International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 12 | Dec 2020 www.irjet.net p-ISSN: 2395-0072

SENTIMENT ANALYSIS OF ONLINE CUSTOMER REVIEWS FOR HOTEL INDUSTRY: AN APPRAISAL OF HYBRID APPROACH

Siew Theng Lai[1], Mafas Raheem[2]

[1]*Student, School of Computing, Asia Pacific University of Technology & Innovation, Malaysia*
[2]*Acadamic, School of Computing, Asia Pacific University of Technology & Innovation, Malaysia*

Abstract - *The websites, social networking platforms and various business applications have gained a prominent spot as a place for consumers to express their feelings and emotions on certain topics or aspects. Similarly, sentiment analysis is also becoming popular among businesses due to the vast amount of opinionated data available in the digital platforms written by consumers. The written narratives or reviews can be sentimentally categorized as positive, negative, or neutral. Customer reviews have a crucial role in influencing the potential and existing customers regardless of the industries among which the hotel industry is one of the prominent. It makes the hotels understand their value in a competitive environment and to understand their customer views about the hotels' attributes, which allow them to improve themselves to play a strong and competitive role in the market. The main focus of this research is to propose a hybrid method to analyze the sentiment of hotel reviews. The hybrid method is proposed from the merging of the lexicon-based technique with a machine learning technique that would produce an optimum result. It performs the analysis in separate levels, where firstly by using predefined words in the lexicon dictionary to search the polarity of the words, and by using the result from the first level, finally the machine learning algorithm would be trained. The study concludes by discussing the limitations and future enhancements of the research.*

Key Words: **hotel review, hybrid approach, lexicon, machine learning, sentiment analysis**

1. INTRODUCTION

The Internet is becoming a significant place which offers various social media platform for users to express their feelings and emotions about certain topics that capture their interest. Businesses are turning towards social media platforms to get involved in online activities due to the availability of a large amount of consumer-generated data which acts as a source of input for making business decisions. One of the businesses that are adopting social media is the contenders in the hospitality industry, where the hotels need to use online guest reviews to perceive guests' needs, enhance the operation of the hotel, and match up with other hotels as the industry is very competitive. Similarly, customers are resorting to finding hotel rooms through the internet as it is more convenient than the conventional method, and they tend to judge hotels based on the comments shared by other guests.

The sentiment analysis gained growing interest due to the availability of the massive amount of consumer-generated data on the Internet. The sentiment is also known as a feeling, opinion, or emotion, made by a person whereas, sentiment analysis includes the classification of reviews as positive, negative, or neutral. The sentiment analysis has become an important area of research to process the textual data and to extract actionable insights such as understanding customers' feelings towards the hotels and their attributes. The feedback from the guests allows the hotel owners to make actionable decisions from this sentiment analysis.

The unsupervised approach, or also called the lexicon-based method for sentiment analysis is generally used to deduce the sentiment expressed by words/lexicons by specifying the polarity and subjectivity. While, the supervised approach, also denoted as machine learning-based approach for sentiment analysis requires to build a classification model by training the classifier based on a set of labelled data, where the dataset contains positive, negative and neutral class levels and the features or words from the dataset are then extracted for proper training of the algorithm.

This study focuses on the online reviews generally shared by the hotel customers which reflects the experience and the satisfaction via the sentiments such as positive, negative, or neutral. In this line, a hybrid approach combining Lexicon-based and Machine Learning was proposed to predict the sentiments more accurately.

The remnant of this paper is arranged in the following manner. The second section provides a brief literature review or research background related to sentiment analysis on online hotel reviews, the different levels of sentiment analysis, and the evaluation of sentiment analysis using a hybrid approach. The problem statement and, aim and objectives in conducting the study are detailed in Section 3 and section 4 respectively. In section 5, research questions are formulated and explored with the focus on understanding the meaning of hotel reviews by customers. Section 6 discusses the significance or contribution of the research to hotel guests and players in the hospitality industry. The methodology section details the proposed way of the data collection and the method for conducting the analysis, and then, the overview of the proposed solution is shown. Finally, the implications and limitations of this study, as well as the future enhancements for the research are presented.

