First Review Document

Fantasy Premier League Assistant

Rajarshi Saha
20BCT0163
8017329178
rajarshi.saha2020@vitstudent.ac.in

Ankan Ray
20BCE2304
9064278580
ankan.ray2020@vitstudent.ac.in

Dr. Ebenezer Juliet
Associate Professor Senior
9443452364
ebenezer.juliet@vit.ac.in

B.Tech.

in

Computer Science and Engineering

School of Computer Science & Engineering

®

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

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 favoritism 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 4 special game-chips: 1. Wildcard, 2. Bench Boost, 3. Free Hit, 4. Triple Captain. 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 visualizations, 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.

Introduction

In the realm of Fantasy Premier League (FPL), the blend of data analytics and strategic decision-making is pivotal for success. Recognizing the challenges FPL managers face in player selection, fixture planning, and maximizing points within budget constraints, our platform aims to revolutionize team management.

By leveraging data engineering, machine learning, and artificial intelligence, our platform provides FPL managers with invaluable insights and predictive analytics. Our model meticulously plans for each game week, prioritizing fixtures based on team form, fixture difficulty ratings, and player fatigue.

We calculate predicted team scoring and conceding indices, enabling precise player point projections. For attackers and midfielders, factors such as team scoring indices, match fitness, and player form are considered. Defenders and goalkeepers are evaluated based on predicted team conceding indices and save performance.

Additionally, our platform incorporates a player popularity model, utilizing data from FPL Analytics and Twitter sentiment analysis to gauge player popularity and sentiment within the FPL community.

To optimize team selection, a statistical model predicts the best permutations of the playing XI and bench players, maximizing points while adhering to budget constraints. Time-series modelling ensures accurate predictions of player points over successive game weeks.

Through the integration of advanced analytics and intuitive interfaces, our platform aims to empower FPL managers, redefining the FPL experience and enabling managers to navigate the complexities of the Premier League with confidence and success.

Objective

The primary objective of the Fantasy Premier League (FPL) Assistant project is to develop a comprehensive platform that leverages data engineering, machine learning, and artificial intelligence to provide FPL managers with actionable insights and tools to optimize their team management strategies. The key objectives include:

- To analyse and predict player performance and points for each game week based on historical data, fixture difficulty ratings, and other relevant factors.
- To assist FPL managers in making informed decisions regarding player selection, transfers, and team formation.
- To prioritize fixtures and plan strategies for maximizing points while adhering to budget constraints.
- To enhance user experience and engagement by providing intuitive interfaces and personalized recommendations.
- To continually improve and refine the platform through feedback and iteration, ensuring its relevance and effectiveness for FPL managers.

Problem Statement

Fantasy Premier League (FPL) managers face several challenges in effectively managing their teams and maximizing their points throughout the season. These challenges include:

- Limited insight into player performance and points potential for upcoming fixtures.
- Difficulty in prioritizing fixtures and planning strategies to optimize team selection and transfers.
- Lack of personalized recommendations and insights tailored to individual team compositions and preferences.
- Inefficient use of budget constraints, leading to suboptimal team configurations.
- Time-consuming and manual processes for analysing data and making decisions, detracting from the overall enjoyment of the FPL experience.

Literature review

1. Player Recommendation System for Fantasy Premier League	Vimal Rajesh, P Arjun, Kunal Ravikumar Jagtap, Suneera C M 2022	The paper aims to enhance decision-making for Fantasy Premier League (FPL) players by providing data-driven recommendations.

| using Machine Learning | 19th International Joint Conference on Computer Science and Software Engineering (JCSSE) | 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. |
|---|---|---|
| **2.**

Sports Analytics algorithms for | **Konstantinos Apostolou, Christos Tjortjis** | The dissertation is divided into two major parts. The first part comprises a **literature review** of existing technologies related to |

performance prediction	2019, 10th International Conference on Information, Intelligence, Systems and Applications (IISA)	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.
3. **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.
4. **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
5. **Football Player Value Assessment Using Machine**	**Ahmet Talha Yiğit, Barış Samak and Tolga Kaya**	**Sports analytics** is a growing field worldwide, and one of its open problems is the valuation of football players.

