Decoding Customer Feedback: Analyzing Hotel Reviews with Deep Learning

Group 8: Gan Tian (A0231897R), Nithesh Reddy (A0232779R), Song Kaiyue (A0177341E), Xu Ruilin (A0157242J), Zhang Jiayue (A0148005N)

GitHub link: https://github.com/KaiyueSong/BT5153_Group8

Abstract

This report investigates the application of machine learning models in sentiment analysis on hotel reviews. We first collect a dataset of hotel reviews and perform preprocessing steps to refine the data. These steps include the removal of noise and irrelevant information, text normalization, and feature extraction. We then experiment with various machine learning models such as Multinomial Naive Bayes, CatBoost, LSTM, and RoBERTa. Our evaluation results show that RoBERTa outperformed the other models, achieving an accuracy of 0.9486, precision of 0.9855, recall of 0.9201, and F1 score of 0.9517.

Furthermore, we discuss potential areas for future research, including the incorporation of topic modeling and utilizing the GPT3 API to improve the accuracy of predictions. Our findings highlight the importance of preprocessing in achieving high accuracy in machine learning models and demonstrate the effectiveness of RoBERTa in sentiment analysis tasks. The results of this study can be valuable for hotels in understanding their customer feedback and improving their services to enhance the customer experience.

1. Introduction

1.1 Problem statement

In the hotel industry, customer reviews play a crucial role in shaping customers' experiences and influencing their decisions when choosing hotels for their stays. However, with the proliferation of online review platforms, hotels are faced with the challenge of analyzing a large amount of textual data to understand their customers' feedback and improve their services. Reading through a large number of reviews to draw accurate conclusions on how customers value different traits of the hotel's facilities, staff, and services can be time-consuming. Moreover, newly established hotels may not have any customer reviews yet, making it difficult for hotel managers to gain insights into their customers' experiences.

For newly established hotels, the lack of customer reviews can be a significant challenge. Without feedback from customers, it can be difficult for hotel managers to know where to focus their efforts to provide the best possible experience for guests. In this case, hotels can turn to social media listening tools to monitor mentions of their brand online. By analyzing social media posts that mention the hotel, hotel managers can gain valuable insights into how customers perceive their brand and what they might be looking for in their hotel experience.

To address these challenges, this report proposes applying deep learning techniques to sentiment analysis on hotel reviews. The goal is to train a sentiment analysis model that can rapidly process a large amount of textual data to provide hotels with valuable insights into their customers' experiences and feedback. By helping hotels better understand their customers, sentiment analysis can improve hotel services and customer review ratings, leading to better customer satisfaction, increased bookings, and revenue for hotels. This, in turn, can help hotels gain a competitive edge in the marketplace. Additionally, sentiment analysis can benefit customers by providing them with more accurate information about a hotel's strengths and weaknesses, helping them make more informed decisions about their stays. Furthermore, the proposed solution can be applied to other fields and industries to improve customer services efficiently, leading to better experiences for customers across various domains. Overall, the proposed solution aims to benefit both the hotel industry and customers by improving the efficiency and accuracy of analyzing customer feedback, leading to better services, increased revenue, and enhanced experiences for all parties involved.

1.2 Objective

This project has two main objectives. Firstly, it aims to develop a deep learning model for sentiment analysis on hotel reviews, which can accurately categorize reviews as positive, negative, or neutral. The model will also be able to identify specific features of the hotel that are most closely associated with the customer's rating and experience, such as the environment, service, or food quality. This approach is crucial as it allows hotels to understand what aspects of their service need improvement, and what customers value the most.

Secondly, this project seeks to evaluate the performance of the proposed model in terms of various metrics such as accuracy, precision, recall, and F1 score. These metrics will help determine how well the model performs in identifying customer sentiment accurately.

Additionally, the project will consider potential constraints that may affect the model's performance, such as imbalanced data or strategic decisions made by the hotel board. For instance, if the hotel introduces a new service or feature, the sentiment analysis model may need to be retrained to account for changes in customer feedback.

1.3 Data Collection

The dataset used in this project is the "515k Hotel Reviews Data in Europe" which is available on Kaggle. The dataset contains over 515,000 customer reviews and ratings of hotels across Europe, spanning a period of ten years from 2008 to 2018. The data was scraped from Booking.com, a well-known online travel agency. The dataset will be pre-processed to remove any irrelevant information and transformed into numerical features suitable for training the deep learning model.

