BT5153 Final Report Group: What's Wrong With The Britannia?
A Text Analysis Case Study of Hotel Britannia Canary Wharf

Group 2

Chen Xianxi	A0079721Y	Nguyen Huy Anh	A0088386L
Ha My Hoa	A0218980W	Pio Lorenzo Peralejo Valdez	A0218839N
Kenneth Tay Hao Quan	A0139786J		

GitHub link: https://github.com/group2BT5153/Group2_BT5153

Executive Summary

In this project, we leveraged upon text analytics techniques to extract business insights and make recommendations to our target hotel, Hotel Britannia Canary Wharf (HBCW), to improve upon customer experience. First, we made use of word cloud visualizations to help the hotel immediately identify the common positive and negative guest feedback based on historical text reviews. Further insights were extracted by analyzing word clouds generated based on time, whether guests were locals or foreigners, purpose of stay, etc. as well as based on competitor hotel guest reviews. Arising from these word cloud analyses, we identified several key areas for improvement as well as identified the hotel's strengths and weaknesses relative to competitor hotels in the vicinity. We then made specific and actionable recommendations to the hotel. Next, to support similar analyses and continuous improvements in the future, we built a text review classification model that achieved an accuracy of 83.09% in classifying review texts into positive and negative classes. Finally, we built a business dashboard which takes in the reviews which have been labeled by our model to facilitate and expedite deep dives by the hotel management team into specific common complaint topics by the hotel guests. This will allow the hotel management team to make quick decisions and respond decisively to customer feedback.

Our efforts are expected to improve the productivity of the hotel while at the same time aid the hotel in achieving enhanced customer experience. Lastly, the tools developed in this project can be applied by any hotels seeking to improve upon customer experience, as long as they have the necessary text data, regardless of whether the data has been labeled or not.

1. Background

Hotel guest reviews are a treasure trove of information for both the consumer and the hotel management. Coupled with the ubiquity of the Internet, it is a powerful means of allowing consumers from all over the world to understand the performance of the hotel without having experienced the hotel first-hand. While guest reviews are a valuable source of information for consumers looking for the "ground-truth" on the performance of a hotel, it is also a rich source of information for the hotel in understanding the expectations of its guests and its performance in meeting those expectations. However, given the potentially countless reviews available, it may not be practical for the hotel management to go through every review to extract insights, especially when the reviews are in the form of unstructured or semi-structured text data. Yet, the advent of data science has provided us with the relevant tools to extract insights from guest reviews to identify common guest feedback, trends in terms of guests' expectations and the hotel's performance as well as areas of improvement.

In this project, we applied text analytics techniques to extract business insights and make recommendations to our target hotel, HBCW, to improve upon customer experience. This hotel was chosen mainly because its reviews (i.e. its guest rating scores on travel review website, Booking.com) were mediocre and therefore, the hotel had significant room for improvement. Other key considerations include the fact that a significant amount of data (i.e. guest reviews) was readily available and there were competitor hotels in the vicinity, which makes for a good case for us to study their relative strengths and weaknesses.

2. Hotel Britannia Canary Wharf (HBCW)

Hotel Britannia Canary Wharf is a modern-day hotel located at London's central business district, which provides convenient access to shopping outlets and iconic tourist attractions, such as Buckingham Palace, London Eye, and the Big Ben tower. The hotel is under the umbrella of Britannia Hotels, a British hotel chain that specializes in low-cost hotel services (Britannia Hotels, n.d.).

To complement our analysis of the strengths and weaknesses of HBCW, we also analyzed the reviews of competitor hotels, namely Hilton and Novotel London Canary Wharf hotels. These hotels are more favorably reviewed and in close proximity to each other at Canary Wharf with approximately 0.3 miles or a 6-minute walking distance between each other, as seen in Figure 1 below.

