IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Special Issue 1, March 2024

1st International Conference on Machine Learning, Optimization and Data Science

Organized by

Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur, India

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

A Sentiment-Based Hotel Review Summarization Using Machine Learning Techniques

Sunder Chaudhary, Prof. Arvind Pandey

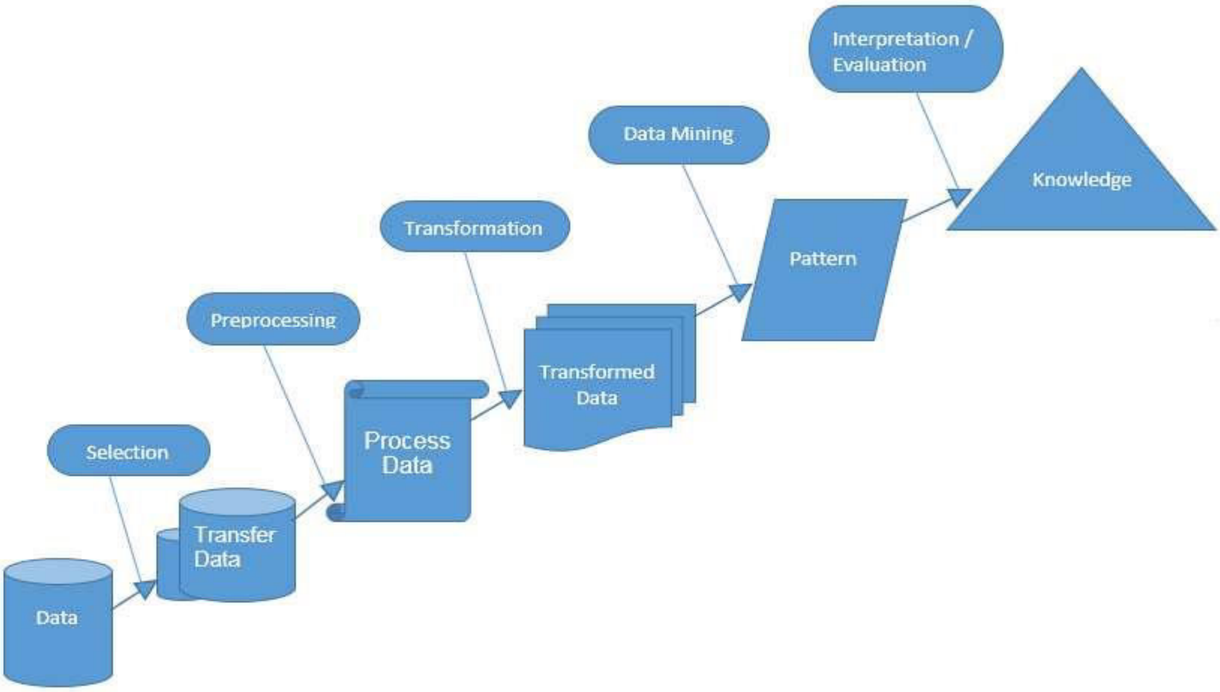
Department of Computer Science & Engineering, Vindhya Institute of Technology & Science, Satna, India

ABSTRACT: The internet revolution has recently brought about a contemporary outlook on individuals' perspectives. It has evolved into a platform where individuals can openly articulate their opinions on a wide range of subjects. This information possesses inherent worth and can be effectively utilized by numerous companies that consistently depend on user input. Opinion mining and its categorization into different emotions are increasingly acquiring significance in decision-making. Sentiment Analysis is a machine learning technology that enables computers to understand and interpret human emotions, as evidenced by the outcomes. This offers a crucial opportunity for numerous organizations seeking to expand and develop. Hotel ratings, whether positive, negative, or neutral, assist customers in discerning their emotional state. The discipline of automatic text processing has extensively examined sentiment classifiers, impact analysis, automated survey analysis, opinion identification, and recommendation structures. The suggested analysis utilizes machine learning algorithms, specifically Naive Bayes and Natural Language Processing (NLP). The system utilizes pre-determined training sets in an effort to minimize the loss of textual knowledge. Moreover, the suggested method allows users to choose between asking or offering a product that meets their precise requirements.

KEYWORDS *:* Sentiment Analysis, Machine Learning, Natural Language Processing, Decision Making, classification, Data Analysis

I. INTRODUCTION

The advent of the internet has recently ushered in a modern perspective on individuals' viewpoints. It has transformed into a platform where individuals may freely express their viewpoints on a diverse array of topics. This information has intrinsic value and can be efficiently exploited by multiple companies who rely on user feedback. The importance of opinion mining and its classification into various emotions is growing in decision-making processes. Sentiment Analysis is a machine learning technology that allows computers to comprehend and analyze human emotions, as demonstrated by the results. This is a pivotal opportunity for several companies aiming to grow and advance. Hotel ratings, regardless of whether they are good, negative, or neutral, help customers determine their emotional condition. The field of automatic text processing has thoroughly investigated sentiment classifiers, impact analysis, automated survey analysis, opinion identification, and recommendation systems. The proposed analysis employs machine learning algorithms, notably Naive Bayes and Natural Language Processing (NLP). The method employs pre-established training sets in order to reduce the loss of textual knowledge. Furthermore, the proposed approach enables consumers to select between inquiring or providing a product that aligns with their specific criteria.



1.1 INTRODUCTION

Technology advances at the same rate as the rapid rise of the internet, and thus travel agents must adapt as well. In these days, people use internet-based travel booking services in a highly integrated ecosystem to book rooms rather than physically visiting the hotel. The feedback left on booking websites, particularly the online booking ones, greatly influences the hotel's long-term sales. Hoteliers have resorted to researching feedback from consumers in order to obtain even the slightest competitive advantage over competitors. It is laborious and time-consuming to review feedback and conduct comparisons when performed manually or by a third party. In order to achieve a competitive advantage, you will need an automated competitor research platform to help you analyse competitor ratings. Unstructured feedback was evaluated using techniques of Natural Language Processing (NLP) and applied to evaluate pre-defined aspects related to the hospitality industry. More analysis was done, such as extracting common citations in reviews, an analysis of reviews, and an analysis of words with regards to their emotions. These types of quantitative and qualitative data were employed by hoteliers to assist them in making assessments of their establishments. Due to outliers in the initial consumer scores, these comparisons ended up being misleading. In order to deal with this tricky problem, a deep learning model was constructed to generate separate ratings for each analysis.

