

Intelligent System for Analyzing and Predicting Football Players using Machine Learning Algorithms

Souheila Boudouda¹, Aya Merouani¹ and Lyna Rayane Meshoul¹

¹ LIRE Laboratory, Abdelhamid Mehri- Constantine 2 University, Constantine, Algeria

Abstract

These days, the margin of error for football is low, thus the ultimate aim of the game is to win the match. The performance of the players in the match affects the results of the game. Due to this, it is very important to evaluate the player and know his weaknesses. Manual evaluation tends to generate many errors and take more time. To create a new aspect for Algerian Football, we propose in this paper an intelligent system which gives solutions to different obstacles and challenges that face the Algerian teams and clubs from the selection to the playing. This system will be able to analyze, predict and recommend players with their performance, as well as predict the matches outcomes in a real-time and a live stream match. The current study has successfully implemented various deep learning models and provided all comparative analysis of the same. The accuracy of these models is compared for further clarification.

Keywords

Prediction, Machine Learning, Deep Learning, Football players, Performance, Real-time, Live Stream, Recommendation System, Analysis, Data Science.

1. Introduction

In recent years, computer science has demonstrated great potential in the sport fields. Sport has become more and more familiar with all artificial intelligence neurons, it started with a revolutionary term in the baseball industry said Money Ball, then Basketball till Football. This boom of explosion in the association of Football made all traditional ways so old fashioned, managers and all kind involved in this industry has become more related to data.

Certainly, Football was and still in the first ranking of the most popular sport around the world, and with this overdue popularity from hundreds of years, it became easier for scientists to build graphs, make reports, resolutions, more and more to understand this sport, explain what happened in the last game, make an assumption of what is going, to predict what will happen next. It made it clearer when data scientists and analysts implied the new terms of artificial intelligence and data science (Machine Learning and Deep Learning) among other tools of the artificial intelligence to make a prediction of performance and outcomes so easy as a one touch or a click. Thousands of actions became so predictable with the use of data, during the selection first, preparation second, till the games at least, players went out from shadows, their targets were tracked from preparations till the 90 pressures, bad deals went off the stadiums with new ways to develop new talents.

Survey on prediction of winning football team is still evolving, due to the difficulty of some data collection. Different scholars and researchers are still working hard to find a full prediction model using multiple techniques. This has been accomplished by various preprocessing, selection of features, statistical model, classification as classifiers, machine learning techniques and exceptional precision in prediction of football. [1]

RIF'23: The 12th Seminary of Computer Science Research at Feminine, March 09, 2023, Constantine, Algeria.

EMAIL: souheila.boudouda@univ-constantine2.dz (S.Boudouda); aya.merouani@univ-constantine2.dz (A.Merouani); lynarayane.meshoul@univ-constantine2.dz (L. Meshoul)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

Our purpose in this current work is to contribute in decision support systems and the most important deep learning algorithms and their integration into football. In this current work, we are studying how can the Algerian selection and prediction of players performance and outcomes come to be planned, out of traditional ways. we will explain the main ideas of our study, where we conduct modified and more developed algorithms and models based on previous studies. As well giving an overview of a newly created program of multi objectives where data can be generalized to serve all Algerian teams. Furthermore, our proposed system can bring a solution to different misguided paths and an easy automatization built with tools of artificial intelligence. This solution gives a new intelligent system to the Algerian Football with One Smart Program, named « TrackThe90 ».

The remaining of this paper is organized as follows. Section 2 discusses some related works. Section 3 presents the proposed intelligent system « TrackThe90 ». Section 4 shows our experimental study. Finally, some conclusions and open research lines are presented.

2. Some research related works

In this section, the literature reviews will be divided for serving 2 tasks, the first one will be: Prediction and Analyzing of the player's performance from the selection till the training cycles. The second one will be the prediction of the main football team's scores in a live video streaming.

2.1. Prediction and analyzing of the player's performance background

One of the most common research projects since the discovery of data science in football associations was how to predict and analyze players' injuries, and their performance, evaluate it, and make built graphs to make at the end players' selection and recommendations. Among the most used artificial intelligence tools is machine learning which was providing easy ways to serve the first task methods as a recommender system, player filtering, talents detection, and so many methods through learning with important models such as RFC (Random Forest Classifier), KNN, SVM, Decision Trees, Logistic Regression...etc.

