Aspect Based Sentiment Analysis and Feedback Ratings using Natural Language Processing on European Hotels

MSc Research Project
Masters in Data Analytics

Princy Dcunha
Student ID: x18135889

School of Computing
National College of Ireland

Supervisor: Dr. Catherine Mulwa

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Princy Dcunha
Student ID:	x18135889
Programme:	MSc Data Analytics
Year:	2019
Module:	MSc Research Project
Supervisor:	Dr. Catherine Mulwa
Submission Due Date:	12/12/19
Project Title:	Aspect Based Sentiment Analysis and Feedback Ratings using Natural Language Processing on Irish Hotel
Word Count:	6340
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	27th January 2020

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Aspect Based Sentiment Analysis and Feedback Ratings using Natural Language Processing on European Hotels

Princy Dcunha
x18135889

12/12/2019

Abstract

The hospitality industry is the most blooming industries in today's date. Around 710 million international tourist arrivals were recorded in the European Union in the year 2008 and the number is predicted to increase greatly in the coming years. Tourism contributes to hospitality opportunities and that serves as a great business for hotel owners. The internet revolution has only made it better for them as it is a platform used by millions of people to express their experience and opinions about a service. These thoughts put forth by the users of the internet serves as a medium to textual data generation, which is helpful to analyse the needs and shortcomings of a business. In this project an aspect based sentiment analysis approach is used to classify user reviews as positive or negative and Machine learning classification models such as Naïve Bayes, Support Vector Machine, Logistic Regression , Random Forest and Decision Tree are used. Naïve Bayes was seen to outperform every algorithm used with a high accuracy of 84.55%. This project would benefit the hospitality industry in a way by knowing what their customers felt towards the services they offered. What remains unsolved is the approach of topic modelling on this particular dataset.

1 Introduction

The Gross Domestic Product (GDP) of Ireland is \$357 billion, and the Gross National Income (GNI) is \$300 billion. Tourism contributes to about 2.3% of the GDP and close to 4% of the GNI. The number of tourists visiting Ireland is growing at a rate of over 2% per year, over the past 8 years, and this growth in number is expected to continue till the year 2025. Currently, there are 820 registered hotels (58,757 rooms) in Ireland, with 33% of rooms located in Dublin. The Dublin market saw an increase of 1,100 new bedrooms (5.7% new rooms added) in 2018, which is the highest percentage increase in rooms since the last 10 years. According to the latest reports, based on the current growth rate, there is a need to construct at least 11,000 additional rooms to accommodate tourists, by 2022. There is a huge growth potential to be exploited in this sector and make tourism an industry which can be a significant contributor to the economy. Ireland has the 5^{th} highest per capita GDP, which means that the people in Ireland have a higher disposable income (above \$25,000), which could be a blessing for the services industry.

The travel platforms available online encourages the service users to help share their thoughts, experience and opinions about the hotels or travel destinations , leading to creation of a large user generated review dataset , says Dundar et al. (2016). At the same time, this information is being consumed by new travellers and tourists to get information about those specific destinations, says Sigala (2009). Hence, this user generated content serves as an electronic Word Of Mouth (eWOM) . This eWOM can be used as an important source of information regarding customer satisfaction with different components of the hospitality services, stated by Chanwisitkul et al. (2018). This research project attempted to summarize and analyze an internal hotel data set using aspect based sentiment analysis and Natural Language processing for qualitative data. Using hotel attributes, this study gives binary sentiment classification of reviews such as positive and negative. Using Hirokawa and Hashimoto (2018) recommended future work , they proposed to implement sentiment analysis data with Naïve Bayes , Random Forest and Decision Trees . Based on Saito and Klyuev (2019) Naïve Bayes was used to check if it really out-performs the other algortihms and it was actually found to yield better results than others, with an accuracy of 84.55% .

The stake holders benefiting by this project are the people linked with the hospitality and tourism industry. The internet revolution has brought about a new way of expressing an individual's opinion as per Ghorpade and Ragha (2012) . Currently, the most common method that tourists make decisions on selecting their preferred hotels is a data analysis from online reviews of experienced tourists to support the decision Kitwatthanathawon et al. (2012). 4.48 billion people worldwide are active internet users, which comprises of 58% of the global population as per[1] .This is one of the major reasons why it felt necessary to examine and perform analysis on the quantitative and qualitative data of such kind.

1.1 Research Question

RQ: *"To what extent can aspect based sentiment analysis of European hotels' review using Natural Language Processing (NLP) approach and machine learning methodologies (Support Vector Machine (SVM), Logistic Regression, Naïve Bayes, Random Forest, Decision Tree) be used to predict customer satisfaction to improve the service delivery?"*

1.2 Objectives and Contributions

Objective 1 Involves investigation and review of literature of hotel reviews research projects from 2009-2019.

Objective 2 Perform aspect based sentiment analysis of hotel reviews using Natural Language Processing (NLP).

Objective 3 Implementation, Evaluation and Results of classification models.

Obj 3.1 Implementation, Evaluation and Results of Naïve Bayes Classifier.

Obj 3.2 Implementation, Evaluation and Results of Random Forest Algorithm.

Obj 3.3 Implementation, Evaluation and Results of Logistic Regression.

Obj 3.4 Implementation, Evaluation and Results of Support Vector Machine (SVM).

[1]https://www.statista.com/statistics/617136/digital-population-worldwide/

Obj 3.5 Implementation of Decision Tree.

Contributions: The major contribution is to classify customer sentiments for satisfaction prediction and to find the hotel rankings based on the score.

The roadmap to the rest of the report is as follows : Section 2 presents Literature Review, Section 3 presents Scientific Methodology, Data Pre-processing and Architectural design, Section 4 discusses about the implementation, evaluation and results of this project and Section 5 states the final conclusion of this project with its recommended future work.

2 Related Work

The internet revolution has brought about a new way of expressing an individual's opinion Ghorpade and Ragha (2012). Currently, the most common method that tourists make decisions on selecting their preferred hotels is a data analysis from online reviews of experienced tourists to support the decision Kitwatthanathawon et al. (2012). 4.48 billion people worldwide are active internet users, which comprises of 58% of the global population as per[2]. These users play the role of data generation by uploading their opinions and reviews about services they have used or are yet to use. In order to analyse this content, which is classified as the one having negative polarity and the rest having positive polarity, it is necessary to use latest techniques and methods in order to achieve this task. Aspect based sentiment analysis is one way of classifying the text on the basis of positivity or negativity. To support the technique of feature extraction done using aspect bassed analysis, a critical and brief overview of related works in this field is done below, wherein section 2.1 is about the research documents that were used as a road map in achieving this project, Section 2.2 is dedicated to various literature that performed similar methods to gain the results related to Hotel Reviews Analysis and section 2.3 is about various models that can be used and are used in machine learning when using Textual data.