Learning Techniques	2019, 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science	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. |
| **6.**

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.
7. **Who Should Be the Captain This Week? Leveraging Inferred Diversity-Enhanced Crowd Wisdom for a Fantasy Premier League Captain Prediction**	**Shreyansh Bhatt, Keke Chen, Valerie L. Shalin, Amit P. Sheth, Brandon Minnery** **2019, 13th International AAAI Conference on Web and Social Media (ICWSM)**	Utilizes the Wisdom of Crowd effect to predict productive players in Fantasy Sports. Proposes the **SmartCrowd framework** to select a small, smart crowd using participants' Twitter posts. Three main steps: **characterizing participants using summary word vectors** **clustering participants based on these vectors** **sampling participants from clusters to form diverse crowds.** Empirical evaluation on the Fantasy Premier League (FPL) captain prediction

		problem shows SmartCrowd outperforming random crowds and crowds consisting of top experts identified from previous performance data. Social media-based diversity supports the sampling of smarter crowds that collectively predict productive players. Studies the assembly of a small subset of the crowd using semantic diversity inferred from **participants' Twitter posts**. Explores the weekly captain selection task in Fantasy Premier League (FPL) starting 11 player teams. Utilizes crowd wisdom for determining parameters influencing captain choice. Introduces the SmartCrowd approach, extending to other prediction problems, based on social media posts. **Mines FPL user tweets** to infer crowd diversity based on topic and communication patterns. Represents each Twitter user by a collection of their FPL tweets and summarizes them using **Word2vec**. **Tests multiple clustering strategies** and selects optimal representatives from clusters using **multi-objective optimization**.

		Evaluates the approach using captain picks, points earned, and participants' previous seasons' performance scores.
		Investigates questions related to the effectiveness of semantic analysis, diversity-based crowd selection, comparison with expert crowds, and the impact of diversity and crowd size.
8. **Time Series Modeling for Dream Team in Fantasy Premier League**	**Gupta, Akhil** **2017, International Conference on Sports Engineering (ICSE)**	Utilizes data from the official Fantasy Premier League website for the past five to six years. Cross-checks historical data using **Kaggle kernels** to ensure accuracy. Employs data scraping techniques in Python 3 to collect two types of datasets: **Master FPL dataset and Points dataset** for each season. Identifies and handles missing values in the datasets. Scales down the cost field in the Master FPL dataset. Creates a new ID field for matching datasets and merges points data into a single dataframe. Removes players with incomplete historical data and those who left the league, ensuring data consistency. Utilizes historical data to predict future player performance.

|		
		Implements two separate models: **ARIMA and LSTM-RNN**, which are later ensembled for improved accuracy.
		Treats each player's performance data as a separate time series.
		Validates models using data from previous seasons.
		Formulates the problem of selecting the dream team as a **linear programming (LP)** problem.
		Seeks to maximize the total points of the selected players within budget constraints.
		Utilizes the Pulp library in Python 2.7 for solving the LP problem.
		Automates the prediction process after setting optimal parameters.
		Validates models using RMSE and compares predicted values with actual values.
		Determines optimal ensemble proportions for ARIMA and LSTM-RNN models.
		Identifies the dream team for the upcoming season based on the optimized predictions.
		Analyzes additional features such as **Points per Match (PPM), Cost per Point (CPP), and Cost-point index (CPI)** to evaluate player performance.
		Discusses factors influencing player performance and team selection.

		Highlights insights gained from the study, including player loyalty, nationality distribution, and team performance analysis. Tracks the performance of the selected dream team for validation and further refinement of the approach.
9. **Football players performance analysis and formal/informal media: Sentiment analysis and semantic similarity**	**Gustavo Henrique de Sousa Silva, Rui Jorge Henriques Calado Lopes** **2021, ISCTE**	The study utilized data from **informal media (e.g., Reddit)** and **formal media (e.g., Live Match commentary, Player Ratings)** to compare semantic similarity with key phrases from **Work Domain Analysis (WDA).** Semantic Analysis involved comparing WDA key phrases with media entries using BERT to generate **vector representations** for textual data was used. Cosine similarity was computed between these representations. Comments Processing: Each comment was subdivided into sentences. Similarity Analysis: **BERT** was used to compare sentences with **WDA key phrases**, producing similarity scores ranging from 0 to 1. Output Generation: **Heatmaps** were created to visualize the similarity values obtained.

		Linguistic Analysis: Named entities and proper nouns were recognized to understand the main subjects of the match. Sentiment Analysis: Polarity and subjectivity of comments were analyzed using TextBlob and Stanza libraries. Output Generation: **Polarity** and **subjectivity** values were expressed in the range of -1 to 1 and 0 to 1, respectively. The study employed Semantic Analysis to compare media entries with WDA key phrases and Sentiment Analysis to analyze the polarity and subjectivity of comments. These methodologies provided insights into the perception of football matches across different media sources.
10. **TF-Pundit: A Real-time Football Pundit based on Twitter**	**Karthik Anantha Padmanabhan**	**Actor Model Implementation**: The system, named **"TF-Pundit,"** is implemented using the Actor model, which facilitates concurrent computation. Actors receive messages and make decisions based on them, allowing the system components to execute concurrently. **Text Processing Components**: **Performance Analyzer**: Analyzes sentiment associated with players based on aggregated tweets. Utilizes a lexicon-based classifier to assign sentiment scores to players, considering words specific to football.

| | | **Summarizer**: Summarizes football events based on tweets, discarding subjective opinions and selecting objective statements.