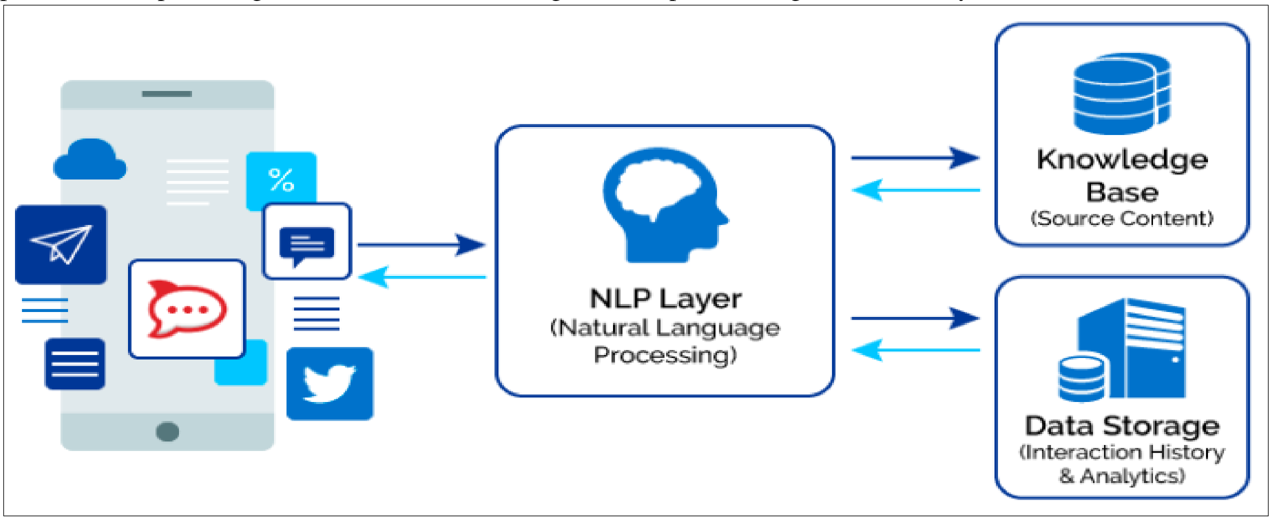


Figure 1.1 Dynamic use of NLP in various forms [30]

The paper delves into the applicability of visual and multimedia analytics to a practical problem, such as hotel reviews. Geospatial analytics, scientific analytics, data management expertise, and many other visualisation and multimedia analytics techniques are employed when processing the volumetric data. A major contribution to the field of travel management could be made with the study's findings. Travel is now a requirement in our daily routine. In the modern era, many people have travelled in order to better themselves and to have experiences from around the world. The international tourism sector has risen at a rate of approximately 4% per year for the past eight years, according to the UNWTO's (UN World Tourism Organization) annual report (2018). In 2017, over 1.3 billion tourists from all over the world travelled to the planet, and this figure is expected to continue rising in the future. The growing number of tourists is evidence of the industry's rising expectations. Over the years, the implementation of information technology has lagged behind the hospitality industry, but this has improved in recent years and science has picked up the slack. As global travel and tourism applications continue to grow exponentially, new possibilities and problems for customers and experts become available, which increases the market size and interest in studying travel and tourism research in all areas, including management science, marketing, and information technology. In order for hotels to be successful in today's evolving market, they must maintain a strong focus on dealing with difficult issues while also searching for new opportunities. It covers the whole topic of decision support and business intelligence for computerised support managers, with the results being an improvement in the way managers make decisions. BI applications are currently scarce in the hospitality industry's most popular topics, motivating new research. The lack of prior studies focusing on using a text mining model exclusively to study consumer attitudes has remained consistent despite the increased study of text mining and sentiment analysis in the last few years. This research is aimed at closing this gap. For small hotels with a different model, such as a "thematic luxury Eco-Hotel," rather than only in large companies or hotel chains, the business intelligence system were used to help with managerial decisions and increase hotel effectiveness. The main goal is to make BI applications known through various marketing channels, and this DSS will make data available quickly in the near future Feedback provided online is important because it opens the hotel's online platforms to visitors, increasing sales and sales performance. Ensuring your brand is effectively managed on the web platforms will help you secure prospective clients and instil in them the need to book your hotel quickly. The evaluation of customer reviews is becoming increasingly important for the hospitality industry. As customers describe it, the storey of how they felt about the hotel's facilities is written in customer feedback. The hotel's activities can benefit from the number of positive reviews it receives, just as bad reviews can be considered when a decision needs to be made. An in-depth study of opinions can help the hotel industry protect against shrinking consumer image, as well as to know what customers like and dislike about their service. Due to the many different websites where consumers can leave feedback, hotels are no longer responsible for evaluating their own performance. These automated systems are required to have meticulous accuracy, rapid response time, and efficiency to enable the companies' decisions. It is necessary to utilise sentiment analysis to better discern whether or not a review is positive, negative, or neutral. Using sentiment analysis, hotels can easily keep track of customer feedback, ratings, and messages on social media websites. In this case, sentiment research is used by hotels to monitor how their brand image is portrayed on the internet, collecting information from customers to create future strategies. When using Natural Language Processing (NLP) techniques, a variety of rule-based methods are used, such as parsing, stemming, and tokenization, in addition to hand-written rules. 2nd, it is critical to first ascertain two sets of various term parameters (For example positive words such as decent, greatest, lovely and negative words such as worse, horrible, poor, etc.). The list of predefined terms and the system counts the number of negative, neutral, and positive feelings in the analysis, after which the system returns negative feelings if the count of negative words exceeds the count of positive words, and vice versa. These results are intended to enable customers and the hotel industry to use the findings to derive relevant information from these reports. It can contribute to the development of new information-related developments, especially in the travel and hospitality industries.

Machine Learning

Machine Learning is the subsection of AI which proceeds with algorithm. An Algorithm is expected to discover an answer for the issue on the PC. It is succession of guidelines which when interpreted or compiled gives us a solution by showing output. Be that as it may, what happens when there is no fix algorithm we have. Least complex illustration is Spam Filtering. Each individual may have an alternate comprehension of what might be viewed as a Spam. We realize that we have emails as information and a yes or no as a yield educating if or not the email is spam. For this situation Data is utilized to compensate for the learning. We gather hundred and thousands of cases about what we consider as spam and gain from them automatically.

∟ Conventional Programming: program is run to deliver the output/comes/result about as appeared in figure 1.3.

Fig.1.3.Conventional Programming

∟ Machine Learning: To make a Program data and Input runs. This program can be utilized as a part of learning and foresee the output appeared by figure 1.4.