2.1 A Review Of Methodologies Used By Other Researchers, Used For This Project

Chauhan et al. (2017) performed sentiment analysis and polarity computation along with spam detection on product reviews. Unlike other researchers, who used the existent ratings to solve the problem, the author proposed to generate a sentiment score based on the reviews and then map it and compare them with the posted ratings. The sentiment score was obtained using the following Sentiment Score(r) = (F)* (W) / L. This technique is used in the performed research, to find the polarity of each sentence based on the aspects. According to Dundar et al. (2016) 's study, sentences were quantified using fuzzy quantification , to create a brief summary of a given review. From the extracted reviews, in order to extract opinion expressions, opinion mining was used. Opinion mining helped to extract categorical emotions such as positive, negative and neutral. A similar approach was used in this research project, wherein the outcomes look similar to the cited paper. Hirokawa et al. (2012) proposed a solution to extract hints and advice from reviews given by guests on their hotel stays, in order to improve the management of that

[2]https://www.statista.com/statistics/617136/digital-population-worldwide/

specific hotel by taking into consideration the featured words from the reviews. This study was solely based on graphical and visual representation. Similar representations is used in this project.Kitwatthanathawon et al. (2012) The fuzzy logic based method used for calculating the satisfaction level of tourists/guests yields an accuracy of 88.91%, at the same time the method of calculating the satisfaction level achieved 0.366 MAE and 0.464 RMSE. This project also uses the Ontology framework and the identified gap is that this project can only extract information using a sentence-by-sentence approach and cannot support complex sentences like *"Some of the rooms smelled like fire and we even had some bugs that joined us in the morning for a shower. Probably I got home with some kind of disease."*

Shi and Li (2011) proposed a supervised machine learning approach with the uni-gram feature to find the polarity of the reviews. The uni-gram feature was used with two types of information, namely Frequency and TF-IDF (term frequency–inverse document frequency). The results of this experiment show that TF-IDF was more effective than frequency, with TF-IDF having the precision of 85.2% and frequency having a precision of 84.5%. The TF-IDF feature is a weight often used in information retrieval and text mining. This weight is a statistical measure used to evaluate how important a word is to a document in a collection. The author believed SVM outperformed NaÏve Bayes on the basis of the related work that they studied as it was highly effective at categorizing traditional text. Recall, Precision and F-score were used as evaluation metrics for text categorization. The proposed future work is to implement Natural Language Processing (NLP) techniques to improve the performance of sentiment analysis , which will be incorporated in the present study. Zhang and Yu (2017) adopted a method to perform sentiment analysis of hotel reviews using Word vector clustering. K-means clustering algorithm and Word2Vec tool was implemented for this task. As K-means clustering was found to be sensitive to initial clusters, they chose Word2Vec and ISODATA clustering algorithm so as to improve the accuracy of text sentiment classification. As Bag-Of-Words and TF-IDF have high dimensionality and high sparseness of feature vector, the technique of Word2Vec was used and LinearSVC classifier was used for sentiment classification and it was found to outperform the above mentioned traditional methods. Finally, eXtreme Gradient Boosting (XGBoost) classifier was used to classify the reviews. XGBoost is faster and more accurate as compared to other boosting tree models. Evaluation metrics used were AUC and accuracy. on applying the combinational method of Word2Vec, ISODATA and XGBoost , the accuracy was observed to increase by 0.25% and ISODATA proved to avoid uncertainty of the results.

The approach used in Ghorpade and Ragha (2012) attempts to eradicate the drawbacks of all the other commonly used techniques, which is loss of text information. This approach will be used as a future work recommendation of this research project. Of the two approaches for sentiment analysis, the author used Semantic Orientation method. JAPE mathematical technique was used to pre-process various words in a given review/sentence. This, with machine learning gave the classification of such reviews as positive and negative. Future work suggests application of Support Vector Machine algorithm for classification.They also used the Ontology learning framework, which will be a task of future work in this research project. This state of art Chin (2009) used the Delphi Method to obtain customer opinions. The PZB is a model used for service quality concept , which is used to find out the reason why the service fails to meet the customer's expectation. The future recommendation suggests updating the equipments as per user expectations and getting an in-depth detailed information for the proprietors to enhance the service quality and to restrain the existing clients.This project uses Aspect Based Sentiment Analysis Omurca and Ekinci (2018) to the user generated reviews and combines it with implicit aspect extraction. A graph based Laplace smoothing is carried out to extract such implicit aspects from the reviews , and the outcomes of this technique are evaluated using F-measure. The evaluation value for F-measure was found to be 0.77 and yielded good results.

2.2 Most Frequently Used Approaches Towards This Topic

Saito and Klyuev (2019) Naïve Bayes is a Generative model used for classification. The author used Naïve Bayes to classify user reviews at sentence level and review level on the basis of positive and negative polarity. The author states that Naïve Bayes outperforms as compared to Support Vector Machine (SVM), K-Nearest Neighbour (KNN) as it uses naïve assumption of independency and gives better results for large size of vocabulary when the class and length of the document are assumed to be independent. This is one reason for choosing Naïve Bayes algorithm for classification. Recommended Future Work: The author proposes to combine opinion mining techniques with collaborative filtering to generate a user-oriented recommendation system, as his future work.Hirokawa and Hashimoto (2018) The author used machine learning approach with SVM and feature selection technique to predict the polarity of any given reviews. Sentiment analysis was done on the dataset having reviews and the prediction performance , that is the F-measure was obtained to be 74% for reviews with positive polarity , whereas only an average of 27% was obtained for negative reviews, which proves that such negative reviews are difficult and hard to predict,making it a gap in this research.Recommended future work : They have proposed to analyse the similar problem using Random Forest, Naïve Bayes and Decision Tree as a future work.

Suzuki et al. (2013) amended a technique of Natural Language Processing (NLP) on hotel reviews. This technique was used to analyse the word-of-mouth communication and to identify the co-occurrence of words. An automated NLP system was used for this purpose, which saved time and cost of gathering data , which was otherwise done manually. The results and outcomes were plotted on a three- axis grid on the basis of services provided by the respective hotels and a comparison and visualization was presented. They also used user ethnicity as a factor influencing the satisfaction ratio. Agarwal et al. (2018) implemented an automated system that could extract the aspect based terms from the textual feedback given by users and perform aspect based sentiment analysis to it. Word2Vec was used to get vector of synonyms of given words, so that they come under the same aspect as their delivered meaning. The vector of similar words was clustered together using K-means. To extract relative terms from a corpus, Stanford dependency parser was used, which returned Amod, Advmod and COP to indicate whether the words were positive, negative or neutral.Chanwisitkul et al. (2018) This study helps find the reason behind the given rating on Hotels by using the textual reviews. Text Mining is used as a technique to analyse and understand the reviews and Topic modelling is used to segregate aspects from the whole set of text. Text mining is used to find patterns and to explore relationships between the textual content of the reviewer's feedback. The gap in this study was homogeneity of the sample , restricting the author to only work on a limited small scale dataset of ten hotels.

Hadoop environment was used with WebCrawler in Python in Jian et al. (2017) to gather reviews form the internet. With the help of MapReduce in Hadoop, in a combination with jieba text segmentation, all the sentences/paragaphs were broken down into single words, as in aspect based modelling which was done in this research project. This paper Kongthon et al. (2010) proposed an approach to study and understand customers' opinion regarding hotels , using Opinion Mining. This approach will help the owners to focus on the necessary opinions and help improvise their service provision. The future work suggests combining data from multiple sources as currently it only used one source and to combine it with Naïve Bayes or Support Vector Machine.

