

CMPT 733 Big Data Programming – Final Project

Call for Project Idea

Title of the Project:

Predicting Soccer games and tournaments

Description:

Predicting soccer games has always been a very interesting challenge for soccer fans. Today, many professional teams use the power of big data and machine learning not only to predict games, but to get the latest health status of players based on the information they get from GPS trackers, predict the game plan that they opponent will play and how they could counter that.

An application that could predict the winner of a game or a tournament (given that world cup is in this upcoming summer and it's the most watched tournament of the world) could certainly be very interesting and helpful at the same time. The challenge between predicting a game winner and a champion of a tournament would be how we could differentiate these two when building our model (What attributes separates them, if any). There are number of articles that explains how big data is changing the world of soccer which I found very interesting:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4996805/>

<http://data-informed.com/how-big-data-is-changing-the-world-of-soccer/>

<https://www.theguardian.com/football/2017/mar/30/expected-goals-big-football-data-leicester-city-norwich>

Datasets

This Kaggle page offers number of datasets which could be used. It's very detailed and contains a lot of information from teams to players and even betting odds from 2008 to 2016.

<https://www.kaggle.com/hugomathien/soccer>

They also offer an application that would scrape the latest results and players statistics which could be very helpful for the model, however using it depends on how much you have worked with crawlers. You could also use free crawler websites like apfiy.com to get the data that you need. The github that contains the code is here:

<https://github.com/hugomathien/football-data-collection>

Contact person:

If you have any questions, please reach me at: Saeeds@sfu.ca or soltani@live.com

Contributor of the Project Idea:

Saeed Soltani
Data Engineer
Best Buy Canada