Fig.1.4.Machine Programming

It is the branch of Artificial intelligence (AI) which plan and create algorithms and help in developing the behavior and information of PC in light of some genuine information as learning is the principle source to secure knowledge or Intelligence.

ML utilizes information and experience from past to fabricate a machine which consequently finds and learn essential and applicable patterns. [5]

Hotel Business will benefit from business intelligence.

For the first time in history, customers have an unprecedented amount of influence on how they express their thoughts and feelings about products and services. As customers, our voices have a huge impact on the experiences of others. Recent advances in opinion service availability are yielding new opportunities and complications for hotels. The hotel industry is drowning in details due to the amount of information available. There is far more information to be found when data is presented in an unstructured or semi-structured format, which prevents quantitative and qualitative data from being mixed together. Efforts that can automatically gather market information from large volumes of text and merge disparate information into business intelligence databases are therefore essential. Consumers' desires and viewpoints will be better taken into consideration with the use of opinion mining and sentiment analysis tools. Because text mining techniques can be used as a component of business intelligence applications, hoteliers can use them to create strategic competitive intelligence. It's now widely accepted that consumer-generated data is helpful in assessing analysis gaps that weren't previously well known.

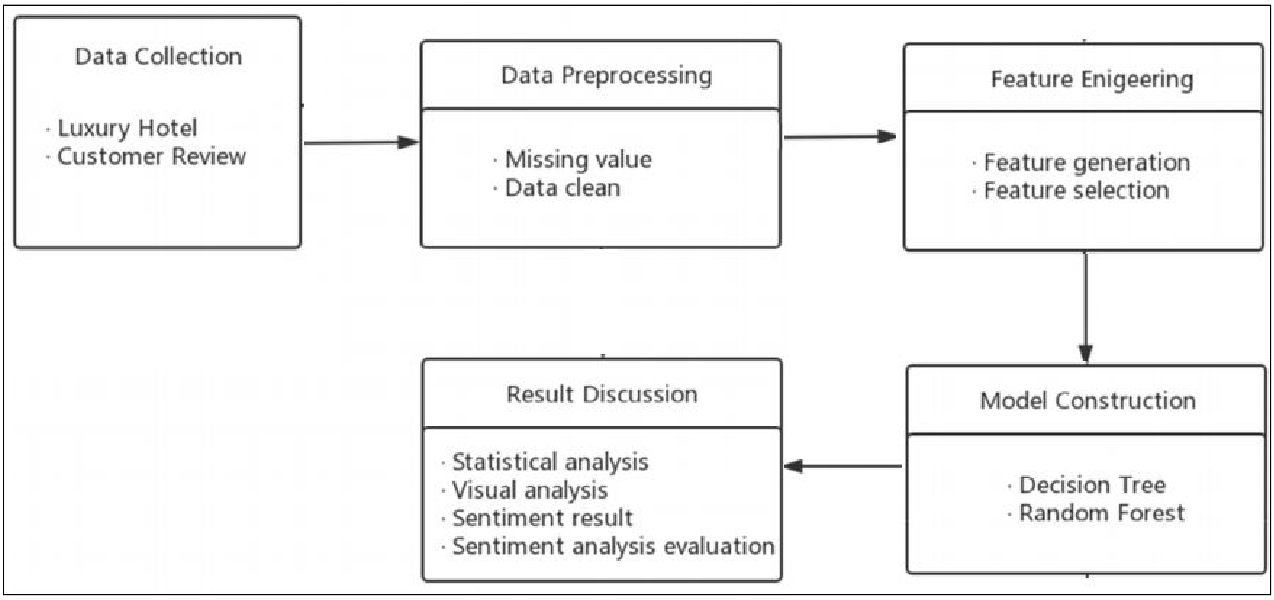


Figure 1.2 Complete Processes of Data Analytics

Data collection serves to improve our understanding of our business and allows us to make more informed decisions. Hotels can take advantage of BI techniques such as identifying common patterns in numeric data, to find more efficient ways of doing things. Customer insights gained from the review of customer newsgroups, online boards, travel journals, and online customer surveys can be applied in environmental screening. Hoteliers can achieve up-to-date knowledge of potential guests by examining corporate records, patent files, and correspondence, which gives them a competitive advantage in hosting business. Business intelligence (BI) data is always up-to-date and current, which results in enhanced hotel productivity and, consequently, a greater customer loyalty. Ignoring customer sentiments could result in businesses being outpaced by competitors, hurting brand recognition, and can even lead to poor market share.

II. REVIEW OF LITERATURE

This section discusses sentiment analysis work done in different zones, Sentiment Analysis. This section concludes with asking about the project's overall inspiration.

One of the first tasks for sentiment analysis is to set up the foundation.

The term 'sentiment analysis' was created in 2003 by Dave, Alan, and Nan based on their observations.

The articles were published between 2003 and Nasukawa et al 2003; however, the research projects were in progress substantially before that in the 1990s. Other portions of the research during the early years of sentiment analysis included finding analogies and creating a dictionary of relevant terms [34] [35]. M.A. Hearst refined data access mechanisms that were based on frameworks that used content to access the data.

One researcher described sentiment analysis as being "a piece of NLP, content mining, and computational phonetics." Discovering the full general extremity of content enables a better understanding of the main point of view of the client.

Two papers were released in the year 1990 and 1994 by J.M. Wiebe. Prior to that, he outlined a procedure to pinpoint the subjective characters in the story's message by locating recurring patterns in the story's content. In later papers, he presented an algorithm to locate the viewpoints of the authors based on the recurrent patterns in the story's content.

Researchers Hu and colleagues discovered that 53% of survey participants reported unusual dispersion, so they concluded that the surveys could not be used to estimate the mean. To illustrate where mean can be used for items quality portrayals and advertisement techniques, a model was developed [40]. The three corpora which the researchers examined were one which they presented, using SVM using different measuring plans and 3-overlap and 10-crease cross approvals, and two other corpora, each of which had a different measuring plan and overlapping creases but no 3-overlap or 10-crease cross approvals.

By dividing each conjunction into three parts, Hatzivassiloglou and McKeown suggested three levels of analysis to uncover the introduction of modifiers. Prior to that, all the conjunctions had been removed from the reports.

Classifier was connected to enable creation of a diagram that shows the differences in the connections among the list of descriptors Finally, the two groups of descriptors have been clearly distinguished, and it has been accepted that there are higher normal recurrence groups that are definite terms. Titles of poetry may be rhymed or can be expressed in iambic pentameter, but for English poets, iambic pentameter is typically used, for the language of the western poets is composed in iambic pentameter.