1.1 Problem Statement

Sentiment analysis or opinion mining is a branch of data analytics that analyzes what people think or opinion or emotions or their stance, viewpoint, perspective towards goods, services, subjects, business, and their attributes expressed in the form of texts [1]. As mentioned by Gupta et al. [2], satisfying the needs and wants of the guests are the main challenges in operating a hotel and as such, the reviews written by the guests have a huge impact on the success of a hotel. It would be difficult for the hotel management to manually analyze a large amount of data to understand whether the guest is satisfied or not, thus requires a big team of people to work together to analyse the data since the reviews on the Internet are in abundance, and that results in wasting of precious time [3].

Jaswal & Tathgir [4] also stated that different approaches of sentiment analysis need to be incorporated to improve the efficiency of sentiment classification and to resolve the existing problems. Many issues in sentiment analysis were noticed due to the content of unstructured customer-generated reviews such as informal language or grammatical errors, domain unspecific, internet slang, abbreviations, emoticons, hashtags, hyperlinks, locations, sarcasm, word sense disambiguation, and opinion spam [5]. Khan et al. [6] argued that the algorithms worked well in supervised approaches, but it is lack of labelled training corpus for the particular area including the data sparsity, whereas, the unsupervised approaches do not show the effects of the same problem, yet the performance levels are not up to par.

1.2 Aim

The main aim of this appraisal is to explore and propose the application suitability of the hybrid approach of sentiment analysis in the hospitality industry.

2. RESEARCH BACKGROUND

One of the major sources of success for hotels is perceiving guests' wants and needs [7]. According to He, et al. [8], automated computational methods like sentiment analysis are becoming popular, to process huge amounts of data written by consumers on social media. While, sentiment analysis, in another word, opinion mining is the study of analyzing people's judgement, thinking, and perspective concerning certain entities [9]. In this section, the latest and relevant pieces of literature regarding opinion mining in online hotel reviews, the distinct levels for sentiment analysis, and the need evolution of opinion mining using hybrid approach are briefly reviewed and discussed.

2.1 Opinion Mining in Online Hotel Reviews

There is an increase in embracing social networking platforms in the hospitality industry and because the players in the industry run in an ambitious environment, the players must use online guest review constructively to have a better understanding of their guests, increase hotel standards and match up with other players in the field [4]. This is supported by [5] which states that customer reviews are becoming crucial in determining the standard of the hotel because the assessments play a huge role in giving influence on other potential guests. Due to the increase of information written in texts available on many social media platforms posted by users, sentiment analysis is becoming popular to analyze the vast amount of customer-generated data to get meaningful insights [10]. Normally, it is practical to indicate the sentiment's polarity in a written piece of text whether they are positive, negative, or neutral, as well as the strength of sentiment, conveyed because texts usually hold a mixture of positive and negative sentiments [11].

As such, in the hospitality industry, hotels use sentiment analysis to discover the truth behind the attitude of customers. [4] utilized the sentiment analysis method to obtain written opinions of 58, three to a five-star hotel in four major cities in China which were selected through TripAdvisor and it helps to give them a fundamental thought of the hotel industry. On the other hand, the research led by Barbosa et al. [12] used different sentiment analysis techniques to evaluate the reliability of numerical ratings of hotels in seven cities in four countries which were withdrawn from TripAdvisor website as well which concludes that sentiment analysis of hotel reviews does correspond with the general ratings on the website. These signify that opinion mining for online reviews in the hospitality industry is very important for not only the customers but also the hotel management.

