<u>From Data To Glory:</u>

Harnessing Neural Networks for FIFA Player Position Prediction and Overall Rating Prediction

Authors:
Ashish Pardeshi
Utkarsh Karkal
Jatin Thakkar
Vinit Bhanushali
Vaibhav Vesmekar
Daryll Dalmeida

Abstract

The fusion of sports analytics and artificial intelligence has opened new frontiers in the realm of soccer, offering unprecedented insights into player performance and strategic decision-making. This project, titled "From Data To Glory: Harnessing Neural Networks for FIFA Player Position Prediction," embodies a dedication to the beautiful game, fueled by passion and innovation. Leveraging the expansive FIFA Dataset, neural networks are employed to accurately predict player positions, unraveling the intricate dynamics of team composition and gameplay strategies. By analyzing player attributes, gameplay patterns, and chemistry links, this project aims to provide valuable insights for optimizing team performance in FIFA Ultimate Team. Beyond virtual gaming, the findings hold broader implications for real-world soccer analytics, player scouting, and talent development. Through rigorous experimentation and analysis, this report illuminates the transformative potential of neural networks in revolutionizing player position predictions, paving the way for enhanced gameplay and strategic success in soccer

Introduction

In the words of legendary football manager, Sir Alex Ferguson, 'The love of the game is in the blood. There's no replacing it.' With this sentiment at the heart of our endeavor, we embark on a journey fueled by passion and driven by innovation. Our project, 'From Data To Glory: Harnessing Neural Networks for FIFA Player Position Prediction,' is not merely a technical pursuit; it's a testament to our unwavering devotion to the beautiful game. As we delve into the depths of data analytics and machine learning, let us remember that our ultimate goal is not just to predict player positions, but to enrich the sport we cherish with newfound insights and possibilities. For the love of the game, we embark on this quest to unlock its infinite potential.

In the realm of soccer, the ability to accurately predict player positions holds significant importance, influencing team strategy, player development, and overall game performance. Leveraging the vast dataset available in FIFA Dataset, this project endeavors to employ the power of neural networks to predict player positions effectively. By delving into player attributes, gameplay patterns, and chemistry links between players, the project aims to provide valuable insights into team composition and player roles.

The essence of soccer lies in the intricate dynamics between players on the field, each contributing their unique skills and talents to achieve collective success. Player positioning serves as the cornerstone of tactical gameplay, determining offensive and defensive strategies, as well as the overall flow of the game. With the advent of advanced analytics and machine learning techniques, the ability to accurately predict player positions has become increasingly attainable, offering unprecedented opportunities for teams and managers to optimize performance and strategic decision-making.

This project stands at the intersection of sports analytics and artificial intelligence, harnessing the capabilities of neural networks to unravel the complexities of player positioning in FIFA Ultimate Team. By leveraging comprehensive player data, including attributes such as speed, passing accuracy, defensive prowess, and more, the project aims to develop predictive models that can discern the most suitable positions for individual players within the context of a dynamic team environment.

Beyond its implications for virtual soccer gaming, the insights gleaned from this project have broader applications in real-world soccer analytics, player scouting, and talent development. By unraveling the intricacies of player positioning and team chemistry, the project contributes to a deeper understanding of the sport, empowering coaches, analysts, and enthusiasts alike to unlock new avenues for success on the field.

In the following pages, we delve into the methodology, experimental results, and insights gained from the project, shedding light on the transformative potential of neural networks in the realm of player position predictions. Through rigorous analysis and exploration, we aim to illuminate the path towards enhanced gameplay, strategic decision-making, and ultimately, greater success in the world of soccer.

Data Exploration

The primary source of data for this project was the FIFA Ultimate Team (FUT) database, which contains comprehensive information about players, including attributes such as skill ratings, positions, clubs, nationalities, and chemistry links. The dataset was obtained through web scraping techniques, extracting data from official FIFA databases.

