Analysis of TripAdvisor Hotels' Customer reviews using Sentiment Analysis

BIA 660-B: Web Mining

Instructor: Jingyi Sun

Group 4

Ronak Kachalia (10459133)
Devanshi Mehta (10459130)
Krina Shah (10459200)
Sacheth Shetty (10459184)

1. Good intro, but please include references. When writing a paper, don't make any claim without evidence. Elaboration of these references can be left to the lit review section, but in the intro we still want to see how you position your study in a field of research. 2. Because you haven't referred to cutting edge studies, your claim of "gap" is not very convincing. Please provide more explanation on your research question. I can guess what you are trying to do, but this is not clear to other readers. 3. Lit review only contains one study. The idea presented is good. So you have the argument why text mining of review is essential even with numerical rating. What you are trying to do is to identify additional aspects and analyze sentiment for those. Then you will need relevant studies on those as well. 4. You may combine research area and objective. 5. Please exclude code screenshots in the final report.6. Good EDA. In your final report, please provide more connection between your EDA and your research question. For example, how's seasonal change and super contributors can help you formulate a better analytical plan for your research question?

Introduction:

Analysis of user comments and reviews can help businesses in understanding how their customers are feeling about their products and services, which in turn provides deep insights to major stakeholders in the business on how to improve specific areas of products and services.

TripAdvisor is a travel company that assists its customers in finding the best rates for their hotel stay as well as booking tickets for their trip. One of the services it offers is their comprehensive hotel booking suite which enables its users to not only view hotels based on location, cost, cleanliness, and various other factors but also review the stay of other travelers at those hotels. The users are prompted to write a text-based review of more than 200 characters and provide an overall rating as well as a rating for cleanliness, rooms, and location as part of their review. Users can read thousands of reviews left by other users for a specific hotel before making their choice. These reviews are not only useful for other users, but they provide several insights to major stakeholders for the hotels which might help them improve the quality of their services.

TripAdvisor sticks to five main ratings for a specific hotel, namely cleanliness, rooms, value or price, services and location of the hotel, along with an overall rating for the hotel. However, it is not necessary that the guests are always looking for these specific services in the hotel. Adversely, the review left by the user might include more details about services which they might be unhappy about, however the overall numerical rating does not provide any information regarding the details of those services.

For example, A guest might be satisfied with the cleanliness of the hotel, their room size as well as the location, but they might be extremely unhappy with other services such as food or value for money. The guests might express these concerns in their text review and change the overall rating for the hotel, but this numerical rating does not provide enough information to the Hotel’s management team to make changes or improve their services.

Our project aims at bridging the gap between these text-based reviews using Sentimental Analysis as well as identifying certain other categories from popular words used in the review text which users have left for specific hotels. These new categories not only help the user’s narrow down their search for their perfect stay, but also helps the businesses to ascertain which services need to be improved in order to increase customer satisfaction, and bring in more business into their respective hotels.

Literature Review:

We reviewed the work detailed in the paper by Hsiu-Yuan Tsao and Ming-Yi Chen "The asymmetric effect of review valence on numerical rating", where the authors have conducted a sentiment analysis via text mining, using self-developed computer programs to retrieve a data set from the TripAdvisor website. This study finds there is an asymmetric relationship between review valence or the verbal review text and numerical rating. The authors further find brand strength to have an important moderating role. For a stronger brand, negative review content will have a greater impact on numerical ratings than positive review content, while for a weaker brand, positive review content will have a greater impact on numerical ratings than negative review content.

Therefore, the overall rating that is provided to a hotel is not a reliable measure of services offered by a specific hotel branch or customer satisfaction. The authors mention that assuming verbal review text is symmetrically related to the numerical rating might be a false one, since brand image is a significant factor that customers consider while writing these reviews on TripAdvisor. Similarly, other factors or services offered by a specific hotel might not be considered while providing their independent overall rating to the hotel. The authors further conclude that marketers could adopt sentiment analysis via text mining of online reviews as a valid measure or predictor of consumer satisfaction or numerical ratings. Strong brands should direct more attention to negative reviews, because in such reviews the negative impact transcends the positive. In contrast, weak brands should aim to exploit as many positive reviews as possible to minimize the impact of any negative reviews.

We noted that part of the "Brand Image" of the Hotel is simply just one of the factors that might affect the Review valence and overall rating. Other factors would include the services offered by the specific Hotel Branch, such as the quality of food and dining services, gym and fitness services, staff politeness, etc. All of these keywords can be identified and a sentiment analysis would provide us with more insights as to whether the customer reviewing the hotel had a positive or negative experience on these specific factors. This might in turn help us to bridge the gap between the review valence and the overall rating provided by TripAdvisor.

We also reviewed the work by Kudakwashe Zvarevashe and Oludayo O. Olugbara "A framework for sentiment analysis with opinion mining of hotel reviews". Because people frequently express their thoughts in complex and sometimes difficult-to-understand ways, computerized text data labeling is tough. The labeling procedure is time-consuming, and mislabeled datasets frequently result in wrong choices. In the paper, authors provide a framework for sentiment analysis with opinion mining for hotel customer feedback in this research. The majority of hotel review datasets are unlabeled, which creates a lot of effort for researchers in terms of text data pre-processing. Furthermore, sentiment datasets are frequently domain sensitive and difficult to construct, as sentiments include feelings such as emotions, attitudes, and opinions that are frequently rich with idioms, onomatopoeias, homophones, phonemes, alliterations, and acronyms. The framework proposed suggests that sentiment polarity is a suggested system that automatically prepares a sentiment dataset for training and testing in order to extract unbiased hotel service judgments from reviews. To find an appropriate machine learning algorithm for the framework's classification component, a comparison analysis was established using Naive Bayes multinomial, sequential minimal optimization, complement Naive Bayes and Composite hypercubes on iterated random projections.