2.2 Different Levels for Sentiment Analysis

Analysis of sentiment can be investigated at three levels subjected to the granularities required such as whether the target of the research is a whole text or document, one or several combined sentences, or one or a few entities or features of those entities [13]. The different levels that will distinctly dictate the sentiment analysis tasks are namely document level, sentence level and feature level [14], [15], [16] which are further discoursed as the following:

2.2.1 Document Level

This is the most straightforward structure of classification where the task is to classify the review from the entire document relays a positive, negative, or neutral sentiment. For instance, when users review a product, the system decides if the general view is positive or negative regarding the product. It is presumed that the document reveals opinions about a single object only and so, it is not an appropriate approach to documents that contain or compare multiple objects. In the study [17], the researches incorporated two different machine learning algorithms by using the review data corresponding with movies to classify the sentiments which have shown that the exactness gained in this approach to outperform other researchers' results.

2.2.2 Sentence Level

Based on the name itself, the task implies that sentiment analysis is performed on sentences and determines if every sentence conveys a positive, negative or neutral sentiment, while neutral generally means no opinion, for a particular product or service. Each sentence is considered a separate entity, and each can have a conflicting opinion. For example,

Jaitly & Ahuja [18] researched on increasing the efficiency of sentiment analysis at the sentence level by classifying tweets based on the sentiment expressed in them which produced a successful outcome to better understand people's opinions.

2.2.3 Feature Level

The analyses on two levels and they are, the document level and the sentence level do not precisely describe what people like, or do not like which makes feature level more suitable. This level is also known as aspect level and it provides a more accurate analysis. The task involved is to identify and extract opinions regarding specific features of a product from the source data. Recent studies [19] suggest that the results obtained in proposing two novel sentence-based features and creating algorithms for identifying sentiments have shown the state of being superior to the proposed features.

2.3 Evaluation of Sentiment Analysis using Hybrid Approach

The traditional way of classifying sentiment can be conducted by using supervised or unsupervised methods and the success depends on the suitable removal of the set of elements applied to identify sentiments [20]. With the advancements of technology and the large amount of data collected on reviews written by customers online, there is a need to implement a hybrid approach to process the sentiment in a given text to provide a better result in a short time. This is because, in general, a hybrid approach to perform sentiment analysis yields better performance and results. A hybrid approach of combining lexicon-based technique and machine learning technique has significantly increased the performance of classification compared to lexicon-based technique and machine learning technique alone [21], [22]. This is supported by the research conducted by [23] which concludes that the hybrid approach results in a better performance compared to the conventional method. Similarly, Lalji & Deshmukh [24] states that the lexicon-based technique results in high precision but low recall, and to increase the performance measurements, the machine learning technique is used along with a lexicon-based approach. As there are different techniques under supervised and unsupervised methods, it can be argued that there are many ways or methods of combining both supervised and unsupervised methods which give out different results. This is asserted by the review done by Ahmad et al. [21] which states the various hybrid tools and techniques for sentiment analysis along with the strengths and limitations for each tool or technique. In a nutshell, it is proven that a hybrid approach is the most suitable method to perform sentiment analysis.

3. METHODOLOGY

A research methodology is a process of how research is being led. The methodology chosen is CRISP-DM, which stands for Cross-Industry Standard Process for Data Mining, and the approach follows data collection/ acquisition, data preparation/ pre-processing, sentiment/ polarity detection, sentiment classification, and display of data, which will be further explained in Section 5, where an overview of the proposed solution will be discussed. It is important to choose the right method that suits to solve the problem statement and to address the aim analysis and objectives of the study, where the aim is to analyze online hotel reviews and to use a hybrid approach to perform sentiment. In the following section, the components are broken down into two for collecting data and analyzing data.

