Sentiment analysis on hotel reviews about staff using Machine Learning and Lexicon-based methods

Marion Bonnard
Student X19161964
Higher Diploma in Data Analytics
National College of Ireland

***Abstract*— How is a hotel's service quality perceived? How do hotel staff interact with and assist guests? Nowadays, people tend to check reviews online before making a purchase; this behaviour also applies when booking a hotel. This paper aims to provide insights on hotel reviews to understand the overall opinion of guests on hotel experience through a sentiment analysis on a dataset containing 515,738 guest reviews. Lexicon-based and Machine Learning (ML) approaches were applied on reviews containing the term "staff". The objective was to understand the overall sentiment that arose from these reviews and compare classifier models to predict the polarity of the reviews. R Studio and Python programming tools were employed to complete this task. Results show that "rude" is the most frequent terms used in the negative reviews by applying the Bing lexicon, whereas "friendly" and "help" are positively associated with "staff". The ML approach has confirmed that the Support Vector Machine (SVM) classifier has proven a better accuracy with 91.55%. This analysis demonstrates how the hospitality industry can benefit from data and text analytic techniques to find the root cause of issues arising from guest reviews.**

I. Introduction

Following the pandemic, most hotels have and will be suffering from the impact of the restrictions. According to McKinsey's research [1], it is forecasted that hotel business recovery might not happen before 2023 in the United States. In the "new normal", guests will have increased expectations on hygiene [2]. Hence, it is crucial to consider their feedback on social media, mainly when electronic Word of Mouth (eWOM) influences guests when they book a hotel [3].
Hotel directors must be aware of any issues happening in their establishment. The solution offered in this project helps hotel managers save time by highlighting the overall sentiment and indicate what term is associated with a feature such as "staff". Front line employees can be the reason for a weak or excellent customer experience. It is more frequent for dissatisfied guests to leave a review than satisfied guests. A bad review can impact reservations by losing 30 potential bookings. [4] The need for this analysis is to highlight the importance of staff attitude towards guests.
Programming tools such as RStudio and Python have been used to clean, pre-process and complete the sentiment analysis with Machine Learning (ML) and lexicon-based approach.

Sentiment Analysis uses Natural Language Processing (NLP) techniques to extract insights from text. It is a technique that determines if a phrase is either positive, negative or neutral (polarity) [5]. Additionally, this analysis can be done with a lexicon to determine the polarity and emotions [6]. In this project, two methods have been applied to carry out sentiment analysis. With the ML method, three algorithms (SVM, Naïve Bayes and Logistic Regression) have been trained to predict the polarity of the reviews. A second method is a lexicon-based approach that determines if each term is positive or negative and categorises terms by emotion.

Staff is the focus because, in some literature, the term "staff" obtained the highest frequency and revealed having a significant impact on guest dis/satisfaction [7], [8] and [9]. Additionally, the term "staff" represents a high proportion of cost in hotels. As per [10], 42.8% of hotels' operating costs are allocated to labour costs. Selecting one specific feature brings some novelty since rare literature focuses on the "staff" feature to conduct sentiment analysis.

This paper has three main objectives: 1) to discover what are the most frequent terms/features used in the overall reviews; 2) to classify hotel reviews by polarity and define what the best prediction model is; 3) to emphasise what is the main sentiment related to the feature "staff".

II. Related Work

Many studies have been conducted to demonstrate the application of Sentiment Analysis on reviews. Some have used classifiers; others have used lexicon and even hybrid approaches. The first part of this literature review explains the reason behind choosing a staff subset. The second part illustrates the various inspiration from other works that have used machine learning to perform sentiment analysis. The third part highlights how lexicons are used to classify reviews.

A. Influence of staff on guest experience

In the study of [7], the word "staff" obtained the third highest frequency, and by using factor analysis, they found that "helpful" and "friendly" had the highest negative correlations (-0.51 and -0.52, respectively). This study demonstrated that a low satisfaction rating is associated with hotel staff. Moreover, [8] highlighted that staff-related attributes have a strong influence on guest satisfaction and reviews. Negative attitude and rudeness were seen as dissatisfying by guests. Staff attitude was the most impactful aspect of customer satisfaction and dissatisfaction, as described in the study [9].