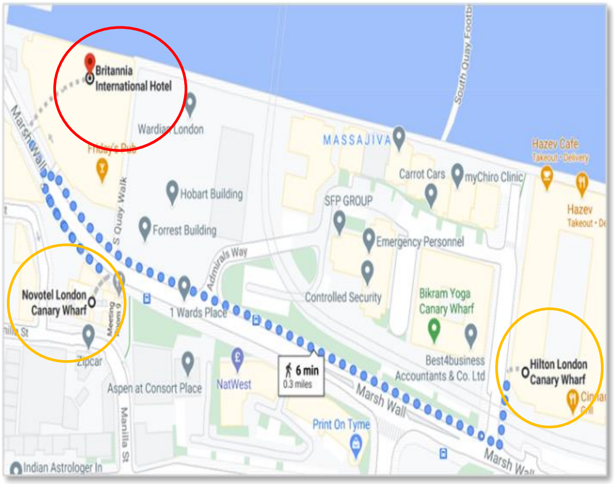

Figure 1. Hotel Locations at Canary Wharf

Based on our research, as of December 2020, Britannia Hotels has been infamously labelled as *"UK's worst hotel chain"* for the past 8 years. This has been primarily due to substandard hygiene practices, improper maintenance, and inept customer service (Sky News, 2020). HBCW therefore appears to be in a serious need of assistance to improve its customer ratings.

3. Project Objectives

We aim to provide HBCW with:

a) specific insights on the areas to be improved to enhance customer experience;
b) a text analytics tool to aggregate and generate insights from future customer reviews; and
c) a business dashboard to support deep dives into problematic areas.

3.1 Project Approach

To achieve objective 3. a), we will evaluate:

- the common feedback provided by guests of HBCW;
- how the feedback has evolved over the years from 2015 to 2020;
- how the feedback varied across customer segments such as guest nationalities (i.e. locals vs foreigners) and purpose of stay; and
- the common feedback provided by guests of competitor hotels in the vicinity (i.e. hotels of the same "star rating" located along the same road and within 300 m of the target hotel).

To achieve objective 3. b), we will build a review text classification model by consolidating guest reviews from the Booking.com and TripAdvisor websites which will be used for model training/validation and testing, respectively.

Objective 3. c) will be achieved by taking in the results of the review text classification model and transforming the results into a user-friendly and interactive business dashboard for quick decision-making.

3.2 Business Value

Our work is expected to bring business value to HBCW as it will help the hotel better understand which are the areas it needs to work on. By evaluating common customer feedback, feedback based on time, whether the customers are locals or foreigners, etc. as well as reviews for competitor hotels, we will provide HBCW with clear directions towards providing an enhanced experience for their customers.

In addition, our review text classification model can help the hotel automatically classify future text reviews from any sources. This will support further review analyses and continuous improvements in the future as well as feed labeled text data into the business dashboard for quick decision-making.

4. Data

4.1 Data Sources

Our project started out with a dataset retrieved from Kaggle, which was in fact extracted from the Booking.com website by a Kaggle member (Liu, 2017). The Kaggle dataset's time frame was from 2015 to 2017.

To collect more data, we proceeded to scrape data from Booking.com using Selenium. The scraped data from Booking.com time frame was from July 2018 until February 2021; this is due to the fact that Booking.com archived all of the reviews so that viewers can only go as far as 36 months back.

Furthermore, we also scraped more reviews from TripAdvisor using Octoparse to increase our data robustness. The details are further illustrated in Figure 2.

	Labelled	Unlabelled
Source	B.	Tripadvisor
Obtained by	kaggle Se	Octoparse
Number of samples	Apr 2015 to Feb 2021* 18,933 rows	Jun 2015 to Sep 2020 1,650 rows
Wordclouds	Yes	No
Classification models	Yes (train-validation set)	Yes (test set)

* Reviews between Aug 2017 and Jul 2018 were not available for our analysis

Figure 2. Table of data sources

4.2 Data Pre-Processing

For all labeled and unlabeled data, the following pre-processing steps were performed:

- **Language Detection:** This was implemented using the "langdetect" package to filter out non-English reviews.
- **Text Cleaning:** Before standardizing text content for modelling, we first filtered out meaningless content such as "There are no comments available for this review", "This review is hidden because it doesn't meet our guidelines", etc. Subsequentially, the following cleaning steps were applied on all rows: convert all alphabets to lowercase, remove numbers, remove special characters, expand contractions, remove punctuations, remove whitespaces and lemmatization.
- **Tokenization:** Where applicable, tokenization was performed to support model building.

Table 1. Data pre-processing steps for word cloud generation and modeling

	Language Detection	Customized Cleaning	Tokenization
Word Clouds	**No**	**Yes**	**No**
Modelling	**Yes**	**Yes**	**Yes**

- **Relabeling of text reviews**: Since data retrieved from the TripAdvisor website was not classified into positive and negative classes, relabeling was required.