As the paper mentions, datasets are domain specific. We decided to scrape data from the relevant source instead of training our models on existing datasets. Our approach uses Vader to generate labels and use BERT on top of it. It also uses different models on dataset in order to get better results that are more accurate and relevant. We also use compound scores to determine the polarity of the tweet.

Objective and Research Question

The three main research questions of your project is:

1. Are we able to predict overall rating based on the sentiment of the review text?
Given the user review text, is it possible to determine the sentiment of the overall text and does the sentiment score matches the overall rating value or not

2. What are the different categories/services in hotels that customers generally mention in their review for a Hotel?
There are various categories many user would like to know about a particular hotel, such as hows the services, hows the location, So we would like to find out what the the common categories guest generally mentioned about their stay in the hotel

3. Are the user reviews representative of the overall rating of the hotel on TripAdvisor?
Some guests might like almost everything about the hotel but had a bad experience, like the pool was closed. Thus, giving an overall rating of less than 3. On the contrary, Some guests didn’t have a good experience but were given a rating of 4.5 out of 5. Will it be possible to generate individual category scores based on the sentiment of the review text and categories mentioned in that text.

Methodology:

A. Data Crawler:

We scraped the first 20 pages of user reviews from TripAdvisor for 5 Hotels in New York City using python web mining libraries like selenium. Below shown is the user interface of Tripadvisor Website.

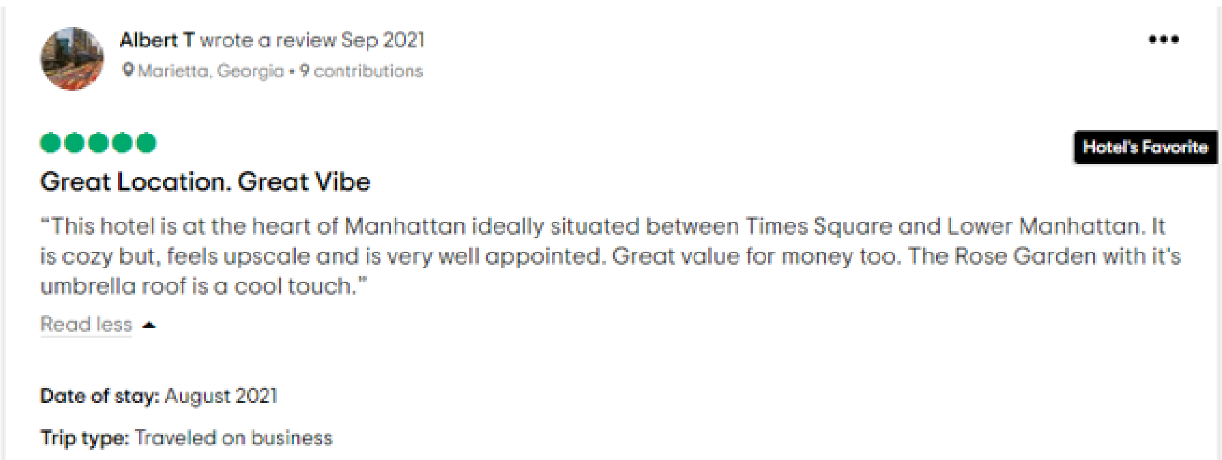

Figure 1: TripAdvisor user reviews

We are extracting the following data from the website:

- Name of the hotel
- Overall ratings
- Number of reviews
- Username of reviewer
- Review date
- No. of contributions
- No. of votes review received
- Reviewer's overall ratings
- Review title
- Review text
- Date of stay
- Individual category ratings (if any)

B. Description of the data:

Following is the snippet of our dataset:

hotel_n	overall_	date_of_revi	user_name	review_ratin	review_title	review_te	date_of_s	contributi	helpful_v	value_rati	rooms_rat	location_r	clean_rati	service_ra	sleep_rati
Innside	4.5	22-Mar	Kim C	5	Great hotel!	We stayed	March 202	0	0	0	0	0	0	0	0
Innside	4.5	21-Jun	Peggy M	5	The Innside	I just want	May 2021	8	11	0	4	0	5	5	0
Innside	4.5	04-Mar	Imran	5	NYC gem	I go to NY	June 2021	2	1	5	0	5	0	5	0
Innside	4.5	02-Mar	Jay B	5	Great Experi	I don't usu	February	0	1	5	5	0	0	5	0
Innside	4.5	22-Feb	Jeweliana15	5	Amazing Ho	I've stayed	February	25	4	5	0	0	0	5	0

Figure 2: First 5 rows of scraped data from TripAdvisor

Column	**Description**
Hotel Name	Name of the hotel
Overall Rating	On the scale of 5, the overall rating of the hotel.
Date of Review	Date on which they posted the review
User Name	User name of the guest that posted the review
Review Rating	On the scale of 5, rating provided by the user
Review Title	The title of the review posted
Review Text	Describes the experience of the user stayed in a particular hotel
Date of Stay	Date when the user stayed in the hotel
Contribution	Number of reviews provided by the user
Helpful Votes	Number of likes a review has received by other users
Value Rating	Rating about how does user felt about the price for a stay on scale of 5
Room Rating	Rating about how the room was in a hotel on the scale of 5
Location Rating	Rating about the location of the hotel on the scale of 5
Clean Rating	Rating about the cleanliness of the hotel on the scale of 5
Service Rating	Rating about the service provided of the hotel on the scale of 5

Sleep Rating	Rating about the sleep quality on the scale of 5

C. Analytical Strategy:

1. Data Pre-processing:

The scraped data from the TripAdvisor website included a number of columns where the data provided was inconsistent. Many columns in the scraped data contained both an object of strings, single string values as well as integer values in the same column. Therefore, we used certain pre-processing steps in order to clean the data and make it a little more consistent in order to perform exploratory data analysis. This will provide us with useful insights into the scrapped hotel review data.

- Dropped column 'link' which is not useful for our analysis.The 'link' column contained a URL link which redirected to the actual review in the TripAdvisor website. We concluded that this data was not useful to us in order to perform any analysis.

