Hotel Cancellation Prediction

Maanik Gupta

Problem Statement

- The seeming randomness of hotel cancellations are a problem that all hotels face. As such I have decided to examine some datasets which have user metrics to see if there is some way to predict what attributes of a customer are most likely to cause them to cancel.

Data

- The hotel dataset used covers all customer reservations from July 2017 through all of 2018. Of these reservations around 30% were cancelled.
- One major datapoint missing from the dataset is the hotel brand. This is a datapoint that could potentially affect the results, but without it there's no way to know either way if that does affect how often reservations are cancelled.

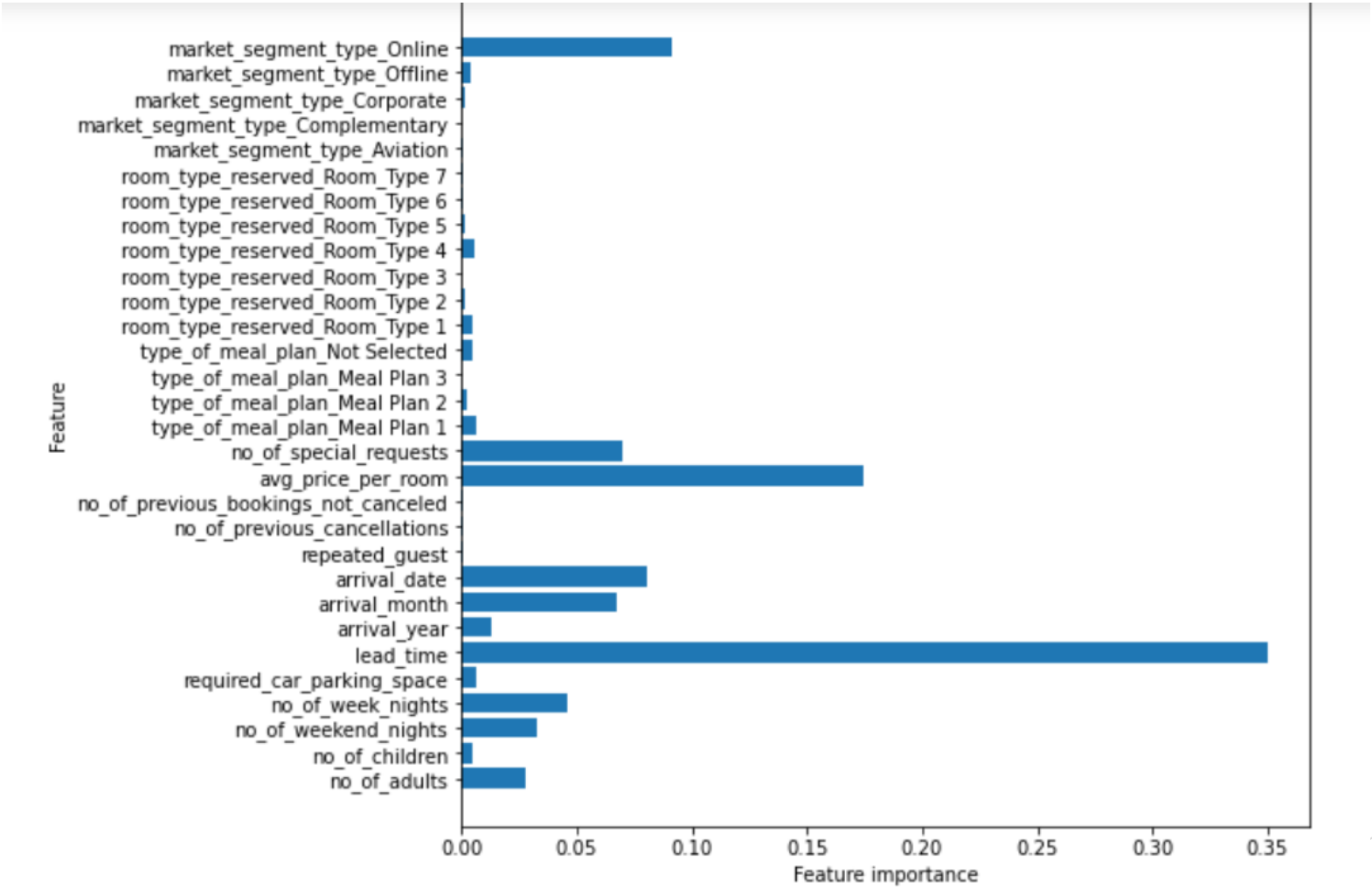
Approach

- Because of the binary result of reservations either being cancelled or booked, I decided to use a Classifier model. Since the dataset is so big and has many possible datapoints, I am using a random forests classifier.