The reviews obtained from TripAdvisor was not labelled as positive or negative, but instead were rated based on a 5-tier system. The 5 tiers comprised of Excellent (5 stars), Very Good (4 stars), Average (3 stars), Poor (2 stars) and Terrible (1 star). We decided to classify Excellent and Very Good reviews as positive reviews and reviews that were average and below as negative. A more conservative approach was taken by deciding to classify Average reviews as negative because there are several instances of Average reviews still containing negative sentiments. An example of such a case is shown in Figure 3 below.

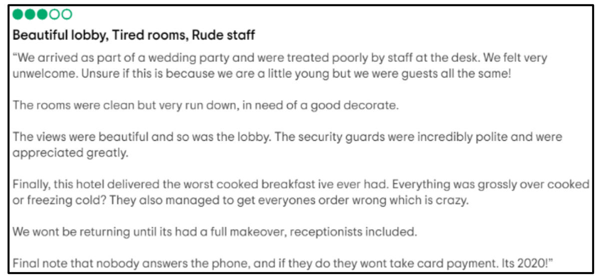

Figure 3. An Average review of the HBCW from TripAdvisor

For reviews from Booking.com, as seen from Figure 4 below, visitors to HBCW can leave both positive and negative comments in a single submission. Thus, a simple additional step was taken to separate the reviews into individual rows and to label them.

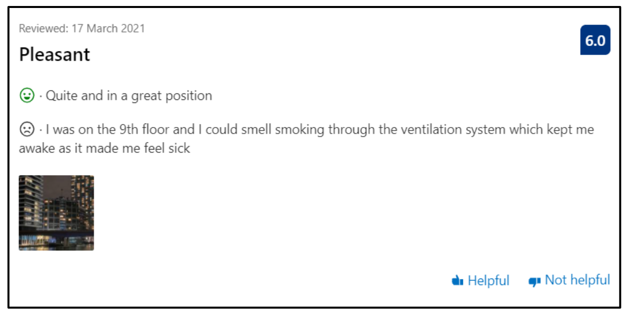

Figure 4. A review of the HBCW from Booking.com

4.3 Exploratory Data Analysis

Before deep diving into the model section, we need to have a basic understanding of our dataset. Hence, several EDA steps were performed.

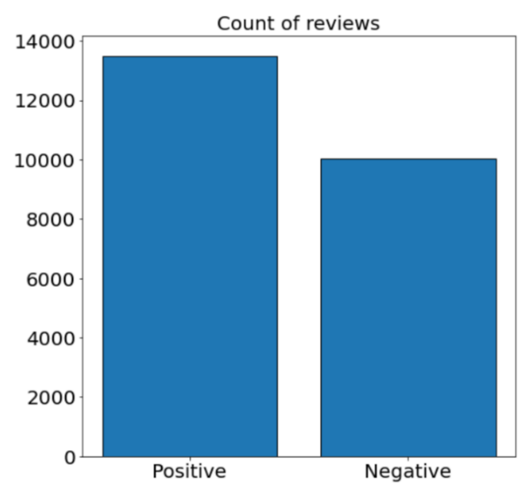

Figure 5. Count of Positive and Negative Reviews from Booking.com (training data)

First, we wanted to check whether our data is imbalanced. The count of positive and negative reviews have been illustrated in Figure 5. Upon a closer inspection, we observed that there are 13,501 and 10,023 positive and negative reviews, respectively (a ratio of 1.35:1). Even though the number of positive reviews is slightly more than negative reviews, we can see that the dataset is quite balanced. Therefore, no oversampling or undersampling steps were taken.