2.3 A Crtique of Text Classification Models in Machine Learning

This paper was used to illustrate about the existing Machine learning Methods. It shows importance of Deep Learning and how it is used in Big Data and in Artificial Intelligence. Simon et al. (2015) states that deep learning techniques have been criticised as there is no way of representing casual relationships between the variables. Zhang and Li (2007) talks about a Naïve Bayes model built to classify spam emails. It yielded good results and was recommended by the author. Naïve Bayes is an easy to use algorithm and works well with binomial data as compared to numeric data. Luo et al. (2019) proposed two ways to enhance the efficiency and predictivity of Random Forest and Logistic Regression, while working with imbalanced data. Cost-sensitive Logistic Regression (CS-LR) and Cost-sensitive Random Forest (CS-RF) were the two methods that helped reduce the performance degradation of RF and LR while working on imbalanced datasets. Momin et al. (2017) made a web application that displays the reviews on basis of features essential to the differently-abled group. The portal was used to put forth the features on the basis of ranks , done using opinion mining. Naïve Bayes was used for classification of the reviews. The classified reviews were then extracted on the basis of their features and put forth on the portal to help the disabled group find the best hotel that fulfils all their disability needs. A portal like this can be put into implementation for the case study hotel being used.

Based on the above critiqued literature articles, this project intended to use a few of them as guidelines to performing the work. From Shi and Li (2011), it was decided to use Term Frequency - Inverse Document Frequency (TF-IDF) with topic modelling Chanwisitkul et al. (2018) as an experiment to see if it yields results and it did well, but did not give the exact desired outcome as the reviews could not be segregated into different topics. Saito and Klyuev (2019) was used to check Naïve Bayes' performance as stated and Kitwatthanathawon et al. (2012) is undertaken as a future work recommendation from this project.

3 Scientific Methodology, Data Pre-processing and Architectural design

3.1 Methodology

This project aims at collecting reviews from a source and processing its textual and numeric data of reviews and ratings , so as to get sentiments of users towards that hotel or a particular service provided by them. Since it involves a hospitality industry, it is considered to be a business oriented project and the most appropriate methodology to use to go about this project is the Cross-Industry Standard Process for Data Mining (CRISP-DM) method. Figure 1 shows the way CRISP DM methodology is used in this project. It follows a circular process flow and its processes is explained briefly below :

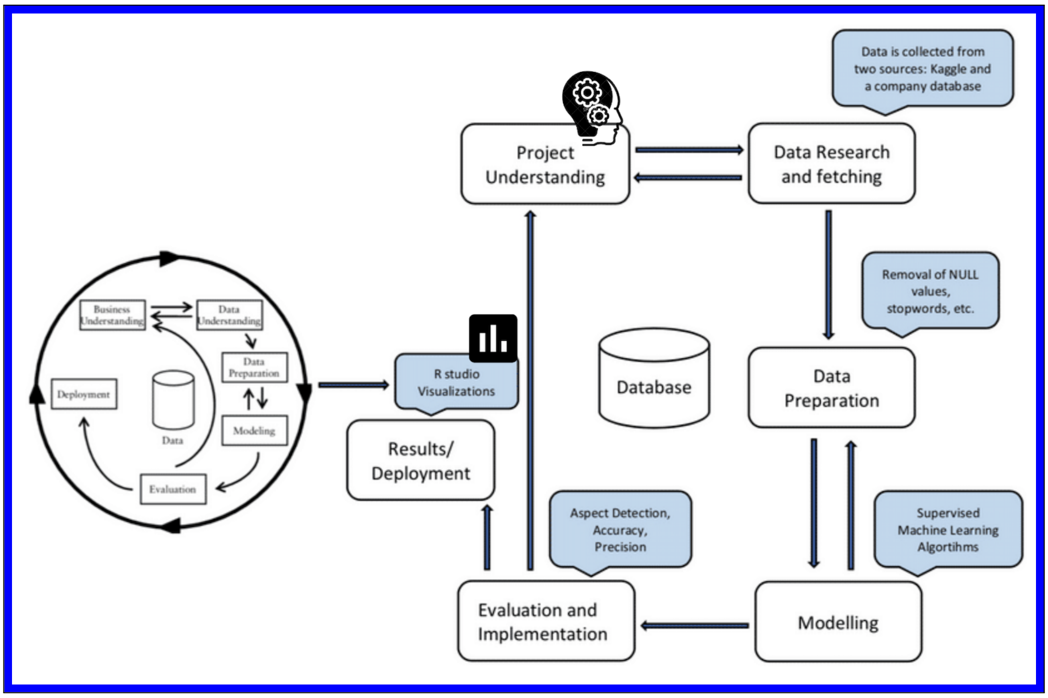

Figure 1: Methodology for Hotel Reviews Analysis

Project Understanding:

As the name suggests, the very first task of this project was to actually understand what the project was, so as to understand the requirements and need to do the project. As mentioned above, the task was to find aspects affecting the review score and to take preventive measures.

Data research and Gathering:

Initially, the data was tried to be extracted from tripadvisor, using webscrapping, but due to ethical guidelines, it failed. Later a similar data was obtained from kaggle where it was previously scrapped from `booking.com` and was publicly available to use. Another dataset for the case study was obtained from an Irish Hotel, by following all the ethical guidelines.

Training Dataset: This data has been scraped through booking.com and was obtained from Kaggle[3]. This dataset contains 515,000 customer reviews and scoring of 1493 luxury hotels across Europe. Meanwhile, the geographical location of hotels are also provided for further analysis. This data contains the following attributes like the hotel name, address, ratings, latitude and longitude and feedbacks in textual format.

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

Testing Dataset: This dataset was obtained from a hotel in Ireland using ethical methods. It is a typical review database that contains only text comments and numeric points for several hotel attributes given by customers, this dataset also contains similiar attributes as that in the test dataset.

Data Preparation:

After the data was acquired, data cleaning was performed in order to prepare the data for modelling and analysis. The data was brought in a standard form which the machine understands, such as binary values.

Data Preprocessing:

NULL values were removed and other unnecessary data like the stopwords, non-english characters and special characters were removed. Variables having categorical data were encoded using categorical encoding technique in R. Figure 2 shows the process flow of the project.

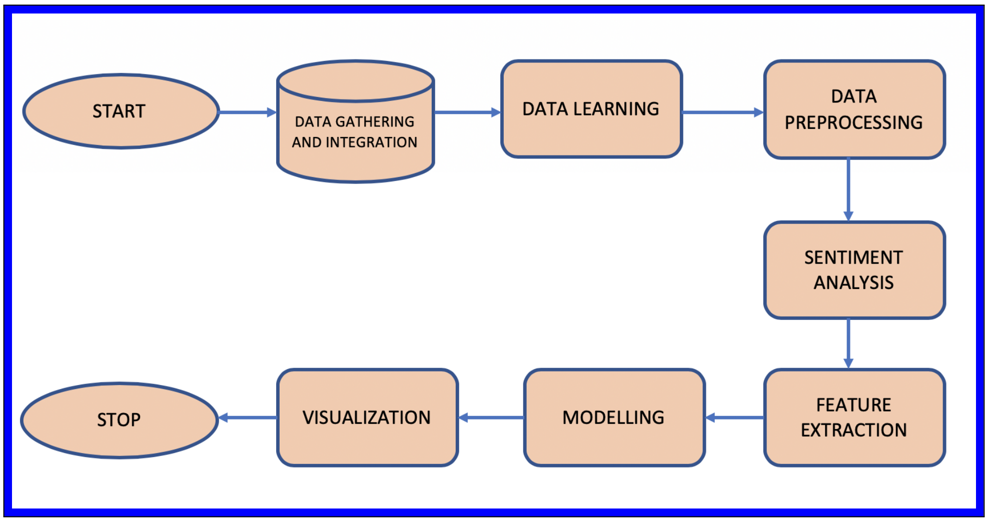

Figure 2: Process flow

Data Exploration:

Various ways to explore the dataset were used. One of it was plotting and locating the hotels on a world map using the 'leaflet' library. The visualization in Figure 4 shows the plotting of all the EU- hotels present in the dataset and Figure 3 shows the most frequently used words and Figure 5 shows the distribution of the dataset, which is balanced.

