

# Football Match Prediction and Outcome Simulation using Machine Learning

*Predicting football matches outcomes using different machine learning techniques with derived performance evaluation metrics and learning the sustainability of team's performance through match simulations.*

## File Contents

### Datasets (raw and generated)

- `matchID\_plEventsData.csv` : Raw event tracking data across 380 matches in the EPL 2020/21 season by Opta
- `all\_shots\_model.csv` : Data of all shots taken across 380 matches in the EPL 2020/21 season
- `shots\_model.csv` : Data of all shots taken excluding penalties and own goals across 380 matches in the EPL 2020/21 season (used to train xG model)
- `xg\_all\_shots\_model.csv` : xG values applied to all shots taken across 380 matches in the EPL 2020/21 season
- `withoutxA\_pred.csv` : Match data fitted with total xG generated and Elo ratings for respective teams
- `prediction\_data.csv` : Match data fitted with total xA generated
- `final\_pred\_data.csv` : Complete match data to be used in prediction models

### Code / Notebooks

- `requirements.txt` : List of all packages with respective versions used in the entire work
- `FCPython.py` : Pitch visualisation functions
- `xG\_Model\_Viz.ipynb` : Visualisations on extracted shot data
- `xGModel\_build.ipynb` : Expected goals (xG) model development
- `genNewDS.ipynb` : Data handling and preprocessing for prediction models
- `SVM\_prediction.ipynb` : Prediction model using Support Vector Machines (SVM)
- `RF\_prediction.ipynb` : Prediction model using Random Forest (RAF)
- `xGB\_prediction.ipynb` : Prediction model using Extreme Gradient Boosting (XGBoost)
- `MCSimulation\_xG.ipynb` : Match outcome simulation

## Guide

### Prerequisites

1. `Python 3.7.13` is used. Any later versions should work.
2. Jupyter Notebook is used to run all the following codes.
3. Download and unzip the repository.
4. Run `pip install -r requirements.txt` to install all required packages.

### Expected Goals (xG) Model

1. Make sure `FCPython.py` is in the same directory
2. Run `xG\_Model\_Viz.ipynb` to build visualisations and extract shots and goals data
3. Observe shots and goals on pitch visualisations in the `ShotsModel\_Output` folder.
4. Run `xGModel\_build.ipynb` to build the xG model and fit it into our event data.

5. Observe xG model visualisations in `xG\_Output` folder.

### **Prediction Models**

1. File loading instructions stated inside respective notebooks depending on the notebooks used. (cloud-based or local PC)
2. Run `SVM\_prediction.ipynb` to build prediction model using SVM.
3. Run `RF\_prediction.ipynb` to build prediction model using Random Forest.
4. Run `xGB\_prediction.ipynb` to build prediction model using XGBoost.

### **Match Outcome Simulation**

1. Run `MCSimulation\_xG.ipynb` to create match simulations using Monte Carlo simulation.
2. Select a match to run the simulation by inputting your desired match ID from [Understat](https://understat.com/league/EPL) when prompted.

! [Browser URL from Understat](https://www.linkpicture.com/q/understat\_ins.png)

The numbers displayed after "match/" is the match ID and it can be obtained from your browser's url after selecting on your desired match from [Understat](https://understat.com/league/EPL).

3. Observe results and visualisations generated.