- Replaced all values in the 'date_of_review' column which contained string values such as 'reviewed today' or 'reviewed yesterday' with consistent date value.

- The column 'date_of_review' contained values such as 'reviewed today' & 'reviewed yesterday' which was inconsistent with the other values in the column which were in the format of a data value. (21-Mar or YY- shorthand month name). We replaced these values with the current month and year value of '22-Mar'.

- We noted that the columns containing integer review ratings for the below given rating columns, contained an integer score out of a 50 which we reduced by a factor of 10 in order to keep it consistent with the reviews present in the Trivago website.

1. Overall rating
2. Value rating
3. Location rating
4. Clean Rating
5. Service Rating
6. Sleep Rating

- We noted that columns of 'date_of_stay' and 'date_of_review' contained different date formats.We noted that the above two columns contained values in 2 different date formats.In order to perform our EDA over the months in a year, we extracted only the month parameter in the date values from both the columns and stored them in new columns 'month_of_review' and 'month_of_stay'.

- Based on the months extracted from the ‘date_of_review’ & ‘date_of_stay’ column values, we grouped the reviews based on the quarter of the year.We also clubbed the reviews based on the financial quarters in a year (Q1,Q2,Q3,Q4) based on the values in the above columns and stored these values in 2 different columns, ‘quarter_of_review’ & ‘quarter_of_stay’. We will further use these new columns to perform EDA in order to understand if there are any insights we can gain based on quarter wise review distribution.

2. Exploratory Data Analysis (EDA)

- Correlation

Initially to get a better understanding about dependent features in our data set we plot the correlation between our variables using heat map. Using the correlation matrix we noted that the rating values are the very closely related.

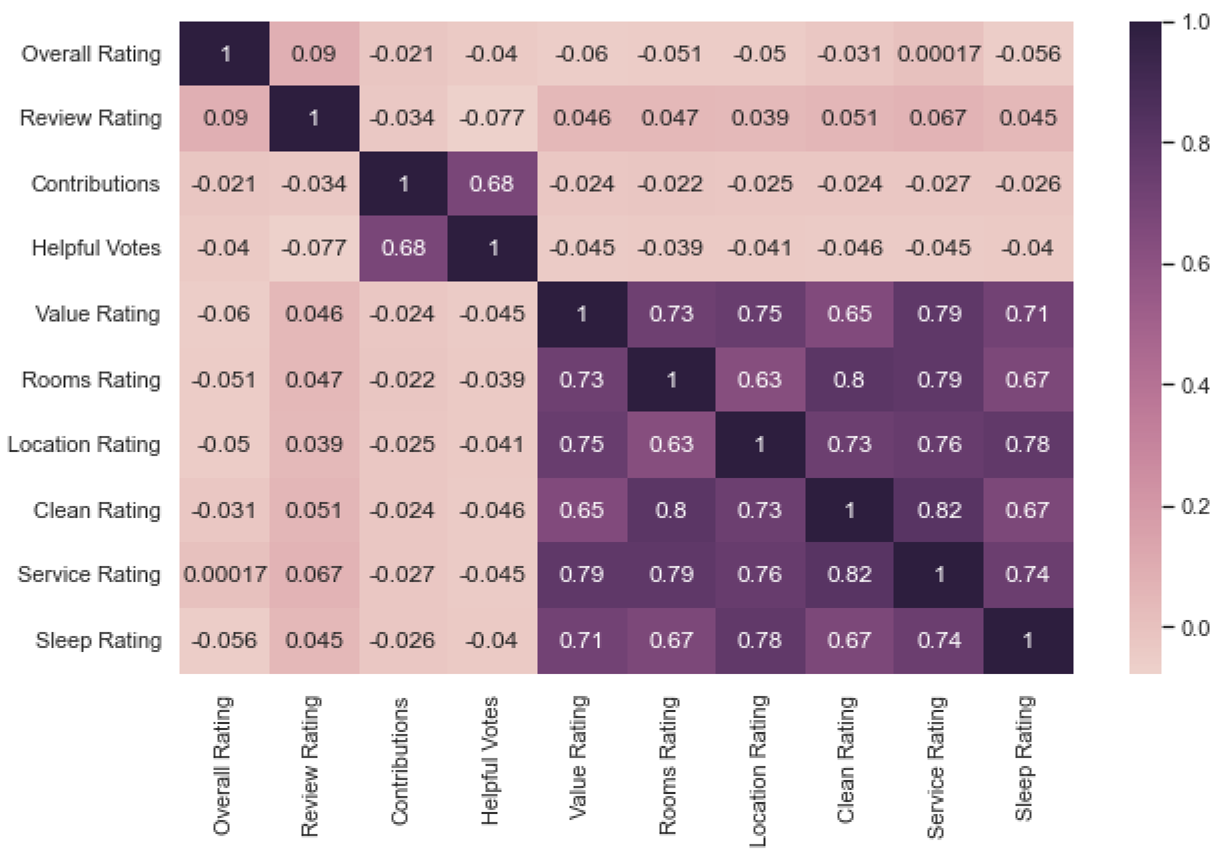

Figure 3 : Correlation matrix

- Hotel Name v/s Review rating

We plotted a bar graph that displays the count of unique review ratings per hotel.Here, we can observe that most of the hotels have a high density of 5 rating. This means that the majority of users have reviewed most of the Hotel and provided them with a rating of 5. Also, hotels like ‘Motto by Hilton New York Chelsea’ do not have any reviews which have 1 or 2 stars rating.