Poor interaction with a hotel employee could have a high impact on customer experience. Moreover, people tend to remonstrate online rather than in-person; this might significantly impact the hotel perception. It is described that women tend to "find it embarrassing to complain in person" [11].

B. Classifiers predictive model approach

The predictive model approach was first inspired by [12], who have used binary and multiclass classifiers to extract maintenance related problems from the hotel reviews. Their goal was to create an automatic system to identify the types of maintenance issues pointed out by the guests. The authors used TF-IDF as term frequency to extract the features, and after comparing several classifiers, they obtained the best performance with the SVM classifier with a 92.17% accuracy. These excellent results motivated the choice of using the SVM algorithm as a classifier. In the study from [5], the authors gathered reviews on Yelp and used the Naïve Bayes algorithm to classify reviews polarity. Their main scope was to define the hotel features (bedrooms, staff, cleanliness) that satisfy guests. They used the TF-IDF technique to highlight the relevance of terms and build a document term matrix (DTM) to measure the terms per document's frequency. The model obtained 89% accuracy. The pre-processing steps of [5] have greatly influenced our project. [13] described Naïve Bayes as a suitable model for feature selection and reached an accuracy of 85.62%. In [14], the authors have developed a hybrid sentiment classification using a boosted SVM model. They underlined the weakness of a single SVM, knowing that it was time-consuming when dealing with big text data. They concluded that boosting improved classification and their highest result was 93%. By comparing Random Forest (R.F.), SVM and Multinomial Naïve Bayes, [15] demonstrated that R.F. by using TF-IDF with unigrams and bigrams improved accuracy to obtain 90%. In [16], Multinomial Naïve Bayes (MNB) classifier was used with feature extraction. Their goal was to create a model that will automatically classify reviews. To extract the features, the authors used the "bag of word" technique; this method was used in the exploratory analysis and lexicon approach of our paper. Also, the authors evaluated the performance of the MNB model with two scenarios. They found that the performance increased when the data was pre-processed and when using the frequency-based technique. Using a Bayesian method seemed to bring a high F1-score of 91.41%, especially when using a frequency-based feature selection method. On the other hand, [17] obtained a higher accuracy when using the TF-IDF technique instead of frequency. They demonstrated that with the TF-IDF technique, the SVM algorithm got a higher F-score with 87.2% compared to using a frequency approach with 86.4%. This paper has inspired our decision to use this method. Reference [6] also used Machine Learning to predict the helpfulness of the model by using Random Forest (R.F.) and SVM techniques. The dataset was divided into 70:30 for training and testing dataset respectively, and the SVM algorithm obtained a higher performance than R.F. A similar split has been performed in our ML model.

C. Lexicon-based approach

The lexicon approach is applied on a pre-processed corpus; it assigns and sums up the score of each term to find the polarity of the text [18]. [6] used the AFINN lexicon to categorise terms by polarity by assigning a score between -5 (negative) and 5 (positive). This author also used NRC Word-Emotion Association Lexicon from Mohammad and Turney (2013)[1], which considers feelings and associated the terms used to each emotion. This technique provided the overall sentiment score and gave more insights about people's emotions retrieved from the reviews. In [19], the author used a combination of Lexicon-based and ML methods to discover the most effective approach for classifying text. The outcome has shown a better performance was achieved for the Lexicon-based approach with 89% accuracy than ML algorithms (KNN, SVM and Decision Tree). This hybrid approach has motivated our methodology. In [20], the authors described three sentiments lexicons such as AFINN, NRC and Bing that can be performed on text analysis to find the overall sentiment or emotions related to terms. The authors criticised the fact that the lexicon-based approach considered only unigrams (unique words) and did not take into account bigrams such as "not clean" or "not great". In this case, the word clean and great would probably be considered as positive and be misleading. [21] also combined Lexicon and ML-based methods on an unlabeled dataset. The authors claimed that by using lexicon, the results obtained a low recall. Indeed, since the lexicon could not read a smiley, abbreviation or informal expressions, an improved lexicon had to be created to consider the specific element of the text. [22] conducted a sentiment analysis on a TripAdvisor dataset and found that the Bing method emphasised negative opinions. The Bing Lexicon has been applied in this project's analysis to provide another perspective on opinions.