Performance Analyzer Evaluation: Compares ratings assigned by TF-Pundit with those from Goal.com, using a scaling formula. Scores are scaled between 0 and 5.

Summarizer Evaluation: Manually assigns scores to summaries on a scale of 1-10 based on how well they summarize one-minute intervals of football events.

The current implementation assumes that sentiment expressed in a tweet is solely for the mentioned player, which may not always be accurate. Tweets often express sentiments about multiple players simultaneously, leading to inaccuracies in sentiment analysis.

Generally, TF-Pundit's ratings were not significantly different from Goal.com ratings, indicating reasonable accuracy in assessing player performances.

Instances where discrepancies occurred were analyzed:

For example, a significant difference in the rating for "Shaarawy" was attributed to a specific event during the game that led to an inflated score due to numerous positive tweets. |
| --- | --- | --- |

|		This highlights the system's sensitivity to specific events and the need for context-aware analysis. Prominent games, characterized by a higher volume of tweets and more noise, posed greater challenges for accurate player performance assessment.

System Architecture
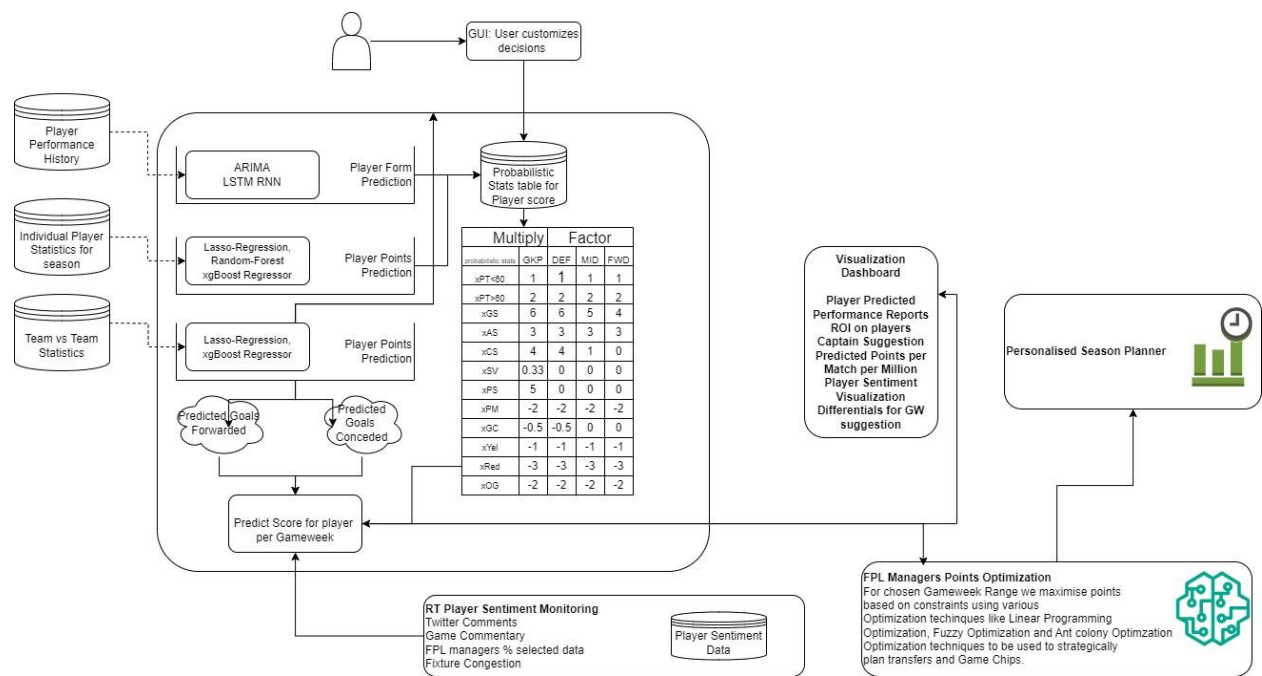
Before each Game week, the Model plans for future Game week fixtures with decreasing order of priority to each successive fixture, and for each player calculates Team Form and Predicted Team Scoring Index (xGfor) and Predicted Team Conceding Index (xGagnst) by using Form, Attack Index, Defence Index, 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 FPLAnalytics 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)
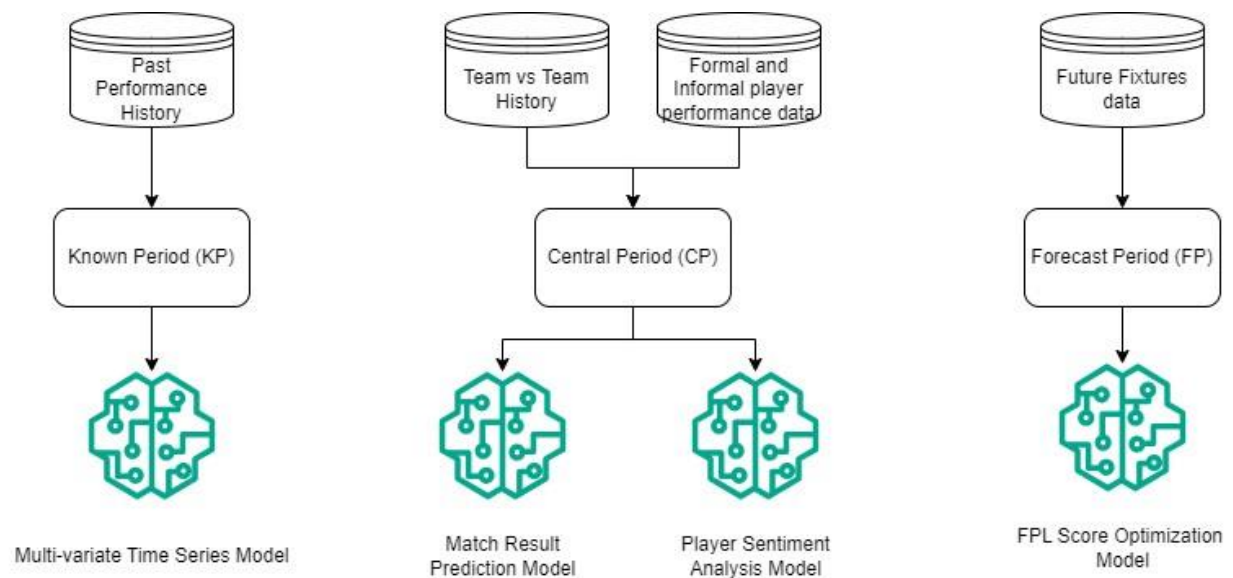
We then design a statistical model that predicts the best permutations of team (playing XI + 4 bench players) with formation using Predicted points per Cost Unit (Million) and create best possible team that maximizes the score for the upcoming fixtures.

