1. **Player Recommendation System for Fantasy Premier League using Machine Learning**

Vimal Rajesh, P Arjun, Kunal Ravikumar Jagtap, Suneera C M, Suneera C M

2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)

* + The paper aims to enhance decision-making for Fantasy Premier League (FPL) players by providing data-driven recommendations.
  + FPL is a popular online game where participants create virtual teams of real-life football players and earn points based on their performance in actual matches.
  + The authors recognize the favouritism bias, where players tend to select players from their favourite teams rather than making informed choices.
  + They propose a recommendation system that leverages machine learning to suggest optimal player combinations.
  + The system uses data extracted from the FPL API.
  + The testing period corresponds to the English Premier League 2021–22 season.
  + The paper takes into consideration the Form, Return on Investment, Fixture Difficulty Rating, Bonus Points System, Points Per Game, and Influence Creativity Threat to determine the points of a player.
  + The paper proposes Random Forest and Gradient Boosting Machines to generate expected points for a player.
  + The paper uses Mean Absolute Error (MAE) is used as a metric to evaluate the model.
  + It explores statistical analysis and data science techniques to generate better recommendations.

1. **Sports Analytics algorithms for performance prediction**

**Konstantinos Apostolou, Christos Tjortjis**

**2019,** [**10th International Conference on Information, Intelligence, Systems and Applications (IISA)**](https://sci-hub.se/10.1109/IISA.2019.8900754)

* + The dissertation is divided into two major parts. The first part comprises a **literature review** of existing technologies related to sports analytics. The second part focuses on **experiments** conducted primarily using football data.
  + The first step involves gathering relevant data. In the context of football, this data could include player statistics, match results, team formations, and more.
  + Once collected, the data needs preprocessing. This step includes handling missing values, normalizing features, and ensuring data consistency.
  + Feature engineering is crucial for building effective predictive models. It involves creating new features or transforming existing ones to improve model performance.
  + For player performance prediction, features might include player position, playing time, goals scored, assists, passes completed, and defensive actions.
  + The model tries the following algorithms Naïve Bayes, Decision Tree, Random Forest, KNN, SVM (rbf kernel), SVM (poly kernel), XGBoost
  + The selected algorithm(s) are trained on the preprocessed data. The model learns patterns from historical data.
  + Evaluation metrics (e.g., accuracy, precision, recall) assess how well the model performs.
  + Once the model is trained, it can predict player performance based on input features.
  + Insights gained from predictions can guide team management decisions, such as player selection, substitutions, and tactical adjustments.

1. **Game Plan: What AI can do for Football, and What Football can do for AI**

Karl Tuyls, Shayegan Omidshafiei, Paul Muller, Zhe Wang, Jerome Connor, Daniel Hennes

May 2021, Journal of Artificial Intelligence Research

* + The paper is a theoretical discussion that sets the stage by emphasizing the unique synergy between artificial intelligence (AI) and the world’s most popular sport, football.
  + It acknowledges that football analytics provides a rich playground for AI research due to its complexity, real-world dynamics, and massive data availability.
  + **Data-Driven Insights**: AI techniques enable data-driven insights into player performance, team strategies, and match outcomes.
  + **Predictive Models**: Algorithms predict player ratings, injury risks, and even transfer market values.
  + **Computer Vision**: AI-powered systems track player movements, ball trajectories, and tactical patterns during live matches.
  + **Fan Engagement**: Personalized content, interactive apps, and augmented reality experiences enhance fan engagement.

The paper discusses the following Challenges in Football Analytics:

* + **High Dimensionality**: Football data is multidimensional, including player attributes, match events, and contextual factors.
  + **Temporal Dependencies**: Understanding how events unfold over time is crucial.
  + **Noise and Uncertainty**: Real-world data is noisy, incomplete, and subject to various uncertainties.

**Statistical Learning Techniques**:

* + **Regression Models**: Predicting continuous variables (e.g., player performance metrics).
  + **Classification Models**: Identifying player positions (e.g., forward, midfielder, defender).
  + **Clustering**: Grouping similar players based on playing style.
  + **Time Series Analysis**: Capturing temporal patterns in match data.

