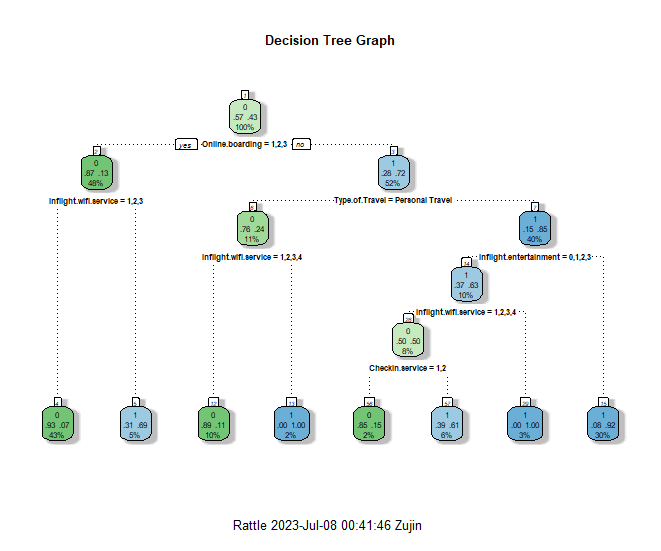
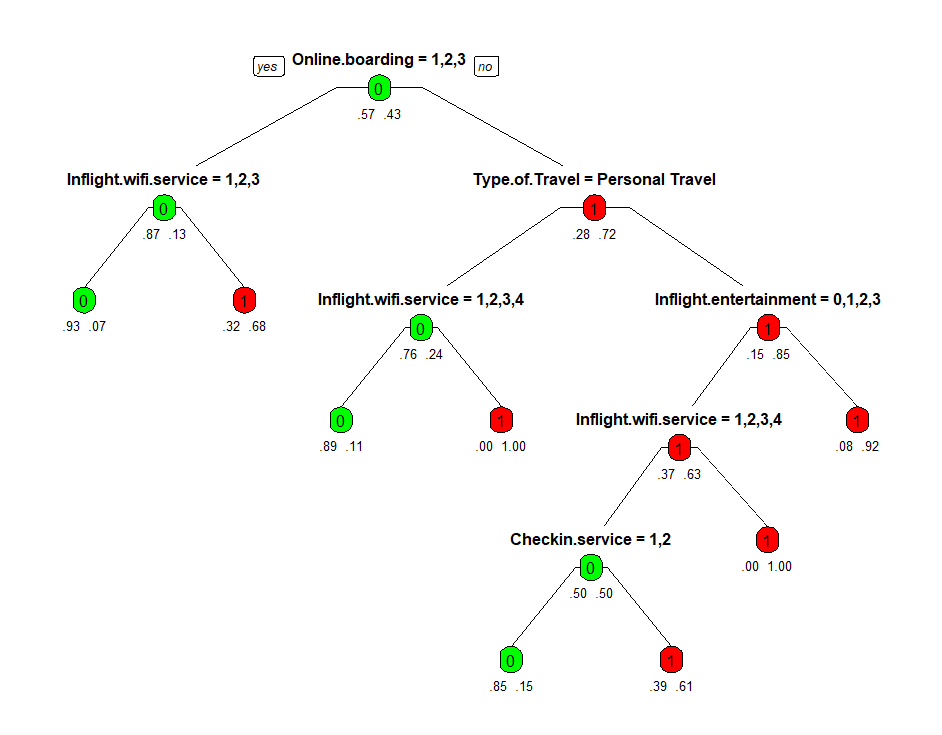
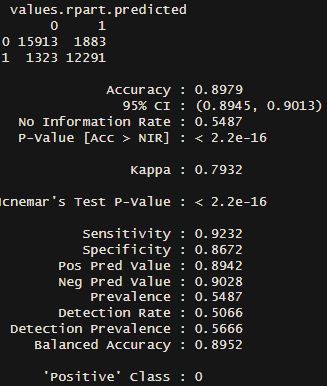
# Decision Tree Model

Any NA values were omitted from the data set and it did not make a difference if the zeros in the data were imputed or not. The overall model accuracy is 89.8% with a sensitivity of 92.3% and specificity of 86.7%. This model only contains the factors Inflight.wifi.service, Checkin.service, Inflight.entertainment, Type.of.Travel, and Online.boarding. Further model tweaking and pruning will be explored at a later time. Do note that a value of 0 means the customer was not satisfied and how often a low score in Inflight.wifi.service results in a non-satisfied customer.

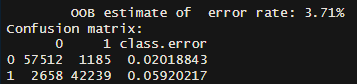
# Random Forest Models

We are using random forests as a machine learning technique to discover which parameters are the most important as well as developing a model to predict customer satisfaction. The model was set to automatically exclude any NA values and any zero values were left as zero. For the second run of models, all values of zero were imputed to the mode of the variable so that the value would be within the Likert scale. The imputation of data caused the error rate to grow overall.

