**Capstone Project Proposal**

**“Predicting the outcome of the football matches”**

1. ***Aims of the Project.***

The aim of the project is to predict the outcome of English Premier League. The work will attempt to rich the same level of accuracy as the betting companies have. Of course, it will be more interesting to set the goal of the project to: bit the bookmaker, but its too ambitious at this point.

The model we are planning to build will produce three outcomes: win, loose or draw. At the end of the work, hopefully we will be able to tell which of the outcomes is most likely to occur with the successful rate of prediction

1. ***Decisions based on the outcomes of the model.***

Based on the results of the model, one can more effectively spread the risk and optimally allocate his/her budget in order to maximise the profit from betting. However, keeping in mind that bookmaker might have more effective model and better strategy.

1. ***Data***

The data on the home and away games needs to be acquired for the base model. Afterwards we will include other variables and the ones which will have the significant effect on the outcome will be included. The data will be used from the following web sites:

[www.championat.com](http://www.championat.com)

[www.statto.com](http://www.statto.com)

Other resources will be explored in the near future.

1. ***Approaches and steps for solving the problem.***

First of all, we need to extract necessary information from the data. Therefore the data extracted should meet the following conditions:

* 1. Does the extracted data seem legit?
  2. Does it help to answer the question?

To this point we are planning to create the table with all the results (home and away) for each season. Afterwards we are going to choose one of the various machine learning algorithms such as (random forest, logistic regression) to work with. Once the method is chosen the base model will be estimated. Next, we are going to add other explanatory variables. Finally, we will try to modify our model until the satisfactory level of the accuracy will be achieved (70%).

1. ***Deliverables.***

a) Final report of the findings about 4-5 pages long.

b) The code and the graphs produced.

c) Presentation of the project and the discussion of the results