According to a number of articles that propose answers to the question of how to get the best player? Machine Learning was highly used to solve this problem, we name among those research works [2], [3], [4], [5].

Garcian Aliaga et al. [2] have proposed another model based on research to determine the on-field playing positions of a group of football players based on their technical-tactical behavior using machine learning algorithms, the procedure consisted of preparing the data and standardizing it according to the steps described in the Materials and Methods section. The resulting dataset was high-dimensional with more than 50 variables. The two Machine Learning techniques used to perform the analysis showed similar projections.

Khan et al. [3] have proposed that the player performance while it's predicted can be the main impact and the key success in the match results through a model that may predict the result of a match in a certain format. This graphical model is built in two layers the training layer and the validation layer, both of those layers are built with Machine Learning and ANN.

Zhang [5] has presented another research study of the physical fitness video system of football players based on an artificial intelligence algorithm, based on the concept of detection of moving objects in an image and separating them from a static area. In this research work, Authors replaced the traditional algorithm of this concept and replace it with another algorithm due to the finding problems in the traditional one like the fast and robustness tracking in a real-time video that requires stronger algorithms.

Zhao et al. [6] have proposed a Multi-Objective Optimization in order to select team members for which they elaborate on the ESP algorithm, which is a modification of NSGA-II algorithm.

Yilmaz et al. [7] have proposed another framework based on the recommendation system using data from FIFA 2019 and Network Graphs. For this, authors used tools of Machine Learning as well like: The KNN to generate a high-quality representation for a Football player.

All the related works presented above have built Machine Learning Models using one of the tools: KNN, Bytes Naives, ANN, Decisions Trees, Regression Linear, Random Forest, SVM. Except of the research work presented in [6] which proposed a modification of NSGA-II algorithm.

2.2. Prediction of the main football team's scores Background

Tracking the outcomes of a real-time live match has a higher popularity to data scientists, analyst, hackers, trackers and even managers. The development of its concepts and algorithms is due a long time ago, but with the fact that there is a single label and type of scoring event that can be predicted which is the « Goal », it was a bit difficult to predict. Until the development of new metrics like the « Expected Goals » metrics or so-called « XG » metrics. For this, deep learning, as well as Machine Learning, is widely used to develop models able to predict match results in a real-time video based on multi algorithms of detecting objects and saving their data for future uses.

In this section, we'll present briefly, some approaches where tools of machine learning and deep learning will be used. A lot of research works have not use real-time live or streaming video to detect objects algorithms but they have used traditional ones. Such as the research work presented in [8], where it is mentioned the use of machine learning tools, naming: KNN, Logistic Regression, Neural Networks, SVM, Naives Bayes, and Random Forest, where each model of those was predicting in different situations, as an example, Logistic Linear was prediction as it's the first time for the teams to play. Another Example, the Naive Bayes, was expecting the winning of the second team at a particular venue at different situations of the match.

In [9], authors have also used free statistical data from 1900 matches of The Premier League five seasons from 2018/2019, also information about both teams' performances...etc. those data that have the highest correlations, will be used for the data preprocessing in order to calculate the prediction models (such as: XG Boost, MLR, ANN, SVM, K-NN, Random Forest, Naive Bayes, and Decisions trees).

Almulla et al. [10] have gave machine learning of previous models where data were collected between the years 2012 to 2019 for the seven consecutive football seasons: (a) 12/13 (b) 13/14 (c) 14/15 (d) 15/16 (e) 16/17 (f) 17/18 and (g) 18/19. The dataset contains a record of 2,419 players who played in 18 teams in a total of 1,121 matches. In order to make a summary based on the machine learning technique, the authors implemented multiple models, some feature selection techniques, others as a result of the use of special formulates by Almulla et al. [10], while some others are of Real-time matches of 15 minutes intervals...etc.

Azeman et al. [11] have used another methodology for their project known by the KDD which means Knowledge Discovery in Database passing by 7 steps.

