

This report details all elements of a semester-long project completed by Christopher Barua, Peyton Camden, and Brandon Cook that aimed to find the best predictors and predictor model for hotel scores using a dataset that collects TripAdvisor reviews from Las Vegas Strip hotels. The process is documented through code, presentations, and reports. The process documents the predictive power of three models, which are the decision tree classifier, the random forest classifier, and the linear regression model. The models were run through three different variations of the data, which were the original dataset, an oversampled version of the dataset, and a weighted version of the dataset.

Ultimately, this project found that the random forest classifier was the best model to use to predict the hotel scores, and the original data or the weighted data was the best balance to use by the highest accuracy and F1 score metrics. The features that were found to best predict the hotel scores were 'Dummy Pool', 'Dummy Free Internet', 'Saturday', 'China', 'Business', and 'Excalibur Hotel and Casino'. However, it was made clear throughout the project that the dataset was not complete enough to translate the model results reliably to reality.