This dataset is a valuable resource for sentiment analysis as it contains crucial details about hotels, their services, and customer experiences. The reviews were collected from different online platforms and are associated with various hotels located in various European cities. The dataset includes several columns, such as hotel details like hotel name, address, city, and country, and customer details like reviewer nationality and type of traveler. The reviews themselves contain ratings on several aspects such as cleanliness, comfort, location, staff, and value for money. Furthermore, the reviews may also include free-text comments written by the customers, providing additional information on their experience.

Overall, this dataset provides a comprehensive overview of customer experiences with hotels, making it an ideal resource for developing a sentiment analysis model. By leveraging the dataset's rich information, the model can accurately classify hotel reviews and identify the key features that impact the customer's sentiment towards a hotel. This approach can help hotels improve their services and enhance the customer experience, ultimately leading to higher customer satisfaction.

The table below provides details of the columns and their explanations.

Table 1. Dataset Column Description.

Column Name	Explanation
Hotel_Address	Address of the hotel
Review_Date	Date when the review was written
Average_Score	Average score of the hotel, based on all reviews available
Hotel_Name	Name of the hotel
Reviewer_Nationality	Nationality of the reviewer
Negative_Review	Negative comments of the reviewer about the hotel
Review_Total_Negative_Word_Counts	Total number of words in the negative review
Positive_Review	Positive comments of the reviewer about the hotel
Review_Total_Positive_Word_Counts	Total number of words in the positive review
Reviewer_Score	Score given by the reviewer based on their experience
Total_Number_of_Reviews_Reviewer_Has_Given	Total number of reviews written by the reviewer
Total_Number_of_Reviews	Total number of reviews available for the hotel
Tags	Tags given to the review by the reviewer
days_since_review	Number of days since the review was written
Additional_Number_of_Scoring	Additional number of scores given to the hotel
lat	Latitude of the hotel
lng	Longitude of the hotel

The dataset detailed above will serve as the foundation for the development of the sentiment analysis model. Through the exploratory data analysis of the various columns and their corresponding data, we identify the patterns and relationships between the variables.

2. Exploratory Data Analysis

In this 515k hotel review dataset, Reviewer_Score from each reviewer and Average_Score of each hotel is provided together with the positive and negative reviews. The Reviewer_Score distribution chart shows general satisfactory hotel review scores from reviewers. More than half of the reviews are above a score of 8. The hotel average scores are also concentrated between 8-9 which shows an overall positive sentiment.

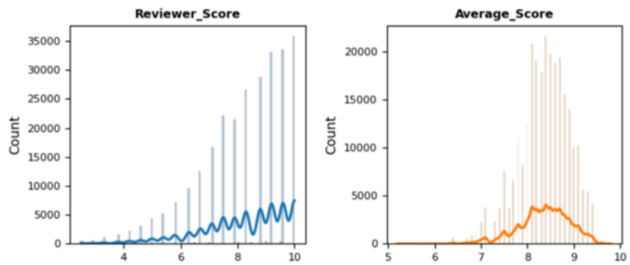

Figure 1. Distribution Chart of Review_Score and Average_Score.

As our objective to perform sentiment analysis on hotel review, we will focus our exploratory data analysis on the text review data. In this dataset, there are Positive_Review and Negative_Review columns which contain the feedback and comments customers have to say regarding their hotel stay experience. We visualize the Positive_Reviews and Negative_Reviews with word cloud.

Figure 2. Word Cloud of Positive_Review and Negative_Review.

In the Positive_Review wordcloud, we see large words like Location, Staff, Romantic, Surroundings Breakfast associated with positive feedback. In the Negative_Review wordcloud, most prominent words are Negative, Room, Elderly, Dirty and Available. We realized in the negative_review column, there are many rows with “No Negative” comment, which means nothing negative on the hotel stay. Such feedback carries on negative sentiment which we think can be removed before further analysis.

There is a ‘Tags’ column in this dataset which contains important information such as Trip Type (Business, Leisure) and Type of visitor (Couple, Solo traveler). In-depth analysis of review from Business travelers vs Leisure travelers can provide us with more insights on different visitor needs and preferences. Therefore, we extract keywords from the ‘Tags’ and create 2 additional columns called ‘Trip Type’ and ‘whom’.