III. Methodology

The below methodology has been applied using tools such as R Studio, Python, Excel and Tableau. Some Exploratory Data and Text Analyses (EDTA) have been performed on the original review dataset. This step used natural language processing techniques to find out about the most used terms. Then, the "staff subset" was created from the initial review dataset to conduct the lexicon and ML approaches.

A. Dataset

The dataset downloaded from Kaggle [2] was originally scrapped from Booking.com; it contains 515,738 records and 17 columns. The reviews have been labelled as "Negative" and "Positive", which presented an opportunity to apply supervised learning.

[1] https://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

[2] https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe

B. Cleaning and pre-processing

The dataset had to be prepared to perform the sentiment analysis. The following cleaning and pre-processing steps have been made:

1) Appended the negative and positive reviews and selected only word count, labels, and text to obtain a combined dataset of 867,640 records.
2) Removed the fields considered as noisy data such as "No negative" and "No Positive."
3) Reduced the dataset by randomly selecting 50,000 observations due to size reasons in R Studio. (Example of the reduced dataset in Appendix 1)
4) Converted the dataset into a corpus, which is an extensive collection of documents. The corpus had to be cleaned to remove noisy data and unnecessary words. The corpus cleaning steps have been done with the "tm" package in R Studio by transforming letters to lowercase, removing numbers, stop words, whitespace and stemming words. Stemming is a type of text normalisation that keeps the root of a word and removes the suffix or prefix.
5) Converted the cleaned corpus to a document term matrix which broke down each document into words. This step is also called tokenisation; it is the process of separating sentences into tokens, and a token is a "meaningful unit of text" [20].

The matrix was then sorted by frequency and transformed into a data frame where the most frequent words would appear first; this is also called Bag of Words.

C. Exploratory Text and Data Analysis

Exploratory Data and Text Analyses (EDTA) has been performed on a reduced review dataset of 50 thousand records using Excel for initial descriptive statistics, Tableau for visualisation and RStudio for deeper insights. Missing values were inspected, some descriptive statistics and polarity observations were completed. Bar charts, word clouds and frequency association have also been prepared to provide valuable insights to answer the research questions.

D. Lexicon-based approach

The lexicon-based approach has been performed on the staff subset containing 178,884 reviews. As described in the previous sections, staff was one of the most important aspects of a hotel and had a high impact on customer satisfaction. Moreover, this subset represented 35% of the total reviews from the original dataset. The same pre-processing steps have been taken, from corpus creation to tokenisation. At this stage, the Tidy Text function called "unnest_token" has been applied to obtain 2,990,058 tokens. Three lexicons have been pulled to convert each token or unigram into a score or an emotion. The NRC Emotion lexicon associated words to one of the eight basic emotions: joy, surprise, trust, anticipation, fear, anger, sadness, and disgust, as well as a positive or negative sentiment. The AFINN lexicon from Finn Arup Nielsen[3] contains over 3000 words associated with a score from -5 to 5. Finally, the Bing lexicon from Bing Liu[4] lists around 6800 words by polarity. The "text data" and "syuzhet" packages have been installed to retrieve these lexicons. They were applied to the tokenised staff data frame, and the parameters were set to sort the tokens by frequency (Bag of Word). The three lexicons were plotted and compared. For NRC, only the "negative" and "positive" sentiments were selected to provide an equal comparison of the lexicons. This step has contributed to discovering the main sentiment related to staff.

E. Polarity Classification

Binary classification techniques using ML models have been applied to predict if reviews were positive or negative (polarity).