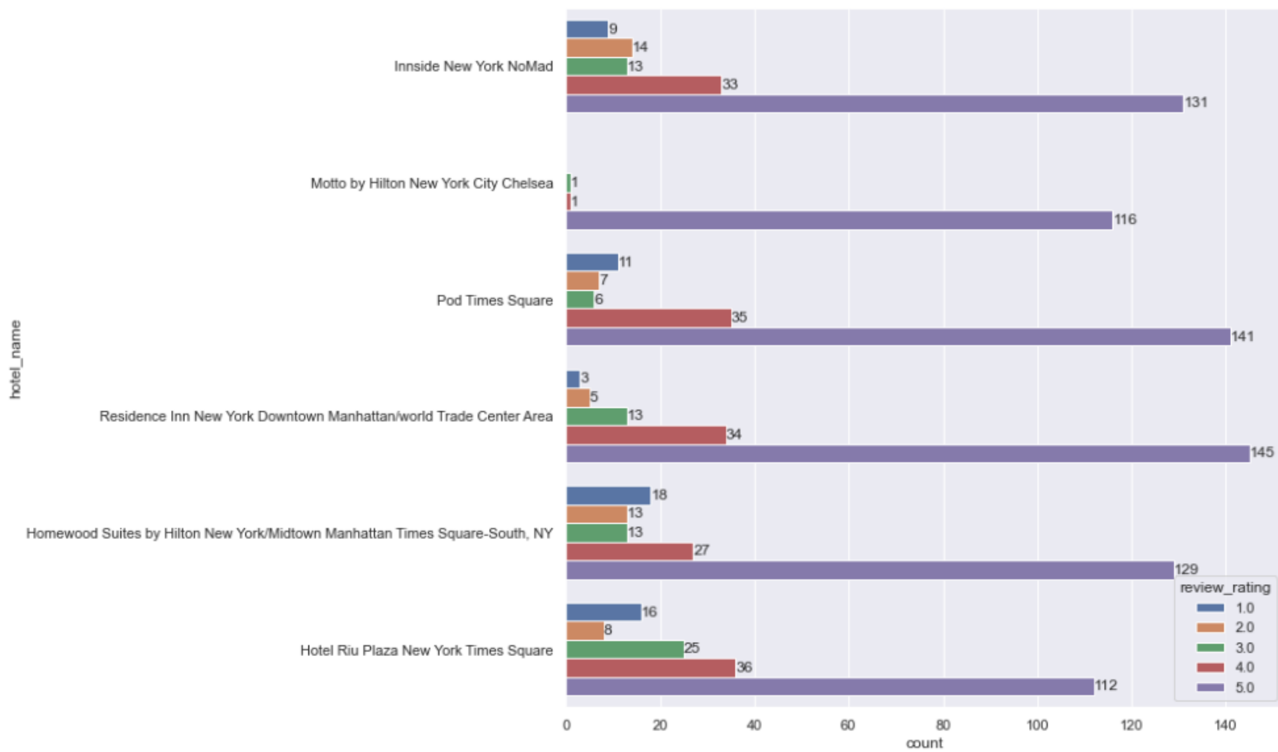

Figure 4 : Plot of Hotel_name v/s Review Rating

- Users who have provided a review 5 times before(contribution>5)
  Here we will say that assuming that the reviewers who have provided a review on TripAdvisor before give better reviews in terms of the “review_text” quality. Therefore we specifically targeted users who have provided more than 5 reviews.

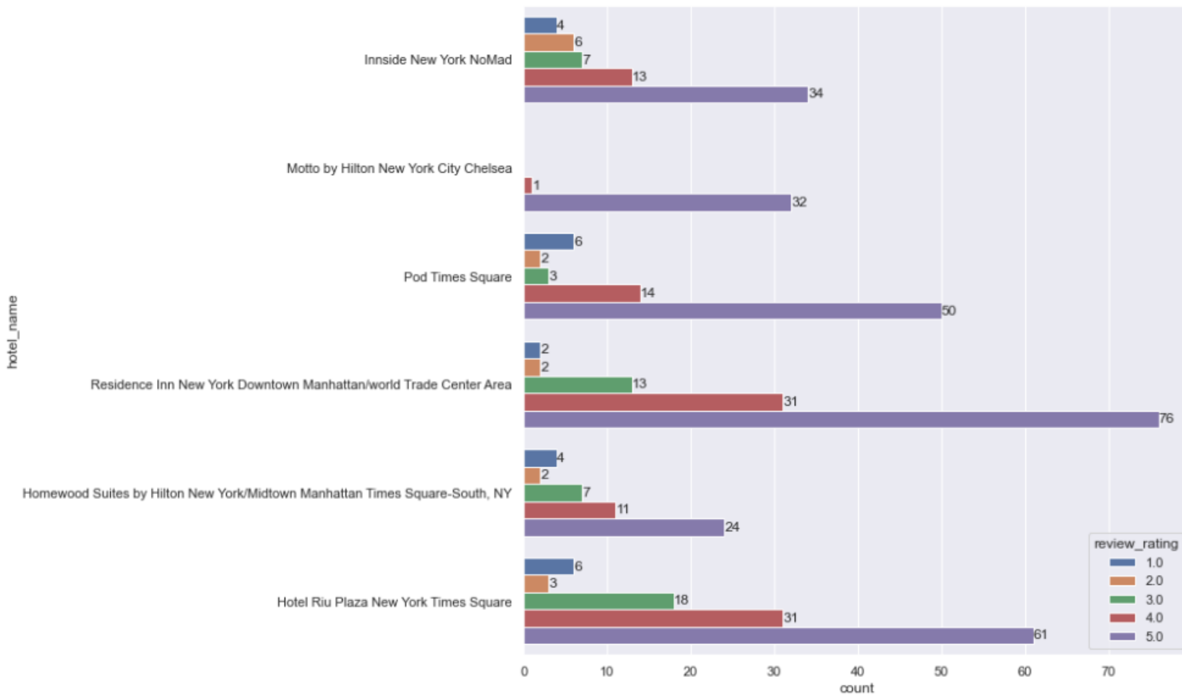

Figure 5: Plot of Hotel_name v/s Review Rating for user who have provided 5 or more reviews

- Word cloud of review title
  Further we made a word cloud of reviews in order to gain some more insight as to the most commonly used words used to describe a hotel as well the most commonly used words to describe synonyms of the services provided in the hotel such as “room” and “bedroom”. Furthermore we noted that certain words are repeated by the user’s who provided these reviews such as “location”, “room”, “service”,etc which coincided with our ratings we extracted as part of the data scraping procedures. This further goes to show that the choice of words used by the reviewers might be useful when assigning a numerical rating to specific services in the hotel. Furthermore we can also say that based on the choice of words used in the review, through sentiment analysis we might be able to provide a more accurate user rating.

