iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50228

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Impact on Performance in Football Game by Players Position using Machine Learning Techniques

Swathi Gadipudi[1], Vyshnavi Kesari[2], Sai Bharath Chitluri[3], Venkata Naga Satya Swaroop Akkinapalli[4], Dr. Ram Prasad Reddy Sadi[5]

[1, 2, 3, 4]*Department of Information Technology, Anil Neerukonda Institute of Technology and Sciences,Sangivasala, Visakhapatnam, Andhra Pradesh, India.*

[5]*Associate Professor, Department of Information Technology, Anil Neerukonda Institute of Technology and Sciences, Sangivasala, Visakhapatnam, Andhra Pradesh, India.*

Abstract: ***For many people who have aspired to be successful in their occupations, sports has emerged as a fantastic new field. One of the sports that has gained popularity is football. Better results are obtained when we are aware of and focus on our strengths. The same is true for a player who is conscious of his skills and makes a concerted effort to develop them. A coach must find a player who suits a given position, but doing so might be difficult for a variety of reasons. In the current study, the binary relevance and random forest have been used to find a solution to the problem of determining a football player's optimal position. Test results show that the strategy is effective.***

Keywords: Football, Machine Learning, Random Forest, SVM

I. INTRODUCTION

Sports provide a sufficient amount of statistical data on each player, club, game, and season. Earlier, it was thought that experts, coaches, team managers, and analysts owned sports science. Sports companies have recently become aware of the science included in their data and wish to utilize that research by using various data mining approaches. Sports data mining helps coaches and management in a variety of ways, including with result prediction, player performance, talent appraisal, and player identification. Prediction aids clubs and managers in making the best choices to win leagues and tournaments. In all the types of sports, this project deals with the game of football. Football is a sport which involves two teams with eleven players of each team without using their hands or arms, move ball into opposing team's goal. Team with most goals scored wins. The best method to develop intuition is through practice and self-assurance in one's position. But it can be challenging to hone these instincts given the range of places on a soccer field. The team's success, though, is influenced by the players' abilities as well as the roles they are given.In this case, we will train the model using different classification techniques to determine which is the best.The player's ideal position is predicted using the trained model. Additionally, rather than using their score in the model training, their skills are used. Take a former forward player as an illustration. A striker is in the optimal position after a rigorous appraisal of the player's abilities, meaning he can play as a striker more effectively than as a forward. However, other elements, such as experience, may have an impact on older players who have been playing in the position for a while. New players won't be affected in the slightest. Simply put, it fits the new model the best.

A. Project Scope and Direction

It is helpful for the players to improve their skills and also used for coach to put the player in the best position Based on the predictions it can give us a detailed report of the player's positions based on their skills.By this way the winning accuracy can be increased and the player can focus on their best skills.

B. Impact, Significance and Contribution

Choosing suitable position is critical for player in any game, but it's slightly more difficult in football due to large number of positions. Likely, they will simply accept the position that is suggested. If they do this, they may or may not be successful. To prevent this confusion, we will provide idea of their position based on their skills.Identifying a player who is suited for a particular position is crucial for a coach or player, but it can be difficult for a number of different reasons. So if we predict it beforehand it will be easier for practice and performance.

II. RELATED WORK

1) Pantzalie Victor Chazan The model was developed using naive bayes, random forests, KNN, and decision trees by the Data Mining and Analytics research team from the School of Science and Technology International Hellenic University. The initial goal of the plan was to separate the clubs into those that would perform better than last season in terms of points earned and those that would perform worse. The results are good but not particularly noteworthy. In order to predict team performance, it was necessary to recreate each game of the season and categorize the outcomes as home wins, draws, or away wins. This technique produced 14 exceptional results. The expected league final standings were then determined by adding up the points earned by each team.
2) Javier Fernandez, Daniel Medina, Antonio Gomez Diaz, Marta Arias, and Ricard Gavalda conducted a study on training and player physical performance. It comes to the conclusion that not every attribute depends on the result. Before training the model, they utilize feature selection and PCA (Principal Component Analysis) for dimensional reduction to eliminate the irrelevant and undesired attributes. The findings demonstrate that deleting the irrelevant attributes improves the model's accuracy.
3) According to research by Verstraete, Kenneth, Decroos, Tom, Coussement, Bruno, Vannieuwenhoven, Nick, and Davis, Jesse, a player's performance is influenced by a variety of factors, including age and experience in addition to their skill level.
4) The Pima-Indians Diabetes dataset was the subject of an experiment by Westin and Lena, and the results show that the missing values had an impact on the outcome. They came to the conclusion that better results were obtained by deleting the missing numbers.
5) An study of the advantages of binary relevance over alternative multi-label classification algorithms has been presented by Luaces, Oscar, Dez, Jorge, Barranquero, Jose, del Coz, and Bahamonde, Antonio.
6) Algorithms like naive bayes, baysen networks, logit boost, and knn are employed by OsipHucaljuk and Alen Rakipovi of the Faculty of Electrical Engineering and Computing at the University of Zagreb. the challenge of predicting football game outcomes. Due of the popularity and ubiquity of the sport, it presents an interesting problem. But because there are so many variables that must be considered that cannot be quantified or described, predicting the results is likewise a challenging task. Because the teams in the Champions League are evenly matched, selecting one team as the "favourite" of a game won't necessarily provide favourable outcomes. The expert who labelled the second batch may not have been all that knowledgeable.
7) Sports analytics have become significant in recent years. Players can record their position and mobility information while playing thanks to recently developed tracking systems. These data can be used to improve performance of both teams and individual players. For advanced tactical analysis, we provide various machine learning techniques toforecast spatial placements of players.