The staff subset was composed of 16% negative and 84% positive reviews; this is called a class imbalance problem. Over and under-sampling had to be completed to obtain an even polarity distribution and avoid any bias in the performance. Additionally, the dataset has been reduced to 50 thousand observations due to big data size not being processed by RStudio.

The corpus cleaning step has been done using the same method as described in the cleaning and pre-processing section, although the stop words "hotel", "staff", "location", and "room" have also been removed. These terms have appeared to be the most frequent and would not bring additional insights as they are aspects of a hotel. Once the corpus was cleaned, a term-document matrix tokenising the documents was created. While this step was being processed, a TF-IDF condition was applied. TF-IDF highlights the importance of a word in a document rather than its frequency (i.e. Bag of Word). It assigns a weight to each term according to the frequency in a document rather than the frequency across all the reviews. The main objective was to assign a numerical value to each feature to run the classification algorithms. This TF-IDF matrix was then transformed into a data frame that has been used to run algorithms of the prediction models for classification. This data frame has been loaded on Python since R Studio could not run code with a big vector size. The data have then been partitioned as 70% training and 30% testing data, as it has been completed in [6] and has proven better accuracy. After this step, the labels were separated from the features.

Binary classifiers such as Logistic regression, SVM and Naïves Bayes have been applied to create the predictive models. These methods were selected as they were all supervised and could be used for text classification; they have also been highly used in the literature described in the related work section and shown satisfactory accuracy. Naïve Bayes uses training data to forecast the probability of an event. This approach was inspired by the Bayes Theorem (1763) using the following formula:

$$P(A|B) = P(B|A) * P(A) / P(B)$$

[3]http://www2.imm.dtu.dk/pubdb/pubs/6010-bibtex.txt

[4]https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

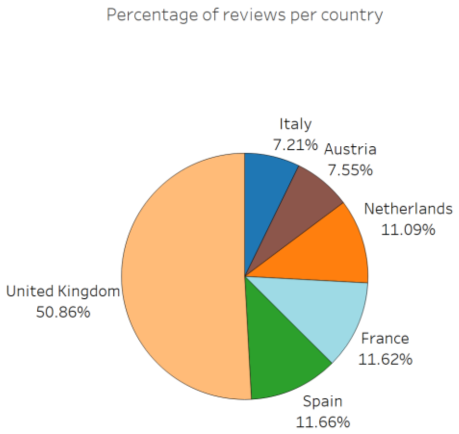
Support Vector Machine or SVM[5] is a linear classifier that locates optimal hyperplane to separate data into two classes and which principle is to maximise margins[6]. For logistic regression, the model was trained on the features to predict the polarity of a review. The Sigmoid function was used to transform an input into a value between 0 and 1.

Following the partitioning, the training dataset was fitted into the algorithms, and the predictions were produced using the models generated. Finally, a confusion matrix was printed to measure and compare the performance of each method.

IV. Results

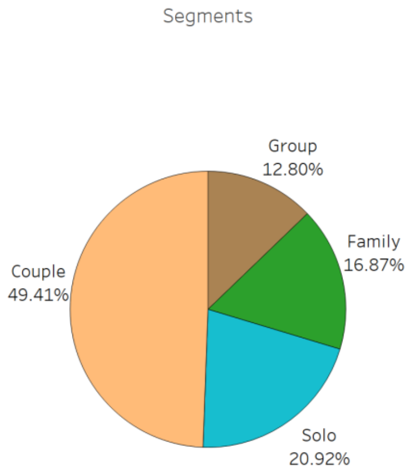
The EDTA answered the first research question to define the most frequent terms used in the overall reviews. The 1492 hotels were located in six European destinations where most reviews (51%) were about hotels in the U.K. (see Figure 1).

