Model Metrics - “Happier Passengers“

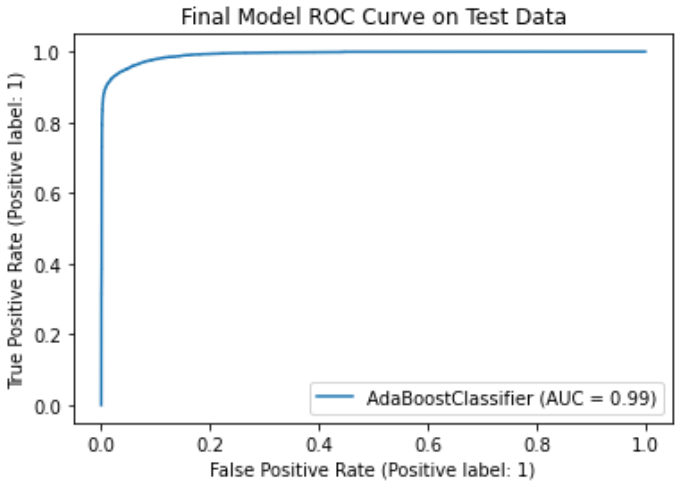
The final model is the following:

AdaBoost model based on a decision tree classifier with a max tree depth of 3 and 100 estimators.

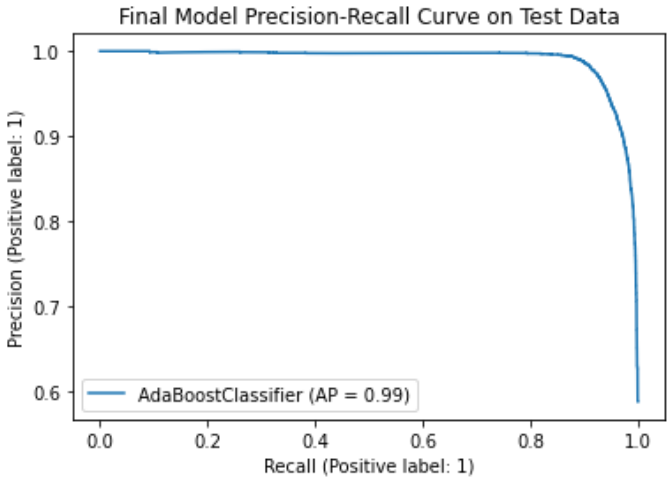
The model trained on all of the training dataset and applied on the test set has the following scores,

Final Model (AdaBoost) Scores	
Accuracy	0.956
F1 Score	0.949
Precision	0.956
Recall	0.941

ROC Curve,

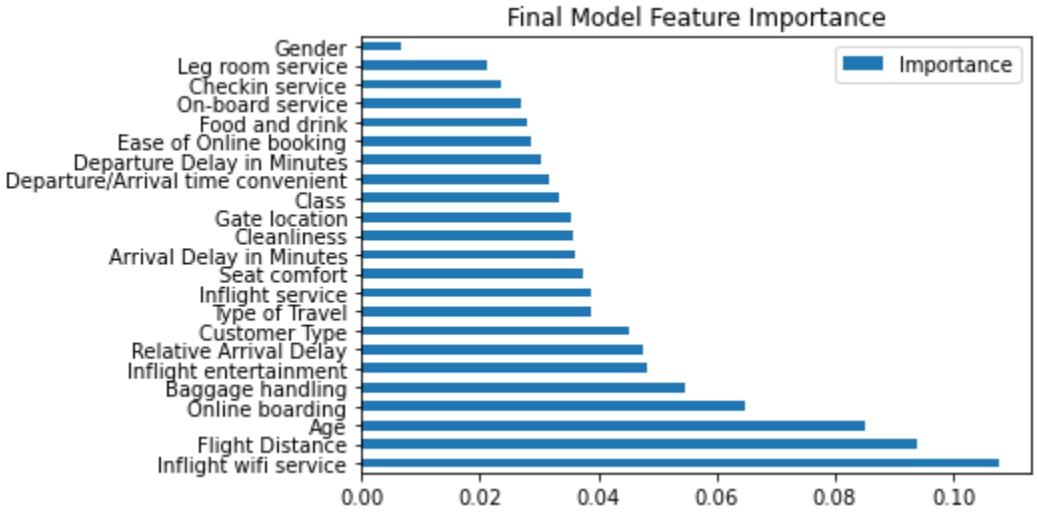


and the Precision-Recall Curve:



Both curves above show a great fit. [Perhaps, too good]

Finally, I close the model section with the feature importance graph.



Regarding the above figure, observe that among the top five “most important” variables, Inflight wifi service, ‘Online boarding’ and ‘Baggage handling’ are survey questions. This is good news as the airline has some direct control over these flight quality factors.

On the other hand, while the remaining variables in top five (‘Flight Distance’ and ‘Age’) are not directly controllable, observe that they moved up in ranking when compared to the prior correlation figure. This gives one hope that there might be a more interesting story there. Namely, it might mean that decision tree structure has found some structure in the data that it mined and that this model could be split into sub-models perhaps based on these variables, in particular ‘Flight Distance’.