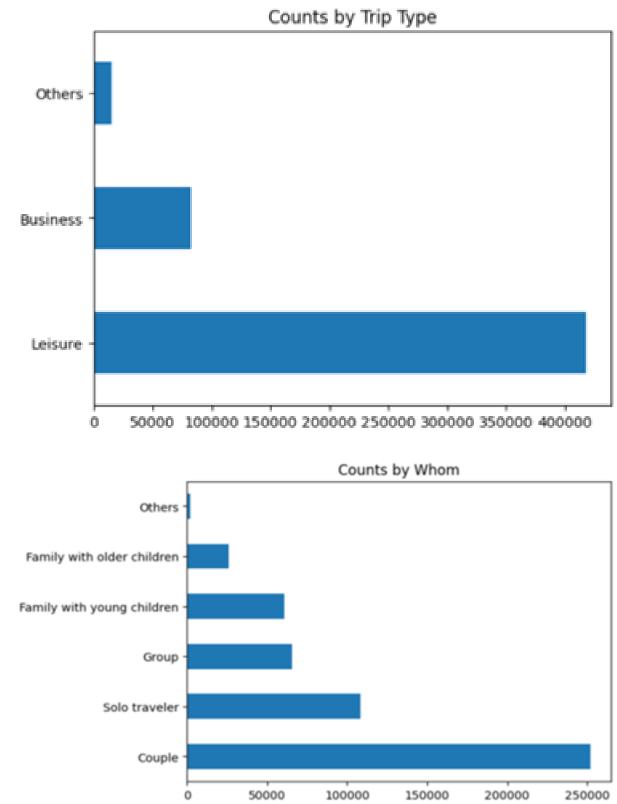

Figure 3. Travelers Count by Trip Type and by Whom.

From the bar chart above, we see more than 80% of reviews were provided by Leisure travelers, and around 15% by Business travelers. Whereas looking at reviews by ‘whom’, we see 50% reviews were from couples, 20% from solo travelers, around 12% from group, and 15% from family with children. The fact that half of the reviews were from couples probably explains why we see ‘Romantic’ as one of the top mentioned words in the positive review.

More insights could be drawn if we look at the frequency word count charts for example by looking at business and leisure travelers separately. Location is the top words with highest counts for business travelers, as they need to travel to offices and meet clients. So convenient location is critical to business travelers. For leisure travelers, friendliness and good service from staff is more prominent in the positive reviews. Both locations and staff are the top 2 aspects appear in the positive reviews. In the frequency word count of negative reviews, breakfast and small appear to be the top 2 negative aspects for business and leisure travelers. Wifi and noisy appear in business travelers top 15 negative aspects, good internet speed and quiet environment is important for business travelers to work in the hotel. Expensive is one of the top negative aspects of leisure travelers. It is understandable that leisure travelers are more cost conscious than business travelers.

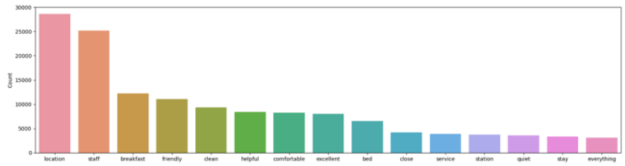

Figure 4. Frequency Word Count by Business Travelers on Positive Reviews.

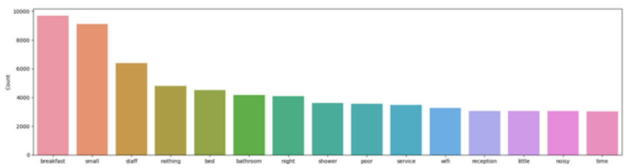

Figure 5. . Frequency Word Count by Business Travelers on Negative Reviews.

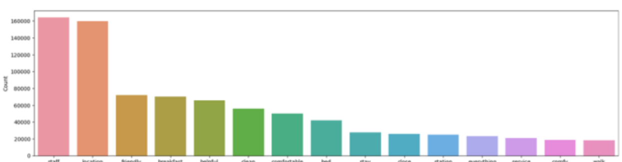

Figure 6. Frequency Word Count by Leisure Travelers on Positive Reviews.

