Guided Capstone Project Report
Big Mountain resort

Vitaliy Shepel

vitaliyshepel20@gmail.com

Abstract

As one of the Montana largest skiing resorts, Big Mountain resort has a sales volume of millions dollars and as a result many facilities to review.. These features are great learning samples which could help support increasing a ticket price.

Our project aims to build a multiclass classifier based on the skiing resort industry market. Different feature selections and models are tested when building this classifier. As we will show below, the random forest classifier performs the best. We evaluate the performance of various supervised learning algorithms for regression followed by ensemble methods, with feature and model selection considerations being treated in detail. We further evaluate all methods on accuracy.

I. Introduction

Key Problem:

Big Mountain Resort has recently installed an additional chair lift to help increase the distribution of visitors across the mountain. This additional chair increases their operating costs by $1,540,000 this season. There's a suspicion that Big Mountain is not capitalizing on its facilities as much as it could. The strategy is to implement a more data-driven business approach.

Our goal:

To make predictions to select a better value for Big Mountain Resort ticket price. Considering some changes that will either cut costs without undermining the ticket price or will support an even higher ticket price.

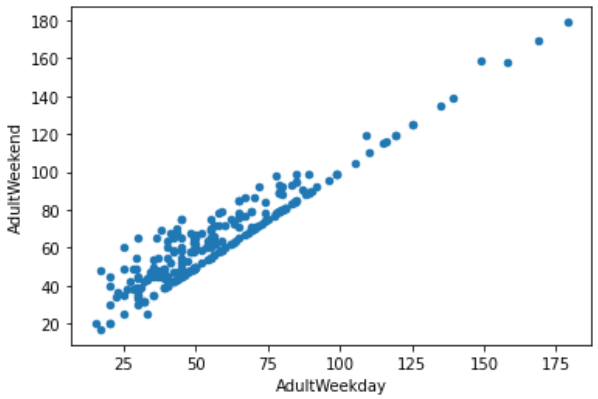
Strategy and Methods:

Feature Selection: An important aspect of this task is creating class-balanced training and test datasets while iden-tifying appropriate metrics for assessment of prediction success. Critical features are analyzed and extracted using low order modeling techniques like Principal Component Analysis (PCA) and correlation analysis. .

II. Dataset

We have metadata that contains information from 330 resorts in the US that can be considered part of the same market share. We also obtained some additional US state population and size data with which to augment the dataset,

III. Exploratory Data Analysis



A couple of observations can be made. Firstly, there is a clear line where weekend and weekday prices are equal. Weekend prices being higher than weekday prices seem restricted to sub $100 resorts. The distribution for weekday and weekend prices in Montana seemed equal. Big Mountain resort is in Montana, so the relationship between these quantities in this state are particularly relevant.

All our data were numerical until we started doing PCA and we have gotten categorical components. Later, we had a quartile category. AdultWeekend ticket price, you see quite a few reasonable correlations. fastQuads stands out, along with Runs and Snow Making_ac. The last one is interesting. Visitors would seem to value more guaranteed snow, which would cost in terms of snow making equipment, which would drive prices and costs up. Of the new features, resort_night_skiing_state_ratio seems the most correlated with ticket price. We can see a positive correlation between chairs and ticket price, so our new implemented machine allows us to increase the ticket prices. The states' labels we can use as an index again to simplify calculations.

IV. Models

We have built 2 multiclass classifiers: Random Forest Classifier (RF), Decision Trees Classifier (DT).

Decision Trees Classifier builds a decision tree according to the information gain for each attribute. It chooses the attribute with the largest information gain at each node, and stops adding nodes when every attribute has already been included along this path through the tree, or the training examples associated with this leaf node all have the same target attribute value.

Random Forest Classifier repeatedly selects, by bagging, a random sample with replacement of the training set and fit 1000 trees to the samples. With each tree-fitting with an individual

.

V. Results

Refitting Model on all available data excluding Big Mountain resort and calculating expected Big Mountain ticket price from the model using training data lead to modeled price $95.87 With actual price $81.00 and even with the expected mean absolute error of $10.39 we can see that price can be increased. By creating a plot_compare function we visualizing our Big Mountain sit in the distribution. When we make scenario's prediction adding 1 run, 150 feet of vertical drop and installing an additional chair lift gives us possibility to increase price by $1.99. Over the season, this could be expected to amount to $3474638. Increasing a snow making area along with mentioned features together does not make a difference. No difference in adding longest run 0.2 miles and snow making by 4. Longest run is the least important in our randam forest model in opposite to linear model.