Sahar Hihath
06/04/21

Guided Capstone Project Report

This report is aimed at investigating whether the Big Mountain Ski Resort in Montana is over-charging or under-charging skiers. The tickets at the resort are being sold at $81.00 currently. The Mountain Ski Resort in Montana has main feature including vertical drop of 2353 feet, a summit elevation of 6817 feet, a total of 14 chairs, and 105 runs. Other features include 3000 acres of skiable terrain, 600 acres of snow making, and an average snowfall of 333 inches. The average ticket price in Montana is $52 whereas for California is $81. First, the data was pre-processed by removing features that have missing data such as 'fast Eights' and reports with no information for ticket prices. To better understand the correlation between the features, a heat map was generated (Figure 1) showing that for adult weekend ticket price, there is a correlation between fastQuads, Runs, Snow Making (acres), Total Chairs, Resort Night Skiing Per State (Ratio), and Vertical Drop. To predict the Weekend Ticket Price, a Dummy Regressor modeled the Mean, which was $64. Models worse than the Mean give negative R2 and models that are better belong to the interval [0, 1]. The mean absolute error for training data was $18 and $19 for test data. The root mean square error was $25 for training data, and $24 for test data. The missing values were replaced by medians. The model considered for prediction of ticket prices was linear regression model. The "make_pipeline function" used the SimpleImputer to replace the missing values for each feature with the median of that feature. The make_pipeline function had two methods, that are "fit" and "predict," to fit a model to the training data, and to predict ticket price based on the model.

Additionally, "SelectKBest" selected the k best features for training the model. Cross-validation was used to find the k best features for model's best performance. Figure 2 showed that k=8 gives the optimized value for achieving the best model performance. The eight best features were Vertical Drop, Snow Making (acres), Total Chairs, Fast Quads, Runs, Longest Run, Trams (negative coefficient) and Skiable Terrain (negative coefficient). Then, Random Forest Regression was used instead of Linear Regression to improve the model. Figure 3 shows the best features to use with Random Forest Regression. It is noteworthy to mention that Mean Absolute Error for the Linear Regression was $11.8, whereas the one for the Random Forest was $9.5. The Random Forest model predicted $95.87 for the Big Mountain Resort for the Weekend Ticket with the actual price being $81, while the mean absolute error is $10.39. For a different scenario, the Big Mountain Resort added a run and increased the vertical drop by 150 feet followed by installation of another chair lift, which resulted in an increase in the ticket price by $8.61 and a new revenue of $15,065,471. For scenario 3, which is the same as Scenario 2 with the addition of Snow Making (acres), the ticket price increase was $9.90 resulting in a new revenue of $17,322,717. For scenario 4, 0.2 miles were added to the Longest Run (miles) and 4 miles to Snow Making (acres) resulting in no increase in ticket price.

Sahar Hihath
06/04/21

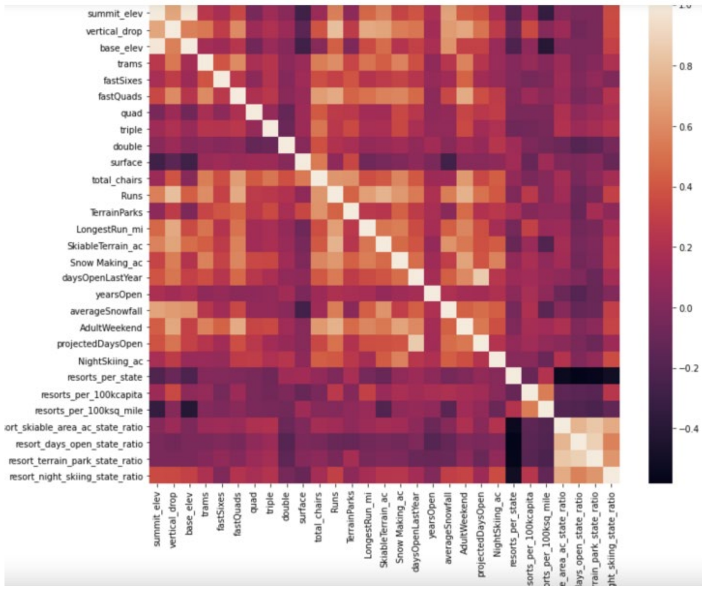

Figure 1. Heat map

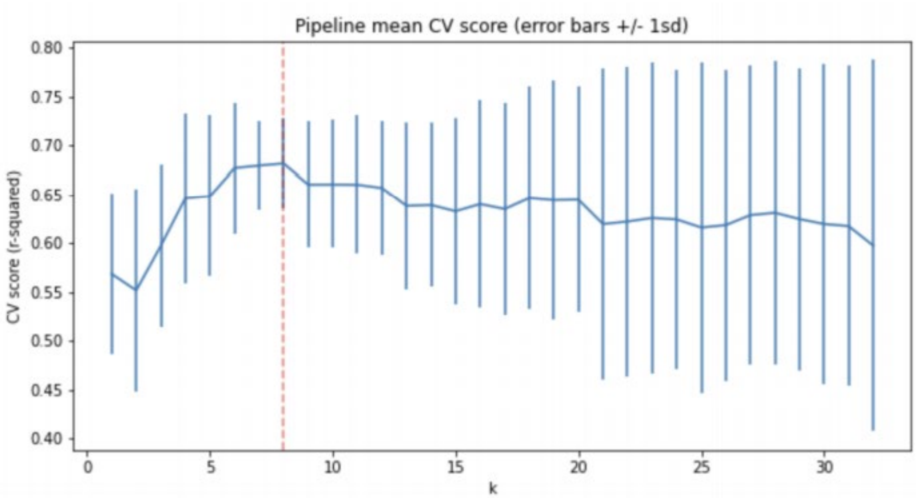

Figure 2. K best features graph

Sahar Hihath
06/04/21

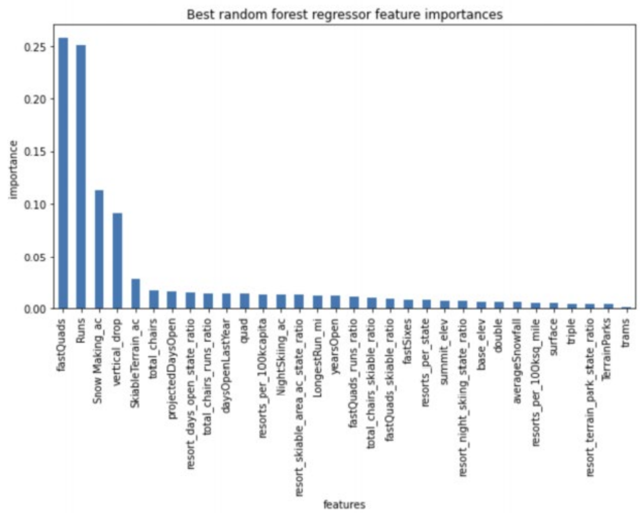

Figure 3. Random Forest Regression