**Predicting Hotel Booking Cancellations**

Student’s Name

Professor’s Name

Institutional Affiliation

Course Code

Due Date

**Executive Summary**

**Data Set Overview**

The dataset selected for analysis is the "Hotel Booking Demand" dataset, obtained from the MISCADA ASML Classification summative coursework page. This dataset contains information about hotel bookings, including features such as booking dates, number of adults, children, and babies, as well as special requests made by guests. Each entry represents a unique booking, detailing factors like the number of guests, meal preferences, market segment, and previous booking history It also includes a binary target variable indicating whether a booking was canceled or not.

**Real-World Objective**

Our study aims to predict hotel cancellations, which is fundamental for management to use resources best and increase revenue. Cancellations influence occupancy rates and income projections, so making exact plans for strategies that utilize assets shrewdly and increase income is essential. Hotels can avoid losing money on cancellations and keep operations running smoothly by planning; this increases the quality of administrations conveyed to customers, therefore increasing customer fulfillment.

**Model Performance**

Our chosen model gets an exactness of approximately 75.36% on the test set after a parcel of research and preparation. For genuine life, this implies that our demonstration correctly guesses the state of roughly 75 out of each 100 bookings as cancellations. This level of exactness could be better, but it gives hotel supervisors valuable data to assist them in arranging for and handling cancellations well.

**Technical Summary**

**Problem Description**

The dataset has many different types of information about hotel bookings, such as booking times, customer demographics, and ticket details. It's your job to guess whether a reservation will be canceled.

**Initial Data Summary**

The primary step in our think about was to figure out how the information was organized and how complicated it was. To begin with, we utilized rundown insights to discover the means and ranges of many diverse components. We paid close attention to finding any numbers that were lost and got distant better, much better, higher, stronger, an improved">an improved sense of how the information was spread out. We considered what these misplaced numbers might be cruel to think and came up with great ways to handle them. Box plots, histograms, and other sorts of charts made a difference in us seeing how the imperative components were spread out and spot any patterns or stands out. We know everything about the data after this close examination. Ready to, at that point, make savvy choices amid the arranging handle, and this moreover sets the organization up for encouraging inquiries.

**Simple Visualizations**

To show how factors and their ranges were associated, we utilized different chart sorts, including histograms, bar plots, and relationship networks. That was a basic way to get it the patterns and designs within the data.

**Model Fitting**

We chose a random forest classifier to prepare the model since it can bargain with massive datasets and avoid overfitting issues. Methods like train/test part and cross-validation made the model more solid and valuable in more extensive circumstances. The random forest algorithm's outfit learning method, which combines predictions from a few choice trees, discovered complicated associations within the information, bringing down the chance of overfitting. Using these methods, we formed an expectation show that might correctly learn patterns in inn booking information and help hospitality management make vital decisions.

**Model Improvements**

Different methods were used to improve the model's performance, counting highlight designing, changing hyperparameters, and cautious show choice. Using feature designing, we added more parts to induce more valuable data. To urge the leading comes about, hyperparameter changing meant carefully progressing demonstrate parameters. We also tried them to see how the model's execution changed when different designs and parameter sets were utilized. We tried to find the most accurate way to foresee when hotel reservations will be dropped by making these programs superior and superior over time. We got the most excellent forecasts, and the finest came from the irregular forest classifier by repeatedly utilizing this approach.

**Performance Report**

The conclusion shows that about 75.36% of the test cases are correct, which shows how well it can foresee when lodging reservations will be canceled. After training the model, we did in-depth considers, such as calibration and post-model audits, to urge a complete picture of how solid it was. We learned about possible false or untrue positives by looking at measures other than exactness, like precision, review, and F1-score. We could make strides in the model's exactness and address particular issues by giving it a full test. This made it better at making predictions and affirmed that it could assist in making wise choices within the hospitality industry.

**Conclusion**

The study of the inn booking dataset gives valid data around foreseeing ticket cancellations, which helps executives make excellent choices on utilizing their assets and making the most cash. Using machine learning methods and careful show review, we made a reliable predictive model to help individuals within the hotel trade make decisions.