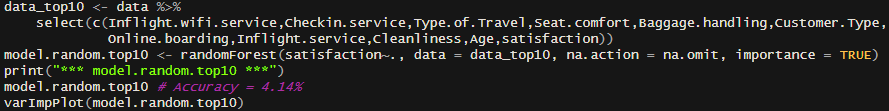
The error convergence was tested to ensure that random forests of 500 trees was sufficient and it was found that the errors generally converged around 150-200 trees. The number of variables to test at each split was also validated that the default number of 3 was sufficient. There was a small amount of reduction in error at higher values but these were not significant enough to change the default value.

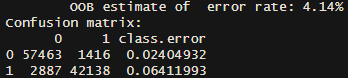
Using all the variables except for Arrival.Delay.in.minutes, an overall error rate of 3.1% was achieved on the training data set. It should be noted that the error rate was 2.02% for Neutral or Dissatisfied customers while the Satisfied customers was 5.92%. The model seems to perform a very good job at predicting customers who will not be fully satisfied with their care but the company must focus their resources to achieve their goal with the least amount of expenditures.





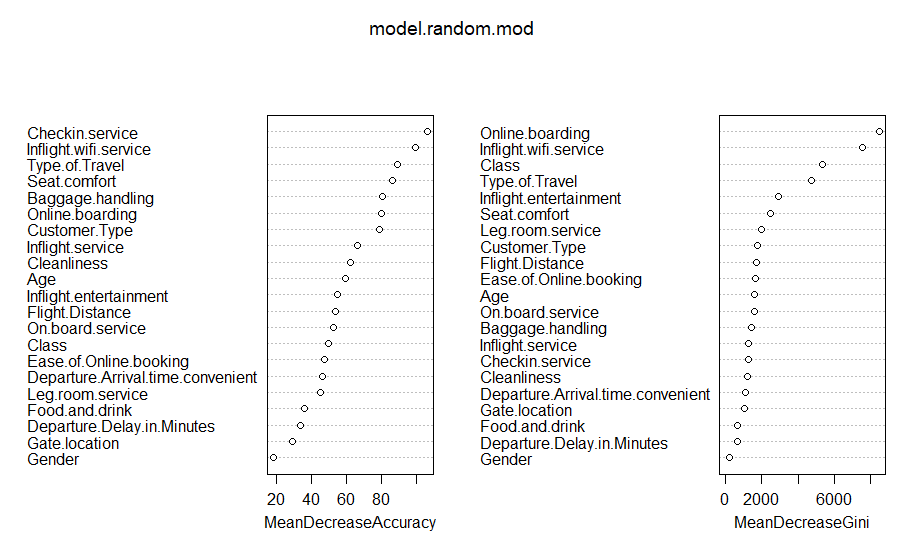
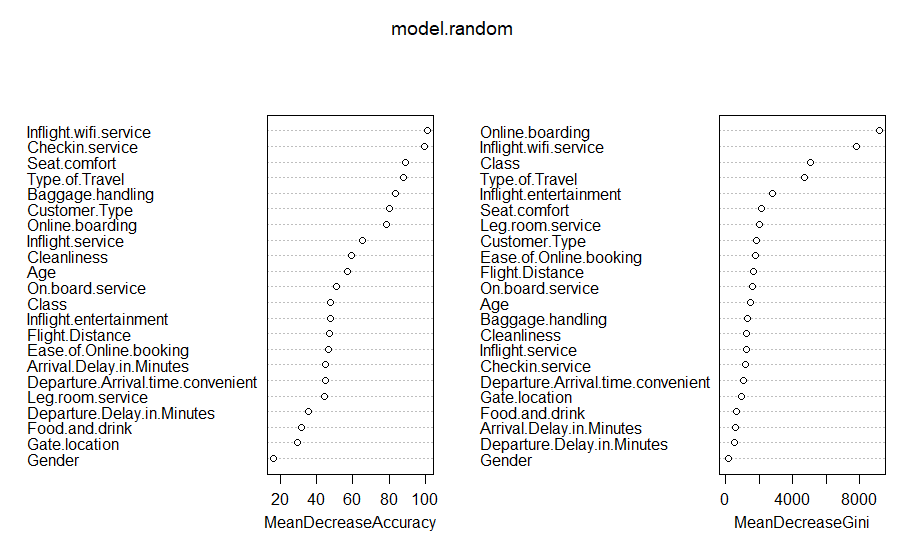
Only taking into account the top 10 factors by important yields an error rate of 4.14%. The Neutral or Dissatisfied classification has an error rate of 2.4% with the Satisfied customer error rate rising to 6.4%. This still demonstrate an extremely high amount of predictability within the top 10 factors.





The results of the importance indicate that the two most important factors are Inflight.wifi.service and Checkin.service with the top 10 being:

|  |  |
| --- | --- |
| Inflight.wifi.service | Customer.Type |
| Checkin.sevice | Online.boarding |
| Seat.comfort | Inflight.service |
| Type.of.Travel | Cleanliness |
| Baggage.Handling | Age |



Sources:

<https://www.youtube.com/watch?v=6EXPYzbfLCE>