STAT 5650 Statistical Learning and Data Mining 1 Spring 2020

Names: Greg Hoffmann, Kristen Sohm, Michael Huber, Varsha Reddy Mandadi

Submission Date: 04/25/2020

Final Project: Applying Prediction Methods to Hotel Data

210 Points — Due Monday 04/27/2020 (via Canvas by 11:59pm)

(i) **Prediction Methods:** After working with the data and reviewing the notes from Dr. Cutler on our project proposal we had to edit what methods of prediction we were going to apply against the data. Below is a list of methods that we were able to get to run against our hotel data set.

- Gradient Boosting Machines (GBM)
- Support Vector Machines (SVM)
- Random Forests
- Adaboost
- Classification Trees

(ii) **Classification Trees:**

To prepare the data for the classification tree we subset the data and removed the following columns from the data set: country, agent, company, reservation status, reservation status date.

(iii) **Untuned Support Vector Machines:**

```
      0    1
 0 2027  505
 1  607  861
```

```
[,1]                              [,2]
"Percent Correctly Classified = " "72.2"
"Specificity = "                  "80.06"
"Sensitivity = "                  "58.65"
"Kappa ="                         "0.3928"
"AUC= "                           "0.7689"
```

(iv) **Untuned Gradient Boosting Machines:**

```
      0    1
 0 1954  578
 1  711  757
```

```
[,1]                                [,2]
"Percent Correctly Classified = " "67.78"
"Specificity = "                  "77.17"
"Sensitivity = "                  "51.57"
"Kappa ="                         "0.293"
"AUC= "                           "0.6886"
```

(v) **Untuned Gradient Boosting Machines:**

General Instructions