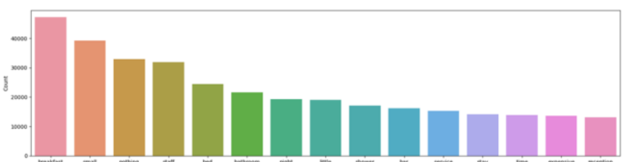

Figure 7. Frequency Word Count by Leisure Travelers on Negative Reviews.

We will share more insights derived from the exploratory data analysis in the business implication section of this report.

3. Data Preprocessing and Feature Engineering

In order to prepare the hotel review dataset for input to the model, several preprocessing steps were undertaken. These steps were designed to clean and transform the raw data into a format that could be more easily analyzed by the model. The following paragraphs will outline each of these preprocessing steps in detail and explain their usefulness in improving the accuracy of the model.

Remove reviews lesser than 2 words

To ensure that the data being fed to the model is of high quality and relevance, we removed all reviews with two or fewer words. This is because such reviews do not contain enough information to provide any valuable insights for hotels or customers. By removing these reviews, we were able to ensure that the remaining data was more focused and relevant, which is crucial for the accuracy and effectiveness of our model.

Removing of stop words for wordCloud visualization

Stop words are words that are commonly used in a language and do not provide any valuable information in sentiment analysis. Examples of stop words include "the," "and," "a," etc. We removed stop words to reduce the dimensionality of the data and improve the performance of our model. Additionally, we used the resulting data to create a word cloud visualization to gain a better understanding of the most commonly used words in the reviews.

Tokenization

Tokenization is the process of splitting text data into individual words or tokens. We used this technique to split the reviews into individual words, which could then be processed and analyzed by our model. Tokenization helps to reduce the complexity of the data and improves the efficiency of the model. By representing the data in this way, our model was better able to identify and analyze the sentiment of each individual word, which ultimately contributed to the overall accuracy of the model.

Vectorization

Vectorization is the process of converting text data into numerical vectors that can be used as input for machine learning models. We used vectorization to convert the tokenized reviews into numerical vectors, which could then be processed by our sentiment analysis model. Vectorization allows us to represent the data in a way that can be easily understood and processed by the model. By converting the text data into numerical vectors, we were able to improve the accuracy and efficiency of our model.

Padding

Padding is the process of adding zeros or a specific value to the end of sequences to make them a consistent length. In our case, we used padding to ensure that all reviews had the same length, which is necessary for training our model. By padding the reviews to a consistent length, we were able to standardize the input for our model, which improved the accuracy and consistency of our results. Additionally, padding helped to prevent overfitting, which is a common problem in machine learning where the model performs well on the training data but poorly on new, unseen data.

In conclusion, the data preprocessing steps undertaken for this project were vital in ensuring that the model was trained accurately and effectively. The removal of reviews with two or fewer words allowed for more meaningful data to be used in the model, and the removal of stop words helped to eliminate commonly used words that were not useful in determining sentiment. Tokenization of the data allowed for the creation of a sequence of words that could be fed into the model, and vectorization of the data enabled the words to be represented as numbers for further analysis. Finally, padding of the data allowed for the input to the model to be of the same length, which improved the accuracy of the model. These preprocessing steps were critical in ensuring that the model was effective in analyzing sentiment in hotel reviews, and can be used as a template for similar projects in the future.

4. Modeling and Hyperparameter Tuning

4.1 Sentiment Analysis

Various models were explored to classify hotel reviews into positive or negative sentiment, namely Multinomial Naive Bayes, CatBoost. LSTM and RoBERTa. For robustness, K-Fold methodology was adopted to validate the train data and to perform hyperparameter tuning. The list of models and the corresponding optimal parameters can be found as below.

Table 2. Summary of Model Hyperparameters

MODEL	PARAMETERS
MULTINOMIAL NAIVE BAYES	alpha = 1, fit_prior = True
CATBOOST	iterations = 1000, learning_rate = 0.1, depth = 4, loss_function = 'Logloss'
LSTM	dropout = 0.2, epochs = 10, batch_size = 32,
ROBERTA	callbacks = [EarlyStopping (monitor = 'val_loss', patience = 2)]

4.1.1 MULTINOMIAL NAIVE BAYES

Multinomial Naive Bayes is a variant of the Naive Bayes algorithm that works well with text data, which makes it a popular algorithm for sentiment analysis (Wang et al., 2015). The algorithm calculates the probability of each word in a document belonging to a specific sentiment class (i.e. positive reviews or negative reviews) (Jiang et al., 2016). These probabilities if individual words are then combined using Bayes' theorem to calculate the overall probability of the document belonging to each sentiment class (Jiang et al., 2016). The class with the highest probability is then assigned as the sentiment label for the document.

	precision	recall	f1-score	support
0	0.91	0.91	0.91	1323
1	0.93	0.93	0.93	1769
accuracy			0.92	3092
macro avg	0.92	0.92	0.92	3092
weighted avg	0.92	0.92	0.92	3092

Figure 8. Classification Report of Multinomial Naive Bayes.