**Game Theory Applications**:

* + **Penalty Kicks**: Applying game theory to analyze optimal strategies for penalty takers and goalkeepers.
  + **Player Interactions**: Modeling interactions between players as strategic games.

**Computer Vision Challenges**:

* + **Player Tracking**: Extracting player trajectories from video feeds.
  + **Pose Estimation**: Inferring player body positions and orientations.
  + **Event Detection**: Recognizing goals, fouls, and other key events.

1. **Multi-stream Data Analytics for Enhanced Performance Prediction in Fantasy Football**

Nicholas Bonello, Joeran Beel, Seamus Lawless, Jeremy Debattista

2019, 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science

* + **Fantasy Premier League (FPL)** performance predictors often rely solely on historical statistical data to predict player performances.
  + However, this approach overlooks external factors such as injuries, managerial decisions, and other tournament match statistics.
  + Traditional statistical predictors lack the ability to incorporate real-time information and human feedback.
  + Predictions based solely on historical data may not account for dynamic changes during a season.
  + The authors introduce a novel method that enhances performance prediction by **automatically incorporating human feedback** into the model.
  + They consider multiple streams of data beyond historical statistics to improve predictions.

**Data Streams Used**:

* + **Previous Performances**: Historical player performance data.
  + **Fixture Difficulty Ratings**: Upcoming fixture challenges.
  + **Betting Market Analysis**: Insights from betting odds.
  + **General Public and Expert Opinions**: Social media discussions, web articles, and expert insights.
  + The proposed model was tested on the **English Premier League 2018/19 season**.
  + It outperformed regular statistical predictors by over **300 points**, averaging **11 points per week**.
  + The model ranked within the top **0.5%** of players, achieving a rank of **30,000 out of over 6.5 million players**.
  + By incorporating diverse data streams and human feedback, this approach provides more robust and accurate predictions.
  + It demonstrates the potential of combining statistical analysis with real-world insights for fantasy football enthusiasts

1. **Football Player Value Assessment Using Machine Learning Techniques**

**Ahmet Talha Yiğit, Barış Samak**and**Tolga Kaya**

**2019,** [**27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science**](https://link.springer.com/chapter/10.1007/978-3-030-23756-1_36)

* + **Sports analytics** is a growing field worldwide, and one of its open problems is the valuation of football players.
  + The aim of this study is to establish a **football player value assessment model** using machine learning techniques to support transfer decisions by football clubs.
  + The proposed models are mainly based on **intrinsic features** of individual players, as provided in the **Football Manager** video game.
  + The individual statistics of **5316 players** active in **11 different major leagues** from Europe and South America serve as the dataset.
  + The study employs advanced supervised learning techniques:
    - **Ridge and Lasso Regressions**: Regularized linear regression models.
    - **Random Forests**: Ensemble models combining decision trees.
    - **Extreme Gradient Boosting (XGBoost)**: Gradient boosting algorithm.
  + All models are built in the **R programming language**.
  + The models’ performances are compared based on their **mean squared errors**.
  + An **ensemble model with inflation** is proposed as the output.
  + The goal is to value players based on their **normalized abilities**, relatively free from environmental variables.
  + The model aims to provide better results than current models relying solely on in-field game statistics.
  + Considering the expansion of the football industry, a model using the latest developments in data analytics and machine learning addresses a significant problem for the industry.
  + It can be a valuable financial leverage for clubs seeking to expand their successes and profits.
  + This research contributes to the world of football by providing a data-driven approach to player valuation.
  + By leveraging machine learning techniques, clubs can make more informed transfer decisions.