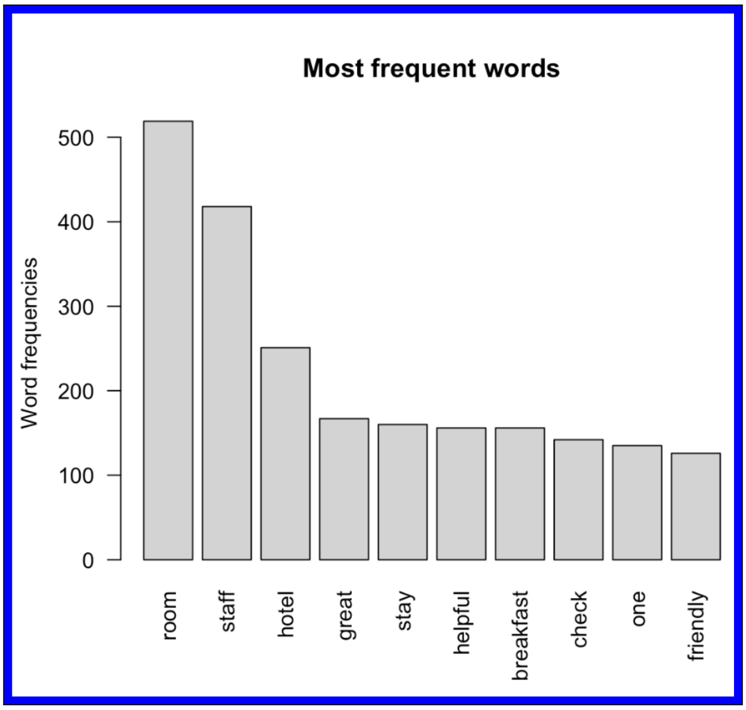

Figure 3: Frequently used words

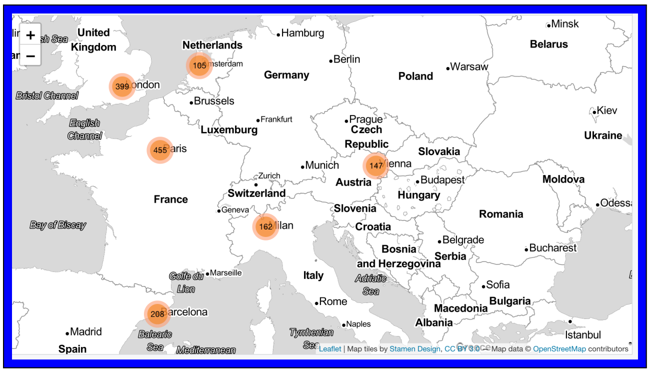

Figure 4: Hotels Location Plot

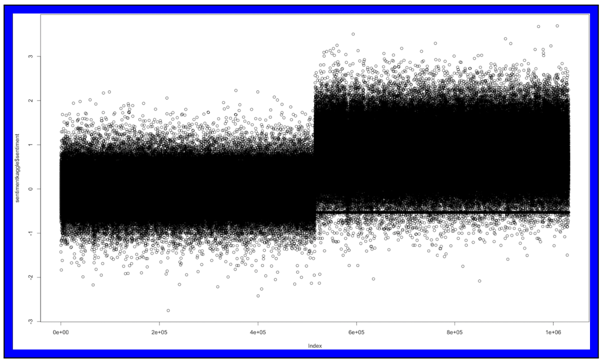

Figure 5: Representation of Balanced dataset

Modelling:

The process of modelling is explained as the act of training a ML algorithm to predict the labels(values) from the features , adjusting it for the business need and then validating and testing it on a heldout data [4]. Various Machine Learning models could be applied to this data , of which Random Forest , Decision Trees, Logistic Regression and Naïve Bayes were implemented. Logistic Regression is typically used when we have dichotomous(binary) data in the dependant variable. Naïve Bayes has proved to be an outperformer in various past researches and hence it was used to check the results. It is observed to work better than logistic regression and can work on a small data too, discussed in Section 4.3.

[4]https://towardsdatascience.com/modeling-teaching-a-machine-learning-algorithm-to-deliver-business-value-ad0205ca4c86

Implementation:

This is the final step of the project, where the models are implemented on the feature that is selected and the models are validated and tested based on Accuracy, Precision and Recall. This will be elaborated in Section 4.

Results and deployment:

The results of the above steps are then displayed and represented in a visualization. This will be discussed in Section 4.

3.2 Design Specification

This project followed a Two-Tier Architecture, Figure 6. A two-tier architecture is a software architecture in which a presentation layer or interface runs on a client, and a data layer[5]. The Figure 6 illustrates the way this project went through. The client in this architecture is the hotel whose dataset was used for performing the tests. The datasets are made ready and prepared for processing. The cleaning, visualizations and sorting was done using R studio and MS Excel VLOOKUP(). After the data was ready, it was sent to the models for training and testing. The outputs received on running the models were evaluated using Accuracy, Sensitivity and Specificity. These evaluation results were then finally represented and can be given to the client for implementation.