4.1.2 CATBOOST

CatBoost is an extension of gradient boosting algorithms that are designed to handle categorical features, which makes it useful for text data. The algorithm works by creating decision trees that optimize the objective function through gradient descent (Hancock & Khoshgoftaar, 2020). In addition, CatBoost includes a range of advanced techniques such as bagging, boosting and feature selection, which makes it a powerful algorithm for sentiment analysis that works effectively even with large and complex datasets (Hancock & Khoshgoftaar, 2020).

	precision	recall	f1-score	support
0	0.89	0.93	0.91	1341
1	0.95	0.91	0.93	1751
accuracy			0.92	3092
macro avg	0.92	0.92	0.92	3092
weighted avg	0.92	0.92	0.92	3092

Figure 9. Classification Report of CatBoost.

4.1.3 LSTM

LSTM (Long Short-Term Memory) is a type of neural network architecture that is commonly used for sentiment analysis. It is especially useful for analyzing sequences of text data, as it captures long-term dependencies between words and phrases (Greff et al., 2017).

In this project, the LSTM model was designed to contain one input layer (word embedding), two hidden layers (SpatialDroupout1D layer and LSTM layer) and one output layer (using sigmoid activation function). The hidden layers contain LSTM cells, which are designed to selectively remember or forget information from previous time steps in the input sequence (Korstanje, 2021).

In order to prevent overfitting and to improve generalization of the model, as well as to assess the impact of different hyperparameters on model convergence time, early stopping was experimented on the LSTM model. To decide when to early stop, another 2 parameters were introduced: patience and min_delta. Patience (default=20) is the number of epochs with no improvement after which training will be stopped, while min_delta (default=0.0005) is the minimum improvement in the monitored quantity to qualify as an improvement. For example, if the validation accuracy did not improve by min_delta of 0.0005 for 20 consecutive epochs (patience), the training would be early stopped.

	precision	recall	f1-score	support
0	0.93	0.93	0.93	1354
1	0.95	0.95	0.95	1738
accuracy			0.94	3092
macro avg	0.94	0.94	0.94	3092
weighted avg	0.94	0.94	0.94	3092

Figure 10. Classification Report of LSTM.

4.1.4 RoBERTa

RoBERTa (Robustly Optimized BERT approach) is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model that was introduced by Facebook AI in 2019. As an improved recipe for training BERT models, RoBERTa is able to match the performance of post-BERT methods with several modifications in the training process, including training the model longer with bigger batches, removing the next sentence prediction objective, training on longer sequences, and dynamically changing the masking pattern applied to the training data (Yinhan et al., 2019). Such modifications allow RoBERTa to learn high-quality representations of words and sentences, resulting in excellent performance on natural language processing tasks, including sentiment analysis in this project.

In this project, the RoBERTa pre-trained tokenizers and models were used. In additional, several hyperparameters were fine-tuned to ensure the best performance, including learning rate, batch size and number of epochs. However, as the GPU used was not able to handle large batch size, a maximum batch size of 16 was then selected.

	precision	recall	f1-score	support
0	0.9095	0.9834	0.9450	1389
1	0.9855	0.9201	0.9517	1703
accuracy			0.9486	3092
macro avg	0.9475	0.9518	0.9484	3092
weighted avg	0.9514	0.9486	0.9487	3092

Figure 11. Classification Report of RoBERTa.

4.1.5 Model Evaluation

It is observed that generally, deep learning models namely LSTM and RoBERTa outperformed conventional machine learning models on this sentiment analysis task. Different from conventional machine learning, deep learning models are able to learn complex and abstract features automatically from the raw data, without the need for hand-crafted feature engineering. This is particularly useful for sentiment analysis, where features such as tone, sarcasm, and context can be subtle and difficult to identify manually. Also, deep learning models are often able to handle noisy and unstructured data, which is common in hotel reviews.