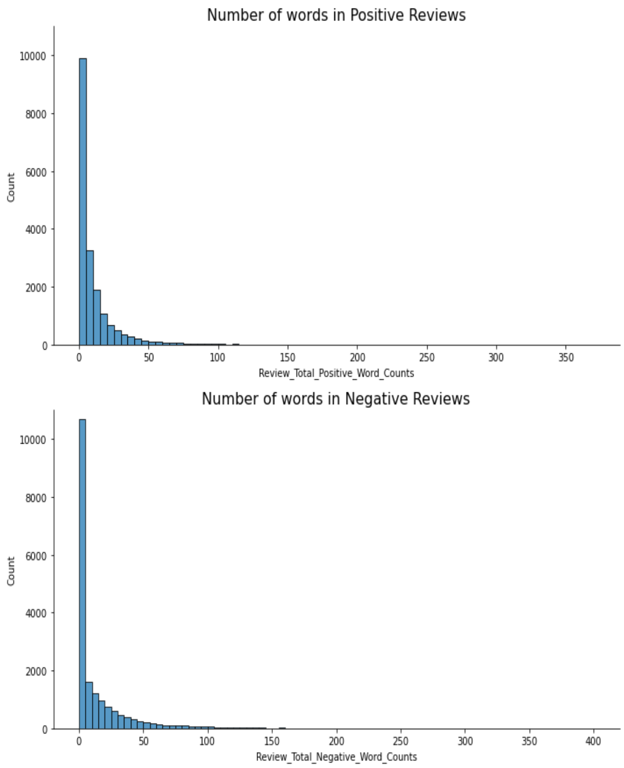

Figure 6. Histograms of the number of words in the positive and negative reviews

Secondly, we calculated number of words in each positive and negative reviews. From Figure 6, we can observe a similar distribution from negative reviews' word count and the positive reviews' word count. As a result, we assessed that engineering a word count feature in modelling is unlikely to yield significant performance improvement.

Another interesting point discovered during EDA (as shown in Figure 7) is that most of HBCW's customers were from the United Kingdom, which can imply that either a large portion of HBCW's customers were indeed from the United Kingdom or that guests from the United Kingdom are more vocal and tend to post online reviews than guests of other nationalities.

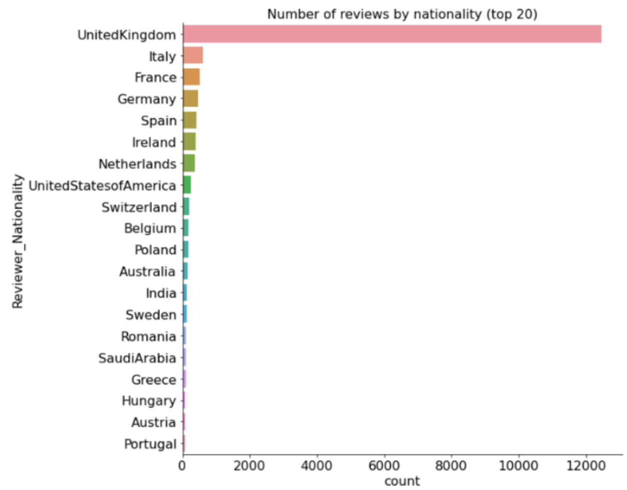

Figure 7. Number of reviews by nationality

5. Word Clouds

In order to quickly generate insights from the reviews we had at hand, we had chosen to use word clouds. Word cloud is a data visualization tool used to represent the frequency of words in a document. The more prominent a word appears in a word cloud, the more common it is, implying greater importance. It is a great tool to help hotel management team, who are typically non-technical oriented, quickly identify key strengths and weakness of HBCW.

We generated the word clouds based on the top 5 most common words for each review. This would help to give a more general picture which focuses on the topics being discussed and avoid the word clouds being skewed towards or dominated by any particular review (which may be heavily biased). The word clouds were created using Andreas Mueller's word cloud library in Python. To be able to look deeper into the word clouds, we also utilized NLTK universal parts-of-speech (POS) tag set to perform POS tagging on the positive and negative reviews corpora. Furthermore, as there were several words which always become prominent despite their inability to offer meaningful insights including “hotel”, “room”, “positive” and “negative”, we decided to drop them from the word cloud generation process.

5.1 General Reviews

Figure 8 shows the general word clouds for positive and negative feedbacks HBCW received.

Figure 8. Word clouds for positive and negative reviews

From the figure above, we can observe that “location” is magnified in the positive word cloud, which suggests that its location is perceived as its key strength. “Staff” were magnified in both word clouds, which could imply that different staff may be delivering varying levels of customer service. This is further reinforced as “clean” and “dirty” also appeared in the positive and negative word clouds, respectively. Key areas for improvement include “breakfast”, “window” and “bed”, which require hotel management’s attention as customers were not satisfied with these items.

5.2 Reviews Over the Years

We looked at the word clouds for each year from 2015 to 2020. 2021 was excluded because there were only 22 reviews for the year at the time of this study. While the positive word clouds for nouns remained relatively unchanged, the word clouds for negative nouns suggested a significant improvement for “breakfast” after 2017, since “breakfast” was no longer magnified after 2017. On the other hand, “staff”, “window”, “bed” and “wifi” continued to be magnified after 2017.