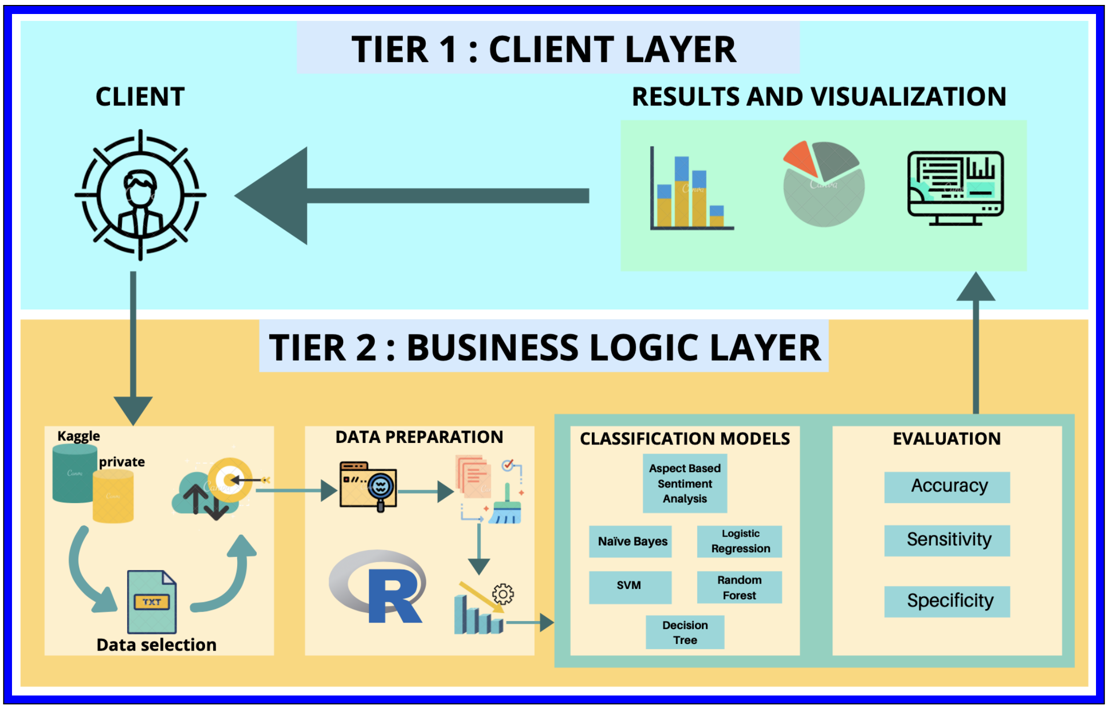

Figure 6: Two-Tier Architecture

[5]https://www.techopedia.com/definition/467/two-tier-architecture

4 Implementation, Evaluation and Results of Classification Models

Implementing the classification models was a major task of this machine learning based project. The Various Techniques used were, aspect based sentiment analysis Section 4.1, Naïve Bayes Classification Section 4.3, Random forest Section 4.4, Support Vector Machine(SVM) Section 4.5, Logistic Regression Section 4.6 and Decision Tree Section 4.7. On successful implementing these models, they were taken to the evaluation step, where they were check for their performance. Evaluation matrices are those that we use in checking how good a model performed.The formula for these evaluation matrices are given in equation 1, equation 2 and equation 3. The results of these Evaluation Matrices are presented in Table 6.

Accuracy:

Accuracy in that case is the percentage of correctly classified instances , which can be calculated using the formula,

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

Sensitivity/ Recall:

Sensitivity is what portion of actual positives was identified correctly. It was calculated using the formula,

$$Recall = \frac{TP}{(TP + FN)} \tag{2}$$

Specificity/ Precision:

Specificity is what rate of positive identifications were actually correct. It was calculated using the formula,

$$Precision = \frac{TP}{(TP + FP)} \tag{3}$$

where,

TP = True Positive , sentiments that are positive and are actually classified as 1

TN = True Negative , sentiments that are negative and are actually classified as 0

FP = False Positive , sentiments that are negative but are classified as 1

FN = False Negative , sentiments that are positive but are classified as 0

F1 Score:

To find an optimal blend of precision and recall, the two metrics are combined using what is called the F1 score. The F1 score is the harmonic mean of precision and recall. It is given by the formula :

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{4}$$

4.1 Aspect based Sentiment Analysis

In aspect-based sentiment analysis, every sentence was broken down into aspects and a sentiment analysis of those aspects was done, which gave decimal values to each word in the review. The sentiment score was calculated and assigned to every aspect, which helped in finding the most important positive and negative aspects. All the words in the dataset were represented visually using a **wordcloud** as shown in Figure 7

Figure 7: Representation of aspects in a Word Cloud

4.1.1 Feature extraction

After finding the important aspects based on the reviews, the feature was extracted, named Sentiment_Score, which lead to categorizing the sentiment type as positive or negative. The score lying between 0 to +5 was taken to be positive and that between 0 to -5 was taken as negative.

4.1.2 Implementation of results of Aspect based Sentiment Analysis

After finding the sentiment type and sentiment score of all reviews, they are incorporated with the existing numeric dataset having ratings and other variables, including the categorical variable that is the Sentiment_Type that was used as the dependent variable for all machine learning algorithms. The Table 1 shows a sample of the breakdown of a single review sentence into single relevant words.

Table 1: A small sample of extraction of words from a single review/sentence

	Hotel_Name	Reviewer_Score	word
1	Hotel Arena	2.9	angry
2	Hotel Arena	2.9	post
3	Hotel Arena	2.9	sites
4	Hotel Arena	2.9	planing
5	Hotel Arena	2.9	trips
6	Hotel Arena	2.9	mistake
7	Hotel Arena	2.9	booking

The Table 2 shows the summary of each word chosen as aspect, with its score and relevance.

Table 2: A small sample of words summarisation

	word	mean_rating	score	count_word
1	perfect	9.219381	3	87489
2	nice	8.487266	3	79340
3	clean	8.544312	2	75426
4	recommend	8.869814	2	32379
5	comfort	8.684292	2	23504
6	free	8.505985	1	23494
7	love	9.262678	3	23230

The Table 3 shows the hotel names with their equivalent calculated scores, which can be useful in plotting the hotel ranking.

Table 3: A small sample of hotel scores

	Hotel_Name	mean_rating	sentiment
1	11 Cadogan Gardens	8.723558	1.6875000
2	1K Hotel	7.662598	0.5826772
3	25hours Hotel beim MuseumsQuartier	8.956293	1.6162162
4	41	9.597714	1.5542857
5	45 Park Lane Dorchester Collection	9.410345	0.7586207
6	88 Studios	8.379763	1.1396450
7	9Hotel Republique	8.893831	1.5844156

The Table 4 shows the plotting of a small sample of sentiment type across the hotel name. This is done on distinct hotels and the overall sentiment score is converted into a category of positive or negative and is then allotted to the respective hotels and Figure 8 shows the plot of sentiment score across the hotel names, where the sentiment score above 1.2601 is termed as positive and below 1.2601 is termed as negative.

Table 4: Table representing sentiment type for each hotel

	Hotel_Name	Category
510	Hidden Hotel by Elegancia	1
511	Hilton Amsterdam	0
512	Hilton Barcelona	0
513	Hilton Diagonal Mar Barcelona	0
514	Hilton Garden Inn Milan North	0
515	Hilton Garden Inn Vienna South	0
516	Hilton London Angel Islington	1
517	Hilton London Bankside	1
518	Hilton London Canary Wharf	0

