**MSc. in Computing**

**Practicum Approval Form**

# Section 1: Student Details

|  |  |
| --- | --- |
| Project Title: | Predict match outcome in English Premier League Football |
| Student ID: | **21260470, 21261256** |
| Student name: | Aneeta Charly,Binit George |
| Student email | binit.george2@mail.dcu.ie |
| Chosen major: | Data Analytics |
| Supervisor | Andrew McCarren |
| Date of Submission |  |

# Section 2: About your Practicum

Please answer all questions below. Please pay special attention to the word counts in all cases.

**What is the topic of your proposed practicum? (100 words)**

The aim of our practicum is to predict match outcomes with respect to English Premier League Football. By using data available about the yearly match statistics and player related data we aim to use a suitable model to predict the match outcome in EPL. The important features which contribute to the match outcome will be identified and using those features a highly accurate predictive model will be determined and experimented as part of our practicum.  
  
**Please provide details of the papers you have read on this topic (details of 5 papers expected**).

1. Alfredo, Yoel Frans and Sani Muhamad Isa. “Football Match Prediction with Tree Based Model Classification.” *International Journal of Intelligent Systems and Applications* (2019): n. pag.

Rank: H-Index 17, Q3 (2019)

1. Baboota, Rahul & Kaur, Harleen. (2018). Predictive analysis and modelling football results using machine learning approach for English Premier League. International Journal of Forecasting. 35. 10.1016/j.ijforecast.2018.01.003.

Rank: H-index: 96, Rank: Q1

1. R. Pariath, S. Shah, A. Surve and J. Mittal, "Player Performance Prediction in Football Game," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 1148-1153, doi: 10.1109/ICECA.2018.8474750.

Rank: H-index: 8

1. D. Prasetio and D. Harlili, "Predicting football match results with logistic regression," 2016 International Conference On Advanced Informatics: Concepts, Theory And Application (ICAICTA), 2016, pp. 1-5, doi: 10.1109/ICAICTA.2016.7803111.

Rank: H-index: 6

1. Razali, Nazim et al. “Predicting Football Matches Results using Bayesian Networks for English Premier League (EPL).” (2017).

[IOP Conference Series: Materials Science and Engineering](https://iopscience.iop.org/journal/1757-899X), [Volume 226](https://iopscience.iop.org/volume/1757-899X/226), [International Research and Innovation Summit (IRIS2017) 6–7 May 2017, Melaka, Malaysia](https://iopscience.iop.org/issue/1757-899X/226/1)

Rank: H-index: 44

**How does your proposal relate to existing work on this topic described in these papers?** (200 words)

All the literature is related to predicting match outcome by first identifying the variables that can be reliable from the data available and then applying various models to get the highly accurate prediction.

In [1], the author identifies the reliable variables by feature selection using backward wrapper model and then uses supervised machine learning for prediction. The author uses models – random forest,c5.0 and extreme gradient boosting algorithms and compares the results. The author points out that logistic regression is not effective in predicting match outcomes as it can give only two outcomes-win or lose whereas in a match there are three outcomes- win, lose or draw.

[2] uses feature engineering and exploratory analysis to identify the important variables that aid prediction of a match.

[3] uses player attributes to predict the performance of the player and the market value of the player. The author uses data visualisation and Principle component analysis to identify the required player attributes that contribute to identify the players performance. They use linear regression model to predict the player performance.

Similar to the existing work, we aim to identify those features or variables that affect the outcome of a match and evaluate how our variables contribute to the accuracy of prediction of a match outcome.

**What are the research questions that you will attempt to answer? (200 words)**

Our practicum aim to achieve the following:

* The most important factors that influence the outcome of a match in EPL.
* Derive a model that can be highly accurate in predicting the match outcome and validate the results with the models used in the published literature.
* Do player parameters influence the match outcome?

**How will you explore these questions?** (Please address the following points. Note that three or four sentences on each will suffice.)   
  
- What software and programming environment will you use?   
- What coding/development will you do?   
- What data will be used for your investigations? .

* Data related to various matches over the years in EPL from <https://www.football-data.co.uk/data.php>

This site has information on various home and away matches and related match stats of the respective games.

* Player related data is planned to be taken from fifa website – sofifa.com.

In this website various player related factors and respective players fifa ranking is available year wise.

- Is this data currently available, it not, where will it come from?

Historic match statistics of EPL is available as downloadable csv files on the website(https://www.football-data.co.uk).

Fifa rankings and other player related data need to be web scraped from the website(sofifa.com).

- What experiments do you expect to run?   
- What output do you expect to gather?

Predict match outcome for the last two seasons based on the historic data we collected.

- How will the results be evaluated?

Cross validate the prediction result to the actual match results to identify the accuracy of prediction models and the relevancy of the variables selected.