Figure 1. Percentage of reviews per country



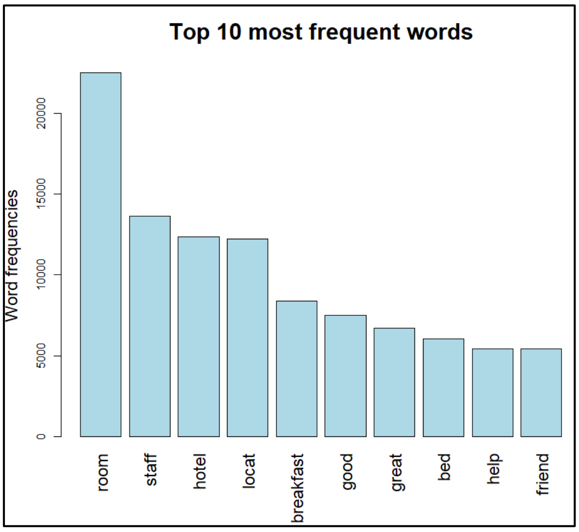
When looking at the distribution of the reviews, 55% were labelled as positive. The score was distributed from 2.5 to 10; the average displayed 8.39, and the mode was 10, which explained the majority of positive reviews. Reviews per nationality showed the highest contribution from U.K. guests with 47.55% reviews. Regarding the guests' segments, it was found that over 80% of the guests were staying for leisure purposes, mainly couples (49%) and the remaining were solo travellers, families and groups (see Figure 2).

Figure 2. Guests segments



The results obtained on the text data using RStudio, exhibited in figure 3 the ten most frequent words. It was found that "room" appeared 45% of the time in the reviews. Additionally, staff was the second most used term appearing 13,637 times (27%) in the reduced dataset. A word cloud (Figure 4) was created to represent the most frequent words displayed with a more prominent font size appearance. The word cloud displayed the top 100 words with a minimum frequency of 500. "Friend" and "help" were the two words associated with staff, with a correlation of 0.46 for both terms. "Friend" is the stem for "friendly", whereas "help" is the stem term for "helpful". It means that the staff were seen as friendly and helpful, which helped understand how guests perceived the staff and partially answered the 3rd research question answer.

Figure 3. Top 10 most frequent words

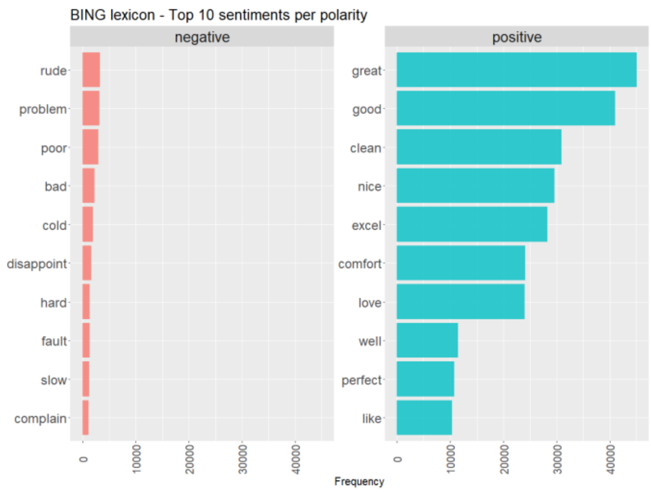


[5] https://cml.rhul.ac.uk/svm.html

[6] https://link.springer.com/referenceworkentry/10.1007/978-0-387-73003-5_299

Figure 4. Wordcloud

Figure 5. Bing Lexicon – Top 10 sentiments per polarity



A. *Lexicon based approach*

Overall, the results showed that positive sentiment emanated from the reviews as the 20 most frequent terms of the Afinn lexicon were positive (Appendix 2). It was also the case for the Bing lexicon, where only 13% of the terms identified were negative. The NRC lexicon categorised the terms by emotion; however, we could see that some terms (i.e., "friend", "good", "excel") were duplicated and spread across several sentiment categories, which made them challenging to characterise. (See Appendix 2). Similar to the EDTA results, the term "help" appeared as the most frequent in the Afinn lexicon and "friend" in the NRC lexicon, as shown in the table in Appendix 2. This outcome emphasised the similarity of the results by using two different techniques. In the NRC lexicon, "breakfast" appeared to be the third most mentioned term, similar to the EDTA results where "breakfast" was in the top five words. This result highlighted another aspect that was highly mentioned and would deserve further investigation.