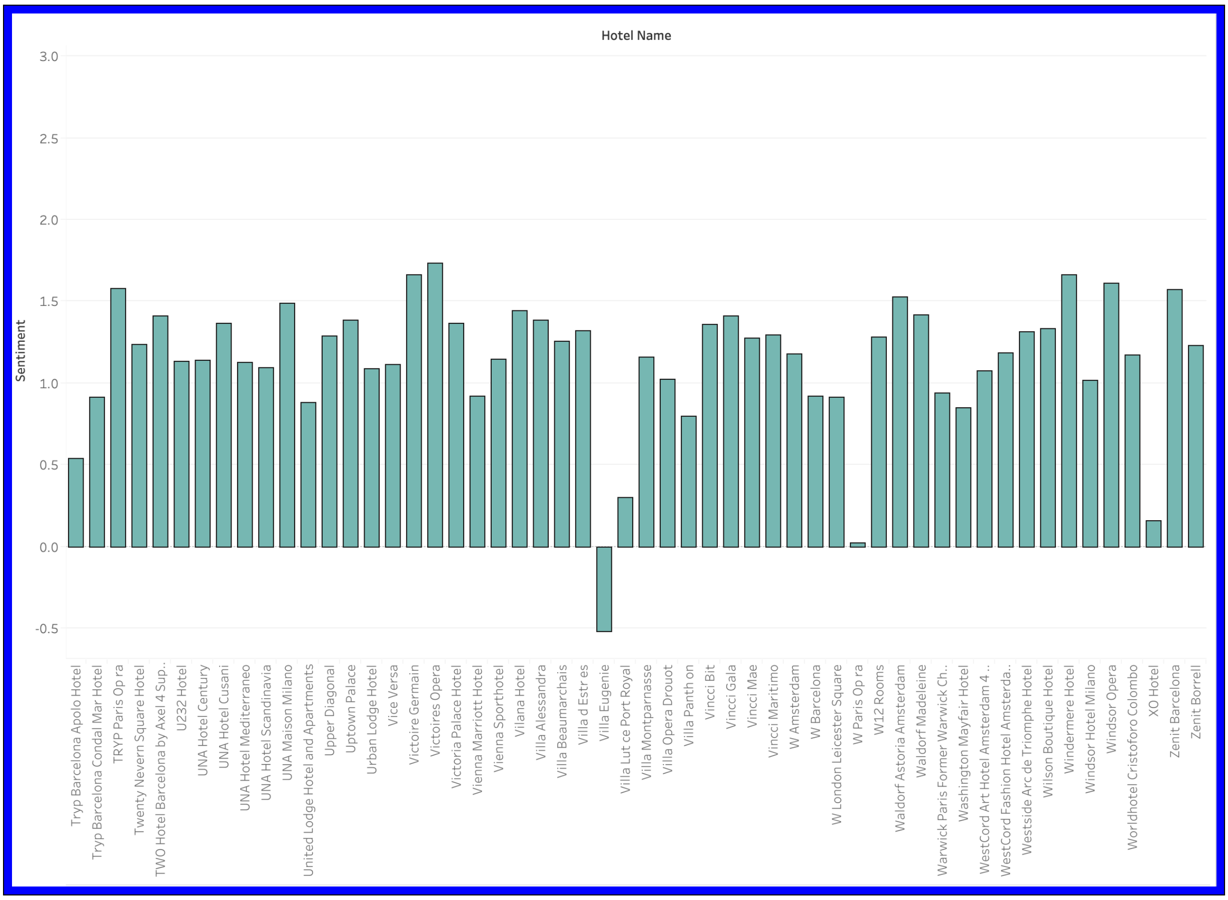

Figure 8: Plot of sentiment score across hotel names

4.2 Attempted work on Topic Modelling and TF-IDF

The author of Shi and Li (2011) suggested to work on hotel reviews using Topic Modelling with TF-IDF. This approach was attempted but could not give the desired outcome as this project could not successfully implement topic modelling on the dataset of mixed reviews. Table 5 shows the TF-IDF values across each word. Which shows how much a specific word is important to that specific document.

Table 5: Term Frequency-Inverse Document Frequency

```
# A tibble: 10,533 x 6
   aspect                                                                        word       n     tf   idf tf_idf
   <fct>                                                                         <chr>  <int>  <dbl> <dbl>  <dbl>
 1 "This was the absolute worst hotel experience of my life.  Upon check i… bed       15 0.0221 2.01  0.0444
 2 "This was the absolute worst hotel experience of my life.  Upon check i… ready     15 0.0221 2.29  0.0504
 3 "This was the absolute worst hotel experience of my life.  Upon check i… family    12 0.0176 3.25  0.0574
 4 "This was the absolute worst hotel experience of my life.  Upon check i… son       12 0.0176 4.27  0.0753
 5 I have no problem with the room   it was lovely. I requested a room ove… staff     11 0.0917 0.460 0.0421
 6 I didnt receive any benefits Staff on reception were both trainees and … staff     11 0.0764 0.460 0.0351
 7 "Great customer service. Very friendly and helpful staff. Front desk st… staff     10 0.0714 0.460 0.0328
 8 "This was the absolute worst hotel experience of my life.  Upon check i… blank…     9 0.0132 5.65  0.0748
 9 "This was the absolute worst hotel experience of my life.  Upon check i… broug…     9 0.0132 3.35  0.0443
10 "This was the absolute worst hotel experience of my life.  Upon check i… call       9 0.0132 3.01  0.0399
# … with 10,523 more rows
```

4.3 Implementation, Evaluation and Results of Naïve Bayes Classifier

Naïve Bayes Classifier was trained on the hotel review dataset as suggested by Hirokawa and Hashimoto (2018) and Saito and Klyuev (2019).Naïve Bayes classifier is much faster than any other discriminative models and can work very well on a small amount of data. On implementing and evaluating this model, it was found to have an Accuracy of 0.845% which is good enough for a new dataset like this one. The ROC curve is shown in Figure 9 and Figure 10 shows the evaluation results of the model.

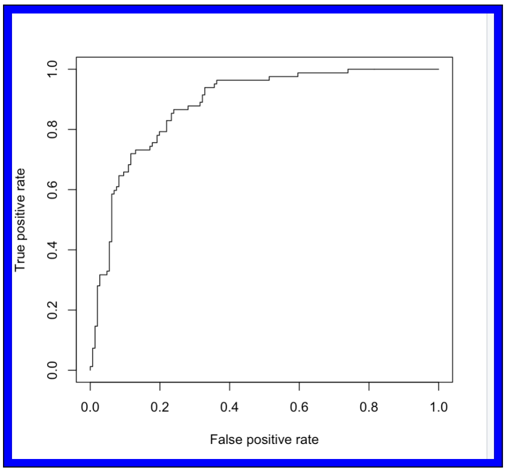

Figure 9: ROC curve

```
> #checking accuracy, precison, recall
> accuracy <- sum(diag(mat$table)) / sum(mat$table)
> accuracy    #0.8455446
[1] 0.8455446
> precision <- diag(mat$table) / rowSums(mat$table)
> precision
        0         1 
0.9035088 0.7578773 
> recall <- diag(mat$table) / colSums(mat$table)
> recall
        0         1 
0.8494845 0.8385321 
> #sensitivity, Specificity, F1 score
> sensitivity(tab1_n)
[1] 0.8459684
> specificity(tab1_n)
[1] 0.8499342
> posPredValue(tab1_n)
[1] 0.9022857
> negPredValue(tab1_n)
[1] 0.7710987
```

Figure 10: Naïve Bayes Evaluation Matrices

4.4 Implementation, Evaluation and Results of Random Forest

Random Forest Algorithm was also used on the training dataset with the recommendation of Hirokawa and Hashimoto (2018). This model has the ability to restrict overfitting without increasing the error, which makes it a very powerful model. Figure 11 represents the data distribution used in the model. The evaluation outcomes showed an accuracy of 0.839.

