**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

|  |  |
| --- | --- |
| **Module Title:** | Strategic Thinking |
| **Assessment Title:** | CA 2 |
| **Lecturer Name:** | James Garza ([james@cct.ie](mailto:james@cct.ie)) |
| **Student Full Name:** | Salvador Navarro Montaño |
| **Student Number:** | [2023212](mailto:%320%32%33%32%31%32@%73t%75%64en%74%2e%63c%74.%69e) |
| **Assessment Due Date:** | 18/12/2023 |
| **Date of Submission:** | 20/12/2023 |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

Report CA#2

Objective

The main objective of the project is to review our internet reviews, insights that might be either positive or negatives comments by customers who previously reserved with us, this will lead us to a new road to improve the marketing strategy. We know that we cannot use comments as data but we will clean our data and transform it into sentiments to do a better understanding on what guest have experienced in the hotel. furthermore, the following experiment with the insights might help in the understanding of what customers have in their mind after they finish conclude their stay with us.

Problem Definition

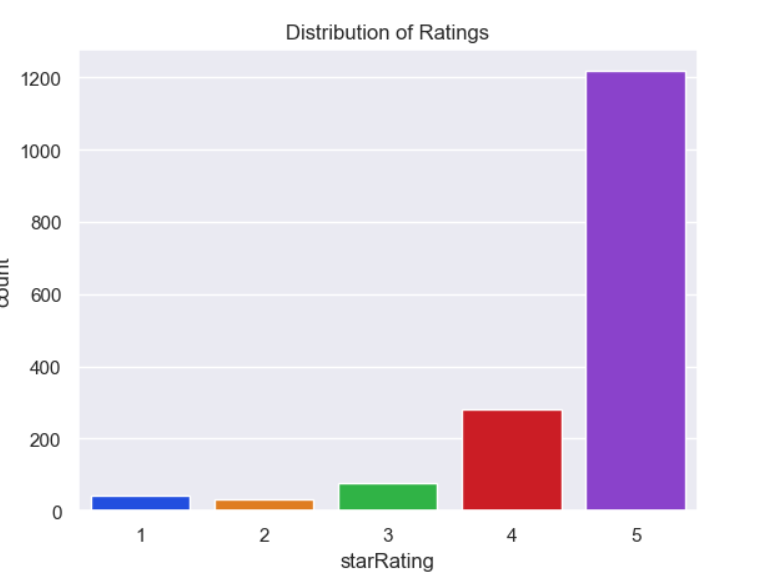
Minimizing cost while optimizing revenues is one of the main objectives for any company around the world, nowadays with the huge users using internet, social-creator, influencers any review from them can affect the status and marketing from the hotel, either they are given a favorable image or bad reputation, we should consider any kind of review with the same valuable, then recognizing the crucial insights where the clients experience were not the expected or discontent. Recognizing and addressing these concerns is priority for the company to improve the organizational culture.

SCOPE

When we began this project, the scope of this project were amend to understand the business of Food and Beverage/sales and weather from the hotel where we are currently working but due the unavailability of necessary data requested to the General Manager made us to shift our data, therefore the target persist to continue with a marketing case study in the hospitality sector, a change from food and beverage, sales and weather was needed to a new data set from reviews, insights taken from internet from our guests.

At first glance it was clearly that there were a lot of text we need to process in order to be able to turn our text into a sentiment analysis for a good application of any Machine Learning model, once started the CRISP-DM on the evaluation it was easy to visualize how the rating were distributed, most of them gave 5 starts which mean a positive review.

Most of the reviews were in Portuguese and they needed to be translated to English using Google Translated, we focus into the English text analysis but keeping the original text to keep working on in next semester if needed.