3.1 Data Collection

The general methodological procedure for investigating the problem of the research is a qualitative approach. This research approach is usually used to perceive meanings, describe and understand experience or beliefs, or gaining insights into specific concepts. This approach fits the overall research design because the study requires online reviews written by hotel guests describing their experience when staying at a particular hotel. For this qualitative approach, existing data is used for the research. Usually, hotel guests are given open-ended surveys and asked to write a review online to know about their satisfaction or dissatisfaction during their stay by the hotel themselves to gain a better insight to improve the hotel's services and compete with other contenders in the industry. The review normally takes up less than 10 minutes of the customers' time. Whereas, some guests prefer to express their feelings on other social media platforms. The reviews will then be stored in the hotel online databases which can be retrieved or can be downloaded through certain websites and it is easier when the website has an application programming interface (API). Open-ended surveys are difficult to quantify the results, but it would allow customers to express or show their feelings and thoughts truthfully.

3.2 Data Analysis Proposal

Data analysis is another important area in the research which has an impact on the overall research. In qualitative research, negative reviews are more common since data found in the form of true feelings and opinions regarding a particular area. Data analysis is the process of searching and arranging the data for the researcher to gain extra knowledge regarding the data and to showcase the findings to others. This study emphasis to analyze the opinions from the hotel guests and the analysis includes data preparation, data reduction, sentiment detection, presentation of data, and conclusion drawing, to create new findings. As the data would be too large to be analyzed, a random sampling method can be used to select the data or review to perform the analysis. The data gathered would be arranged, analyzed, and summarized according to the sentiment drawn from the texts whether the sentiments are either negative, positive, or neutral sentiment. Then, the analysis and the interpretations of the findings concerning the aim of the study would be completed. Open-ended surveys usually give out a result that provides a more in-depth understanding of customers' feelings and emotions towards a particular element. A potential limitation in the analysis of the data would be that some reviews written by customers may not be related to the domain concerned.

4. OVERVIEW OF THE PROPOSED APPROACH

Figure 1 shows the workflow for the hybrid approach proposed in the study to increase the performance of sentiment analysis which would yield a better result.

The sentiment analysis of hotel reviews includes the steps as stated as the following:

4.1 Data Collection/Acquisition

The data acquired is from the particular hotel database itself or through the hotel's online website API. It is easier to get the reviews from the API because the user can download the data according to the user's wants and needs. Data can also be collected through forums, blogs, or several social media sites.

4.2 Data Preparation/Pre-processing

Review data obtained online can contain noisy data such as abbreviations, hashtags, emoticons, slangs, and so on. In short, it includes indicating and excluding non-textual content and subject from the data that is not relevant to the research area. Therefore, it is essential to filter or pre-process the extracted data before analyzing the data to polish the data and eliminate the noise which makes it easier to perform analysis on the cleaned data.

4.3 Sentiment/Polarity Detection

In this stage, every sentence in the review is inspected to find the polarity by using the lexicon dictionary to search the occurrence of the word, and the polarity value shown by the lexicon dictionary will replace the word position. Sentences with subjective statements are kept while objective statements are removed.

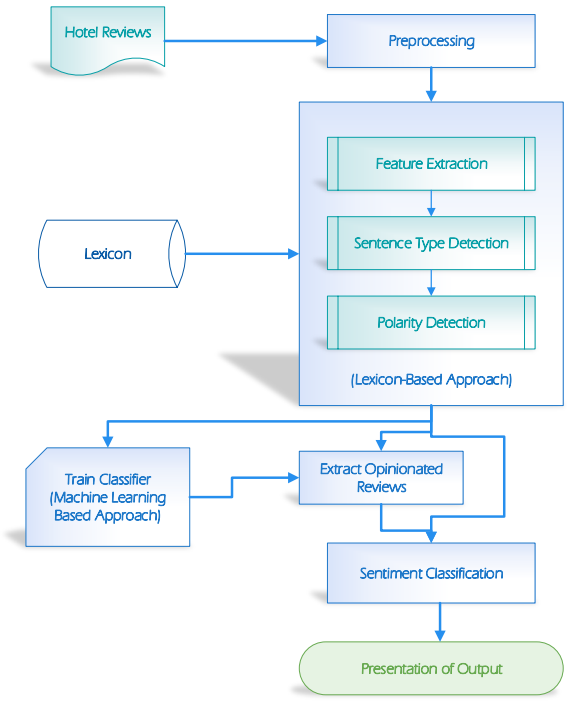

Fig -1: The Workflow for using Hybrid Approach to Perform Sentiment Analysis

4.4 Sentiment Classification

Sentiments can mainly be separated into two groups and they are positive and negative. The polarity obtained by the lexicon dictionary for every phrase is thought to be the training data and then it is employed by the machine learning algorithm to train the classifier. At this phase, every subjective phrase detected is separated into different groups.