Even if the negative terms are less frequent, the Bing lexicon managed to highlight the most frequent negative term: "rude" (see figure 5). This term means "impolite" and could only be associated with a human. This insight suggested that it was due to a rude attitude when a guest was not happy with the staff.

When looking at the comparison between the three lexicons in appendix 3, a similar pattern was observed and revealed that the last reviews (as from review 150,000) looked torn between negative and positive sentiments. The explanation comes from the fact that the staff subset has not been randomly shuffled; indeed, the positive reviews were on top of the negative reviews due to the appending step in the cleaning phase. Thus, the same order was kept when creating the staff subset, which explained why the sentiment score appeared as negative from review number 150,083. This result proves that the lexicon technique approach can successfully differentiate terms polarity.

B. *Polarity Classification*

The re-sampling performed has proven successful when reducing the staff dataset, with 50.3% positive 49.7% negative labels. The confusion matrix in Table 1 assesses the classification models' performance, and results have revealed SVM as the best predictive model. SVM method obtained the best accuracy with 92.03%, followed by Logistic regression with 91.85% and Naïve Bayes with 87.99%. However, Naïve Bayes performed faster with less than 1 second compared to Logistic Regression with 9 seconds and SVM with nearly 7 minutes. Precision is the percentage obtained when the model has correctly predicted that a review is negative. SVM has obtained a precision score of 91%, which showed that the negative reviews were correctly predicted over 90% of the time. The recall measures if the prediction correctly predicted that the review was positive. In this case, the recall of the SVM model appeared to be 92.13% which means that the prediction showed a high recall rate compared to precision. Therefore, the model performed better when predicting a wrong "positive" review than predicting an incorrect "negative" review. The recall score was the most important because it needed to be accurate when it came to wrongly predicting that a review was positive while it was actually negative. This result has a high impact as hotels are interested in knowing what aspects are not satisfying. Missing 504 negative reviews (SVM) means that some potentially helpful constructive feedback has not been considered.

Table 1. Polarity classification results

Algorithms	**Accuracy (%)**	**Error Rate (%)**	**Precision (%)**	**Recall (%)**	**F-score (%)**	**Time taken to run models (seconds)**
Naïve Bayes	87.99%	12.01%	85.78%	91.06%	88.34%	3.85
Logistic regression	91.85%	8.15%	90.53%	93.46%	91.97%	7.17
SVM	92.03%	7.97%	91.00%	93.27%	92.13%	702.23

V. CONCLUSION AND FUTURE WORK

In this project, sentiment analysis has been performed on hotel reviews using two main techniques such as Lexicon-based and Machine Learning and some additional EDTA using natural language processing techniques. The objective has been reached by answering the three research questions described in the introduction.

The most frequent term used in the overall reviews was the word "room". However, it was decided to focus the analysis on the second term, "staff", to understand what sentiment was related to that crucial aspect of a hotel. "Helpful" and "friendly" were highly correlated with "staff"; also showing in the lexicon approach, Afinn and NRC displaying these terms as most frequent and positive sentiments. On the other hand, the Bing lexicon proved to bring meaningful insight by drawing attention to the term "rude" in negative reviews. The term "rude" can only be associated with a person; therefore, it can be concluded that staff attitude needs to be closely monitored. Additionally, it has been found that SVM obtained the best performance in predicting the classification of the reviews by polarity with 91.55%. Although the model took more time than the other models, the recall of 92.94% appeared to be satisfying, with only 314 negative reviews misclassified.

This paper looked at two different ways to analyse over 180 thousand reviews without reading any review. It showed how ML and Lexicon approach could classify sentiment and express an opinion from relatively big data using NLP.