III. METHODOLOGY

A. *Multi-label Classification*

Algorithms that can be used for multi-label classification include:

1) *Random Forest:* most accurate method for predicting "win" and "loss" is random forest. By averaging the outcomes from various trees, it generates forecasts. As the number of trees grows, the accuracy of the outcome improves. In a random subset of features, it looks for the best feature. Therefore, the random forest method will be useful in determining the position of players.
2) *Binary Relevance:* Multilabel classification problems can be solved using binary relevance. The most logical approach to learning from examples with many labels is binary relevance. It functions by breaking down multi-label job into several separate binary tasks.

Figure 1: Diagrammatic representation of the methodology

IV. PROPOSED FRAMEWORK

In this system rather than directly predicting the position by eliminating the multi label tuples in the data set here, it is dealt with the binary relevance methodology i.e., the conversion of multi labeled data in 14 multi class classification models. Unlike the Existing system that can predict only 4 positions but here it can predict 14 different positions. Based on the predictions it can give us a detailed report of the player's positions based on their skills.

V. RESULTS

1) *Data Set:* Players' skills are included in a dataset that was gathered from the Kaggle website. It has 75 columns and 17981 rows. These columns provide information about the player, such as name, country, and skills. The input for our model is used in this.

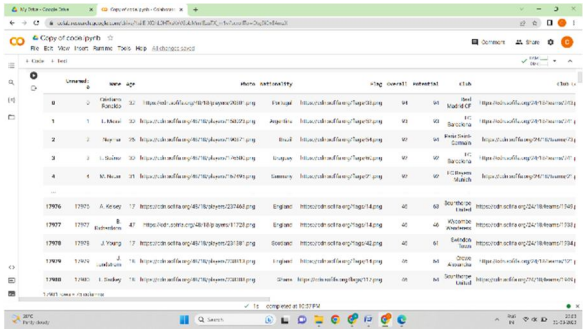


Figure 2: Dataset of the players skills

2) *Pre-processing:* Importing the data set and removing the rows with null values during pre-processing are the first steps because they affect the model's accuracy. To improve the performance of our model, we will eliminate the useless columns from our data set.

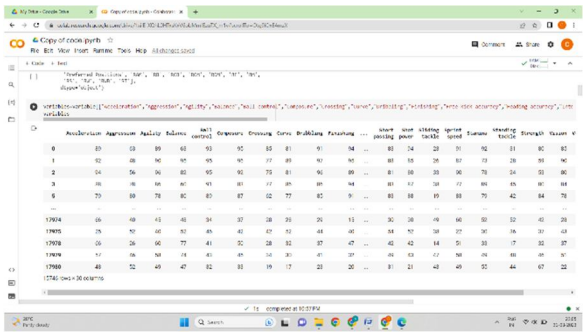


Figure 3: Dataset after removing the null valued columns

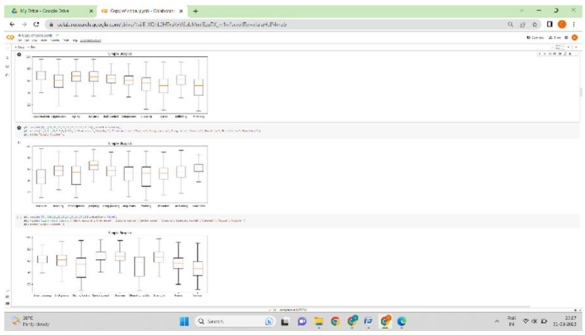


Figure 4: Outliers detection is done on the dataset

3) *Splitting Training and Testing Data:* Thirty percent of the data will be used for testing, and seventy percent will be used for training.
4) *Fitting the Model:* We will take into account a variety of machine learning models before choosing the one that best fits our data.

5) *Model Comparison:* For every position all the models are compared and the best of all are taken.

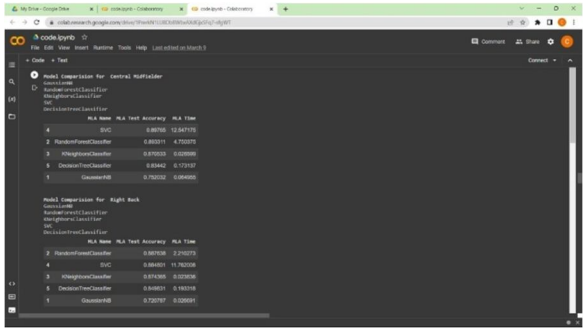


Figure 5: Model comparison is done by taking Randomforest, naivebayes, svm, decision tree and kneighbors

6) *Results:* Finally, we will determine the player's best position based on their skills to maximize win accuracy. If the input is given with the skills of the player in this format we get the following output.

