CSC 466 - Project Warmup

Summary Statement:

We want to create a model that takes in baseball player statistics and predicts specific values that we choose. Given this model, we should be able to choose a few statistics that we want to predict and, based on the other statistics that we have, predict them accurately. For example, we want to be able to predict the batting average that a player will have in the upcoming MLB season.

Specific Aims:

- A neural net with player statistics as the input and a specific statistic as the output.
- A linear regression model that predicts future performance statistics based on the input statistics.
- A gradient boosted model that predicts future performance statistics based on the input statistics.
- A presentation that highlights the most effective model and important features and displays the results of that model in an easy-to-understand fashion.

Roles:

Kyle Jennings -

- Project Manager
    - Designate specific locations for the code to be stored and procedures for working collaboratively.
    - Work on presentation material finalization.
    - Help with all aspects of the project that need extra work

Dhaval Panchal -

- Machine Learning Engineer
    - Develop a model using Jupyter notebook to predict a given statistic given the set of features

Ryan Zhang -

- Data Analyst
    - Wrangling initial data into a form that is easy for the Machine Learning Engineer to work with.
    - Take initial findings from the preliminary model and use them to improve the model's accuracy.

Timeline(w/ deliverable due dates):

Week 0 (4/19-23):
- Project Pitch due 4/21
- Weekly Flipgrid

Week 1 (4/26-30):
- Finding a dataset
- Weekly Flipgrid

Week 2 (5/3-7):
- Preprocessing the data set
- Weekly Flipgrid
- Data visualization and preliminary analysis

Week 3 (5/10-14):
- Create code to train based on the data set
- Weekly Flipgrid
- Begin training

Week 4 (5/17-21):
- Analysis of training results
- Weekly Flipgrid
- Creation of models to visualize the training results

Week 5 (5/24-28):
- Finalize training
- Weekly Flipgrid
- Create a final report and presentation based on findings

Week 6 (5/31-6/4):
- Finalize report and presentation
- Final Project Due Date #1

Week 7 (6/7-11):
- Final Project Due Date #2

Final Deliverables:

We intend to deliver a Jupyter notebook that contains the previously mentioned models used to predict statistics for a given baseball player in a specific year. We will also deliver a slide deck that explains the differences between the models and highlights which model performed the best.