Some limitations were encountered, such as not considering bigrams ("not good", "not clean") when using the lexicon-based approach, which could potentially mislead some of the results. Also, due to a vector size problem in RStudio, the remaining part of the ML section had to be done using Python. Additionally, when selecting the feature "staff" while creating the "staff subset", other hotel employee-related terms have not been considered such as "waiter", "concierge", "receptionist", "bartender"; therefore, in future works, all the people related features should be included.

Finally, other works could dive deeper into the "breakfast" feature, one of the most frequent terms to appear. By extracting a subset on breakfast, we could use the lexicon-based approach to reveal additional insights.

REFERENCES

[1] V. Krishnan, R. Mann, N. Seitzman and N. Wittkamp, "Hospitality and COVID-19: How long until 'no vacancy' for U.S. hotels?," 10 June 2020. [Online]. Available: https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/hospitality-and-covid-19-how-long-until-no-vacancy-for-us-hotels#.

[2] D. Zemke, J. Neal, S. Shoemaker and K. Kirsch, "Hotel cleanliness: will guests pay for enhanced disinfection?," *International Journal of Contemporary Hospitality Management,* pp. Vol. 27 No. 4, pp. 690-710, 2015.

[3] Y. Yang, S. Park and X. Hu, "Electronic word of mouth and hotel performance: A meta-analysis," *Tourism Management,* pp. 248-260, 2018.

[4] R. Patel, "How to Prevent Guests from Leaving Unhappy," [Online]. Available: https://smartguests.com/blog/turning-rants-to-raves-how-to-prevent-hotel-guests-from-leaving-unhappy/#.YB7Wfej7Sbg.

[5] M. J. Sánchez-Franco, A. Navarro-García and F. J. Rondán-Cataluña, "A naive Bayes strategy for classifying customer satisfaction: A study based," *Journal of Business Research,* pp. 499-506, 2019.

[6] S. Chatterjee, "Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach," *International Journal of Hospitality Management,* 2020.

[7] Z. Xiang, Z. Schwartz, J. Gerdes and M. Uysal, "What can big data and text analytics tell us about hotel guest experience and satisfaction?," *International Journal of Hospitality Management,* pp. 120-130, 2015.

[8] P. Padma and J. Ahn, "Guest satisfaction & dissatisfaction in luxury hotels: An application of big data," *International Journal of Hospitality Management,* p. 84, 2020.

[9] B. Kim, S. Kim and C. Heo, "Analysis of satisfiers and dissatisfiers in online hotel reviews on social media," *International Journal of Contemporary Hospitality Management,* pp. 1915-1936, 2016.

[10] R. Mandelbaum, "2016 Trends in the U.S. Hotel Industry," CBRE Hotels' Americas Research, 2016.

[11] I. Steiner, "People Prefer to Complain Online Rather than in Person," 20 May 2016. [Online]. Available: https://www.ecommercebytes.com/2016/05/20/people-prefer-complain-online-rather-person/.

[12] B. Athuraliya and C. Farook, ""Revyew" Hotel Maintenance Issue Classifier and Analyser using Machine Learning and Natural Language Processing," in *IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON),* Vancouver, Canada, 2018.

[13] J. Chen, H. Huang, S. Tian and Y. Qu, "Feature selection for text classification with Naïve Bayes," *Expert Systems with Applications,* pp. 5432-5435, 2009.

[14] A. Sharma and S. Dey, "A boosted SVM based ensemble classifier for sentiment analysis of online reviews," *ACM SIGAPP Applied Computing Review,* pp. 43-52, 2013.

[15] R. Talekar, "Classification of Customer Satisfaction for the Development of Hospitality Businesses," Research Project, Dublin, 2020.

[16] A. Farisi, Y. Sibaroni and S. Al Faraby, "Sentiment analysis on hotel reviews using Multinomial Naïve Bayes classifier," *Journal of Physics: Conference Series,* pp. 1-10, 2019.

[17] X. Li and H. Shi, "A sentiment analysis model for hotel reviews based on supervised learning," in *International Conference on Machine Learning and Cybernetics,* Guilin, China, 2011.

