

Mike Wang

MAIS 202

Jan 29<sup>th</sup>, 2019

[siyu.wang5@mail.mcgill.ca](mailto:siyu.wang5@mail.mcgill.ca)

## Data Selection Proposal

Idea: Base on the past game result and use machine learning to predict the winning (with highest points) rate of *FC Barcelona* in *Liga de Primera division* in the 2019-2020 season.

Data Set: <https://datahub.io/sports-data/spanish-la-liga>

I chose this dataset is that it not only contains 10 datasets of all clubs game result from 2010 to 2019, but also game factor data such as number of shoot, number of pass, number of critical save, etc, It would be more than enough to train my model. I will dis-assemble the dataset in to different columns individually, since each factor have different impact on the final result.

The methodology I want to use in this project is gradient decent with regulations. Since this project meanly focusing on how to estimate the winning rate of *FC Barcelona*. Since I have all data for each team I can even further expend my idea to let the model to predict the winning rate of the each team in the Division.