The research work presented in [12], describe the use of a novel application of Stream Team in Football that allows to analyze Big Data on football matches fully automatically and in real-time on the basis of tracked player, ball positions and additional match metadata such as the field dimensions.

As a result of literature review, machine learning has been successfully applied in football field. It has been used to assists coaches and managers in many topics in football such as: result prediction, evaluation players & select best players for formation, predicting of player skills, and football analytics. Previous presented works on predicting outcomes of team sport matches using machine learning have mainly focused on team data. The focus also lies on a single tournament, league or team. In our research work we focus on the tow tasks presented in literature review: Prediction and Analyzing of the player's Performance and Prediction of the main football team's scores.

In the next section, we will present our proposed intelligent system that offers to managers and players the vision to discover the available opportunities and possibilities automatically. This system will be a multi task smart software before and after the game.

3. Analyzing and predicting football players System

In our proposed system, analysis is divided in two parts. The first one clarifies how to evaluate a performance of a player, how to predict it, and how to recommend players through the recommendation system concept to serve the first part of our program's task. The second one describes how to predict scores of a football match in a real live stream video which will be a live prediction while the run of the 90 minutes, these indicators will be serving the second part of our program's mission.

3.1. Football player's performance

There are many ways to select or recommend players far from the traditional ways, using different algorithms such as Optimization Problem Algorithms, but in the journey of this task, we choose to work with what artificial intelligence offers to us, mostly with its tool machine learning. The selection of a player as an amateur player will be different from the selection of a qualified player, for why we choose an already collected and free generalized dataset that has all correlated columns that either a football club or a team will be able to use "TrackThe90".

Through different intelligent activity trackers, the data collected by those will be analyzed by our system to evaluate and make a fully analyzed and predicted player's profile based on his level. For the analysis of this study, we will be working with multiple machine learning models of built graphs based mostly on a supervised type of problem, these models will be compared at last in their score's accuracies in both the categorical and regressions features.

The recommendation system will be playing a huge part as well, there will be advising players based on the preferred physical fitness and mental health indicators, with the consideration of their positions requires where the Goalkeeper profile is different from the one of a Forward. This state of the art was offered before by the famous industry "Stats Bomb", where this industry come up with the idea of using IQ for a player requirement, bringing the example of a forward player.

In the case of our system "TrackThe90", it will be working on the functionality of an unsupervised problem that allows us to bring players according to their similar profiles.

In the next section, we discuss the second task of our project: The prediction of scores in a real-live video or stream match.

3.2 Prediction of scores in a real-live or stream match

In the specific selection step, web service requesters often have many different technical requirements aspects when many services provide similar functions. Therefore, in order to distinguish services with similar functions, the non-functional properties must be taken into account.

As wash shown in the last World Cup of Qatar 2023, multi models was made and predicted in order to give prediction about the winner of the Cup at the final match. This one is yet just another example about the prediction world in Football games, yet there are so many other models like the ones used in gambling based on mathematical formulas.

In our research work, we are among who change the path to a prediction in the moment automatically by Data Science Analyzing a football match consists make prediction about the next goal, based on the concept of tracking the moving ball's position, while tracking the player in a

video or said dynamic image in the given stadium's dimension or the field of the video automatically. As an additional part, we add tactical values like the speed of the ball, the position of the player, the timing of the movement...etc.

The deep learning technologies are needed in this task in a neural network layer, like presented in the Figure 1.

