have a software that understands the reviews and give us the percentage of positive reviews and the negative reviews which help usin improving the customer satisfaction and make them happy which eventually increases the customers for thehotel.

We may also analyze hotels with the help of reviews, i.e., how is the hotel in terms of location and other features, which can assist consumers in making their own decisionof which hotel to choose as their criteria. This also applies to other services like movies, universities, and online goods, and it aids service providers in improving their resources through improvedfeedback.

Most importantly opinion mining and summarization process can be used to extract the useful information from the reviews and help to increase the accuracy in classifying the reviews which helps the web users to review the huge number of comments in a short period of time. Users can save time and acquire all the necessaryinformation from reviews about hotels. This can also be

used to generate more customizedreviews about a particular subject, such as a room service, or meal, as well as a list of hotels organized by words. The model tries to discover the characteristics or aspects u sing the gradient descent method, predict nearby words. Summarization is automatic.

Sentiment analysis is getting more and more popular as a way to evaluate the massive volume of customer-generated data to gain insightful knowledge.

The idea behind these studies is that if the usefulness of reviews can be assessed after they have been posted, then travel websites can assist online users in assessing the review quality and thereby shorten their time spent looking for travel-related information. Prediction models that use content features perform better than those that use semantic or review quality features.

III. Significance

Since the invention of the Internet, the number of people searching the web for a decent spot to stay while on vacation has increased. Users frequently move towards helpful techniques to fulfill their desires. These user reviews have greatly aided the development of today's modern e-commerce and digital world. However, customers are still forced to make an informed choice to choose wisely from a wide range of reviews that include frequent opinions about the market. Businesses must create cutting-edge marketing tactics and techniques centered on the demands and serenity of their customers in this modern generation of e-tourism if they want to stay ahead of the escalating competition.

The volume of records/knowledge available online has grown so much that it is becoming more and more difficult for users to read and comprehend all of the reviews. It is hard for customers to pursue and figure out the items from an enormous number of reviews.

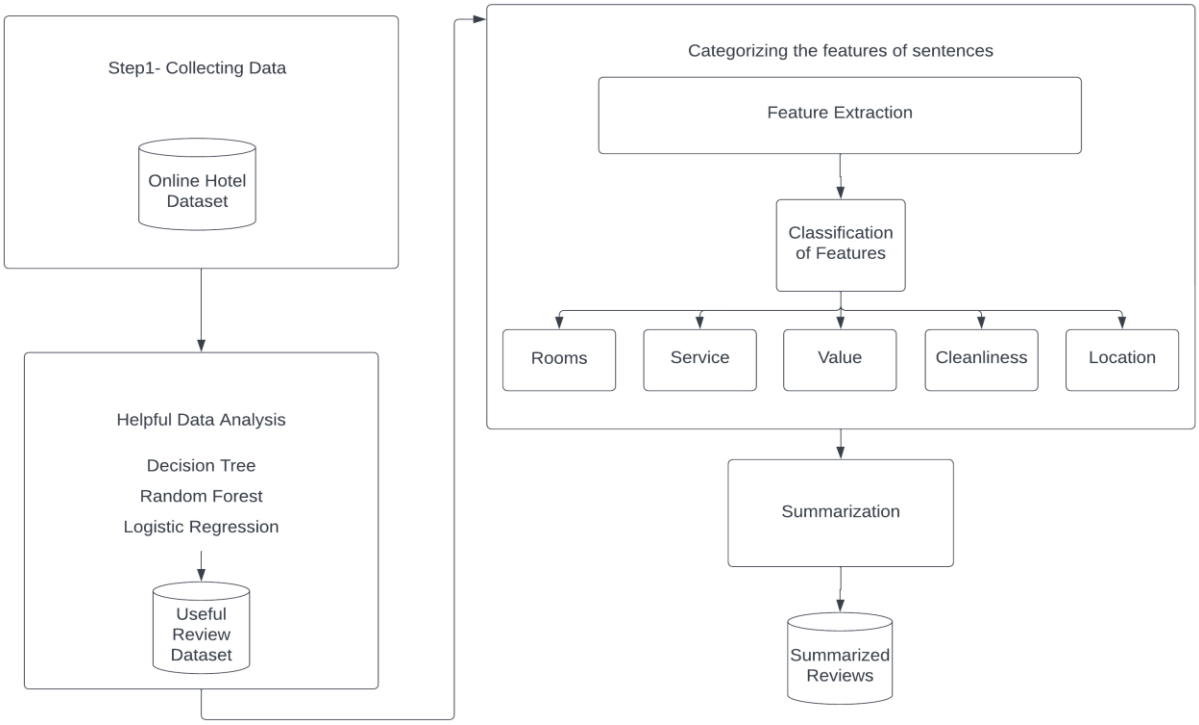
I had an experience of staying at a hotel back in India based on online reviews through various websites. I could not read all the reviews on all the different websites, so I ended up staying in a hotel that was an average place according to the price I paid to stay for a night. The hotel industry's foundation is built on the success of its guest's experience and service satisfaction. A consumer's service experience and satisfaction can be expressed simply as the correlation between his expectations and post buy assessment. Because no two users have the exact needs and requirements, sculpting the visitor experience is challenging and complex.