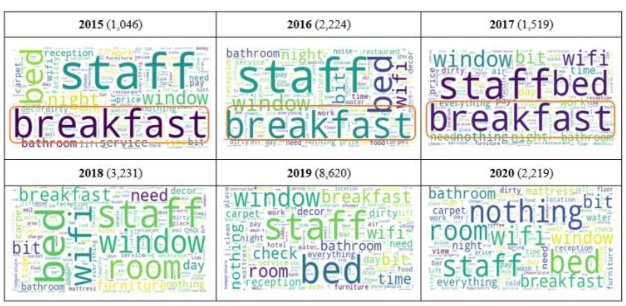

Figure 9. Word clouds for negative nouns 2015 – 2020
**number in brackets indicates number of reviews (inc empty)*

5.3 Reviews by Customer Segments

Next, we looked at how reviews differed for different customer segments. This could be helpful for the hotel if it wants to devise strategies or promotional campaigns targeting specific customer group.

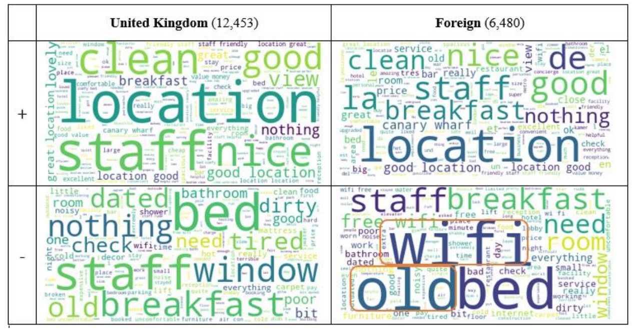

Figure 10. Word clouds for UK vs foreign guests
**number in brackets indicates number of reviews (inc empty)*

From Figure 10, we can interpret that Wi-Fi could be a more important factor in influencing hotel choice for foreign guests, likely because data roaming mobile plans may be costly for them. In addition, foreign guests also appear to perceive the hotel as older and more dated compared to domestic ones. It was also noted that the number of reviews for domestic guests almost doubled that of foreign guests. However, it is not sufficient to conclude that the hotel has much more domestic guests, because it could be due to foreign guests are less likely to leave reviews.

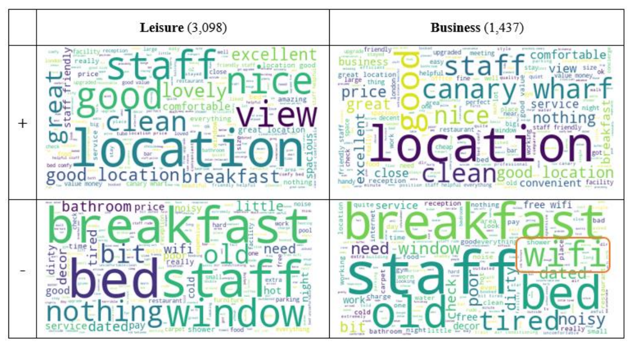

Figure 11. Word clouds for leisure vs business guests
**number in brackets indicates number of reviews (inc empty)*

Similarly, Wi-Fi seems to be an important factor for business guests. If a large portion of business guests are regular business travelers, it could be critical for the hotel to improve its Wi-Fi or it might lose these returning guests to its competitors.

5.4 Reviews In Comparison with Competitors

Finally, we examined HCBW’s relative strengths and weaknesses in comparison with its competitors: Novotel and Hilton.

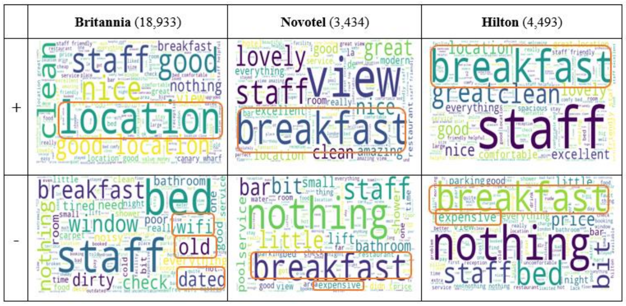

*Figure 12. Word clouds for HCBW and its competitors *number in brackets indicates number of reviews (inc empty)*

From the figure above, there are a few observations to be made. First, "location" is most prominent in HCBW's word cloud. This could indicate that despite being in close proximity of each other, location could still be HCBW's greatest competitive advantage. Unfortunately, it might also be simply due to the hotel not having other good things to offer. Second, "breakfast" has appeared in the positive word clouds of both the competitor hotels, but not in HBCW's. From the word clouds generated, it also appears that breakfast plays an important role in customers' perceptions towards a hotel, while being a relatively simpler issue to address. Consequently, the hotel should seriously review its breakfast menu options. Third, HBCW is the only hotel with "old", "dated" and "wifi" magnified in its negative word cloud, implying that it might be in need of a renovation. Finally, the term "expensive" featured more prominently in the negative word clouds of the 2 competitor hotels, and less so in HBCW's. Therefore, HBCW's relative strength could be that it is more value-for-money.