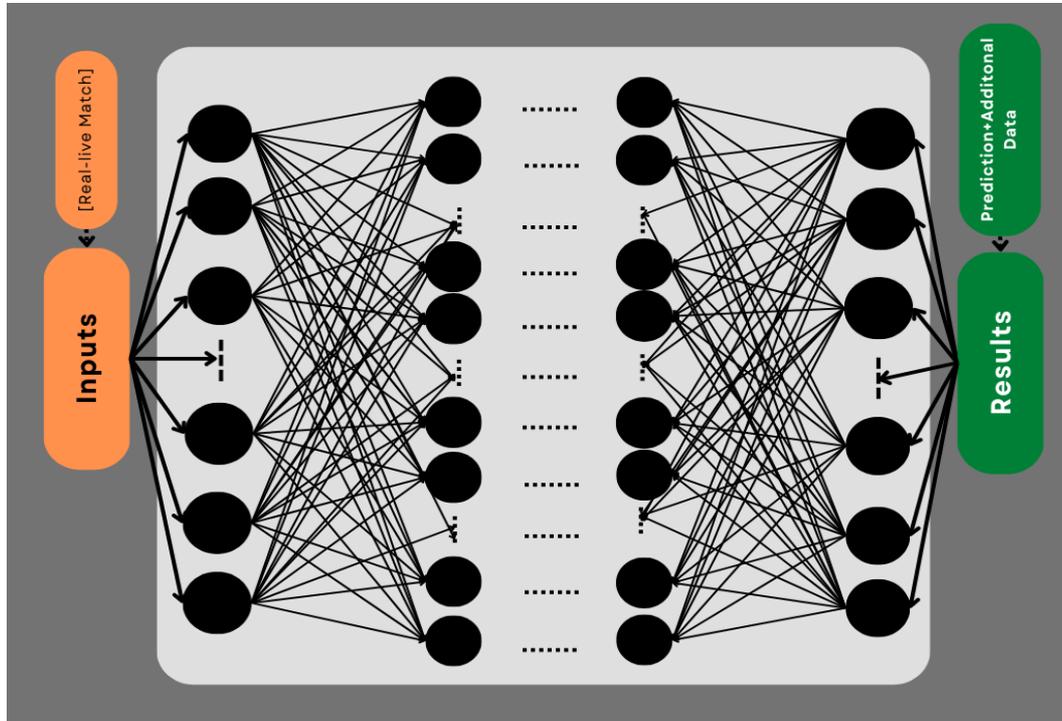


Figure 1: Prediction in real-live match

This task's design will be a 2-tier architecture where there are two known levels: the Client and the Server, the collected data by the Server will be then assigned to the deep learning model to make the possible prediction plus a possible result of the match: Home team, Away Team or a draw game.

The next section will be dedicated to discuss the workflow diagram of both tasks proposed.

3.2. Workflow of our models

In this section, we'll present the design workflow of our models. The tow tasks architectures will be detailed.

3.2.1. Architecture of performance selection, prediction and recommendation of a Football player

Figure 2 represents the proposed workflow of how we can predict and analyze a Football player's performance using machine learning, as well as how we implied the recommendation system in order to give this player a rank among the similar player's profile.

