

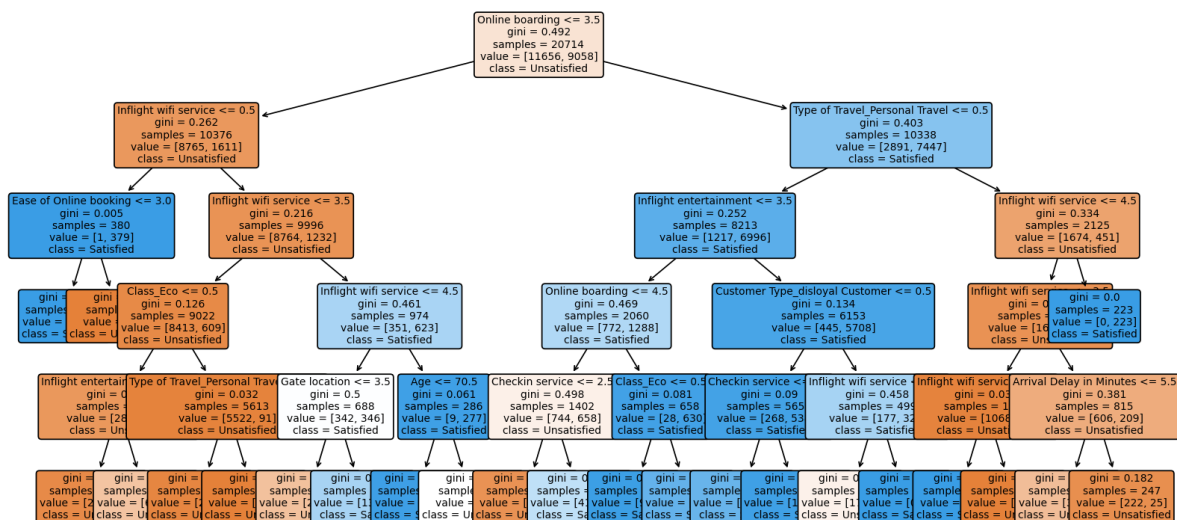
Samuel Octavio González Azpeitia

A01704696

Airline Passenger Satisfaction

I make a decision tree and a random forrest. this to predict the satisfaction of people who decide to fly, this depends on variables such as if they have access to Wi-Fi, age of the people, flight class, distance, among others.

The following shows the decision tree generated where it is searched if the passenger I am satisfied or not.



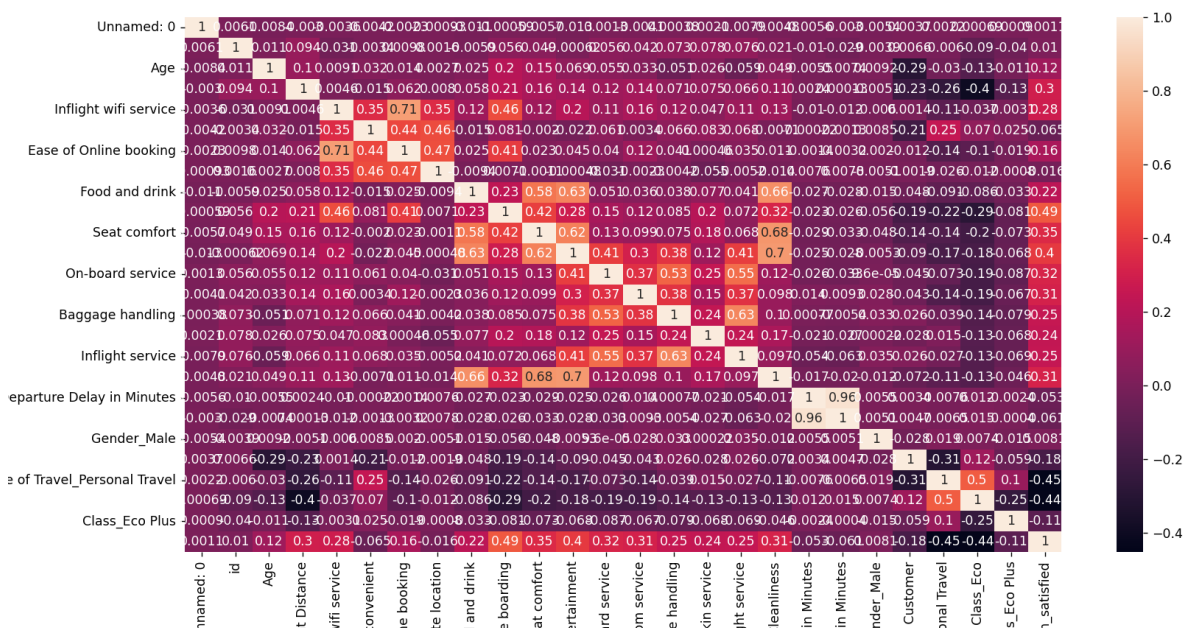
Correlation is a statistical technique that shows how they are related.

two variables in this case the correlation matrix is shown to find by pairs

all columns in the data frame. The values on the diagonal show the correlation

of a variable with itself therefore the diagonal shows the correlation 1

It is graphed on a heat map where it can be better appreciated.



On the other hand, the classification report is presented here is the summary of the quality of the classification carried out, the columns that compose it are first the name, followed by accuracy, F1 score recovery and support.

Precision is calculated with respect to predicted heats

Recovery Calculated with respect to the actual values in the data set

F1 is the harmonic mean of precision and recall

Support These are the total entries of each class in the actual data set.

	precision	recall	f1-score	support
0	0.93	0.94	0.94	2872
1	0.93	0.92	0.92	2307
accuracy			0.93	5179
macro avg	0.93	0.93	0.93	5179
weighted avg	0.93	0.93	0.93	5179

The confusion matrix was used to evaluate the performance of the classifier.

Where true positives, false positives, false negatives and true are evaluated negatives To obtain the precision, divide the true positives by the false ones. positives + true positives.

```
[[2654 218]
 [ 266 2041]]
0.9065456651863294
```

Finally, a Random Forrest was made with 200 estimates in the same way with 5 of max depth after all the estimations, in the same way, its matrix of confusion and its score.

```
[[2714 158]
 [ 192 2115]]
0.9324193859818498
```