6. Review Text Classification Model

6.1 Model Preparation

Data retrieved from Kaggle and Booking.com was chosen to be training data since they have already been clearly labeled (Booking.com collects review through "negative" and "positive" review fields on their website) while data from the TripAdvisor website needed further processing and labeling into "positive" and "negative" reviews. The intuition for this treatment is that upon deployment, our model will be used to classify reviews from different sources or online platforms, which may not offer readily labeled reviews. Hence, our decision to use TripAdvisor reviews as the test data will provide us with a robust assessment of the in-use accuracy of our model.

6.2 Model and Word Embedding

Methods and Libraries	
Rule-based Methods (Baseline)	TextBlob
	Vader
Feature-based Methods	Logistic Regression
	SVM
	Naïve Bayes
	Random Forest
	XGBoost
	Light GBM
Embedding-based Methods	LSTM with GloVe
	FastText

Figure 13. Table of the models used

The rule-based methods were used as a baseline as these models already consisted of in-built sentiment scores for different words. For example, in TextBlob it uses the NLTK Python library which calculates sentiments from 3 scores the polarity, subjectivity and intensity of a string of words. Vader is similar to TextBlob with a focus on social media text, such as short text, repetitive words and excess punctuation, which are similar characteristics of the Booking.com reviews.

The feature-based methods consist of traditional machine learning models: Logistic Regression, Support Vector Machine, Multinomial Naïve Bayes (MNB), Random Forest and boosting methods XGBoost and Light GBM. Such models make use of a document-term matrix generated from the vectorization of the words used in each review for classification.

For the rule based and feature based models, we considered two approaches for building the document-term matrix – CountVectorizer and Term Frequency-Inverse Document Frequency (TF-IDF).

Lastly embedding based methods were also considered in an attempt to use more complex models to achieve a higher level of accuracy. Two methods considered were Long Short Term Memory (LSTM) and FastText. LSTM is a recurrent neural network with GloVe word embedding, which is an extension of the fundamental Word2Vec intuition. FastText, a model created by Facebook's AI Research lab, is in essence it is a shallow neural network with its own pre-trained word vectors.

6.3 Model Evaluation Metric

Since our dataset is relatively balanced, and we wanted to extract insights on both the strengths and weaknesses of the hotel, we did not need to focus specifically on the model performance with respect to any particular class. Therefore, the model's accuracy score was chosen as the evaluation metric for this project.

6.4 Results

5-fold cross-validation was performed on the train dataset for us to evaluate the model performance in a more robust manner. Thereafter, each model was evaluated on a hold-out test set. The cross-validation and test accuracy scores of each model are shown in Figure 14 below.

Methods and Libraries		Cross-Validation Accuracy	Test Accuracy
Rule-based Methods (Baseline)	TextBlob	61.91%	55.27%
	Vader	63.72%	65.03%
Feature-based Methods	Logistic Regression	82.76%	75.58%
	SVM	83.11%	80.90%
	Multinomial Naïve Bayes	84.25%	83.09%
	Random Forest	82.59%	80.54%
	XGBoost	84.11%	78.48%
	Light GBM	83.65%	76.12%
Embedding-based Methods	LSTM with GloVe	75.12%	79.00%
	FastText	56.78%	78.12%

Figure 14. Table of cross-validation accuracy and test accuracy results from various models built

We can see that, in general, comparing within the rule-based and feature-based methods, the models achieved similar cross-validation accuracy scores, although the feature-based methods proved to be the superior methods, with all models achieving accuracy scores above 80%. Also, it appears that the two embedding-based methods used in this study has quite a significant difference in performance. Lastly, some models appear to have overfitted, since their cross-validation accuracy scores were significantly higher than their test accuracy scores. These models include TextBlob, Logistic Regression, XG Boost and Light GBM, which all showed at least a 7% higher cross-validation accuracy scores than the test accuracy scores.