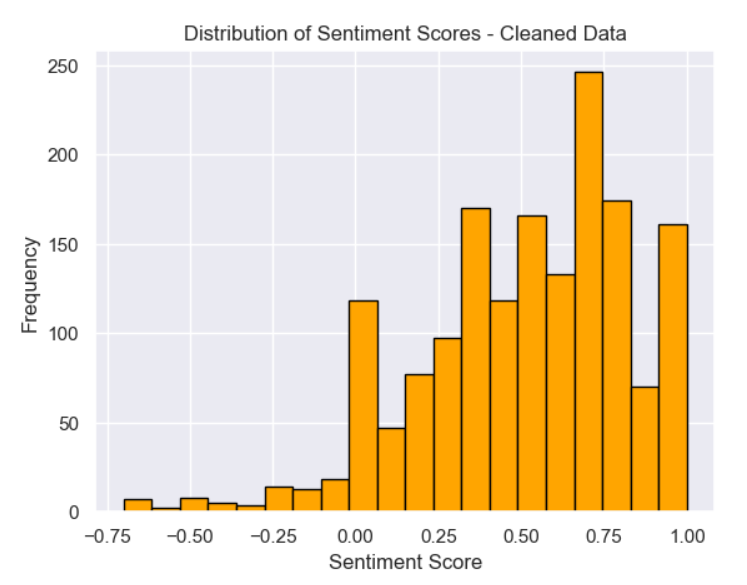


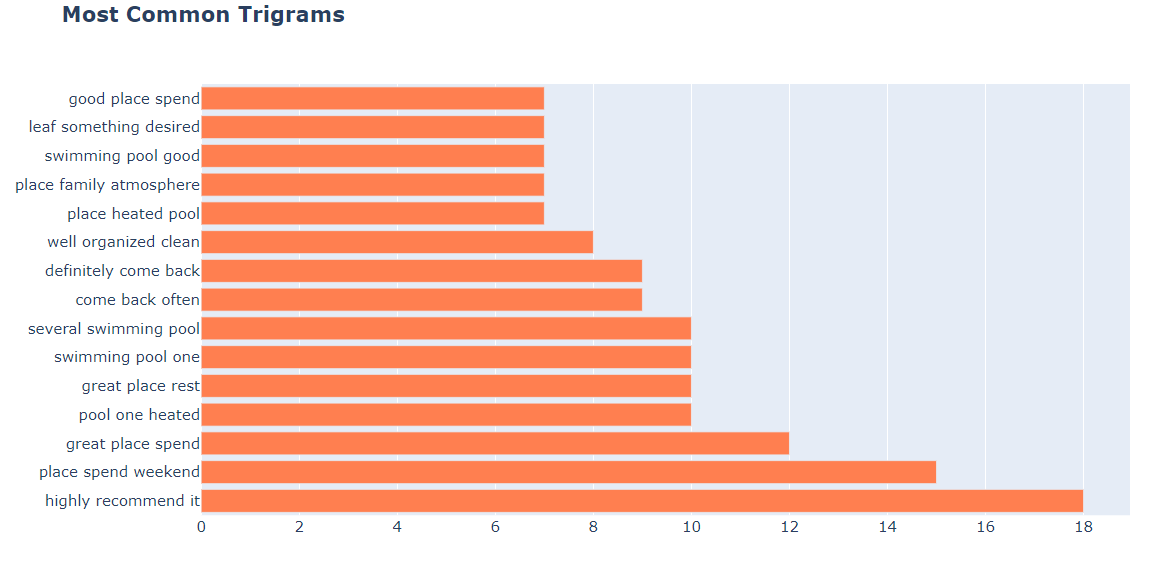
While in the sentiment score from -1 taken as negative reviews up to 1 which is a positive review clearly most of the data are distributed positive, exploring the most common words we have in the seniment analysis the #1 were ‘’highly recommend it’’ given us a total of :

Number of positive reviews: 1492

Number of negative reviews: 60

Number of neutral reviews: 96





Based on the sentimental score from the confuxion matrix It calculated the ratio of true positive predictions to the sum of true positives and false positives.