4.5 Display of Data

The main goal of this investigation is to perform analysis on online hotel reviews and to convert those unstructured data into meaningful information for potential customers and hotel management. At the end of the analysis, several graphs will be used to show the results.

5. CONCLUSIONS

It is becoming a norm for people to publicly express their thoughts and feelings on social media platforms, whereby businesses, especially hotels should be able to use their customers' opinions and its analysis to make effective decisions. Hotels must be provided with technological approaches which would help them to efficiently process a large amount of data and improve their services, especially via the sentiment analysis. This paper proposed a hybrid approach for classifying the hotel customers' online reviews to perform sentiment analysis for the hotel to elevate the competitive values and customer relationship. A sentiment analysis performed using a lexicon-based approach will show high precision but low recall which results in poor performance. A hybrid approach of combining both unsupervised and supervised is proposed to improve the performance.

The hotels can follow the approach proposed in the study to help them track, collect, and analyze guest reviews on the Internet. A constraint of the research is that it analyzes reviews written only in English and not in other languages. Thus, the research needs to be further measured with reviews written in other languages too. One of the important directions for future research would be summarizing the predicted positive and negative hotel reviews, and to show them in a comprehensible format for end-users to aid hotel in management of online hotel image. It is also important to building a system for hotels to track and monitor the opinions due to its dynamic aspect on the internet. Besides, a comparison of using online reviews written in different languages can be expected to see crucial differences in the results.

REFERENCES

[1] N. Divyashree, K. L. Santhosh, and J. Majumdar, "Opinion Mining and Sentimental Analysis of TripAdvisor.in for Hotel Reviews", International Research Journal of Engineering and Technology, vol. 4, no. 11, pp. 1462-1467, 2017.

[2] S. Gupta, S. Jain, S. Gupta, Shruti, A. Chauhan, "Opinion Mining For Hotel Rating Through Reviews Using Decision Tree Classification Method", International Journal of Advanced Research in Computer Science, vol. 9, no. 2, pp. 180-184, 2018.

[3] T. Najam, K. B. Nowshath, and Vazeerudeen, "Sentiment Analysis of tweets using Supervised Machine Learning", Journal of Applied Technology and Innovation, vol. 3, no. 1, pp. 30-32, 2019.

[4] X. Tian, R. Tao, W. He, and V. Akula, "Mining Online Hotel Reviews: A Case Study from Hotels in China", in Americas Conf. on Information Systems, San Diego, CA, 2016, pp. 1-8.

[5] P. Choudhari, and S. V. Dhari, "Sentiment Analysis and Machine Learning Based Sentiment Classification: A Review", International Journal of Advanced Research in Computer Science, vol. 8, no. 3, pp. 1051-1056, 2017.

[6] F. H. Khan, U. Qamar and S. Bashir, "A Semi-Supervised Approach to Sentiment Analysis using Revised Sentiment Strength based on SentiWordNet", Knowledge and Information Systems, vol. 51, no. 3, pp. 851-872, 2017.

[7] P. Phillips, S. Barnes, K. Zigan, and R. Schegg, "Understanding the Impact of Online Reviews on Hotel Performance: An Empirical Analysis", Journal of Travel Research, vol. 56, no. 2, pp. 235-249, 2017.

[8] W. He, X. Tian, R. Tao, W. Zhang, G. Yan, and V. Akula, "Application of social media analytics: A case of analyzing online hotel reviews", Online Information Review, vol. 41, no. 7, pp. 921-935, 2017.