Another observation we made was that, while typically, in-sample performance is higher than out-of-sample performance, some models, namely Vader, LSTM and FastText, were observed to have higher cross-validation accuracy scores than test scores. This can stem from the fact that our training data is retrieved from Booking.com while test data is from TripAdvisor. Potentially, data extracted from different sources can have different levels of noise which may allow certain models to perform better on the test data than others.

6.5 Model Selection

From Section 6.4, we can observe that the MNB model produced the best results on the test set. The use of more complex models through neural networks and different word embeddings did not yield better accuracy scores. Between TF-IDF vectorizer and CountVectorizer, model accuracy is roughly the same with difference in range of 0.1%. Hence, we decided to choose CountVectorizer as the vectorizer of choice due to the shorter time it requires to run the models.

In conclusion, the model of choice moving forward is MNB with CountVectorizer due to highest level of accuracy and shorter runtime.

7. Solution Deployment – An Interactive Business Dashboard

After having all reviews classified into positive and negative category, to demonstrate how we can apply the business dashboard on negative reviews of HBCW, we applied Topic Modelling LDA to extract all the main topics revolving around our negative review corpus. Subsequently, LDA will assist us in generating a condensed list of all important topics. Using spacy phrase matcher will now allow us to assign reviews to their respective topics. Note that one review might contain different topics as reviewers might be complaining about more than one topic, e.g. room condition and staff attitude in the same review. For the prototype dashboard, the topic list is not fully expanded and detailed yet. Hence, if the topic is not listed in the topic list, it will be listed under category "Others". Figure 15 provides an illustration of our interactive business dashboard (implemented via Tableau).

Figure 15. Illustration of the interactive business dashboard

8. Recommendations to HBCW Management

8.1 Short to Medium Term Recommendations

Playing to its strength – location: As discussed in paragraph 5.1, based on our word cloud analysis, it was very evident that the key strength of HBCW was its location. This is an area which we strongly recommend to the hotel to highlight to its prospective guests. In addition to the current descriptions on the HBWC website (Figure 16) (Britannia Hotels, 2021) on distances to places of interest around London, the hotel can further emphasize on its strength by highlighting the short travel time it takes to travel around Inner London from the hotel. For example, it can mention that, for those without private transport, the hotel is just a short 6 minutes’ walk to the nearest light rail station, the Heron Quays light metro station, which is also connected directly to the Jubilee Place Mall, a shopping mall which offers shoppers a choice of fashion, pharmacies, cafes, and restaurants (LondonTown, 2021). For guests who drive, the hotel can highlight that the hotel is just a short 8 minutes’ drive to attractions such as the Tower of London and the Tower Bridge.

> “Plenty of London’s main attractions are only a few miles away… Visit the Tower of London, 3.5 miles… Buckingham Palace is only 7 miles from the hotel… Westminster Abbey, 6.5 miles, is 700 years old and a must-see living pageant of British history… The British Museum is just 5.9 miles from the International Hotel… Why not break up the sight-seeing and take the kids to London Zoo? It’s only 8.9 miles away…”

Figure 16. Current location description on HBCW’s website.

Improving the quality of its breakfast: As highlighted in paragraph 5.4, the term “breakfast” was magnified in both Novotel’s and Hilton’s positive word clouds, but it was not the case in HBCW’s positive word cloud. This suggests that HBCW was losing out to the competitor hotels in terms of the quality of its breakfast. We strongly recommend for HBCW to review and improve upon its breakfast menu.

Feedback on room windows: We observed that the term “window” was magnified in HBCW’s negative word cloud (general text and noun). Upon a detailed review, guests were generally complaining that the room windows were dirty, not serviceable and some of the rooms were without windows. To address the first two points, we strongly recommend for the hotel to engage a contractor the clean and repair the windows. On the last point, we recommend for the HBCW to highlight specifically (possibly with larger or different coloured fonts) on its own and partners’ booking websites that some rooms come without windows, to manage its guests’ expectations.

Feedback on beds: As the term “bed” was magnified in the negative word clouds of HBCW, we studied the reviews more carefully and observed that guests were complaining that the beds were very uncomfortable to sleep on. Some guests even complained that it felt like sleeping on a “chain link fence” and that the bed springs were “digging” into their backs. We recommend for HBCW to consider changing the bed mattresses so that guests can have better sleep quality**.** Since that may take time, the hotel may wish to provide guests with mattress toppers for the time being, to improve guest experience.