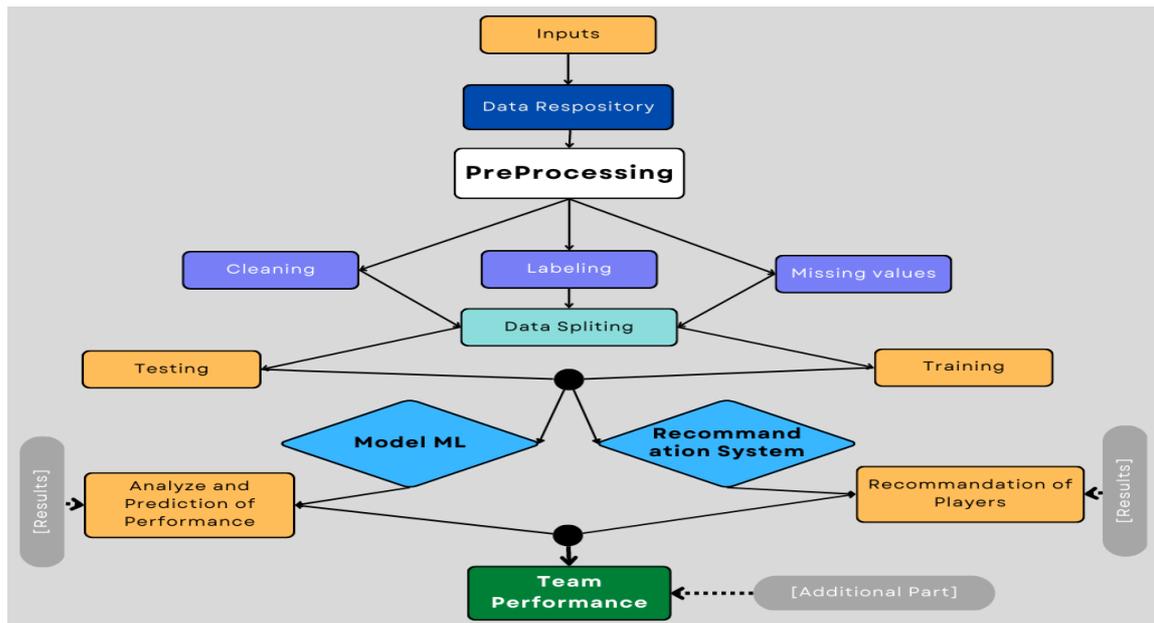


Figure 2: Proposed design of the first Task of TrackThe90

3.2.2. Architecture of a Real-time match goals Prediction

Figure 3 represents the proposed workflow or design of how we can predict a real-time match result from a Client to Server and from a Server to a Client.

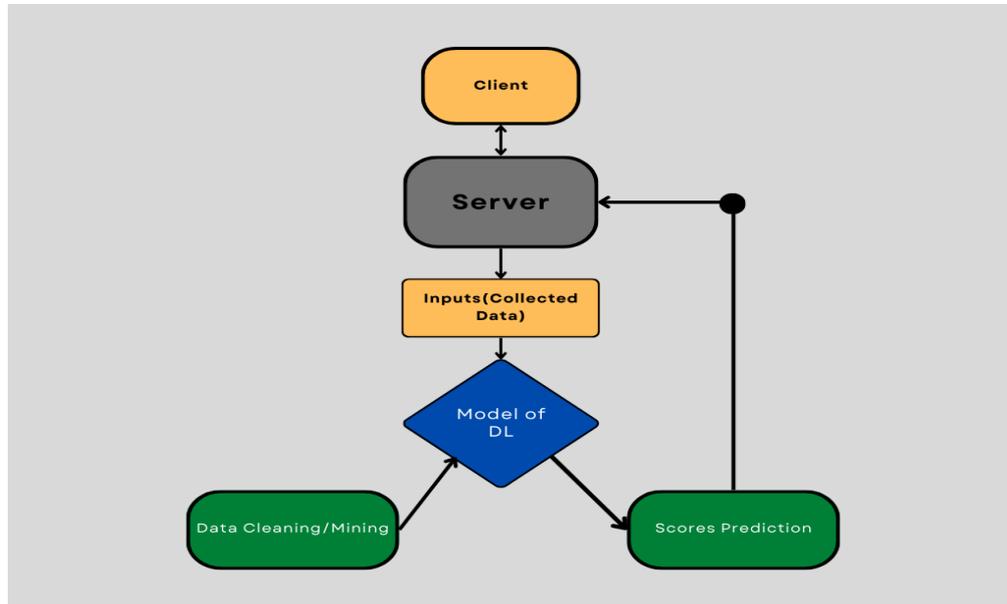


Figure 2: Proposed design of the second Task of TrackThe90

4. Some Implementation Aspects

In this section, an overview of implementation for the first task is presented. In which we will be able to tell the story of a player performance through his Overall Ranking due some features that must be entered to give the prediction of his performance.

the Overall ranking represents numerical values, this will a Regression Problem. Therefore, to test different models, we use one of test algorithms to predict using multiple regression model as the Linear Regression, Decision Tree Regressor, SVR and XGB Regressor, and to compare them with their accuracies scores in 10 cross-validation with the Error possibility.

Table 1 represents the indicators for this comparison.

Table 1
Indictors for comparison

The Model	The Accuracy Score	Error score
LinearRegression	0.87	0.35
DecisionTreeRegressor	0.88	0.37
SVR	0.97	0.14
XGBRegressor	0.96	0.19

From the table above, we can notice that The SVR model has the most accuracy score: 97% with low Error Score: 14%. Thus, as conclusion we will be working with the best model which is the SVR for the rest of this task of the system.