In particular, RoBERTa is the top-performing model with highest Precision, F1 Score and Accuracy on validation dataset. This may be because it is designed to capture the context and meaning of words in a sentence. It was pre-trained on large amounts of text data to learn the language representation in an unsupervised manner, and then fine-tuned based on this specific task. Additionally, RoBERTa is based on the transformer architecture, which is known for its ability to process long sequences of data, making them well-suited for sentiment analysis that involves analyzing entire sentences or even paragraphs.

Table 3. Summary of Model Performance

Model	Precision	Recall	F1 Score	Accuracy
Multinomial Naive Bayes	0.9304	0.9293	0.9299	0.9198
CatBoost	0.9467	0.9126	0.9293	0.9214
LSTM	0.9454	0.9465	0.9459	0.9392
RoBERTa	0.9855	0.9201	0.9517	0.9486

4.2 Aspect-Based Sentiment Analysis

Apart from training sentiment analysis models, we also explored on Aspect-based sentiment analysis, a powerful technique used to analyze reviews in order to identify specific aspects of a product or service that customers are positive or negative about and the sentiment polarity (Zhao et al., 2023). In this section, we demonstrated an ABSA approach that employed Latent Dirichlet Allocation (LDA) to determine the most talked-about aspects in the hotel reviews dataset, and TextBlob to extract the sentiment polarity for each aspect.

4.2.1 Data Preprocessing

Our ABSA approach began with preprocessing the text data, which included tokenizing the text, removing stop words, and stemming the remaining words. These steps ensured that the text data was in a format that can be used for analysis. After preprocessing the text data, we created a bag of words matrix using CountVectorizer, which converted the text data into a matrix of word counts.

4.2.2 Topic Modelling

Next, we trained an LDA model, a widely used probabilistic topic modeling technique, on the bag of words matrix to identify the top topics discussed in the dataset (Blei et al., 2003). The LDA model identified the most common words used in the dataset and grouped them into topics based on their co-occurrence. For each identified topic, we extracted the top 10 most frequently occurring words, which, after preprocessing, were reduced to their root or base form. These words were used to represent the aspect of the product or service that the customers were commenting on.

4.2.3 Sentiment Polarity

After identifying the most important aspects of the product or service, we used TextBlob to perform sentiment analysis on each aspect. For each aspect, we collected all of the customer feedback that mentioned that aspect. We then used TextBlob to calculate the sentiment polarity of the feedback. The sentiment polarity is a score that ranges from -1 to 1, with -1 being extremely negative and 1 being extremely positive. We used the sentiment polarity to determine whether the customers were positive or negative about each aspect.

4.2.4 Conclusion

Based on ABSA approach applied to the hotel review data, the following aspects were identified: room, staff, breakfast, location, bathroom, bed, noise, price, parking, and facilities. The sentiment polarity for each aspect was extracted using TextBlob, and the results showed that the top 3 most positively rated aspect were location, staff/help and breakfast/restaurant. The most negatively rated aspect was noise, followed by room size and price.

5. Business Implication

Applying sentiment analysis on a hotel review dataset can have several business implications for the hospitality industry. By analyzing customer reviews, hotels can gain insights into the specific aspects of their service that are highly praised or criticized by customers. This can help hotels to identify areas for improvement and provide a better overall customer experience. We analyzed positive and negative reviews from the specific type of hotel visitors such as couples, solo travelers, couples with young and old kids. ‘Reception’ appears to be one of the highly-mentioned word in negative reviews for couples with kids. It could be the kids tend to be noisy, which undermines the hospitality and service quality provided by the reception staff. With better training of the reception staff specifically on handling guests with kids, customer experiences can be improved without incurring high cost.

Sentiment analysis can be used to compare customer sentiment across different hotels and competitors, helping hotels to identify areas where they can differentiate themselves and improve their competitive position. We identified the best hotel ‘Ritz Paris’ and the worst hotel ‘Hotel Liberty’ based on their average scores and visualized their positive and negative reviews with word cloud. Ritz Paris earned 9.8 average scores mainly through friendly staff, excellent service, and comfortable beds which is their differentiating factor and competitive advantage. Hotel Liberty was penalized with a 5.2 average score mainly due to poor hygiene of the rooms and bad facilities, with words like dirty, Wi-Fi and conditioning mentioned most frequently in their negative reviews. Hotel liberty should improve on the negatively mentioned aspects and tap on their advantage with good location and delicious breakfast to attract more customers.