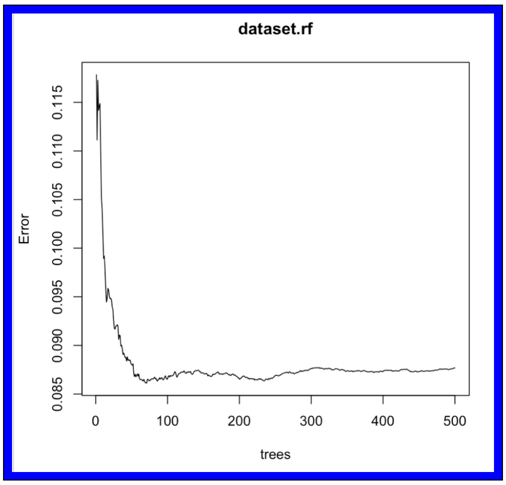

Figure 11: Representation of dataset in random forest

4.5 Implementation, Evaluation and Results of Support Vector Machine (SVM)

Support Vector Machine was implemented with reference to Shi and Li (2011) as it was said to outperform Naïve Bayes in that research. Support Vector Machine (SVM) is very useful when there is no knowledge about the data and can even work on unstructured or semi-structured data like text. Figure 12 shows the results of implementation , that yields a moderate accuracy of 0.813.

```
> caret::confusionMatrix(test$Sentiment_Type,pred_test )
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1024  152
         1  200  506

               Accuracy : 0.813
                 95% CI : (0.7946, 0.8303)
    No Information Rate : 0.6504
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.5956

 Mcnemar's Test P-Value : 0.01224

            Sensitivity : 0.8366
            Specificity : 0.7690
         Pos Pred Value : 0.8707
         Neg Pred Value : 0.7167
             Prevalence : 0.6504
         Detection Rate : 0.5441
   Detection Prevalence : 0.6249
      Balanced Accuracy : 0.8028

       'Positive' Class : 0
```

Figure 12: SVM output

4.6 Implementation, Evaluation and Results of Logistic Regression

Logistic Regression works best with binomial data, which is a perfect fit for this project as its main attribute itself is a categorical variable have two categories. Logistic Regression was also recommended as a future work in Hirokawa and Hashimoto (2018). Figure 13 represents the classification of the predictor variable , containing "positive" and "negative" values. This model gave the accuracy of only 0.780.

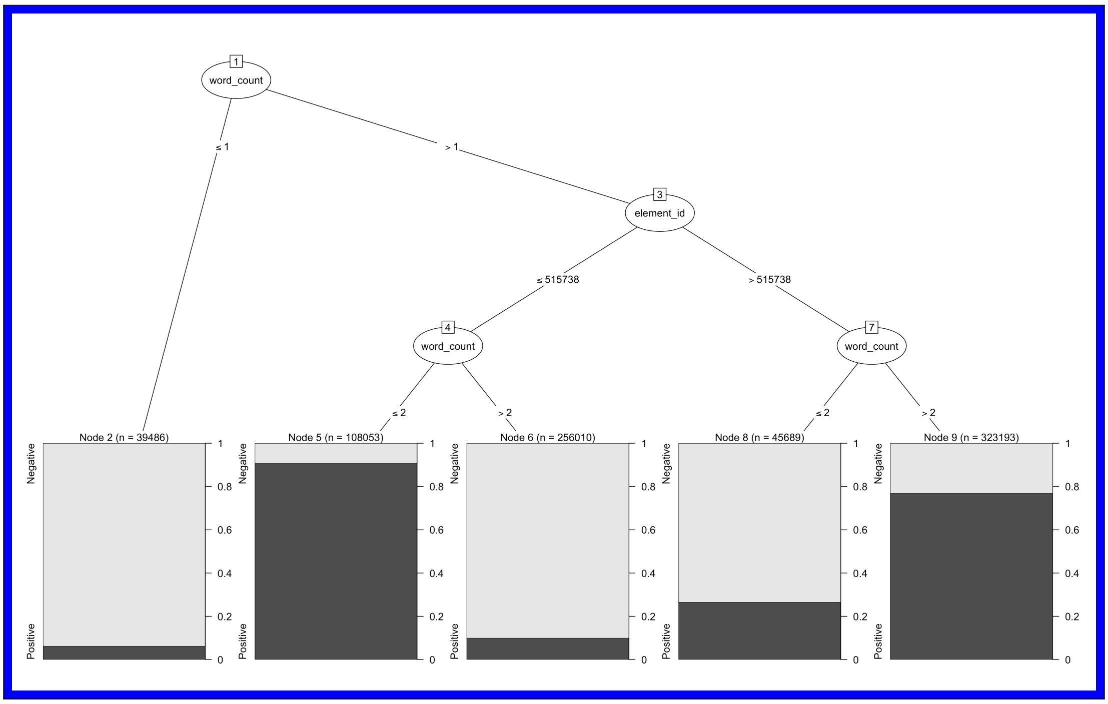

Figure 13: Logistic Regression

4.7 Implementation of Decision Tree

Decision Tree was implemented as it provides an effective method of decision making as they help in clearly laying out the problem so that it is easy to see all the options[6]. The Decision Tree is shown in Figure 14 shows the generated decision tree.

[6]https://www.mindtools.com/dectree.html

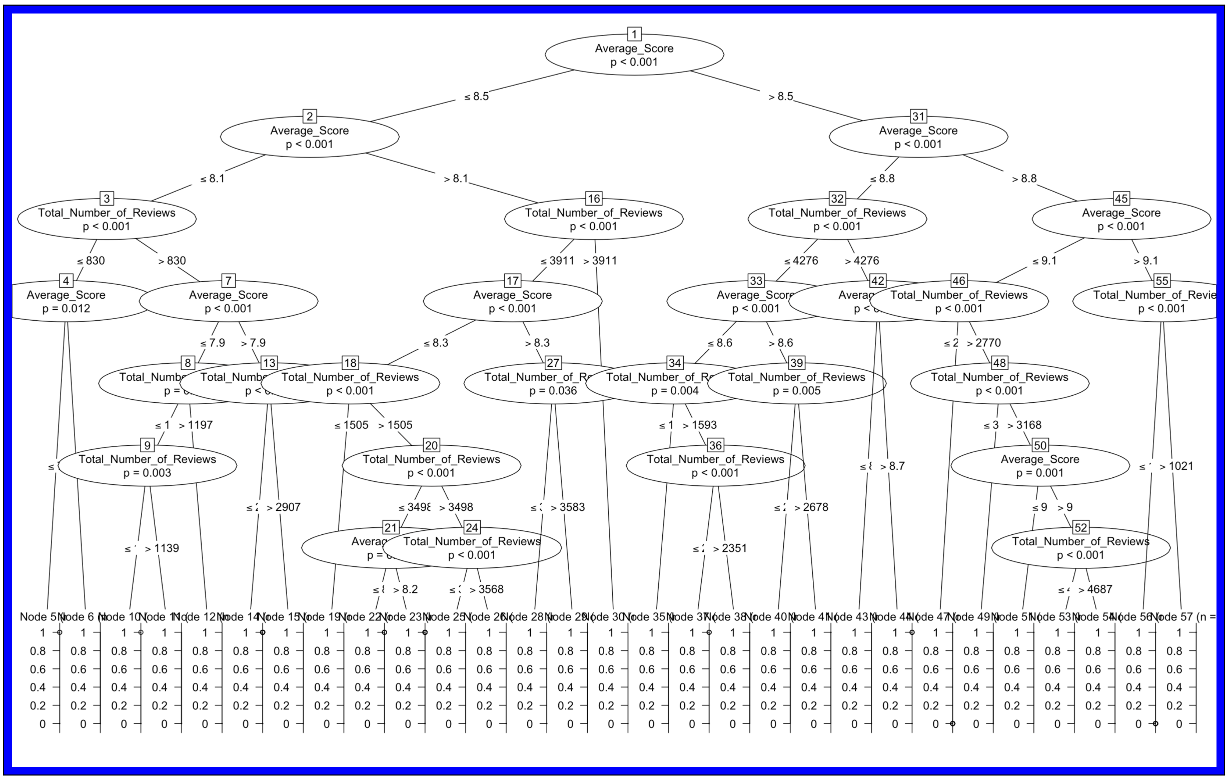

Figure 14: Decision Tree

Results

Table 6 displays the values for Accuracy, Sensitivity(Recall) and Specificity(Precision) and the F1 Score for the applied Models. As it is clearly visible, Naïve Bayes yielded the best performance results with an accuracy of 0.845.