5. Conclusion

With the appearance of different new technologies, and with the help of the collected data over the last decades from different names of Football fields, Data Science reaches the most accuracy scores in football not only with different algorithms but with the main help of Artificial Intelligence tools, such as machine learning and deep learning. In this paper, we have proposed an intelligent system that offers to managers and players the opportunity to discover the available opportunities and possibilities automatically. This system will be a multi task smart software before and after the game. Our proposed system is based on deep learning models. Through different intelligent activity trackers, the data collected is analyzed by our system to evaluate and make a fully analyzed and predicted player's profile based on his level. We have used multiple machine learning models of built graphs based mostly on a supervised type of problem, these models are compared at last in their score's accuracies in both the categorical and regressions features. With the fact that a machine is wildly different from a human player, we considerate the option of the mental health as an important feature to be analysed as well in both tasks, yet this a very important part that should be more developed.

As an addition in the future, we are planning to make this system able to use the recommendation system to give the best team positions, as well as give the prediction of the team performance based on those positions, players ...etc.

6. References

- [1] X. Wang, A. Tapani, N. Jari. Applying CDMA technique to network-on-chip. IEEE transactions on very large-scale integration (VLSI) systems 15.10: 1091-1100, (2007).
- [2] A. Garcia, M. Aliaga, J. Marquina & A. Coteron. Rodriguez-Gonzalez & S. Luengo-Sanchez. In-game behaviour analysis of football players using machine learning techniques based on player statistics, (2020).
- [3] M. A. Khan, M. Habib, S. Saqib, T. Alya, K. M.Khan, M. A. Ghamdi, S. H. Almotiri. Analysis of the Smart Player's Impact on the Success of a Team Empowered with Machine Learning. Lahore, Pakistan; Makkah City, Saudi Arabia, (2020).

- [4] P. Pugsee, P. Pattawong. Football Match Result Prediction Using the Random Forest Classifier. Bangkok, Thailand: Innovative Network and Software Engineering Technology Laboratory, Department of Mathematics and Computer Science, Faculty of Science, Chulalongkorn University, (2021).
- [5] S. Zhang. The Physical Fitness Video Tracking System of Football Players Based on Artificial Intelligence Algorithm. Nanchang, Jiangxi, China: Nanchang Institute of Science and Technology, (2022).
- [6] H. Zhao, H. Chen, S. YU, B. Chen. Multi-Objective Optimization for Football Team Member Selection. Xiamen, China, (2021).
- [7] Ö. İ. Yılmaz & Ş. G. Ögüdücü. Learning Football Player Features using Graph Embeddings for Player Recommendation System. Istanbul, Turkey: Istanbul Technical University, Department of Computer Engineering, Istanbul Technical University, Department of Artificial Intelligence and Data Engineering, (2022).
- [8] H. Mittal, H. Mittal, J. Kumar, A. K. Singh. A Study on Machine Learning Approaches for Player Performance and Match Results Prediction. Kurukshetra, Haryana: Department of Computer Applications, National Institute of Technology, (2021).
- [9] F. Rodrigues, A. Pinto. Prediction of football match results with Machine Learning . porto, Portugal: International Conference on Industry Sciences and Computer Sciences, (2022).
- [10] J. Almulla, T. Alam. Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League. Doha, Qatar: Qatar Football Association and College of Science and Engineering, Hamad Bin Khalifa University, (2020).
- [11] A. Azeman, A. Mustapha, N. Razali, A. Nanthaarnomphong, M. H. Wahab. Prediction of Football Matches Results: Decision Forest against Neural Networks . Johor, Malaysia; huket, Thailand: Faculty of Computer Science and Information Technology; Faculty of Applied Sciences and Technology; Andaman Intelligent Tourism and Service Informatics Center; Faculty of Electrical and Electronic Engineering, (2021).
- [12] L. Probst, H. Schuldt, P. Seidenschwarz, M. Rumo. StreamTeam-Football: Analyzing Football Matches in Real-Time on the Basis of Position Streams. Bern, Switzerland; Basel, Switzerland: Department of Mathematics and Computer Science, University of Basel ; Centre for Technologies in Sports and Medicine, Bern University of Applied Sciences, (2020).