[18] N. Gupta and R. Agrawal, Hybrid Computational Intelligence, Elsevier Inc., 2020.

[19] N. Mukhtara, M. Khan and N. Chiragh, "Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains," *Telematics and Informatics,* pp. 2173-2183, 2018.

[20] J. Silge and D. Robinson, Text Mining with R: A Tidy Approach, O'Reilly, 2017.

[21] L. Zhang, R. Ghosh, M. Dekhil, M. Hsu and B. Liu, "Combining lexicon-based and learning-based methods for Twitter sentiment analysis," H.P. Laboratories, Technical Report HPL, 2011.

[22] V. Luzón, A. Valdivia and F. Herrera, "Sentiment Analysis in TripAdvisor," IEEE, 2017.

VI. APPENDICES

Appendix 1: Reduced dataset

	Review	Word_Count	Review_Type	isPositive
1	close to Wembley stadium good shower pressure	8	Positive	1
2	Location ok rooms ok but basic for price	9	Positive	1
3	Everything was amazing We had imperial room free high wi...	61	Positive	1
4	Best location	4	Positive	1
5	Nice cosy bar areas and lounges	8	Positive	1
6	Greasy breakfast choices and a minimal selection of option...	20	Negative	0
7	Seems should walk for a while for Wiener Stephansdom	10	Negative	0
8	The location is fantastic Would stay there again just for that	13	Positive	1
9	The staff are always helpful the room spacious which is the ...	18	Positive	1
10	A beautiful hotel very elegant Although we booked a triple ...	42	Positive	1
11	I m not sure it was worth 180 for just bed no breakfast noth...	16	Negative	0
12	The location it was only 100 meter far away from the Metro...	19	Positive	1
13	Lovely continental breakfast with a good choice Spacious r...	33	Positive	1
14	Comfortable small hotel with a lived in atmosphere Staff ve...	32	Positive	1
15	I always stay here for Milano trips Perfect location and very ...	20	Positive	1
16	communications when i checked the book by phone	9	Negative	0
17	Room was cold	5	Negative	0
18	Staff were very helpful	6	Positive	1
19	Air conditioner was really poor the hotel should do more wi...	14	Negative	0
20	My wife loved her anniversary stay at the rafayel hotel The ...	25	Positive	1
21	Location	2	Positive	1

Appendix 2: Most frequent terms using lexicon approach

	Afinn			Bing			NRC		
	word	value	n	word	sentiment	n	word	sentiment	n
1	help	2	63019	great	positive	45116	friend	joy	66588
2	great	3	45116	good	positive	41053	friend	positive	66588
3	good	3	41053	clean	positive	30974	friend	trust	66588
4	clean	2	30974	nice	positive	29643	good	anticipation	41053
5	nice	3	29643	excel	positive	28249	good	joy	41053
6	comfort	2	24118	comfort	positive	24118	good	positive	41053
7	love	3	24022	love	positive	24022	good	surprise	41053
8	perfect	3	10731	well	positive	11514	good	trust	41053
9	like	2	10340	perfect	positive	10731	breakfast	positive	40852
10	recommend	2	7203	like	positive	10340	clean	joy	30974
11	free	1	6754	quiet	positive	8255	clean	positive	30974
12	pleasant	3	5497	recommend	positive	7203	clean	trust	30974
13	best	3	5260	free	positive	6754	excel	anticipation	28249
14	kind	2	4719	work	positive	5911	excel	joy	28249
15	super	3	4710	wonder	positive	5828	excel	positive	28249
16	big	1	4329	modern	positive	5776	excel	surprise	28249
17	want	1	4266	pleasant	positive	5497	excel	trust	28249
18	enjoy	2	4135	spacious	positive	5323	comfort	anticipation	24118
19	thank	2	3660	best	positive	5260	comfort	joy	24118
20	better	2	3203	super	positive	4710	comfort	positive	24118

Appendix 3: Lexicon comparison