The perfect food for any study binge, pizza is a favorite for college students everywhere. In fact, one could argue that the only thing more common than pizza in a college students' diet, is the almost certainty that said student is financially unstable. Luckily for college students everywhere, the data I present today will help alleviate some of the stress related to the monetary value of pizza by hopefully relating the price of pizza with a few measurable parameters to be used in the creation of a Pizza price prediction model. The goal of this ProJet is to help alleviate some of the stress a student may face in deciding which unhealthy food to eat by hopefully predicting which pizza is most economical, ensure the only stress factor remaining in their life is the immense social pressures and crippling student loan dept. The data will use records numerous parameters about the pizza, including diameter, size, variant, and toppings, as well as price which I will split into the features that will be used in the creation of my model.

As all the features are categorical, linear models are not perfect for prediction modeling, instead I resigned to using Decision Trees to create my model. Because three of the features I will be measuring can be answered in a True/False form, utilizing a Decision Tree seemed like the obvious choice. The decision trees being a decision tree regressor and a random forest regressor. Random forest is an ensemble of decision trees. This is to say that many trees, constructed in a certain “random” way form a Random Forest. The average makes a Random Forest better than a single Decision Tree hence improves its accuracy and reduces overfitting. For these reasons, I have decided to include a Random Forest model in my model comparisons.

|  |  |  |
| --- | --- | --- |
| Model used: | R2 Score | Accuracy |
| Decision Tree Regressor | .9633 | 69% |
| Random Forest Classifier | .8666 | 87% |

After creating my models and testing the data, I found that while the Decision Tree Regressor held a higher r2 score than the Random Forest Classifier, the latter was significantly more accurate. If I were to redo the project using this data, I would further attempt to correlate the featured parameters to insure a more accurate prediction.