**Final Report**

**Project Title:** **Hotel reservation prediction**

**Problem Statement:**

When we plan a business trip or a family trip the first thing, we think about is where do we stay? It is quite normal that we might make a booking at a hotel for a few days and at the last minute due to a sudden change in plans, we may cancel the booking. Hotels generally come across this behavior and few a times they must drop a customer due to the existing booking due to which their revenue is highly affected. Therefore, using classification models, we are trying to leverage the guest dataset and perform an analysis of the Room Booking Behavior Pattern and then devise a strategy for the Hotel Revenue Management.

**Data Description:**

[**https://www.kaggle.com/code/sanjana08/hotel-booking-cancellation-prediction/data**](https://www.kaggle.com/code/sanjana08/hotel-booking-cancellation-prediction/data)

The Dataset set contains 1.1 million observations and 33 Columns.

**Methodology:**

* Firstly, we obtained the dataset from Kaggle. Extracted dataset into SQL lite database, remove and modify the incomplete, null, and duplicate data.
* Normalized the database and imported the data into the jupyter notebook and start the analysis.
* Perform Bi-variate analysis and generate derived columns and create a new table with updated columns.
* Bi-variate analysis is performed by generating various bar plots to understand the relationship between features.
* Splitting the dataset to the Training set (70%) and testing set (30%).
* Applying machine learning classification models to the dataset. Models used on: Logistic regression, Decision Tree Classifier, and Random Forest Classification.
* Constructed a dashboard with insights into the data.

**Results:**

Firstly, the features are compared with the property variable to find the feature that best fits our model, this comparison is called Bi-variate analysis.

The following bar graphs plotted are used to compare features:

1. Is\_canceled vs meal
2. Is\_canceled vs is\_reapeated guest
3. Is\_canceled vs market\_segment
4. Is\_canceled vs arrival\_date\_month
5. Is\_canceled vs stay\_in\_weekend\_nights
6. Is\_canceled vs stay\_in\_week\_nights
7. Is\_canceled vs previous\_cancellation

On comparing the features with the target variable we see a change of trend between the 3,4,5,6 and 7 from which we can deduce that the above features will be a good fit for our model.

Chart, bar chart

Description automatically generated Chart, box and whisker chart

Description automatically generated

Chart, bar chart

Description automatically generated Chart, bar chart

Description automatically generated

Chart, bar chart, histogram

Description automatically generated Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated Chart, bar chart

Description automatically generated

Classification models are applied to the dataset with new filtered features, and we generated a confusion matrix and a classification report which are the metrics used to test the classification models.

* Accuracy of logistic regression classifier on test set: 0.62
* Accuracy of Decision Tree classifier on test set: 0.84
* Accuracy of Random Forest classifier on test set: 0.82

The decision tree gives us the best accuracy for the given data set whereas Logistic Regression is the worst.

Comparing the models, it can be said that Random Forest has made the best prediction as it prevents overfitting.

The Receiver Operating Characteristics graph is plotted to compare the number of False Positive rates and True Positive rates of all three models. Helping us to under which models has given us the best prediction values.

Chart, treemap chart

Description automatically generated Table

Description automatically generated with low confidence

Chart, line chart

Description automatically generated

**Conclusion:**

The main objective of the project is it generates a report for the Hotel Revenue Management team to understand the trend in the cancelation by the various customer over the period while will help the hotel revenue team analyze and make sure that the customer who has currently made a reservation will be staying in the hotel or not, or is he/she likely to cancel the reservation as before. This way the hotel can take up reservations from other customers and keep the business running.

**Project Files Link:**

<https://buffalo.box.com/s/1jjfgs6x8oqruqn93xifx3vl15ewcbch>

**Members:**

1. Nisarg Negi
2. Shraddha Waphare
3. Siddhi Jadhav
4. Jagath Sai Narayana Kakaraparty