Feedback on Wi-Fi: We observed many reviews complaining about the lack of unlimited in-room Wi-Fi access, and that the term “wifi” featured prominently in the word cloud generated from the reviews of business travelers. Given that Wi-Fi connectivity is one of the top desired amenities for travelers (Bachman, 2016), HBCW should seriously consider revising its Wi-Fi policy, which currently provides unlimited free Wi-Fi only in the common areas and requires its customers to pay if they use the in-room Wi-Fi service for more than 20 minutes. To manage costs, HBCW can build in Wi-Fi charges into the room rates.

8.2 Long Term Recommendations

Staff training: Another concerning observation we made was that the word “staff” was magnified in both HBCW’s positive and negative word clouds. We also observed that “clean” and “dirty” appeared in the positive and negative word clouds, respectively. This suggests that the hotel staff may be providing different levels of service quality to different guests. We therefore recommend for the HBCW management to look into this and implement additional staff training to raise the service quality of its staff.

Refurbishment/renovation: Our analysis of the word clouds also revealed that many guests described the hotel as being “old”, “dated” and even “run down”. To improve upon the appeal of the hotel, we strongly recommend for HBCW’s management to plan for a refurbishment or even a renovation to spruce up the hotel.

9. Conclusion

In conclusion, through this project, we have demonstrated how text analytics techniques can be used to transform large amounts of semi-structured data into business insights. We have achieved this by building an accurate model for the classification of text reviews into positive and negative classes. This can then support the generation of word clouds to provide specific and actionable insights for us to formulate recommendations to help the hotel improve upon customer experience. We have also demonstrated how text analytics techniques can be applied by the hotel to determine its competitive advantage relative to competitor hotels in the vicinity. Finally, we built a business dashboard which takes in the reviews labeled by our model to facilitate and expedite deep dives by the hotel management team into specific common complaint topics by the hotel guests. This will allow the hotel management team to make quick decisions and respond decisively to customer feedback. What makes the review text classification and the interactive business dashboard tools developed through this project especially valuable is the fact that it can be readily used by any hotel seeking to improve upon guest experience by leveraging upon reviews text data, regardless of whether the text data has been labeled or not.

10. Limitations and Recommendations

Lack of domain knowledge: One key limitation we faced in this project was that we did not have the domain knowledge to complement the findings from our text analysis. For example, if we knew constraints faced by the hotel (e.g. cost, manpower limitations, etc.), we could also consider them when formulating recommendations for improving upon customer experience. In a real-world project, this can be achieved by engaging the hotel stakeholders to share and discuss our findings in more detail.

Skewed number of reviews of competitor hotels: As shown in Figure 12, we had a lot less review text data for the Novotel and Hilton hotels (3,434 and 4,493 reviews, respectively, compared to 18,933 reviews for HBCW). This means that we are less confident that the keywords observed in their word clouds are/will be representative of the views of the "population" of the guests who stayed and will be staying at the hotels. To address this, more review text data from the Novotel and Hilton hotels will need to be collected.

POS tagging: In the exploratory stages of the project, we observed that there were some errors in the tagging of verbs and adjectives by the NLTK POS tagging library. While this was not a significant issue in this project (since we did not perform POS tagging for model building or perform extensive analysis based on verb or adjective word clouds), a more accurate POS tagging may help to provide additional insights. This issue could possibly be addressed by using other POS tagging algorithms followed by an evaluation of which performed the best on hotel review text data.

References

Bachman, J. (2016, January 25). *Even the Super Rich Want Free Wi-Fi in Hotels.* Retrieved April 22, 2021, from https://www.bloomberg.com/news/articles/2016-01-25/even-the-super-rich-want-free-wi-fi-in-hotels

Britannia Hotels. (n.d.). A History of Britannia Hotels, About Us. Retrieved April 22, 2021, from https://www.britanniahotels.com/about-us

Liu, J. (2017, August 21). *515K hotel reviews data in Europe*. Retrieved April 24, 2021, from https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe

LondonTown.com. (n.d.). *Jubilee place Mall, Bank Street, London.* Retrieved April 22, 2021, from http://www.londontown.com/LondonInformation/Shopping/Jubilee-Place-Mall/2eb9f/

Sky News (2020, December). *'A filthy hovel' - Britannia named worst UK hotel chain eight years running.* Sky News. https://news.sky.com/story/a-filthy-hovel-britannia-named-worst-uk-hotel-chain-eight-years-running-12151638

The International Hotel. (n.d.). Retrieved April 22, 2021, from https://www.britanniahotels.com/hotels/the-international-hotel