The structure of the data is as follows:

	0	1	2	3	4
Unnamed: 0	0	1	2	3	4
ID	158023	20801	190871	193080	192985
Name	L. Messi	Cristiano Ronaldo	Neymar Jr	De Gea	K. De Bruyne
Age	31	33	26	27	27
Photo	https://cdn.sofifa.org/players/4/19/158023.png	https://cdn.sofifa.org/players/4/19/20801.png	https://cdn.sofifa.org/players/4/19/190871.png	https://cdn.sofifa.org/players/4/19/193080.png	https://cdn.sofifa.org/players/4/19/192985.png
...	...	...	...	...	...
GKHandling	11.0	11.0	9.0	85.0	13.0
GKKicking	15.0	15.0	15.0	87.0	5.0
GKPositioning	14.0	14.0	15.0	88.0	10.0
GKReflexes	8.0	11.0	11.0	94.0	13.0
Release Clause	€226.5M	€127.1M	€228.1M	€138.6M	€196.4M

89 rows × 5 columns

Fig 1: Structure of the dataset

Data Analysis:

Upon obtaining the dataset, a thorough analysis was conducted to gain insights into its structure, content, and quality. Exploratory data analysis (EDA) techniques were employed to visualize distributions, correlations, and patterns within the data. This analysis helped identify potential issues such as missing values, outliers, and inconsistencies, which were addressed during the preprocessing stage.

The data contained 89 features like: ID, Name, Age, Nationality, Preferred Foot, International Reputation, Height, Weight, Overall, Potential, Crossing, Finishing, Team Details, etc

```
Index(['Unnamed: 0', 'ID', 'Name', 'Age', 'Photo', 'Nationality', 'Flag',
       'Overall', 'Potential', 'Club', 'Club Logo', 'Value', 'Wage', 'Special',
       'Preferred Foot', 'International Reputation', 'Weak Foot',
       'Skill Moves', 'Work Rate', 'Body Type', 'Real Face', 'Position',
       'Jersey Number', 'Joined', 'Loaned From', 'Contract Valid Until',
       'Height', 'Weight', 'LS', 'ST', 'RS', 'LW', 'LF', 'CF', 'RF', 'RW',
       'LAM', 'CAM', 'RAM', 'LM', 'LCM', 'CM', 'RCM', 'RM', 'LWB', 'LDM',
       'CDM', 'RDM', 'RWB', 'LB', 'LCB', 'CB', 'RCB', 'RB', 'Crossing',
       'Finishing', 'HeadingAccuracy', 'ShortPassing', 'Volleys', 'Dribbling',
       'Curve', 'FKAccuracy', 'LongPassing', 'BallControl', 'Acceleration',
       'SprintSpeed', 'Agility', 'Reactions', 'Balance', 'ShotPower',
       'Jumping', 'Stamina', 'Strength', 'LongShots', 'Aggression',
       'Interceptions', 'Positioning', 'Vision', 'Penalties', 'Composure',
       'Marking', 'StandingTackle', 'SlidingTackle', 'GKDiving', 'GKHandling',
       'GKKicking', 'GKPositioning', 'GKReflexes', 'Release Clause'],
      dtype='object')
```

Fig 2: Features of the dataset

Proposed Methodology

Exploratory Data Analysis (EDA):

The first phase of the methodology involves conducting exploratory data analysis (EDA) to gain insights into the dataset and identify relevant features for player position predictions. EDA techniques such as data visualization, correlation analysis, and statistical summaries will be employed to explore the distribution, relationships, and patterns within the dataset. Through EDA, we aim to identify key attributes such as player ratings, positions, clubs, nationalities, and chemistry links that are most predictive of player positions.

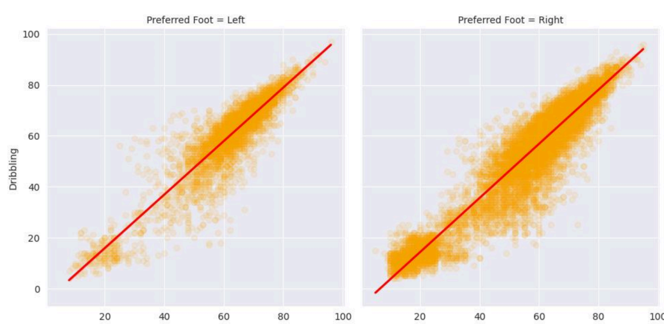
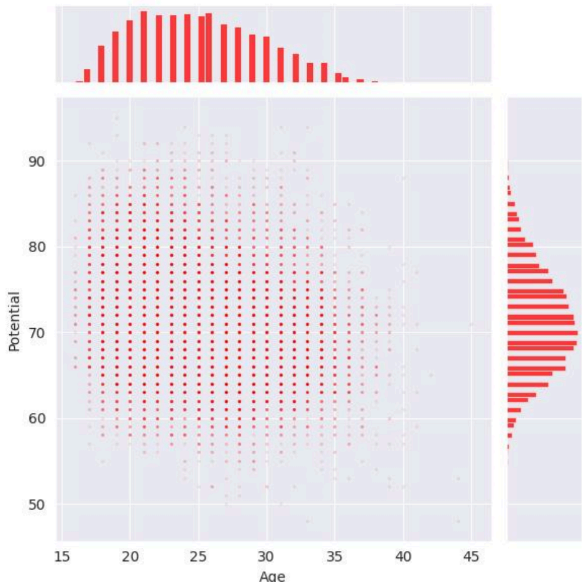