Table 6: Evaluation Outcomes

	Accuracy	**Recall**	**Precision**	**F1 Score**
Naïve Bayes	0.845	0.845	0.849	0.846
Random Forest	0.839	0.902	0.735	0.809
Support Vector Machine	0.813	0.836	0.769	0.801
Logistic Regression	0.780	0.732	0.865	0.792

Discussion

As seen in Table 6, Naïve Bayes outperforms the other models by an accuracy of 84.5% and Random Forest being the next outperforming model with an accuracy of 83.9% . The recall and precision were found to be 0.845 and 0.849 for Naïve Bayes and 0.902 and 0.735 respectively for Random Forest. From the above experimental results, it is safe to say that Text Classification in Sentiment Analysis works best with Naïve Bayes Classifier and Random Forest in this research project.

5 Conclusion and Future Work

The Objectives of this project were met successfully and the Research question stating "To what extent can aspect based sentiment analysis of European hotels' review using Natural Language Processing (NLP) approach and machine learning methodologies (Support Vector Machine (SVM), Logistic Regression, Naïve Bayes, Random Forest, Decision Tree) be used to predict customer satisfaction to improve the service delivery?" has found a solution to it. Using the technique of Sentiment analysis under NLP, it was possible to find out the sentiments of people towards the hotel they mentioned. As seen in Table 6, Naïve Bayes outperforms the other models by an accuracy of 84.5% and Random Forest being the next outperforming model with an accuracy of 83.9% . The recall and precision were found to be 0.845 and 0.849 for Naïve Bayes and 0.902 and 0.735 respectively for Random Forest. From the above experimental results, it is safe to say that Text Classification in Sentiment Analysis works best with Naïve Bayes Classifier and Random Forest in this research project.

The identified gap which is recommended as future work that can leverage this project is the successful implementation of topic modelling on mixed reviews, which can be combined with uni-grams.

Acknowledgement

The accomplishment of this Research Project could only be possible under the guidance and tremendous support of Dr. Catherine Mulwa . She had a key role to play in getting me to the completion of this thesis. I would also like to pay gratitude to my family and loved ones for all the financial and moral support needed to get through my Masters. A special thanks to my fellow colleagues, had it not been for them, I would not know what I know today.

References

Agarwal, V., Aher, P. and Sawant, V. (2018). Automated aspect extraction and aspect oriented sentiment analysis on hotel review datasets, *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, pp. 1–4.

Chanwisitkul, P., Shahgholian, A. and Mehandjiev, N. (2018). The reason behind the rating: Text mining of online hotel reviews, *2018 IEEE 20th Conference on Business Informatics (CBI)*, Vol. 01, pp. 149–157.

Chauhan, S. K., Goel, A., Goel, P., Chauhan, A. and Gurve, M. K. (2017). Research on product review analysis and spam review detection, *2017 4th International Conference on Signal Processing and Integrated Networks (SPIN)*, pp. 390–393.

Chin, J. (2009). Customer requirements assessment model of fitness club in international tourist hotel, *2009 International Conference on Management and Service Science*, pp. 1–4.

Dundar, B., Ozdemir, S. and Akay, D. (2016). Opinion mining and fuzzy quantification in hotel reviews, *2016 International Symposium on Networks, Computers and Communications (ISNCC)*, pp. 1–4.

Ghorpade, T. and Ragha, L. (2012). Featured based sentiment classification for hotel reviews using nlp and bayesian classification, *2012 International Conference on Communication, Information Computing Technology (ICCICT)*, pp. 1–5.

Hirokawa, S. and Hashimoto, K. (2018). Simplicity of positive reviews and diversity of negative reviews in hotel reputation, *2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)*, pp. 1–6.

Hirokawa, S., Okada, M. and Hashimoto, K. (2012). Extraction of hints and advice from hotel reviews for improving small hotel management, *2012 IEEE 14th International Conference on Commerce and Enterprise Computing*, pp. 166–170.

Jian, M., Fang, Y., Wang, Y. and Cheng, C. (2017). Big data analysis in hotel customer response and evaluation based on cloud, *2017 19th International Conference on Advanced Communication Technology (ICACT)*, pp. 791–795.

Kitwatthanathawon, P., Angskun, T. and Angskun, J. (2012). A knowledge extraction system from online reviews using fuzzy logic, *2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)*, pp. 189–196.

Kongthon, A., Angkawattanawit, N., Sangkeettrakarn, C., Palingoon, P. and Haruechaiyasak, C. (2010). Using an opinion mining approach to exploit web content in order to improve customer relationship management, pp. 1–6.

Luo, H., Pan, X., Wang, Q., Ye, S. and Qian, Y. (2019). Logistic regression and random forest for effective imbalanced classification, *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, Vol. 1, pp. 916–917.

Momin, H. G., Kondhawale, C. S., Shaikh, R. E. and Gawandhe, K. G. (2017). Feature based evaluation of hotel reviews and ratings for differently abled people, *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 2452–2454.

Omurca, S. and Ekinci, E. (2018). Using adjusted laplace smoothing to extract implicit aspects from turkish hotel reviews, *2018 Innovations in Intelligent Systems and Applications (INISTA)*, pp. 1–6.

Saito, Y. and Klyuev, V. (2019). Classifying user reviews at sentence and review levels utilizing naïve bayes, *2019 21st International Conference on Advanced Communication Technology (ICACT)*, pp. 681–685.

Shi, H. and Li, X. (2011). A sentiment analysis model for hotel reviews based on supervised learning, *2011 International Conference on Machine Learning and Cybernetics*, Vol. 3, pp. 950–954.

Sigala, M. (2009). Web 2.0, social marketing strategies and distribution channels for city destinations: Enhancing the participatory role of travelers and exploiting their collective intelligence, *Information Communication Technologies and City Marketing: Digital Opportunities for Cities Around the World* pp. 221–245.

Simon, A., Deo, M. S., S.Venkateshan and Babu, D. (2015). An overview of machine learning and its applications, **1**: 22–24.

Suzuki, T., Gemba, K. and Aoyama, A. (2013). Hotel classification visualization using natural language processing of user reviews, *2013 IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 892–895.

Zhang, H. and Li, D. (2007). Naïve bayes text classifier, *2007 IEEE International Conference on Granular Computing (GRC 2007)*, pp. 708–708.

Zhang, X. and Yu, Q. (2017). Hotel reviews sentiment analysis based on word vector clustering, *2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA)*, pp. 260–264.