By analyzing sentiment across a large dataset of reviews, hotels can also gain insights into market trends, such as emerging preferences for certain amenities or services. From the word frequency charts of the positive reviews, we see staff, location, cleanness emerging as the top themes which customers appreciate. Breakfast, small, staff, bed, bathroom are the most frequently mentioned words across all negative reviews. Our ABSA analysis provided slightly different perspectives with staff, breakfast and location as mostly positively rated aspects versus noise, bathroom and parking as most the most negatively rated aspects. Word frequency counts and ABSA can complement each other in identifying the key positive and negative aspects from the hotel reviews. Parking emerges from the negative reviews not due to high frequency of word counts, but negative polarity.

6. Conclusion and Future Improvements

With the successful implementation of data preprocessing steps, we evaluated various models, including Multinomial Naive Bayes, CatBoost, LSTM, and RoBERTa. After careful evaluation, we found RoBERTa to be the best-performing model with an accuracy score of approximately 0.95. In addition to its high accuracy, RoBERTa also showed high precision of 0.9855, recall of 0.9201, and F1 score of 0.9517. This model was chosen due to its ability to better handle complex language patterns, capture the context meaning of phrases in hotel reviews and the large amount of data available for training. By using RoBERTa, we were able to effectively classify the sentiment of hotel reviews with high accuracy, providing valuable insights to the hotel industry.

While we achieved good results with RoBERTa, there are still several areas where we can improve on the accuracy of our sentiment analysis model. For instance, we can incorporate hotel specific domain knowledge into the model through feature engineering or the use of external resources such as ontologies or lexicons which help to detect sentiment in hotel-specific language. Moreover, we can even try combining predictions results from RoBERTa, MNB, Catboost and LSTM using methods like majority voting or weighted averaging and see whether accuracy can be further improved.

In conclusion, sentiment analysis has emerged as a powerful tool for businesses across various industries to gather insights and improve their services. In the hospitality industry specifically, hotel review sentiment analysis has become increasingly important given the rise of online review platforms and the impact that customer feedback can have on a hotel's reputation and success. By analyzing customer sentiment towards different aspects of their experience, hotels can identify areas for improvement and implement changes to enhance customer satisfaction.

Furthermore, as natural language processing models continue to advance, the potential for sentiment analysis to provide even more accurate and nuanced insights is significant. Additionally, sentiment analysis techniques can be applied to larger datasets across multiple industries, allowing businesses to make data-driven decisions and improve their services efficiently. Overall, the use of sentiment analysis in the hospitality industry and beyond has the potential to greatly benefit both businesses and customers alike, leading to improved experiences and enhanced satisfaction.

References

Kaggle. (n.d.). *515K Hotel Reviews Data in Europe*. Retrieved April 19, 2023, from https://www.kaggle.com/datasets/jiashenliu/515k-hotel-reviews-data-in-europe

MLearning AI. (n.d.). *NLP Tokenization, Stemming, Lemmatization, and Part-of-Speech Tagging. Medium*. https://medium.com/mlearning-ai/nlp-tokenization-stemming-lemmatization-and-part-of-speech-tagging-9088ac068768

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993-1022.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2017). *LSTM: A Search Space Odyssey*. Cornell University Library, arXiv.org.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, *7*(1), 94–94

Jiang, L., Wang, S., Li, C., & Zhang, L. (2016). Structure extended multinomial naive Bayes. *Information Sciences*, *329*, 346–356.

Korstanje, J. (2021). *Advanced Forecasting with Python: With State-of-the-Art-Models Including LSTMs, Facebook's Prophet, and Amazon's DeepAR* (1st 2021.). Apress.

Wang, S., Jiang, L., & Li, C. (2015). Adapting naive Bayes tree for text classification. *Knowledge and Information Systems*, *44*(1), 77–89.

Yinhan, L., Myle, O., Naman, G., & Jingfei, D. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. Cornell University Library, arXiv.org.

Zhao, G. S., Luo, Y., Chen, Q., & Qian, X. (2023, January 21). *Aspect-based sentiment analysis via Multitask Learning for online reviews*. ScienceDirect. Retrieved April 17, 2023, from https://www.sciencedirect.com/science/article/pii/S095070512300076X