

Slide 1

Good afternoon. We are group N and we'll be presenting our predictive model for the Hotel Chain C's Revenue Manager Director.

Lilou:

Slide 2

Our team had the task of performing a predictive model to evaluate which are the most likely cases of cancellation for the hotel, using information regarding bookings. We were hired by the revenue manager of the hotel chain to deal with the problem of cancellation, overbooking and excessive policies regarding trying to avoid these problems.

Data Mining goals will be established in the context of data exploration by identifying the variables that are more relevant for the prediction analysis, with the help of data visualization throughout the research.

Slide 3

We used the F1 Score to compare the accuracy of our different models.

As you can see on the graph, our highest score of 0.876 was made by the Random Forest Classifier Model. The second closest with 0.85 is the Stacking Classifier, which ensembled 5 different models, used the Logistic Regression as a meta-classifier.

We used the GridSearchCV on most of our models in order to find the best parameters.

Slide 4

To evaluate our Random Forest Classifier model, we used a confusion matrix. You can see on the slide, the four different metrics of our evaluation model.

Looking closely at the matrix, we can see that there are 658 cases of False Positive, meaning that the clients were predicted as cancelled when in reality they weren't. In such cases it would mean that the Hotel would have an overbooking situation in their hands, and they would be a 4% probability that they would not be able to welcome the guests and thus be required to relocate them to different available hotels at the hotel costs.

Here, we see that there are 1283 cases of False Negative, where the model predicted as not cancelled when it was in fact cancelled. 8% of customers will be cancelled, and not be predicted by the model, which implies that the Hotel will lose money on empty rooms.

But overall we can say that our Random Forest Classifier model properly classified around 87% of the bookings from the testing set, and is able to determine 89% of the cases that were cancelled.

With this model, Michael will be able to forecast a more accurate demand based on the booking reservations, knowing that 89% of the predicted cancelled customers will not come.

Xavier:

Slide 5

If we have this model correctly implemented in the system available for reservations, there must be synchronization between the application of the model and the hotel's reservation system, as well as the most relevant variables, such as BookingChanges and LeadTime. These variables can be changed over time and are important tools in the implementation of the predictive model, and therefore can be constantly updated in the system.

This way there could be a dashboard to better understand customers to easily access relevant information in terms of bookings. When predicting cancellation, the hotel may deliver efforts in the attempt to avoid this cancellation, offering advantages such as discounts, spa treatments, or even local show tickets. The deployment of the model could be applied using the full database of customers, allowing the possibility of predicting the behaviour of future clients, so that the marketing and management strategies to follow are as accurate as possible.

Slide 6

The model used to predict the booking cancelations could be improved if more data was collected. Adding new data may lead to more accurate models, although it is hard to quantify and specify. In order to understand more the topic of booking cancelations in this hotel, different data in the property management system may be incorporated. New features may be added such as price level, deposit policies, weather information, current reputation of the hotel, exchange rates between currencies of the local and the most important nationalities in terms of number of clients are important features that can be added to improve the model.

Slide 7

Thank you! Questions?