DSE 6211 Analytical Plan

Joseph Annand

2024-01-27

To help ABC hotels identify bookings that have a high risk of cancellation, given project data and a supervised classification approach will be used to determine the probability of cancellation for a given booking. Each observation in the given data set is a different booking at ABC hotels. These observations will be used to train the supervised classification model, which will attempt to predict the value of the “booking\_status” variable. This variable of the data set indicates whether the the booking was cancelled or not, so “booking\_status” is the label, or target, in the supervised classification problem.

Data processing will be required for the “arrival\_date” and “lead\_time” features before creating the model. The arrival date can be broken up into month or season. We may also want to trace the day of the week the arrival date is. These details may be more valuable to predicting booking status than the what is provided by the given format of the arrival date feature. Similarly, the lead time feature can be processed to determine the season, month, and weekday of the day that the customer booked their room. All features plus the six new variables for booking and arrival date will be included in the supervised classification model. It may be worthwhile to explore the implementation of a variable subset selection, in which the features most indicative of booking status are used in the model, to achieve the best prediction accuracy possible.

The output of the supervised classification model will be a value between 0 and 1, representing the probability of cancellation for a given booking. The interpretation of and consequent action on the output may vary depending on the company’s available resources and current needs. For example, the model could be developed to classify each booking as “high risk of cancellation” or “low risk of cancellation” to separate the bookings that need some intervention to prevent cancellation and the bookings that do not. The probability threshold to classify the booking as one of the two categories may vary for several reasons. If the company has committed a plethora of money and resources into cancellation prevention, the threshold may be set relatively low like 50% chance of cancellation. If the company has limited capacity for extra advertisements and offers, then the threshold may be set to a higher probability value like 75%. The model could also classify bookings with a high cancellation risk into priority groups. Then, the hotel could easily distinguish between a booking with a 90% chance of cancellation and a booking with a 75% chance so that it prioritizes sending advertisements and offers to the booking at a higher risk. Nonetheless, the model will provide useful information to ABC hotels that will help keep their rooms filled with customers.