Ian Sime

Bridger Fisher

Noah Blumenfeld

Thomas Eliassen

**Support File for Step 2**

During this step of our project we ran our data set through five algorithms: decision tree, random forest, gradient booster, k nearest neighbors, and logistic regression. Using grid search we found that gradient booster is the most accurate with 0.93207 accuracy. The second best was decision tree with 0.92873 accuracy. The worst was logistic regression with 0.628062 accuracy.

We made a bar graph and pie chart of feature importance. The bar graph shows that protoss ground armor 1 was the most important feature using the gradient booster classifier. The pie chart shows that protoss third expansion is the most important feature. The bar graph and chart only rank feature importance of the top 10 features, we did this to simplify the visuals since we have 56 attributes. We used matplotlib for the bar graph and for the pie chart we used plotly. We also used bokeh to make a scatter plot. We plotted staisis information on the y-axis and ground armor on the x-axis, these are the best feature and 10th best feature. The plot showed that there is almost no correlation between the two attributes. From this we can tell that splitting between these attributes would be a good idea because it would help differentiate the classes. We used seaborn to graph a swarmplot. From this plot, we can tell the value of which a midBuild exists at a certain attribute.