Deeba Mumtaza and colleagues, together with the Senti-lexical algorithm, presented sentiment analysis on movie review data. They've also suggested a method to handle words that have the effect of negation in reviews, as well as the role of emoticons. An opinion mining objective is to identify, recognise, and categorise data according to people's opinion polarity. According to the Senti-lexicon algorithm, the machine learning algorithm is simple, versatile, and feasible. While the algorithm does not include sarcasm, complex sentences, spam detection, forged reviews, sensitivity over time, and other similar features, it does contain other important elements.

Though Archive level unsupervised direct data sharing is utilised to help arrange surveys, they are neither required nor recommended. 74% of the results were accurately calculated using this algorithm [44]. Later Turney and Littman subsequently elaborated on this research by using PMI and LSA, and the result was an 82.2% precision. In order to communicate effectively, we must clearly understand what others want, need, and value.

Esuli and Sebastiani utilised definitions from online resources to win the challenge. In the beginning, he employed a seed set isolated into positive and negative classification, using that as information. Afterwards, to train the classifier, he created an online lexicon from scratch, examining every one of the definitions. To prepare and test the model, a double classifier was linked to it. Stop words and the holes between stop words served as the basis for sentiment ordering for the work by V Suresh et al [47]. PHOAKS was presented by L. Terveen et al. Individuals Who Share Knowledge Enabling One Another to Discover Information was a pilot framework which used a sifting technique to help clients locate information on the web.

Pallavi Sharma and Nidhi developed a method for classifying the polarity of movie reviews based on various features. The process involves treating statements containing words like "negation," "intensifier," "conjunction," and "synonym" with preprocessing. The SentiWordNet rating calculation tool has been used for the reviews' score calculations. Features associated with a movie's sentiment will be accounted for with the system.

Negation, intensifier, coordinating conjunctive, and synonyms words are seen in the usage of reviews. SentiWordNet 3.0 software has been used to generate the classifications and scores of feature polarity. Not handling the True Negative reviews correctly is a potential disadvantage of this technique.

Virtual commentators were utilised by J. Tatemura in an exploratory manner for examining film surveys from varied viewpoints. The marketers in charge of advertising and customer relationship management worked with Morinaga et al, who sorted through customers' perceptions of the web to find out the recognition of the products [51]. Furthermore, Bing Liu demonstrated a method of critically examining feelings in simlar sentences, as well as the unjustified use of data obtainable on Wen to investigate sentiment setting. Ding et al demonstrated a method to locate the location of the meaning introduction of a number of different sentiment words, which are found in the same sentence. The Sentiment Observer framework was applied as well on this approach (in this instance).

To make a case for SA as pertaining to NLP using Machine Learning methods, BK et al reviewed NLP as it pertains to the assignment of NLP, and came to the conclusion that machine learning strategies are used to discover if the content contains subjective data, regardless of the mind-set of the content. BK et al used this as the framework for an analysis of NLP, as they found that Machine Learning is used to look for subjective data, regardless of the mental state of the content.

String et al. provided an analysis of a motion picture survey dataset in which they studied the performance of SVM, NB, and MaxEnt. They also analysed motion picture audit datasets, studying the results of supervised machine learning (SVM, NB, and MaxEnt) and distinctive element choice (motion picture surveys). Instead of just using one system, he connected various pre-prepping systems like stemming or lemmatization. He utilised NB and MaxEnt to build the execution using POS Labeling; it also involved POS Labeling. Some researchers proposed an approach where they named the archive into target and subjective, after which they used IMDB film database to discover the least slice in the chart that was accurate to within 5% of the targeted 85% [56].

In their article, S.Ahmed and A. Danti describe Opinion mining as the practise of following people's mindsets when it comes to opinion surveys and opinion polls on an item. The results of analysis are found through

SentiWordNet is beneficial for clients because it gives them information on various options in light of surveys and feedback.

The precision of 88.1% was obtained by utilising n-grams and SVMs on sub trees which chose utilising the second form of the dataset. Another model of SVM was shown by Mullen et al. To assess the scores, they utilised WordNet relations along with subject significance and utilised Pang et al's information collection, which found that of the given number, 86% were accurate.

In order to process movie data set that is available on the twitter website in the form of reviews, comments, and feedbacks, Hadoop Framework was used by Huma Parveen and Shikha Pandey. Twitter has been used here to pull out the sentiments from a well-known microblogging website. Using sentiment analysis on tweets is beneficial to providing business intelligence predictions. In addition to illustrating positive, negative, and neutral sentiments, the results of sentiment analysis on Twitter data will be shown in three different sections: one with positive sentiments, one with negative sentiments, and one with neutral sentiments. Opinion, feedback, reviews, remarks, and complaints on Twitter are seen as big data, and therefore can't be used without treatment. Business productivity will definitely improve thanks to this type analysis of will. It is a fast, easy, and speedy method for analysing Twitter trends. Tweets can also be used to help predict future sales, company service quality, or user feedback.

An untrained and supervised half and half system in which they used a vocabulary to find sentences with feelings and expanded this reference by dividing new audits and afterward. Amazon and Google's proprietary corpora were used in a model that utilises stop words to eliminate words that provide no data and give equal chance of displaying determination. Overall, they found that the model they prepared with the intention of applying to all environments had a precision of 80%, whereas an area-specific model had a precision of 83%.

Combining machine learning techniques with techniques developed for interlanguage (English, Dutch, and French), the authors concluded. Multinomial gentle Bayes, SVM, and extreme entropy are contrasted with get the majority execution (62).

Chi-square was used along with higher request n-gram in the proposal by Nigram et al. The team did an internet learning algorithm known as Winnow, as well as a type of neural network called SVM, and discovered that the best outcomes were realised by using SVM [64]. Raychev et al. examined the relationship between a word's location and how powerful a word can be to other people when situated in different locations.

The online hotel review is a rich research tool due to the abundance of information available and the quality of customer ratings (Boo & Busser, 2018). Hotel reviews are extremely important in guiding hotel guests' decision-making (Chang, Ku, & Chen, 2017).

Based on research conducted by Sridhar and Srinivasan, online user reviews tend to moderate or even counteract the impact of hotel rooms (2012). Customers see a ranking that is higher than the hotel rating, so they are more confident about booking the hotel.