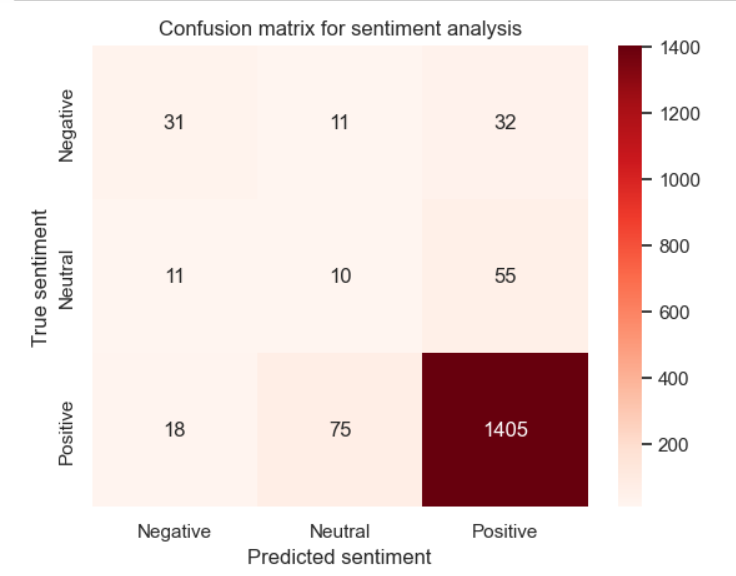
Precision for each class:

Negative: 0.52% of predicted negatives were actually negative (32)

Neutral: 0.10% of predicted neutrals were actually neutral (55)

Positive: 0.94% of predicted positives were actually positive (1405)

Total positives: 1498



total for each class:

Negative: 74 instances

Neutral: 76 instances

Positive: 1498 instances

Accuracy: 0.88 (88% of all predictions were correct)

.

In summary, the model performs well in predicting positive sentiments, with high precision, recall, and F1-Score. However, it struggles with neutral sentiments, as indicated by lower precision, recall, and F1-Score. The classification report provides a comprehensive view of the model's performance for each sentiment class. (Neutral, Positive, Negative)

Random forest with start rating evaluated the performance from the classes denoted as 0, 1, and 2

if rating in positive: 2

if rating in negative: 1

Neutral: 0

the accuracy is approximately 91.82%, indicating that the model correctly predicted the class labels for about 91.82% of the instances in the dataset.

Class 0:

Precision: 0.00

Recall: 0.00

F1-Score: 0.00

Class 1:

Precision: 1.00

Recall: 0.03 (Only 3% of actual instances of class 1 were correctly predicted)

F1-Score: 0.07

Class 2:

Precision: 0.92 (92% of instances predicted as class 2 were correct)

Recall: 1.00 (All actual instances of class 2 were correctly predicted)

F1-Score: 0.96

Total amount of Classes

Class 0: 25 instances

Class 1: 29 instances

Class 2: 606 instances

The model demonstrates high accuracy. The model performs well for class 2, achieving high precision, recall, and F1-Score. However, for classes 0 and 1, the model's performance is notably lower which is notable our low average in neutral and positives ratings.

//////

Random forest with sentiment score

The provided classification report evaluate the performance of a sentiment analysis model with three classes denoted as 0, 1, and 2, where 2 corresponds to "positive," 1 to "neutral," and 0 to "negative."

the accuracy is approximately 81.62%, indicating that the model correctly predicted the sentiment labels for about 81.62% of the instances in the dataset.

Class 0

Precision: 0.50

Recall: 0.05

F1-Score: 0.09

Class 1

Precision: 0.74

Recall: 0.88

F1-Score: 0.81

Class 2

Precision: 0.90

Recall: 0.82 (82% of actual instances of positive were correctly predicted)

F1-Score: 0.86

Class 0 (Negative sentiment): 21 instances

Class 1 ("Neutral sentiment"): 213 instances

Class 2 ("Positive"): 261 instances

In summary, the model demonstrates relatively good performance for classes Neutral and Positive with high precision, with an accuracy on Training Set: 0.997398091934085 and Accuracy on Test Set: 0.8161616161616162

////

To Summary

The decision on use Random Forest it was because we noticed a high accuracy and this model reduce the risk of overfitting. With the goal of exploring novel techniques and conducting a comparative analysis of their results. In the upcoming semester we plan to investigate and use alternative model to make variations in the sentiment analysis

the project is strategically crafted to offer practical insights into guest perceptions by harnessing sophisticated analytical methods within the hospitality sector. The thorough methodology, which involves tasks such as data preprocessing, sentiment analysis, and machine learning, holds the potential to reveal insightful patterns and correlations. This is expected to significantly contribute to the refinement and optimization of the hotel marketing strategy.