Figure 6 : Word cloud containing most frequently used words in review_text

4. Sentiment Analysis

Based on the EDA processes performed on the data we noted that even though most of the reviews are positive based on the ratings provided by the users on TripAdvisor, the review text from the text cloud analysis shows that there are majority of negative words used to describe the different services and aspects of hotels which may or may not be taken into consideration by the users while providing an overall rating for their stay at the Hotel. The hotel might lose valuable customer feedback under the over encompassing nature of the "overall rating". A novel solution to this, would be applying sentiment analysis in order to obtain a score of the positive or negative sentiment each review has on the services of the hotel. Information related to what a guest might have liked or disliked during the duration of their stay at the hotel would be analyzed and assigned a score. This score could ideally provide the hotel useful insights into what services need to be worked on and what services leave a positive impression on the guest.

- Sentiment analysis can help us in identifying what services have had a positive or negative impact on the guest's stay at a particular hotel
- This kind of analysis can also help users identify the pros and cons of a specific Hotel
- This helps hotels to understand their business better and make improvements to services based on the overall sentiment regarding a particular service line

In order to understand what the service or aspect of the hotel a user review is describing in the hotel we would need to perform more text processing on the review itself in order to provide a little more "context" to each review in our dataset. The "context" here would be the services or aspects of the Hotel itself, which are:

1. Room
2. Location
3. Value

4. Clean / Cleanliness
5. Service (room service, restaurants,gym,etc)
6. Other (speciality services)

A. Text Processing Pipeline

PART 1: Prediction the sentiment of the review text using BERT model where Source Variable is review text and Target Variable is Label generated by the Vader Analysis.

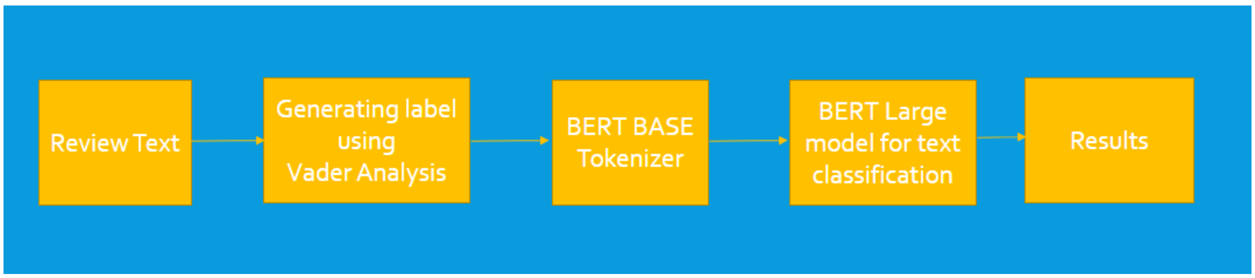

Figure 7 : Text processing pipeline in order to assign context to BERT analysis

PART 2 : Predicting Individual Categorical Value based on review Tex

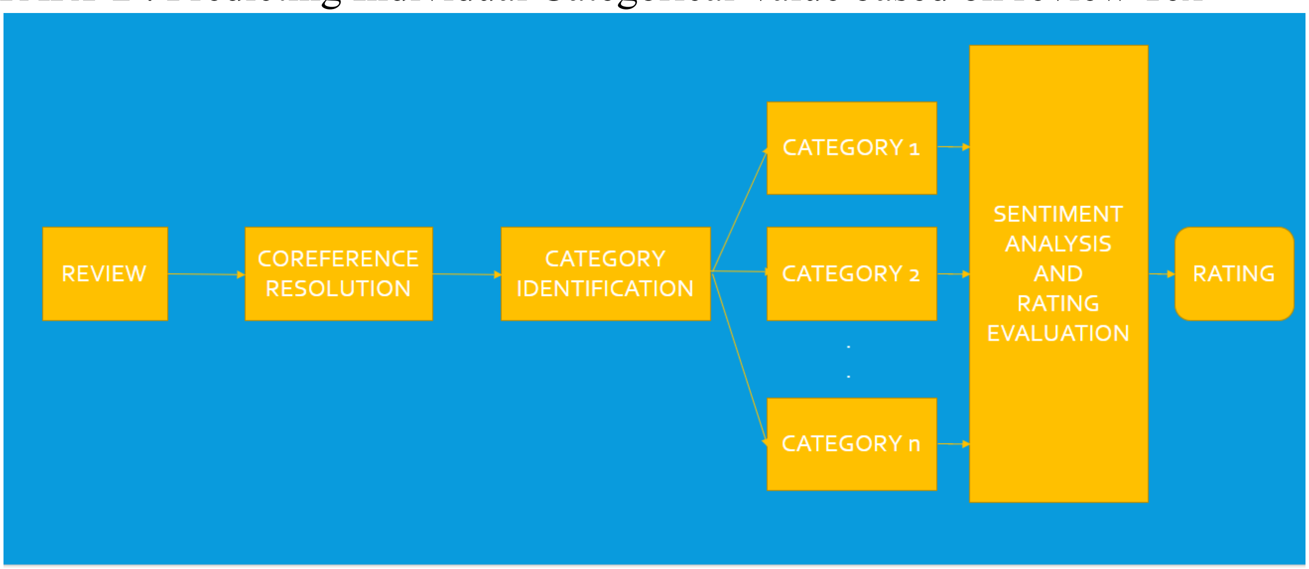

Figure 8 : Text processing pipeline in order to assign context to categories in the VADER model

- Algorithms used for Sentiment Analysis

  Part 1:

  ## VADER Model

  VADER (Valence Aware Dictionary and Sentiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media, and works well on texts from other domains. It analyses a piece of text to see if any of the words in the text is present in the lexicon. Sentiment metrics are derived from the ratings of such words. This form of sentiment analysis fits the use case of hotel reviews perfectly, since all of the user reviews have been scraped directly from the TripAdvisor website and we need to analyze the sentiment where we will make use of VADER model polarity scores (positive, negative,neutral and compound).

