**EXPLORATORY DATA ANALYSIS / FEATURE ENGINEERING**

**Before modeling. Example questions to answer: Are there interesting relationships among predictors? What variables appear to be important? Can you visualize the data in unique ways? Are there interesting clusters? Are there additional variables you should include in the data? Are there variables you should remove? Should you consider interactions, transformations, etc.?**

**MODEL EVALUATION**

**Before and after modeling. Example questions to answer: How did you partition your data for your analyses? How did you know if your model(s) were any good? Did you compare to a baseline, and if so, how did you choose your baseline(s)? How did you choose among models for comparison?**

Partitioned Data into 75:25% (Training and Validation).

We used accuracy and compared it with the baseline.

Checked up with a few other competitors and decided if our model was comparatively accurate or not.

|  |  |  |
| --- | --- | --- |
| **Algorithm** | **Accuracy** | **Package** |
| Logistic Regression | 82.79% | NA |
| GBM |  | gbm |
| AdaBoost |  |  |
| Random forest | 83.03% | randomForest |
| Caret |  | Caret package |
|  |  |  |
|  |  |  |

**MODELING**

**Example questions to answer: What models did you try and why? (Justify your model choices!) How did you tune these models? Which models/type of model had the best performance? How do you know? Important: Make sure you document all models and their performance in this section to support your conclusions/choices!**

**PERFORMANCE**

**This is the only non-subjective part of the assignment. Points will be assigned as follows:**

**Top-scoring team: 10 points**

**Within 1 standard deviation of top team: 6 points**

**Within 2 standard deviations of top team: 3 points**

**More than 2 standard deviations from top team: 0 points**

**EXECUTIVE SUMMARY**

**A short (no more than one page) summary of your overall results. Example questions to consider: What actionable results have you found? Do you have business insights and recommendations? Can you explain your results and analyses in layman’s terms (that is, terms that are easy for a non-technical person to follow)? Why did you do your modeling?**

**Extra Data:**

Our aim was to predict if an Airbnb accommodation has high booking rate or not. We arrive on this result by basing our analysis on 70 features that were provided.   
This algorithm can be used by Airbnb to predict which of their lodging property may have high booking rate.

[OR]

**The algorithm can be used by airbnb to predict and identify accomodations which might not have a high booking rate and suggest things that owners could implement to change that.**

Currently, there is no way of determining if a particular accommodation might lead to high booking rate. With this algorithm, Airbnb can have the upper hand in identifying and implementing features that lead to a high booking rate.