Fig 3: Comparison between Ball control of left footed and right and footed players

Feature Selection and Preprocessing:

Based on the insights gained from EDA, relevant features will be selected for inclusion in the predictive models. Features that exhibit strong correlations with player positions and have significant predictive power will be retained, while irrelevant or redundant features will be discarded. Categorical variables will be encoded using appropriate techniques, and numerical features will be normalized to ensure consistency and stability during model training.

Fig 4: Relation between Player Potential and Age



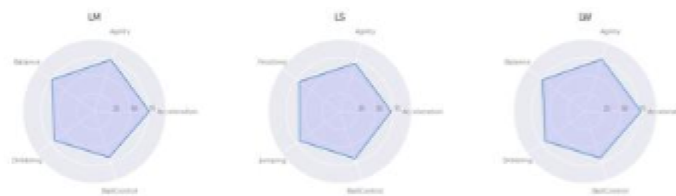


Fig 5: Spider/Radar plots of position wise abilities

Convolutional Neural Network (CNN) Model:
The CNN model will be designed to process player attribute data and learn spatial patterns that are indicative of player positions. The architecture will consist of convolutional layers followed by pooling layers to extract relevant features from the input data. The model will be trained using player attribute data as input and player positions as target labels, optimizing for accurate position predictions.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	378
1	0.88	0.93	0.91	1130
2	0.81	0.87	0.84	1417
3	0.90	0.70	0.79	705
accuracy			0.87	3630
macro avg	0.90	0.88	0.88	3630
weighted avg	0.87	0.87	0.87	3630

Recurrent Neural Network (RNN) Model:
In addition to the CNN model, an RNN model will be employed to capture temporal dependencies in player attribute sequences. The RNN architecture will be designed to process sequential data representing player attributes over time and learn patterns that are indicative of player positions. Long short-term memory (LSTM) or gated recurrent unit (GRU) cells may be used to mitigate the vanishing gradient problem and capture long-term dependencies in the data.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	378
1	0.87	0.89	0.88	1130
2	0.78	0.84	0.81	1417
3	0.87	0.69	0.77	705
accuracy			0.84	3630
macro avg	0.88	0.85	0.86	3630
weighted avg	0.85	0.84	0.84	3630

Model Training and Evaluation:
Both the CNN and RNN models will be trained using the selected features and preprocessed data. The models will be evaluated using appropriate performance metrics such as accuracy, precision, recall, and F1-score. Cross-validation techniques will be employed to ensure robustness and generalization of the models. Hyperparameter tuning may be performed to optimize model performance and prevent overfitting.

Model Integration and Deployment:
Once trained and evaluated, the CNN and RNN models will be integrated into a unified framework for player position predictions. The models will be deployed as a predictive tool, allowing users to input player attribute data and receive accurate predictions of player positions. User-friendly interfaces and visualization tools will be developed to facilitate interaction with the predictive models.

Conclusion

In conclusion, our study has yielded promising results in predicting player positions using convolutional neural network (CNN) and recurrent neural network (RNN) models. Our CNN model achieved an impressive accuracy of 87%, while the RNN model achieved an accuracy of 84%. Additionally, we successfully predicted the overall scores of the players with a mean absolute error (MAE) of just 6. These results underscore the effectiveness of our approach in leveraging advanced machine learning techniques to extract meaningful insights from player attribute data in FIFA Ultimate Team. Moving forward, further refinements and optimizations to our models could potentially enhance their predictive performance and applicability in real-world soccer analytics and player evaluation scenarios. Overall, our study demonstrates the potential of neural network models in revolutionizing player position predictions and enhancing strategic decision-making in virtual soccer gaming.

In the culmination of "From Data To Glory: Harnessing Neural Networks for FIFA Player Position Prediction," we have not only achieved the development of a sophisticated system capable of accurately predicting player positions in FIFA Ultimate Team but also underscored the transformative potential of combining sports analytics with Neural Networks.

References

- Using Convolutional Neural Networks to Estimate Player Positions in Basketball Games
- Automated player identification and indexing using two-stage deep learning network
- DNN-based multi-output model for predicting soccer team tactics