

CSE 482: Project Proposal

Tyler Rozwadowski | CSE 482 | 2/11/19

For my CSE482 project, I want to analyze data from the PGA Tour (professional golf). It will be titled something similar to “Correlations Between Winning and Player Statistics in Golf.” Using Kaggle, I found a dataset giving over 50 data points for nearly 200 professional players. The data set can be found at: <https://www.kaggle.com/grantruedy/pga-tour-golf-data-2017-season>. Using this data, I would be able to find correlation between things such as “driving distance” and “average score.”

This data won’t need to be collected anymore, since the data set contains all the information I need. However, the data will need to be pre-processed. There are unnecessary data columns that can be removed, and I’ll need to handle any null values in the data set. I will have to divide the data into training, validation, and test sets in order to use any regression technique.

By the end of February, I plan to have selected the attributes that I will use to train my model. Starting in March, I will do data preprocessing and begin building the model. April will be used for evaluating and tweaking my model. And finally, creating the final report and presentation.