Feedback with a simple framework gained more focus and intensified the level of reservation intent as claimed by Sparks and Browning (2011). Hospitals studies affect hotel bookings whether they are negative or optimistic, according to several studies. These conclude: Torres, Singh, and Robertson-Ring (2015).

Similar to the previous study, Fagerstrøm and Eriksson (2018) postulated that customer feedback may influence the desire to book. The hotel reviews' response and recovery strategy affect customers' intentions to book, especially when they are negative (Sparks & Browning, 2011).

For client decision making, 830 consumers (Herrero, San Martín, and Hernández, 2015) conducted an online survey and found that online polls were a major reason. The conclusion of a recent study shows that because of customer feedback, hotels should collect customer behaviour data by tracking marketing campaign reviews throughout the decision-making process. According to the results of some research, the financial success of hotels is linked to hotel financial analysis.

An improvement of 1% in online review scores would result in a sales increase of about 1.42% per room available, according to Anderson (2012).

In this study, conducted by Phillips, Barnes, Zigan, and Schegg (2017), it was found that good customer experience has a positive impact on the hotel's market and potential sales. According to Ye, Rule, and Gu (2009), inspecting hotels will increase room sales.

Reservation count and average daily room rate are also influenced by online reviews (Zhang & Mao, 2012). Every word used to calculate hotel financial success, such as profits per room open, the average daily rate, and occupancy rate, is related to hotel revenue. A hotel's financial success is dependent on how it is rated by guests. Content of the online study includes quantitative and qualitative aspects. In quantitative analysis, specific numbers for hotels, such as hotel reviews and the overall hotel experience, are examined. Useful results can be found in online quantitative evaluations.

Melian-Gonzalez, Bulchand-Gidumal, and Lopez-Valcarcel found the number of hotel reviews and ratings to be positive (2013). This section contains text-like information for customers or consumers. A mix of quantitative and qualitative approaches will lead to top-quality outputs. For example, Banerjee and Chua (2016) assert that the rating trends will vary based on the concerns of travellers in various regions. It is used by different agencies to evaluate feedback.

Surveys are commonly used as a method of getting information. To gather data for a client's online reviews, web surveys are set up on the clients' websites (Filieri, Alguezaui, & McLeay, 2015). According to data obtained from a survey, analysts provide feedback to hotels about how hotels can boost quality and sales.

Several studies employ econometric approaches to calculate the commercial value of reviews (Xie et al., 2014). Using statistical and mathematical models, this approach tries to predict how helpful and trustworthy online reviews are. In short, all of the above options provide fresh insights into hotel operations. However, data gleaned from online reviews is hard to make good use of with traditional methods. As a result, there are various opinion mining and sentiment analysis techniques that are applied to handle massive amounts of data (Liu, 2012).

in some research, computer systems are used to gather feelings and to sift through consumer feedback obtained via online platforms (Barreda & Bilgihan, 2013). In 2009, TripAdvisor conducted a study to identify New York City's top three hotel properties, and 15 different hotel characteristics were identified. A study of the clients' complaints found that, by far, cost and workers were the primary gripes. It is implied that computer-assisted tools are devoted to limiting interplay, and to examining the various characteristics of goods and services in depth. Currently, large-scale data is acquired via the Internet and using sentiment analysis processes is used by NLP researchers (Shi & Li, 2011).

"Melville et al use lexical information with content characterization to perform sentiment analysis, using a multinomial innocent Bayes classifier that fuses the concepts of building a case and practising." Their findings build on the work of Liu et al. and also highlight the importance of utilisation named as fundamental information. After recording each class, they register the cosine similarity between each report and word. This allows them to calculate the likely association between each word and a class. Next, a generative foundation learning model gauges the class priors based on how much class preparation information the students are given before they receive vocabulary words that have either positive or negative names. Thus, the two likelihood circulations are connected, which produces enormous outcomes as a result of direct pooling [67].

Sitesh Kumar, Jyoti Gautam, and Ankur Goel ran an experiment in which we used twitter sentiment140 training data, along with a Naive Bayes model, and developed a method to make classification easier. SentiWordNet (along with Naive Bayes) can improve the accuracy of classifying tweets, by providing a score of words that appear in tweets in terms of positivity, negativity, and objectivity. We implement this system in Python with the Natural Language Toolkit (NLTK) and the Python Twitter APIs. Sentiment analysis of tweets for movies yields rating scores that are based on the unbiased opinions of Twitter users. Naive Bayes has been widely adopted because of its speed. Classification with the use of sentiwordnet and Naive Bayes, with an improved accuracy to an appreciable extent, is proposed in the paper

III. PROPOSED WORK

The World Wide Web has made it possible for people to express their views and thoughts in a different manner than before, and it has created a treasure trove of client-created content for experts. The world is rapidly shifting and leveraging late

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

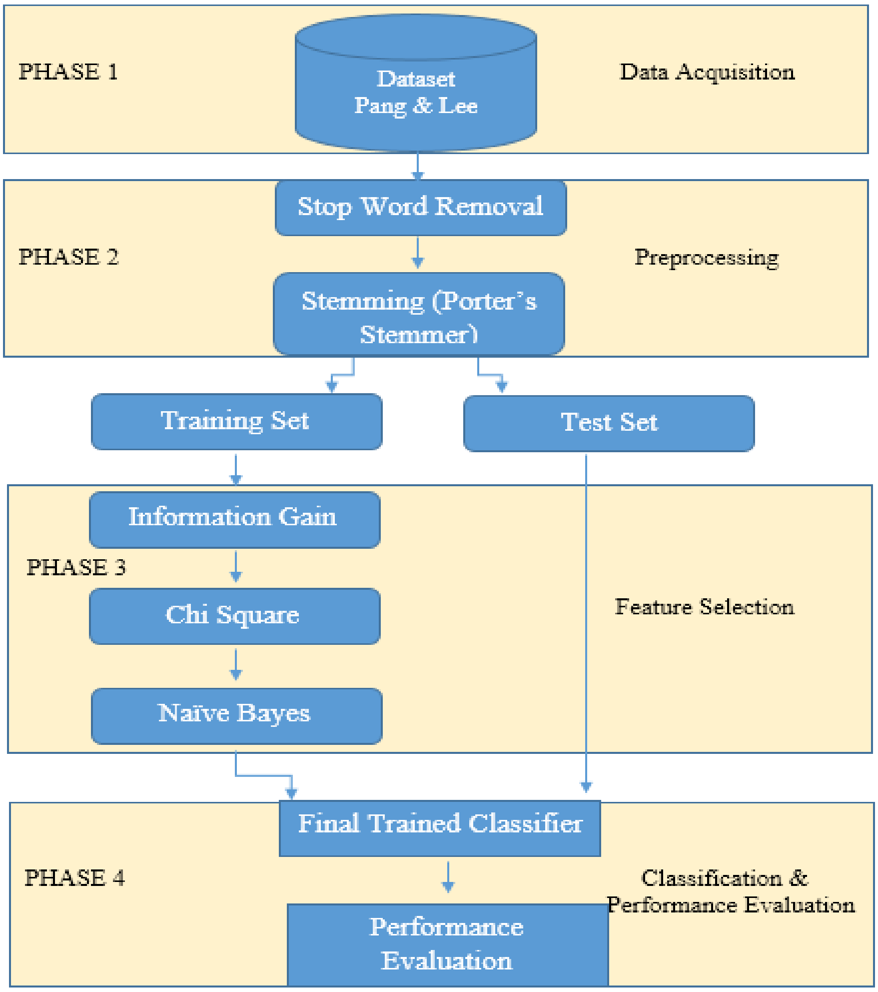
| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

