**Big Mountain Resort Project Report**

Title: Predicting Ticket Prices for Big Mountain Resort using Regression Algorithms

Date: Aug 08, 2023

Author: Hui Zhang

Abstract:

In this project, we explored the application of various regression algorithms to predict ticket prices for Big Mountain resort. We analyzed and preprocessed the dataset, selected relevant features, and implemented different regression models. The performance of each model is evaluated using metrics such as R-squared and Mean Absolute Error (MAE).

1. Introduction:

Problem Statement: Big Mountain resort is a ski resort located in Montana, which has 105 trails and serves 350000 people every year. There's a suspicion that the resort is not capitalized on its facilities as much as it could. The management want some guidance on how to select a better value for their ticket price. They are also considering changes that can either cut costs without undermining the ticket price or support an even higher ticket price.

Objectives: Build a ticket price prediction model and give facility investment suggestions.

Dataset: csv file from data manager

2. Data Wrangling:

Checked the percentage of missing values and outliers, corrected wrong input, removed irrelevant and useless features, analyzed distribution of resorts and ticket prices by states, created several new features and finally selected adult weekend ticket price as the target variable.

3. Exploratory Data Analysis (EDA):

PCA analysis and visualization of high dimensional data.

Visualization of features correlation. Found the features with high importance are: 1) vertical drop, 2) snow making area, 3) number of total chairs, 4) number of fast quads, 5) number of runs, 6) longest run length, 7) number of trams, 8) skiable area.

4. Pre-processing and Training Data:

Split train set and test set before training.

4.1 Linear Regression:

Implementation: created pipeline consist of imputer, standard scaler and linear regression.

Assessment: cross validation with MAE and R-squared as metrics.

Hyperparameter tuning: GridSearchCV and cross validation of best estimator.

4.2 Random Forest Regressor:

Implementation: created pipeline consist of imputer, standard scaler and random forest regressor.

Assessment: cross validation with MAE and R-squared as metrics.

Hyperparameter tuning: GridSearchCV and cross validation of best estimator.

5. Results and Discussion:

Tabular Comparison:

|  |  |  |
| --- | --- | --- |
|  | Linear Regression | Random Forest Regressor |
| cross validation score(R2) | 0.682 | 0.710 |
| test set (R2) | 0.597 | 0.751 |
| cross validation score (MAE) | 10.499 | 9.645 |
| test set (MAE) | 11.793 | 9.538 |

Finally selected Random Forest Regressor, and the features with high importance are almost same as the correlation results in EDA step.

Scenario Modeling: four scenarios analyzed: 1) close up to 10 of the least used runs, 2) increase the vertical drop by adding a run to a point 150 feet lower down, 3) same as 2 but adding 2 acres of snow making cover, 4) increase the longest run by 0.2 mile.

Visualization:

Big Mountain Ticket Price (left: ticket price distribution in US, right: ticket price distribution in Montana):

A graph of a number of tickets

Description automatically generated with medium confidenceA graph with blue lines

Description automatically generated

Big Mountain Facilities (left: fast quads, right: runs)

A graph with numbers and a red line

Description automatically generatedA graph of a number of runs

Description automatically generated

Big Mountain Facilities (left: snow making, right: vertical drop)

A graph of snow making

Description automatically generatedA graph of a vertical drop

Description automatically generated

Pricing Recommendation: based on the market position and facilities of Big Mountain resort, the model suggest the ticket price of Big Mountain should be $95.87, even the MAE is $10.39, there is still room for the increase of ticket price.

Improvement Suggestion: 1) close one least used run, which will not influence the revenue, 2) increase the vertical drop by adding a run to a point 150 feet lower down with installation of an additional chair lift without additional snow making coverage, which will increase revenue roughly $3 million.

6. Conclusion:

The Random Forest Regressor model performed better than Linear Regression with cross validation score (MAE) $1 lower and test set (MAE) $2 lower. Based on this model, the ticket price still has room for increasing and the revenues can be increased by facilities improvement. The primary goal of this object is achieved.

7. Future Work:

The deficiency in the data is that the operation cost is unknown. Cost information will be helpful to improve the performance of the models.

To make machine learning models available for business analysts to use, it is necessary to deploy the trained models in a production environment where they can be accessed by business analysts, such as cloud-based services. Developing a dashboard or web application with an interface that allows business analysts to interact with the machine learning models without needing coding skills is also helpful.