The categorized and shortened hotel review data enables the consumers to quickly and readily comprehend the review's contents. Hence it is essential to find a solution

for this problem of online reviews assessment so that the customers can get a pleasuring experience of their expectations.

IV. Features

This approach includes four stages. Data gathering and preprocessing along with stemming. The next step is the selection of beneficial reviews from the dataset using trained classifier. Categorization based on hotel feature extraction and classification is done in the third step. Summaries are produced at the last phase.



More features might be added in the future based on the time remaining to finish the project.

V. References

1. K. Manguri, R. Ramadhan, P. Mohammed Amin, Twitter sentiment analysis on worldwide COVID-19 outbreaks, Kurdistan J. Appl. Res. (2020) 54–65, http://dx.doi.org/10.24017/covid.8
2. A. Abdi, S.M. Shamsuddin, S. Hasan, Machine learning-based multidocuments sentiment-oriented summarization using linguistic treatment, Expert Syst. Appl. 109 (2018) http://dx.doi.org/10.1016/j.eswa.2018.05.010

3. E. Marrese-Taylor, J. Velasquez, F. Bravo-Marquez, A novel deterministic approach for aspect-based opinion mining in tourism products reviews, Expert Syst. Appl. 41 (2014) 7764–7775, http://dx.doi.org/10.1016/j.eswa.2014.05.045.
4. A. Garain, Hotel reviews from around the world with sentiment values and review ratings in different categories for natural language processing, 2020, http://dx.doi.org/10.21227/8ggw-hm23.
5. P.-J. Lee, Y.-H. Hu, K.-T. Lu, Assessing the helpfulness of online hotel reviews: A classification-based approach, Telemat. Inform. 35 (2018) http: //dx.doi.org/10.1016/j.tele.2018.01.001.
6. Improving text summarization of online hotel reviews with review helpfulness and sentiment Chih-Fong Tsai a , Kuanchin Chen b , Ya-Han Hu a,c,d,* , Wei-Kai Chen a,e
7. An ensemble-based hotel recommender system using sentiment analysis and aspect categorization of hotel reviews Biswarup Ray, Avishek Garain ∗ , Ram Sarkar
8. SENTIMENT ANALYSIS OF ONLINE CUSTOMER REVIEWS FOR HOTEL INDUSTRY: AN APPRAISAL OF HYBRID APPROACH Siew Theng Lai1, Mafas Raheem2
9. J. Yao, H. Wang, and P. Yin. Sentiment feature identification from Chinese online reviews. In H. Tan and M. Zhou, editors, Advances in Information
10. Technology and Education, volume 201 of Communications in Computer and Information Science, pages 315–322. Springer Berlin Heidelberg, (2011)
11. H.-X. Shi and X.-J. Li. A sentiment analysis model for hotel reviews based on supervised learning. In Machine Learning and Cybernetics (ICMLC), 2011 International Conference on, volume 3, pages 950– 954, (July 2011)
12. Aspect based Sentiment Oriented Summarization of Hotel Reviews NadeemAkhtar[a]NashezZubair[a]AbhishekKumar[a]TameemAhmad[a]
13. https://www.springerprofessional.de/en/a-sentiment-based-hotel-review-summarization/16938654
14. https://link.springer.com/chapter/10.1007/978-3-030-49190-1_14

Contributions:

Motivation by Shanmukha Guntuboina

Objectives by Rushitha Chalamasetty

Significance by Vinaykumar Rajeshwar Gadewar

Features by Prashik Waghmare