  - Positive: Review has a positive sentiment based on the lexicon comparison
  - Negative: Review has a negative sentiment based on the lexicon comparison
  - Neutral: Review has neither a positive nor negative sentiment based on the lexicon comparison
  - Compound: It is the normalized sum of all the above scores.

  ## BERT Model

  "BERT stands for Bidirectional Encoder Representations from Transformer. BERT is based on the Transformer architecture. As BERT is Bidirectional it earns information from both the left and the right side of a token's context during the training phase.
  The BERT architecture builds on top of the Transformer. Currently there are 2 models :

- BERT Base: 12 layers (transformer blocks), 12 attention heads, and 110 million parameters
- BERT Large: 24 layers (transformer blocks), 16 attention heads and, 340 million parameters

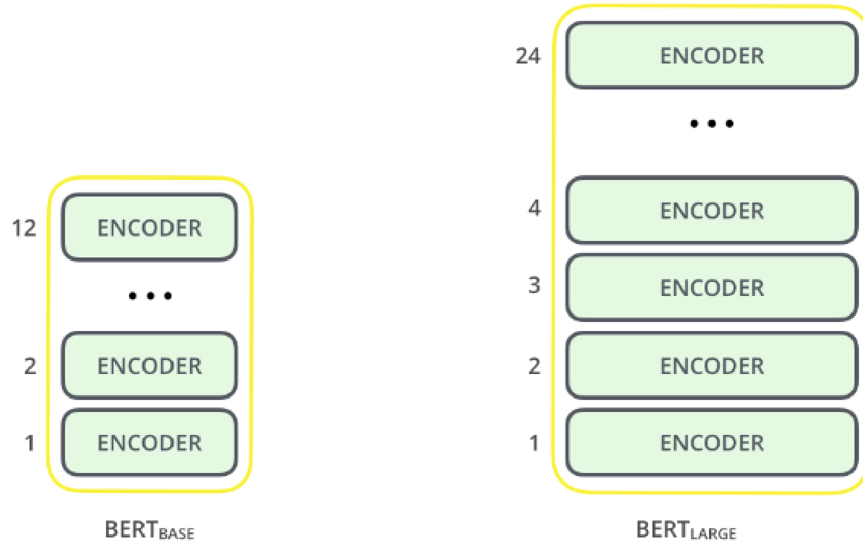

Figure 9 : Bert base and Bert Large encoding layers

Since the BERT model requires labels in order to predict the sentiments for the reviews, we will be using the BERT in conjunction with VADER which is a lexicon based analysis tool which analyses a piece of text to see if any of the words in the text is present in the lexicon. Sentiment metrics are derived from the ratings of such words. These sentiment metrics will function as labels i.e. positive and negative sentiment labels when we feed this data into the BERT model.

We passed review text to our VADER analysis model, which generated polarity scores as positive, negative and compound values. Refer Figure 10. As we can see, We passed the whole review text to our VADER analysis model and the output was generated as a polarity score of positive, negative and compound. Those reviews whose positive polarity score is greater than negative polarity score are labeled as "1" which indicates positive sentiment while those with negative polarity score greater than positive polarity score are labeled "0" signifying negative sentiment.

After Generating the labels, We split our data into training, testing and validation as 60-20-20 percent.

We generated a tokenizer using the AutoTokenizer from the 'bert-base-uncased model. We padded each review with the max length of the review text. We set the parameters as return_token_type_ids = False and return_attention_mask = True, As Attention mask will help in identifying the long term dependencies of the text.
After generating the tokenizer we train 'bert-large-uncased' using BertForSequenceClassification pretrain model. Our Training parameters were num_train_epochs = 10 (Epoches)

per_device_train_batch_size = 16(Training Batch size)
per_device_eval_batch_size = 64 (Evaluation Batch size)
weight_decay = 0.01 (learning rate)

Vader Results:

We use the entire review text for each user review and pass it to the VADER analysis model in order to receive the labeled data which would in turn be required for the BERT model. The below given figure contains review text for a single user review from TripAdvisor which has been passed to the VADER model. We obtained polarity scores from the VADER's lexicon based analysis and have used only the "neg" & "pos" values. In this case we noted that the positive polarity score is greater than the negative polarity score, therefore we assign a label of "1" to this particular user review. This label signifies a positive sentiment.

```
Review:
 We stayed in a family loft which is two connecting rooms, one with a king and one with 2 doubles. Each has a bathroom. So pe
rfect for our family of 5! The hotel and our room is very clean. The beds are so comfortable.  The front desk staff is amazin
g and provides us as much bottled water as we need. The supervisor Kima made sure we were comfortable and gave us an early ch
eck in which was so appreciated. The location is great, very easy to get anywhere we want. I would definitely stay here agai
n!

SIA scores:
 {'neg': 0.0, 'neu': 0.68, 'pos': 0.32, 'compound': 0.9896}
```

Figure 10 : VADER analysis on the entire review text.

