PROJECT PROPOSAL

Project Title: Predicting winner of a Soccer Match

The metrics being used here is: Player Rating, Player Performance, Team Rating, Overall Team Performance and Team's past Performance.

Dataset:

European Soccer Database. The dataset has been scraped from http://football-data.mx-api.enetscores.com/ and https://www.kaggle.com/hugomathien/soccer)

Project Description:

Football or **Soccer**, is a team sport played between two teams of eleven players with a spherical ball. It is played by 250 million players in over 200 countries and dependencies making it the world's most popular sport. The game is played on a rectangular field with a goal at each end. The object of the game is to score by getting the ball into the opposing goal. In this project, we will predict the most probable winner given two teams using the different attributes of this game such as: the players and their performance, overall teams' performance, and whether the team is playing in its home ground or not.

We plan on applying Support Vector Machines, Logistic Regression and K-Nearest Neighbor Algorithms for the prediction. The reason for choosing the above SVM is because its defined by a convex optimization problem (no local minima). The reason for choosing logistic regression is that don't have to be normally distributed, or have equal variance in each group. The maximum-likelihood ratio is used to determine the statistical significance of the variables. The reason for choosing KNN is because the dataset used is very large and since KNN is robust to noisy data, we will be able to achieve a better accuracy with KNN.As a part of the conclusion, we will compare the accuracy of the different classifiers.

Software/Tools: Anaconda (Spyder) will be used to develop the project in python

Teammates:

1. Mohanakrishna Vanamala Hariprasad (mxv160430)
2. Vinaya Shashidhara Ganiga (vsg160130)

References:

[1] Sierra, Adrian, et al. "Football futures." *URL http://cs229. stanford. edu/proj2011/SierraFoscoFierro-FootballFutures. pdf*.