- First, I had to clean up the dataset, there were no NaN values, so all I had to do was ensure that all the columns that produced object classes were turned into multiple dummy columns that were binary.
- For the training and test sets I simply set it as 75/25 respectively.

Current Results

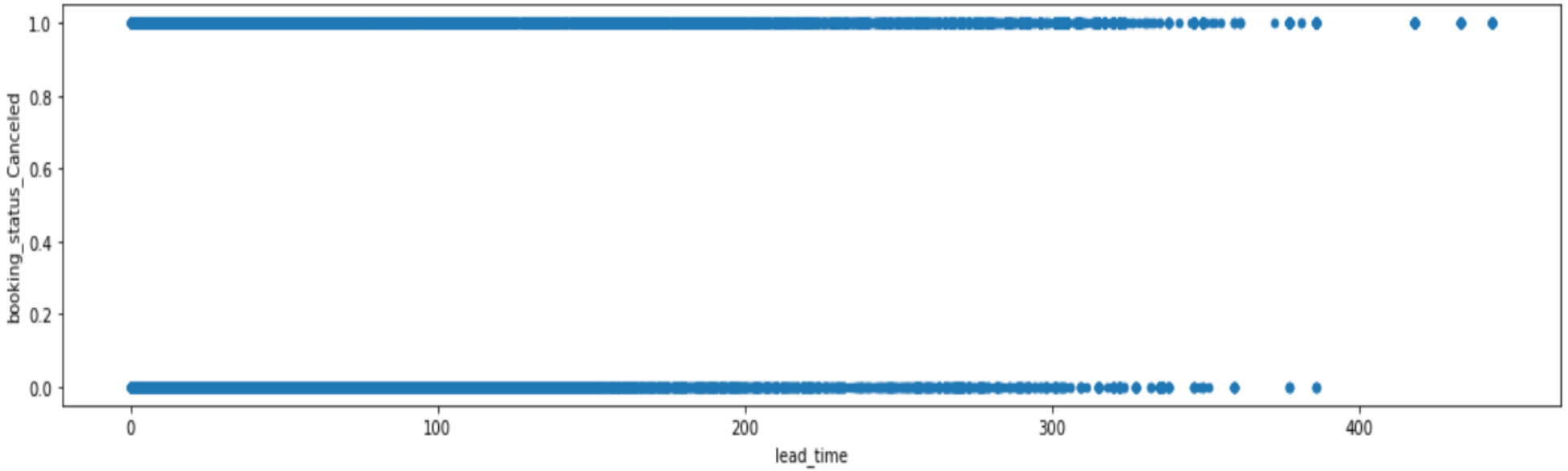
- I used a Random Forest Classifier to create a model that comes at about 91% accuracy.
- I then took a look at how the model ranked the importance of each feature and found this figure ->

Results: Feature importance of the Random Forest



Results

- To verify if this feature importance truly reflects the data I chose to plot the lead time vs the cancellation of the booking (1 meaning it was cancelled and 0 meaning it wasn't)



Results: Uses of this data

- With the model coming out to about 91% accuracy (as shown in the code). I believe it can be of great use to you.
- You can use this to predict which customers need extra enticing to be less likely to cancel their booking, giving them better deals or just focus on the customers who are more likely to stick with their booking instead.