BERT Results:

For testing purposes, we converted those given a review rating of more than 3 as a positive sentiment otherwise a negative sentiment. We noted that the training and testing loss is decreasing over the period of time. This indicates our model fits well.
We got a training accuracy of **93.45%** and that of test accuracy of **91.34%**
Training **loss** was approximately **0.306** and that of test loss was approximately **0.454**.
We also used Vader analysis on the entire text and got an accuracy of **90.23%**

Figure 9 : BERT's training-validation loss graph

Test Results On Bert Model :

Test Loss	0.4542
Test Accuracy	0.9137
Test F1	0.9152
Test Precision	0.8437
Test Recall	1.0

Part 2:

VADER Analysis: Process

In this analysis we will be using VADER to assign a sentiment score to specific categories /services that a particular hotel has and that are mentioned in a user's review and based on these sentiment scores for each of these services we will be finding out the overall score for each review . Sentiment metrics are derived from the ratings of such words. This form of sentiment analysis fits the use case of hotel reviews perfectly, since all of the user reviews have been scraped directly from the TripAdvisor website. Based on the text processing we performed, we now have divided the review text into the following categories:

1. Room
2. Location
3. Value
4. Clean / Cleanliness
5. Service (room service, restaurants,gym,etc)
6. Other (speciality services)

These categories each contain review text, which is specifically talking about the category as a whole. For example, the category "room" will contain all the statements from the user review text which reference the word "room " and its synonyms in the text.

Performing VADER analysis on each and every category would ideally provide us with polarity scores of positive,negative, neutral and compound, thus giving an overall sentiment score to the review based on the specific service for a hotel. For our analysis we have made use of the compound score, which is the normalized sum of the positive, negative and neutral scores that are found for each category review text. We further noted that based on the text processing performed we had certain “NA" values in certain categories, since the user made no mention of that category in their review. We simply assigned a compound score of “0” for these “NA" review texts.

For example, if a user has provided their opinions on the room, location,cleanliness and value in their review text, but has not mentioned anything about the services and speciality services, then our VADER analysis would calculate the sentiment compound score for room, location,cleanliness and value only. The other 2 category compound scores simply get a 0 compound score.

Further in order to get the overall user rating based on the compound scores provided by VADER, we simply average the compound scores obtained in each of the 6 categories.An important thing to note here is that, as previously mentioned if a user does provide an opinion on a certain category, the compound score for that category is not taken in consideration while finding the overall compound score for the review.

For example, if a user has provided their opinion on all 6 categories, the average compound score is calculated using a total of 6 category compound scores.
If a user has provided their opinion on 4 out of 6 categories, the average compound score is calculated based on 4 categories only and not 6.

We then noted the data overall compound score for each user review obtained after a VADER analysis and noted that we could further assign a simple numeric user rating ranging from 1- 5, 5 being excellent and 1 being poor based on the threshold values of the compound score.We will call this rating as VADER rating. While deciding these threshold values we considered that the rating system would still follow a bell curve. The threshold values noted were-

- Average Compound rating less than -0.2 would be considered as VADER rating 1
- Average Compound rating greater than -0.2 but less than equal to 0 would be considered as VADER rating 2
- Average Compound rating greater than 0 but less than equal to 0.4 would be considered as VADER rating 3
- Average Compound rating greater than 0.4 but less than equal to 0.8 would be considered as VADER rating 4.
- Average Compound rating greater than 0.8 would be considered as VADER rating 5.

We applied these threshold limits to all the reviews and compound scores we had obtained in order to obtain the final rating results for VADER.

VADER Analysis: Results

We firstly compared and plotted the graph of the origin TripAdvisor rating provided by users for each review in order to have a base result to which we would compare the results of the VADER analysis. We then plotted the results obtained after text processing and applying VADER analysis in order to obtain the VADER rating for each review.

```
review_rating
5    717
4    151
3     64
1     52
2     40
dtype: int64
```

```
: <AxesSubplot:xlabel='review_rating'>
```

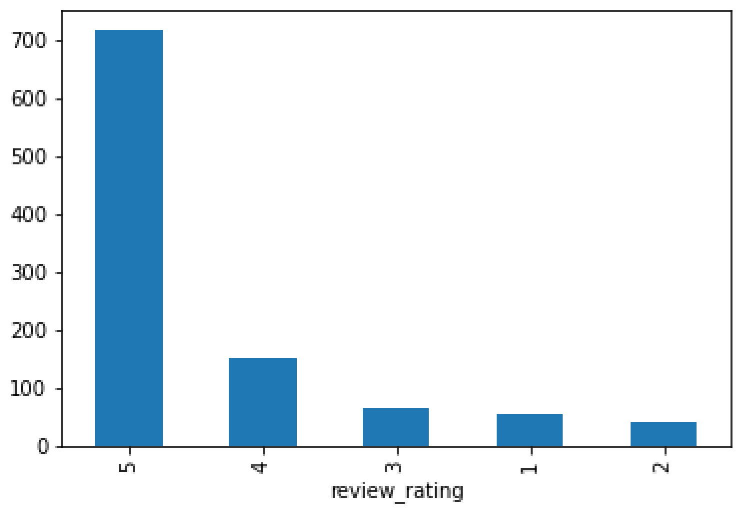

Figure 11: Count of all the review rating provided on TripAdvisor

```
Vader Rating
4    572
5    252
3    101
1     57
2     42
dtype: int64

<AxesSubplot:xlabel='Vader Rating'>
```

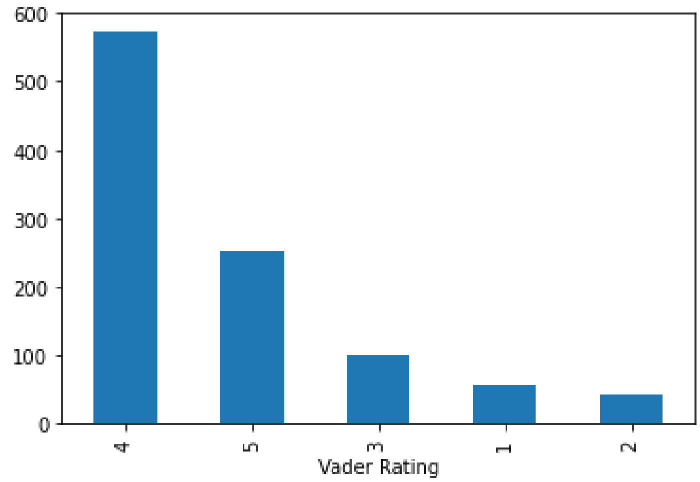

Figure 12: Count of VADER ratings found on the same user reviews.