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 maximization 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.

Conclusion

The Fantasy Premier League (FPL) Assistant presents a pivotal tool for FPL enthusiasts, offering invaluable insights and predictive analytics to optimize team management strategies. By leveraging advanced data analytics and artificial intelligence, this platform empowers users to make informed decisions, maximize points, and enhance their overall FPL experience. With its user-friendly interface and personalized recommendations, the FPL Assistant revolutionizes the way FPL managers navigate the complexities of the Premier League, making fantasy football more accessible, engaging, and rewarding for the public.

References

1. Rajesh, V., Arjun, P., Jagtap, K. R., Suneera, C. M., & Prakash, J. (2022, June). Player Recommendation System for Fantasy Premier League using Machine Learning. In *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (pp. 1-6). IEEE.
2. Apostolou, K., & Tjortjis, C. (2019, July). Sports Analytics algorithms for performance prediction. In *2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA)* (pp. 1-4). IEEE.
3. Tuyls, K., Omidshafiei, S., Muller, P., Wang, Z., Connor, J., Hennes, D., ... & Hassabis, D. (2021). Game Plan: What AI can do for Football, and What Football can do for AI. *Journal of Artificial Intelligence Research*, *71*, 41-88.
4. Bonello, N., Beel, J., Lawless, S., & Debattista, J. (2019). Multi-stream data analytics for enhanced performance prediction in fantasy football.
5. Yiğit, A. T., Samak, B., & Kaya, T. (2020). Football player value assessment using machine learning techniques. In *Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making: Proceedings of the INFUS 2019 Conference, Istanbul, Turkey, July 23-25, 2019* (pp. 289-297). Springer International Publishing.

6. Pokharel, P., Timalsina, A., Panday, S. and Acharya, B., 2022. Fantasy Premier League-Performance Prediction.

7. Bhatt, S., Chen, K., Shalin, V.L., Sheth, A.P. and Minnery, B., 2019, July. Who should be the captain this week? Leveraging inferred diversity-enhanced crowd wisdom for a fantasy premier league captain prediction. *In Proceedings of the international AAAI conference on Web and Social Media (Vol. 13, pp. 103-113).*

8. Gupta, A., 2019. Time series modeling for dream team in fantasy premier league.

9. Silva, G.H.D.S., 2021. Football players performance analysis and formal/informal media: *Sentiment analysis and semantic similarity (Master's thesis).*

10. Padmanabhan, K.A., TF-Pundit: A Real-time Football Pundit based on Twitter.