innovations like the Internet. The web has gone from a luxury to a necessity for everyone, and it's being used in many different areas. For the typical person, browsing the internet is an enjoyable pastime (web based shopping, talking, giving audits). Needing to purchase an item once more doesn't include speaking to companions or relatives, nor having to depend on purchaser opinions about your own merchandise. While it may be so, it is undisputed that the wealth of critical data available on the Internet is unstructural, making slant investigation necessary in order to arrive at beliefs and present them in an organised manner. Now that each person has the liberty to express their ideas on the web, companies utilise this reality as a bountiful ground to implement their company strategies. A large number of audits can be found on the internet, making it difficult for a client to thoroughly search through each and every audit to select a service. Recognizing the client's needs and sorting through that information to identify which perceptions clients have is an essential task.



Proposed Methodology Design

Sentiment analysis is a Natural Language Processing assignment that searches for extremity of opinion in a small quantity of content on a point. It has been acknowledged as a valuable line of research by the academic world, and the interest has inexorably been growing from the routine dialect handling group. The terms "feelings," "judgments," "opinions," and "thoughts" can all be used to describe feelings that prompt them. opinionated and truthful data are divided into two distinct types of data. Articulations that portray individuals' sentiments about elements, occasions, and their properties are called opinions while target articulations are called actualities when talking about substances, occasions, and other subject matters. opinion can be painted as well

Four terms: Topic, Holder, Claim, and Sentiment; this is the main concept of the post. It is possible to accept the claim as a whole on a certain subject. He is even capable of attaching an emotional quality to a feeling of conviction that can be positive or negative.

IV. RESULTS

Train a classifier for sentiment analysis

We will use the classification Naive Bayes, as seen in the lecture; we will train 80% of the data and evaluate 20% of the remaining data.

```
from nltk.classify import NaiveBayesClassifier

#Using 80% of the data for training, the rest for validation:
split = int(len(positive_features) * 0.8)
split : 412590
classifier = NaiveBayesClassifier.train(positive_features[:split]+negative_features[:split])
#check the accuracy on the training and test sets, turning accuracy into percentage:

training_accuracy = None #check accuracy of training set
training_accuracy = nltk.classify.util.accuracy(classifier, positive_features[:split] + negative_features[:split])*100
```

training_accuracy : 93.48687559078019

The training accuracy is around 93.5 percent, which is quite good, as expected since the classifier has seen t he data (I was actually expecting it to be a bit higher)

```
test_accuracy = None #check accuracy of test set
test_accuracy = nltk.classify.util.accuracy(classifier, positive_features[split:] + negative_features[split:])*100
```

test_accuracy :92.54663202388801

This test has a precision of nearly 92.5%, which is very precise. It's also substantially higher than 80% of human prediction precision. This highlights that using Naive Bayes for this analysis is a good idea, as the data collection does well with this data. In contrast to the high-precision collection of data from the lecture, the test used in this experiment collects information more crudely. Some of the new features below will be printed to help understand their rationale. When looking for insightful characteristics, focus on words that have either a positive or negative impact on the accuracy of the forecast.

```
classifier.show_most_informative_features()
```

Most Informative Features:

```
Negative = 1        neg : pos   = 22605.9 : 1.0
Positive = 1        pos : neg   = 11601.8 : 1.0
Comfy = 1           pos : neg   =   234.6 : 1.0
Outstanding = 1     pos : neg   =   211.7 : 1.0
Friendly = 1        pos : neg   =   208.5 : 1.0
Spacious = 1        pos : neg   =   184.5 : 1.0
Brilliant = 1       pos : neg   =   168.8 : 1.0
History = 1         pos : neg   =   154.3 : 1.0
Charming = 1        pos : neg   =   153.7 : 1.0
Beautifully = 1     pos : neg   =   133.4 : 1.0
```

When I read various reviews, not only are there words like "bad" and "positive," but also detailed descriptors, like "gratifying" and "inspiring." To many reviewers, saying "no positive" or "no negative" reflects the opposite emotion because of a dataset (one worded as a questionnaire) (number of reviews versus 1). As a result, I wanted to learn more about these insightful characteristics, since 9 out of 10 accurate forecasts have to do with it.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |
| "1st International Conference on Machine Learning, Optimization and Data Science"|
Organized by
| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|
|| Volume 12, Special Issue 1, March 2024 ||

classifier.show_most_informative_features(50)

Most Informative Features:

Negative = 1	neg : pos	=	22605.9 : 1.0
Positive = 1	pos : neg	=	11601.8 : 1.0
Comfy = 1	pos : neg	=	234.6 : 1.0
Outstanding = 1	pos : neg	=	211.7 : 1.0
Friendly = 1	pos : neg	=	208.5 : 1.0
Spacious = 1	pos : neg	=	184.5 : 1.0
Brilliant = 1	pos : neg	=	168.8 : 1.0
History = 1	pos : neg	=	154.3 : 1.0
Charming = 1	pos : neg	=	153.7 : 1.0
Beautifully = 1	pos : neg	=	133.4 : 1.0
Convenient = 1	pos : neg	=	132.4 : 1.0
Helpful = 1	pos : neg	=	125.3 : 1.0
Excellent = 1	pos : neg	=	121.8 : 1.0
Fantastic = 1	pos : neg	=	116.1 : 1.0
Comfortable = 1	pos : neg	=	114.6 : 1.0
Delicious = 1	pos : neg	=	109.0 : 1.0
unstable = 1	neg : pos	=	108.3 : 1.0
Luxurious = 1	pos : neg	=	108.3 : 1.0
Thin = 1	neg : pos	=	103.7 : 1.0
Conveniently = 1	pos : neg	=	103.7 : 1.0
Beautiful = 1	pos : neg	=	101.0 : 1.0
inconsistent = 1	neg : pos	=	97.7 : 1.0
Hop = 1	pos : neg	=	93.7 : 1.0
Superb = 1	pos : neg	=	91.8 : 1.0
Charged = 1	neg : pos	=	88.3 : 1.0
Quiet = 1	pos : neg	=	87.9 : 1.0
Ease = 1	pos : neg	=	87.7 : 1.0
Efficient = 1	pos : neg	=	87.6 : 1.0
Great = 1	pos : neg	=	86.7 : 1.0
Spotless = 1	pos : neg	=	85.7 : 1.0
unusable = 1	neg : pos	=	85.7 : 1.0
Stylish = 1	pos : neg	=	83.4 : 1.0
Lack = 1	neg : pos	=	81.5 : 1.0
Fab = 1	pos : neg	=	81.4 : 1.0
Exceptional = 1	pos : neg	=	81.0 : 1.0
unreliable = 1	neg : pos	=	81.0 : 1.0
Convenience = 1	pos : neg	=	79.3 : 1.0
Amazing = 1	pos : neg	=	78.2 : 1.0
Lovely = 1	pos : neg	=	76.7 : 1.0
Close = 1	pos : neg	=	76.4 : 1.0
Stunning = 1	pos : neg	=	73.3 : 1.0
Wonderful = 1	pos : neg	=	73.0 : 1.0
Peaceful = 1	pos : neg	=	73.0 : 1.0
damaged = 1	neg : pos	=	73.0 : 1.0
Fabulous = 1	pos : neg	=	72.5 : 1.0
Loud = 1	neg : pos	=	72.2 : 1.0
Noisy = 1	neg : pos	=	67.3 : 1.0
Smelly = 1	neg : pos	=	65.0 : 1.0
Missing = 1	neg : pos	=	63.7 : 1.0
loudly = 1	neg : pos	=	63.7 : 1.0

In this list of most informative features it is important to notice that a range of informative terms from the favorable feedback apply to the hotel staff (friendly, supportive, effective) and place (comfortable, convenient and convenience) (Thin, Charged, Unusable, Lack, unreliable, damaged, Loud, Noisy, Smelly, Missing, loudly)

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

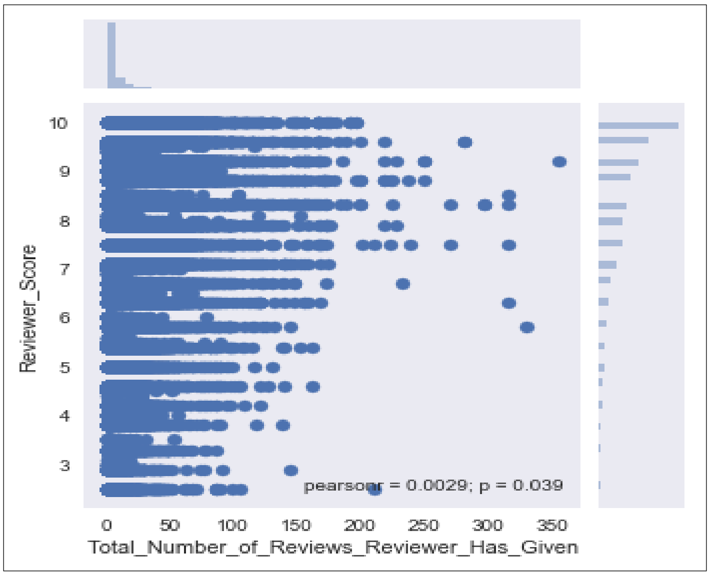


Figure 5.1 Total Number of Reviews Reviewer Has Given

From the above plot, we see no direct connection between the evaluator mark for a specific hotel and the total amount of reviews the evaluator gave. Any outlines are also high and the evaluator score is even higher than 8.5. The evaluator score is high. Most reviewers appear to have given this dataset a limited number of ratings, so I want to change the size of the axes to see this.

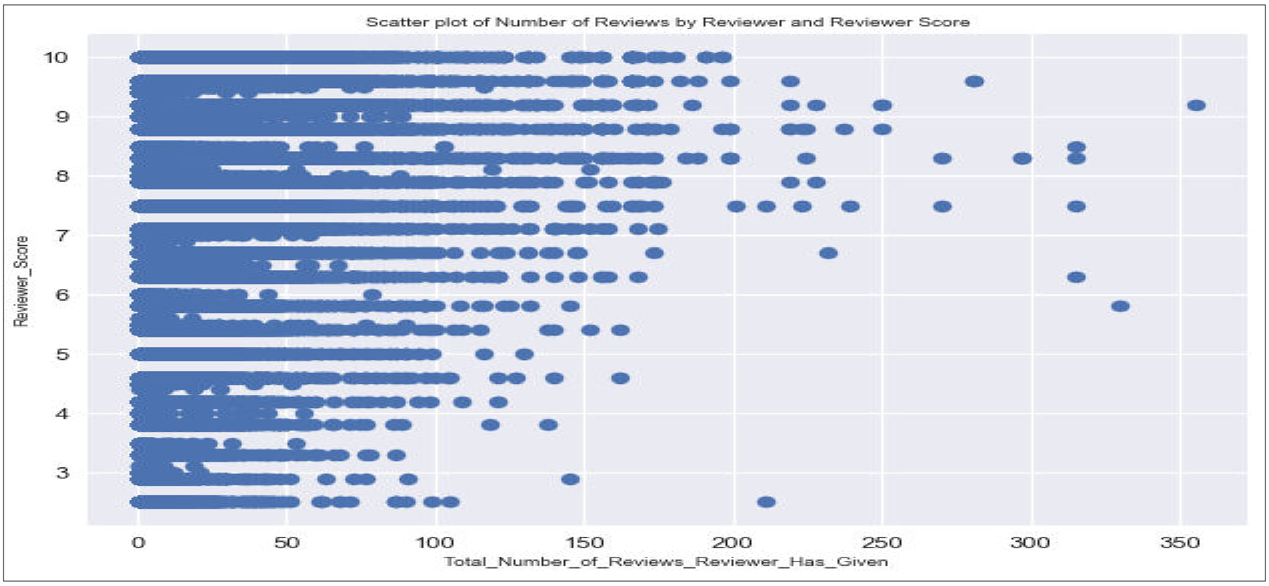


Figure 5.2 Scatter plot of Number of Reviews by Reviewer and Reviewer Score

IV. CONCLUSION

The study and aggregation of opinions into various emotion groups is now playing an increasingly significant role in making important decisions. The results of the automatic text analyses for emotion classifiers, the research into their findings, and the automatic survey analysis have all been thoroughly studied. This methodology usually seeks to identify the average view that is expressed in a sentence or text, regardless of whether or not it is positive or negative. Adopting these strategies, however, has its downside: Information may be lost in text that includes information loss. The proposed strategy aims to deal with text loss by utilising qualified training sets. Additionally, following the