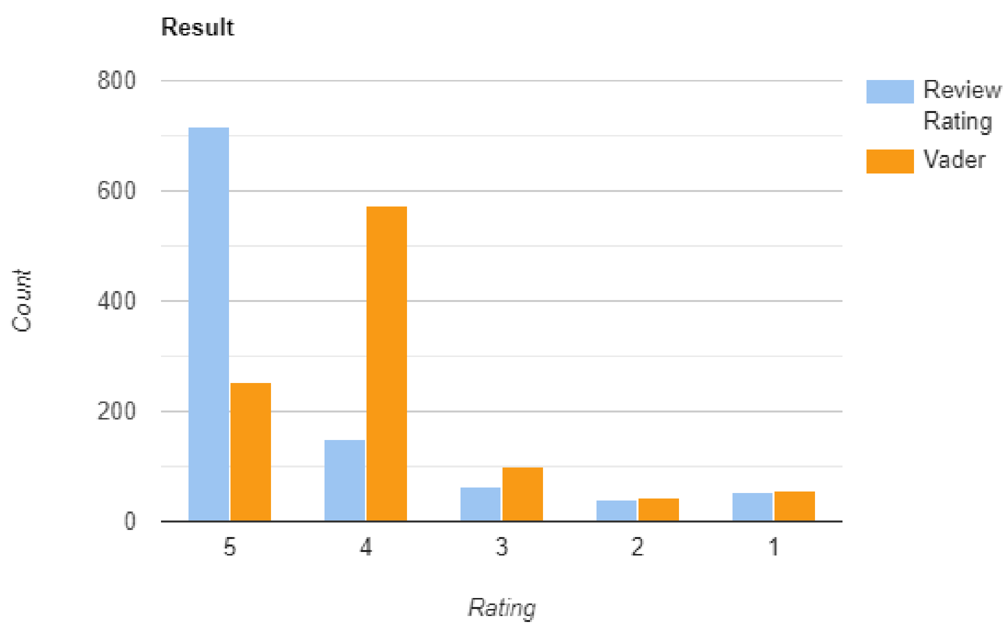

Figure 13: Count of all the review rating provided on TripAdvisor

We noted that 717 users have provided an overly positive rating of '5' on TripAdvisor, but 252 reviews had been placed in the rating 5 group, after the user review text had been analyzed and a sentiment had been assigned to it.

Similarly we noted that 151 users had provided a rating of '4' on TripAdvisor, however the 572 reviews had been placed in the VADER rating '4' group based on the review text sentiment.

We made a side-by-side comparison of the original review rating provided by users in TripAdvisor and the VADER rating (*Please refer to figure 13*). We noted that there was a significant difference in the TripAdvisor and VADER ratings in the range of 3 to 5. From this analysis we can say that users are more inclined to provide an overly positive review on TripAdvisor even though the review might have a few negative points. This feedback is extremely important for any management board of a hotel in assessing its staff and services. VADER analysis provides a novel solution to this problem by assigning a review score to services, based on the sentiment in a user's review.

Analyzing these results more closely we noted that user provided ratings do not correctly reflect what the user has mentioned as part of their review. Therefore user ratings are not a reliable metric for hotels to assess their services. We further looked at an example of user text in order to more closely examine the differences between these results.

Example :
Review Text: "*The hotel in general was very nice, very clean and we appreciated the upgrade, but very disappointed in the extra charge when I booked the hotel with Melia app. It gave me a final price with taxes and I got charged $184 usd extra. I send to emails asking why I never receive a response.*"

	Unnamed: 0	Unnamed: 0.1	review_text	text	review_rating
53	53	56	The hotel in general was very nice, very clea...	The hotel general wa nice, clean appreciated u...	4

Figure 14: Review data from dataset showing TripAdvisor rating of 4.

clean	...	value	other	Room Rating	Location Rating	Clean Rating	Service Rating	Value Rating	Other Rating	Avg Rating	Vader Rating
The hotel in general was very nice, very clea...	...	The hotel in general was very nice, very clea...	I send to emails asking why never recibe a res...	0.0	0.0	-0.0682	-0.0682	-0.0682	0.0	-0.07	2

Figure 15: Category based Review scores obtain through VADER as well as FINAL VADER rating '2'

Analyzing this review carefully we can see that even though the user was still left disappointed due to the extra charge that they were levied. Even though the user has left a "review_rating" of '4' on TripAdvisor, their review itself does not correctly reflect the rating. Our sentiment analysis firstly divided the entire review text based on categories "value", "service" &"clean" . We can see that VADER has assigned a negative sentiment to each of the categories, thus obtaining an average rating of "-0.7" which based on our threshold values is converted to a VADER Rating of 2.

Conclusion:

When comparing the performance of the VADER model and BERT Model on the entire user review text, we noted that BERT model had test accuracy of 91.34% while the VADER model had a test accuracy 90.23%. Therefore we can conclude that the BERT model performs better as compared to the VADER model when applied to the entire user review text.

The VADER model on the other hand is more versatile, as we saw in the PART 2 of report. Breaking the entire user review text based on the services offered by the hotel, provided a score rating for each of the services in the Hotel. Thus providing useful insights into what the hotel should improve on and what services are faring well with the guests. The analysis also showed that most users are more likely to give a positive rating on TripAdvisor even though they express negative feedback in their user review, thus bridging the gap between users reviews and user rating score.

Limitations

- **If more services are to be analyzed or if new services are added to a hotel, we would need to modify the synonym set used by VADER service Rating model in order to include the new or modified services.**
- **Based on when the new service or services, have been implemented by a hotel, our model would require a significant amount of time to in order for users to review the specific service in a hotel**

Future Work

- **The scope of this project can be expanded to sub-categories of an Hotel's services. For example , the Room category can be subdivided into categories related to the room like bed, sheets, furniture, etc.**
- **Clustering using historical data to check if there is a change in the sentiment of people over time.**