

Statement of work



Anmol Panna (100722039)

Sharang Ahluwalia (100700203)

Abdul Rahman Khan (100719011)

Levi Nair (100733132)

Course Name: Capstone Term II

Course Code: AIDI 2005

Facilitator: Marcos

Date: 07 Feb 2020

**Statement of Work**

**Data Requirements:**

The main data requirements for this project is comprised of historical data from games played and the specifics of every match viz.

1. Home Goals
2. Away goals
3. Team form
4. Team line up
5. Team formation
6. Team gameplay
7. Match stadium
8. Match referee
9. Player attributes
10. Match stats
11. Bookmaker’s odds for every match
12. FIFA Stats

Assumptions: We assume the NULL values to be set to 0, the data is normalized, no player injuries, players age is optimal and no fixed matches.

Constraints: The features extracted and excluded the remainder of features, monotonicity of predicted outcome and feature interactions.

Limitations: The limitations for this dataset is if a player is included in the team and we don't have the stats of the player or it may be his debut then it will be difficult to evaluate team and result probability.

**Data Needs**

Data that is needed for our solution to work is understanding goal prediction, team dynamics such as away or home, player attributes and team line up.

**Data Creation**

Data that must be created with regards to the type of passes that were made in game. Having passes data would allow our models to better predict the number of expected goals a team should have scored in a match.

**Unattainable Data**

As away and home goals becomes key feature, we also need to know how many goals a team scored against a team when they play at home or away. We also require who are all the players that involved in team whey they won and lost. These are the features which we are unable to find.

**Data Evaluation**

When we consider player attributes, we do not require goals scored attribute for goalkeeper or tackles for forward. These are the unnecessary attributes. We don’t look at team leagues like whether the team is from England or Spain I.e. Table – League, Attribute – name. We also don’t require players birthday, height, weight to achieve solution.

In contrast, we would consider team dynamics such as formation, past gameplay history and line up since these features would accurately assist in predicting goal predictions as well as geographical positions on the field, we could potentially extrapolate angles and create new columns or features that would further increase predictions.

**Dataset:**

The dataset being used is “European Soccer Database” has been acquired from Kaggle (<https://www.kaggle.com/hugomathien/soccer>) and data.world (<https://data.world/data-society/european-soccer-data>) which contains,

* +25,000 matches
* +10,000 players
* 11 European Countries with their lead championship
* Seasons 2008 to 2016
* Players and Teams' attributes sourced from EA Sports' FIFA video game series, including the weekly updates
* Team line up with squad formation (X, Y coordinates)
* Betting odds from up to 10 providers
* Detailed match events (goal types, possession, corner, cross, fouls, cards etc...) for +10,000 matches.