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |
| "1st International Conference on Machine Learning, Optimization and Data Science"|
Organized by
| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|
|| Volume 12, Special Issue 1, March 2024 ||

recommendation of a product or product request, the approach recommended has adopted the customer's specifications. The framework determines where web analysis exists, locates it, classifies it, and summarises the different perspectives. While there are still issues, we also demonstrated that the system works well for research and classification tasks.

REFERENCES

[1]. Ghorpade, T., & Ragha, L. (2012). Featured based sentiment classification for hotel reviews using NLP and Bayesian classification. 2012 International Conference on Communication, Information & Computing Technology (ICCICT). doi:10.1109/iccict.2012.6398136
[2]. Tsai, C.-F., Chen, K., Hu, Y.-H., & Chen, W.-K. (2020). Improving text summarization of online hotel reviews with review helpfulness and sentiment. Tourism Management, 80, 104122. doi:10.1016/j.tourman.2020.104122
[3]. Chang, V., Liu, L., Xu, Q., Li, T., & Hsu, C. (2020). An improved model for sentiment analysis on luxury hotel review. Expert Systems. doi:10.1111/exsy.12580
[4]. Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. In Sentiment analysis and ontology engineering (pp. 313–340). Cham: Springer.
[5]. Boo, S., & Busser, J. A. (2018). Meeting planners' online reviews of destination hotels: A twofold content analysis approach. Tourism Management, 66, 287–301.
[6]. Victor Chang (2018). An improved model for sentiment analysis on luxury hotel review. Expert Systems. 2020;e12580, © 2020 John Wiley & Sons, Ltd.
[7]. Ana Catarina dos Santos Calheiros (2015), Sentiment Analysis In Hospitality Using Text Mining: The Case Of A Portuguese Eco-Hotel.
[8]. Chang, Y. C., Ku, C. H., & Chen, C. H. (2017). Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor. International Journal of Information Management, 48, 263–279.
[9]. Dredge, D., & Gyimóthy, S. (2017). Collaborative economy and tourism: Perspectives, politics, policies and prospects, New York, NY: Springer.
[10]. Eriksson, N., & Fagerstrøm, A. (2018). The relative impact of Wi-Fi service on young consumers' hotel booking online. Journal of Hospitality & Tourism Research, 42(7), 1152–1169.
[11]. Agorakis Bompotas (2020). A Sentiment-Based Hotel Review Summarization Using Machine Learning Techniques, (IFIPAICT, volume 585) springer.
[12]. Filieri, R., Alguezaui, S., & McLeay, F. (2015). Why do travelers trust TripAdvisor? Antecedents of trust towards consumer-generated media and its influence on recommendation adoption and word of mouth. Tourism Management, 51, 174–185.
[13]. Gu, H., Ryan, C., & Yu, L. (2012). The changing structure of the Chinese hotel industry: 1980–2012. Tourism Management Perspectives, 4, 56–63.
[14]. Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. ACM Computing Surveys (CSUR), 49(2), 28.
[15]. Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. Tourism Management, 61, 43–54.
[16]. Gao, S., Tang, O., Wang, H., & Yin, P. (2018). Identifying competitors through comparative relation mining of online reviews in the restaurant industry. International Journal of Hospitality Management, 71, 19–32.
[17]. He, Y., & Zhu, T. (2018). Research on the haze public opinions based on sentiment analysis and social network analysis to microblogging data. Information Science., 36(7), 91–97.
[18]. Herrero, A., San Martín, H., & Hern ández, J. M. (2015). How online search behavior is influenced by user-generated content on review websites and hotel interactive websites. International Journal of Contemporary Hospitality Management, 27(7), 1573–1597.
[19]. Jani, D. (2014).Personality, satisfaction, image, ambience, and loyalty: Testing their relationships in the hotel industry. International Journal of Hospitality Management,37,11–20.
[20]. Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. New York, NY: Cambridge University Press.
[21]. Liu, M. T., Wong, I. A., Tseng, T. H., Chang, A. W. Y.(2017). Applying consumer-based brand equity in luxury hotel branding. Journal of Business Research, 81, 192–202.
[22]. Li, K. X., Jin, M., & Shi, W. (2018). Tourism as an important impetus to promoting economic growth: A critical review. Tourism Management Perspectives, 26, 135–142.
[23]. Nieves, J., & Segarra-Ciprés, M. (2015). Management innovation in the hotel industry. Tourism Management, 46, 51–58.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

[24]. Phillips, P., Barnes, S., Zigan, K., & Schegg, R. (2017). Understanding the impact of online reviews on hotel performance: An empirical analysis. Journal of Travel Research, 56(2), 235–249.
[25]. Torres, E. N., Singh, D., & Robertson-Ring, A. (2015). Consumer reviews and the creation of booking transaction value: Lessons from the hotel industry. International Journal of Hospitality Management, 50, 77–83.
[26]. Zervas, G., Proserpio, D., & Byers, J. W. (2017). The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry. Journal of Marketing Research, 54(5), 687–705.
[27]. Vaibhav Singh, Aayushi Mahajan, Deepanshi Chaudhary (2020). Sentimental Analysis of Hotel Reviews from TripAdvisor, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 06 | June 2020.
[28]. Barreda, A., & Bilgihan, A. (2013). An analysis of user-generated content for hotel experiences. Journal of Hospitality and Tourism Technology, 4(3), 263–280.
[29]. Banerjee, S., & Chua, A. Y. K. (2016). In search of patterns among travelers' hotel ratings in TripAdvisor. Tourism Management, 53, 125–131.
[30] https://www.upwork.com/hiring/for-clients/artificial-intelligence-and-natural-language-processing-in-big-data/.
[31] https://revenue-hub.com/how-sentiment-analysis-can-help-your-hotel-build-a-strong-online-reputation/.