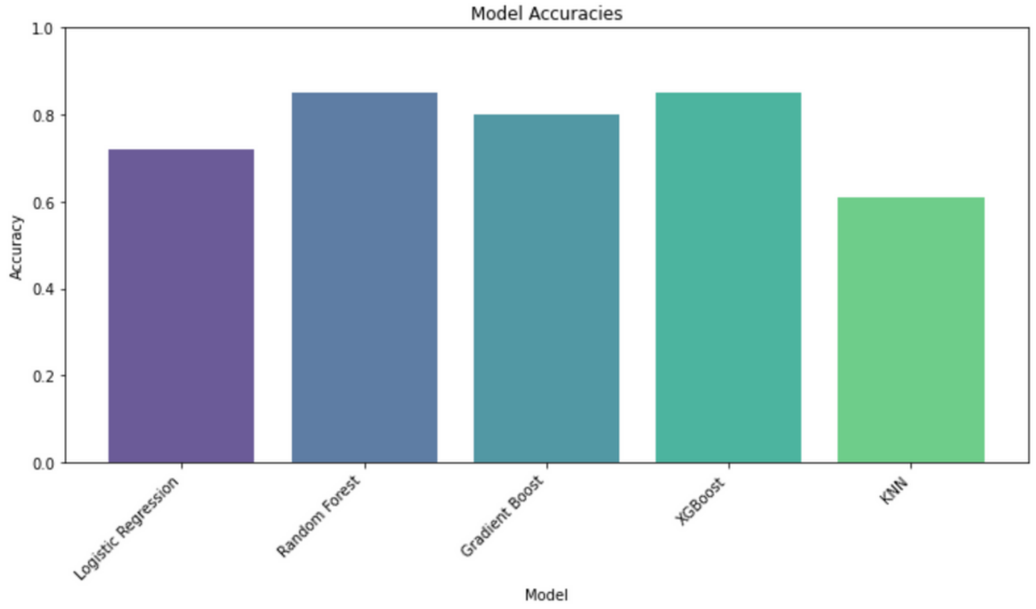
For purposes of Forage Virtual Work Experience program

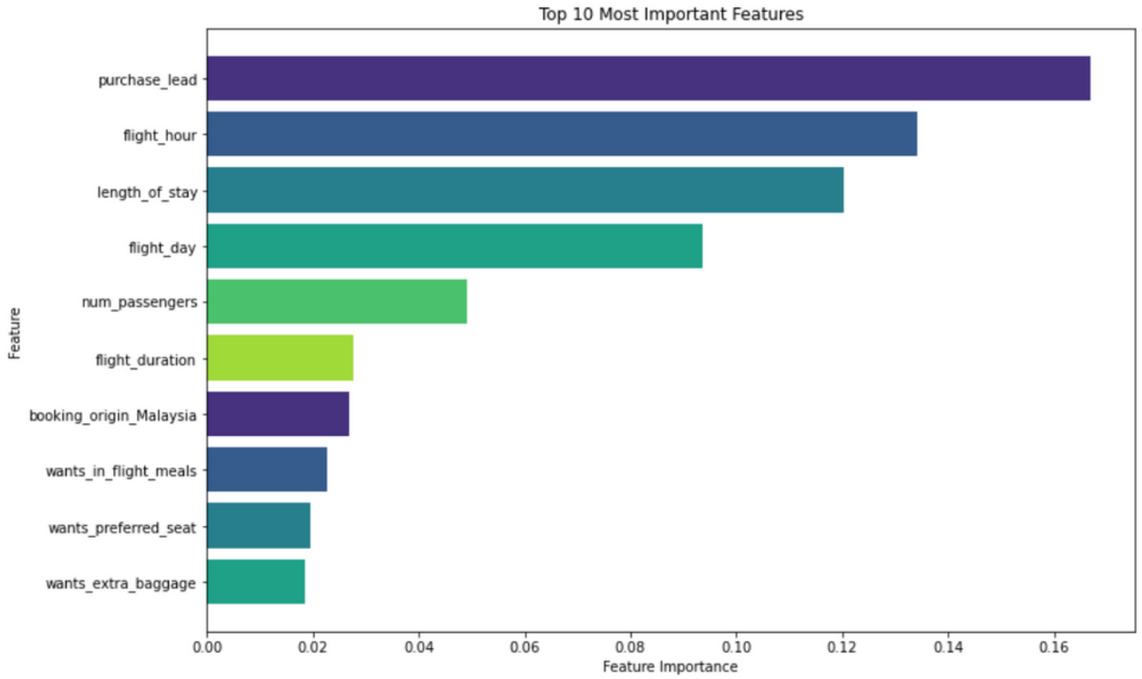
Forage - Task 2: Web scraping to gain insights from customer reviews



Model Choice: RF Classifier has the highest accuracy on our data from these 5 models.

Feature Importance Analysis:

purchase_lead was the most important feature in the model (number of days between travel date and booking date)



RANDOM FOREST CLASSIFIER

85.1%

Mean cross validation accuracy; CV scores were fairly consistent across different folds.

Class 0

Precision: 0.86
Recall: 0.97
F1 Score: 0.91

Class 1

Precision: 0.5
Recall: 0.13
F1 Score: 0.21

These results indicate that the model performs well for class 0 but struggles to predict for class 1. Thus, potential improvements may involve the employment of other class balancing techniques.