[9] S. Yordanova, and Kabakchieva, "Sentiment Classification of Hotel Reviews in Social Media with Decision Tree Learning", International Journal of Computer Applications, vol. 158, no. 5, pp. 1-7, 2017.

[10] X. Zhang, Y. Yu, H. Li, and Z. Lin, "Sentimental Interplay between Structured and Unstructured User-Generated Contents-An Empirical Study on Online Hotel Reviews", Online Information Review, vol. 40, no. 1, 2016.

[11] O. Hoeber, L. Hoeber, M. El Meseery, K. Odoh, and R. Gopi, "Visual Twitter Analytics (Vista) Temporally changing sentiment and the discovery of emergent themes within sport event tweets", Online Information Review, vol. 40, no. 1, pp. 25-41, 2016.

[12] R. R. L. Barbosa, S. Sanhez-Alonso, and M. A. Sicilia-Urban, "Evaluating hotels rating prediction based on sentiment analysis services", Aslib Journal of Information Management, vol. 67, no. 4, pp. 392-407, 2015.

[13] V. K. Bongirwar, "A Survey on Sentence Level Sentiment Analysis", International Journal of Computer Science Trends and Technology, vol. 3, no. 3, pp. 110-113, 2015.

[14] S. Behdenna, F. Barigou, G. Belalem, "Document Level Sentiment Analysis: A Survey", EAI Endorsed Transactions on Context-aware Systems and Applications, vol. 4, no. 13, pp. 1-8, 2018.

[15] S. Kolkur, G. Dantal, and R. Mahe, "Study of Different Levels for Sentiment Analysis", International Journal of Current Engineering and Technology, vol. 5, no. 2, pp. 768-770, 2015.

[16] P. Patil, "Sentiment Analysis Levels and Techniques: A Survey", International Journal of Innovations in Engineering and Technology, vol. 6, no. 4, pp. 523-528, 2016.

[17] A. Tripathy, A. Anand, and S. K. Rath, "Document-level sentiment classification using hybrid machine learning approach", Knowledge and Information Systems, vol. 53, no. 3, pp. 805-831, 2017.

[18] A. Jaitly, S. Ahuja, "Improving The Accuracy For Sentence Level Sentiment Analysis", International Journal of Advanced Research in Computer Science, vol. 9, no. 4, pp. 37-41, 2018.

[19] S. L. Gupta, and A. S. Baghel, "Efficient Feature Extraction in Sentiment Classification for Contrastive Sentences", International Journal of Modern Education and Computer Science, vol. 10, no. 5, pp. 54-62, 2018.

[20] B. Keith, E. Fuentas, and C. Meneses, "A Hybrid Approach for Sentiment Analysis Applied to Paper Reviews", in Proc. of ACM SIGKDD Conf., Halifax, Nova Scotia, Canada, Aug. 2017, pp. 1-10.

[21] M. Ahmad, S. Aftab, I. Ali, and N. Hameed, "Hybrid Tools and Techniques for Sentiment Analysis: A Review", International Journal of Multidisciplinary Sciences and Engineering, vol. 8, no. 1, pp. 28-33, 2017.

[22] C. P. Ashish, and K. M. Patel, "Sentiment Analysis Using Hybrid Approach: A Survey", International Journal of Engineering Research and Applications, vol. 5, no. 1, pp. 73-77, 2015.

[23] P. K. Jaswal, and J. S. Tathgir, "Sentiment Analysis of Social Media Data using Hybrid Approach", Journal of Economic Development, Management, IT, Finance and Marketing, vol. 10, no. 2, pp. 1-6, 2018.

[24] T. K. Lalji, and S. N. Deshmukh, "Twitter Sentiment Analysis Using Hybrid Approach", International Research Journal of Engineering and Technology, vol. 3, no. 6, pp. 2887-2890, 2016.