1. **Fantasy Premier League - Performance Prediction**

Pratik Pokharel, Arun Timalsina, Sanjeeb Panday, Bikram Acharya

2022, 12th IOE Graduate Conference

* This paper proposes a rational approach to player selection, team drafting, and transfers by predicting ROI using xgboost regression.
* The study also evaluates the impact of fixture congestion on FPL points using mid-week cup fixture data.
* Evaluation based on FPL global ranking reveals that initial drafted teams without transfers performed better than those with transfers, which were hindered by dependency on the accuracy of the regression model.
* The mean RMSE score for all players was 2.048, and the effect of cup fixture congestion on FPL points was found to be insignificant.
* The uncertainty of team selection makes it challenging for FPL managers, with Game-week 1 being crucial for laying the foundation of the season.
* The paper highlights the vast number of potential team combinations and the difficulty in making optimal selections.
* FPL analytics traditionally rely on historical statistical data but may overlook external factors such as mid-week fixture fatigue and squad rotation strategies employed by managers due to European and domestic cup competitions.

Abstract:

FPL (Fantasy Premier League) is one of the most popular fantasy leagues based on arguably the most popular sports annual event in the world i.e. the English Premier League. Sports Analytics has been a challenge for even the best of pundits and personal biasness and favouritism and also the human mind’s inability to comprehend and account for the entire season and entirety of the 38 fixtures for each of the 20 teams in the league contribute to the poor performance of an individual and that’s where data engineering, machine learning and artificial Intelligence can assist an individual in setting up the dream team. A Fantasy Premier league season however is not as straight-forward as there’s a budget cap of 100 million for the FPL Managers, a max limit and minimum limit to players of each position and also from each team and a number of specific metrics that award players from different positions points. Moreover, there are Game weeks where a team has no fixtures and game weeks where a team has double fixtures and also there’s a restriction of one free transfer per game week and 3 special wildcard chips. Hence, it is very important when building the team for game week 1 as it can make or break your season. Hence, we aim to build a platform that gives FPL managers insights by using visualisations, data engineering, ml and ai to provide predicted points for each player for each match, create best possible teams for the season and for the game week, transfer tips and build and train an AI model that can compete with other managers on FPL and make decisions.

System Model with Description:

Before each Game week, the Model plans for future Game week fixtures with decreasing order of priority to each successive fixture, and for each player calculates Team Form and Predicted Team Scoring Index (xGfor) and Predicted Team Conceding Index (xGagnst) by using Form, Attack Index, Defence Index, Fixture Difficulty Rating FPL API data, previous history from FPLAnalytics.com and whoscored.com, time-series forecasting from previous season data and team fatigue Index from other Fixtures from whoscored.com.

Now, if the player is an Attacker/ Midfielder, we take the Predicted Team Score Index (xGfor) and the probable minutes per 90 index (xMP) and the match fitness (fatigue/ rest probability) and the Form of the player and the expected Scoring Index (xSc) and the expected Assisting Index (xAst) to calculate the probable points for the game (xPts).

If the player is Defender/ Goalkeeper, we take the inverse of Predicted Team Conceding Index (xGagnst) and the probable minutes per 90 index (xMP) and the match fitness (fatigue/ rest probability) and the Form of the player and the expected save Index (xSv) and the discipline Index to calculate the probable points for the game.

For each player, we also consider a player Selection popularity Model where we using % selected, % captained data from FPLAnalystics and FPL API and using Twitter Sentiment Analysis on tweets on threads related to @FPLtweets for Sentiment analysis for players and formulate Popularity Metric using BERT (Bidirectional Encoder Representations from Transformers)

We then design a statistical model that predicts the best permutations of team (playing XI + 4 bench players) with formation using Predicted points per Cost Unit (Million) and create best possible team that maximises the score for the upcoming fixtures.

For Time-Series Modelling we will be using a hybrid of Autoregressive Integrated Moving Average (ARIMA) and Recurrent Neural Networks (RNNs) for time series prediction of player points and subsequent maximisation of total points using Linear Programming (LPP).

For Prediction of Fixture results and points for same season we will be using Random Forest Regressors with extreme Gradient Boosting (xgBoost).

We aim to feed trained data to an LLM model to create a chat interface for user with the data.