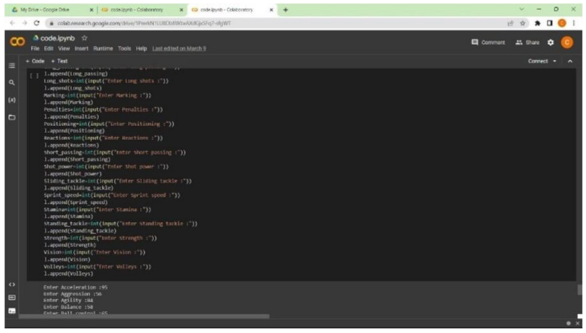


Figure 6: Input of player's skills for the model

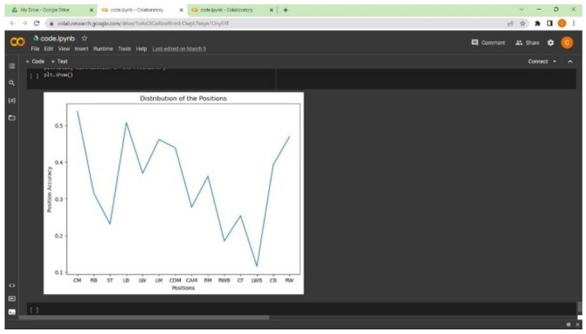


Figure 7: Output of the positions for the player

After the positions are selected in this way winning accuracy is calculated using the logistic regression.

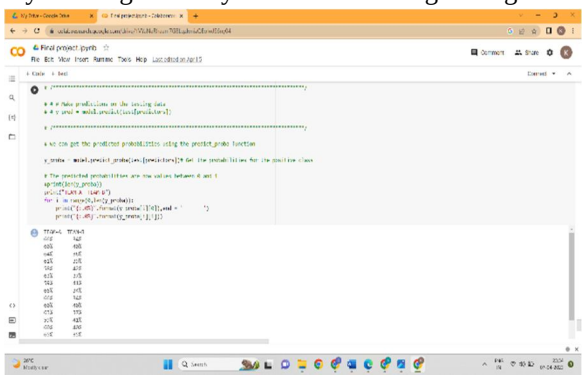


Figure 8: Winning accuracy for the test data

VI. PERFORMANCE ANALYSIS

Confusion metrics is taken into consideration for calculating performance metrics.

The performance metrics that are taken into consideration for this project:

1) *Accuracy:* (TP+TN)/(TP+TN+FP+FN)
2) *Precision:* (TP)/(TP+FP)
3) *Recall:* (TP)/(TP+FN)
4) *F1-score:* (2*PRECISION*RECALL)/(PRECISION+RECALL)

POSITION	ACCURACY	PRECISION	RECALL	F1 SCORE
CM	0.9092987804878049	0.8356164383561644	0.7746031746031746	0.8039538714991762
RB	0.9032012195121951	0.5882352941176471	0.4729729729729729	0.5243445692883896
ST	0.9565548780487805	0.9151785714285714	0.8436213991769548	0.8779443254817987
LB	0.9077743902439024	0.5913043478260870	0.4788732394366197	0.5291828793774319
LW	0.9489329268292683	0.5	0.029850746268656716	0.0563380281690140 86
LM	0.8544207317073171 7	0.5909090909090909	0.4155251141552511	0.4879356568364611
CDM	0.9268292682926832 683	0.8451327433628318	0.7579365079365079	0.7991631799163181
CAM	0.8803353658536586	0.7067669172932330	0.4433962264150943 5	0.5449275362318884
RM	0.8620426829268293	0.6666666666666666	0.3084112149532710	0.4217252396166335
RWB	0.9885670731707317	NAN	0.0	0.9885670731707317
CF	0.9824695121951219	0.8571428571428571	0.0214285714285714 27	0.3428571428571428 6
LWB	0.9916158536585366	1.0	0.0833333333333333 3	0.1538461538461538 5
CB	0.9542682926829268	0.9469964664310954	0.8562300319488818	0.8993288590604026
RW	0.9527439024390244	0.6666666666666666	0.0317460317460317 44	0.0606060606060606 061

Table 1: Performance analysis for all the 14 positions

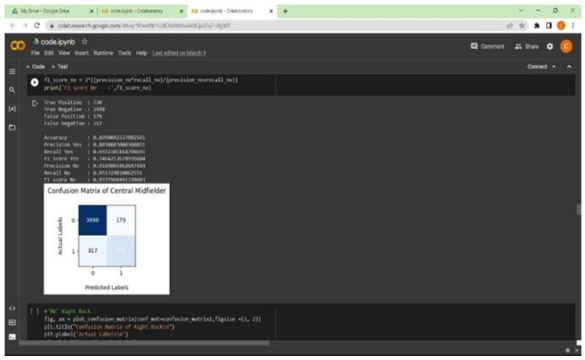


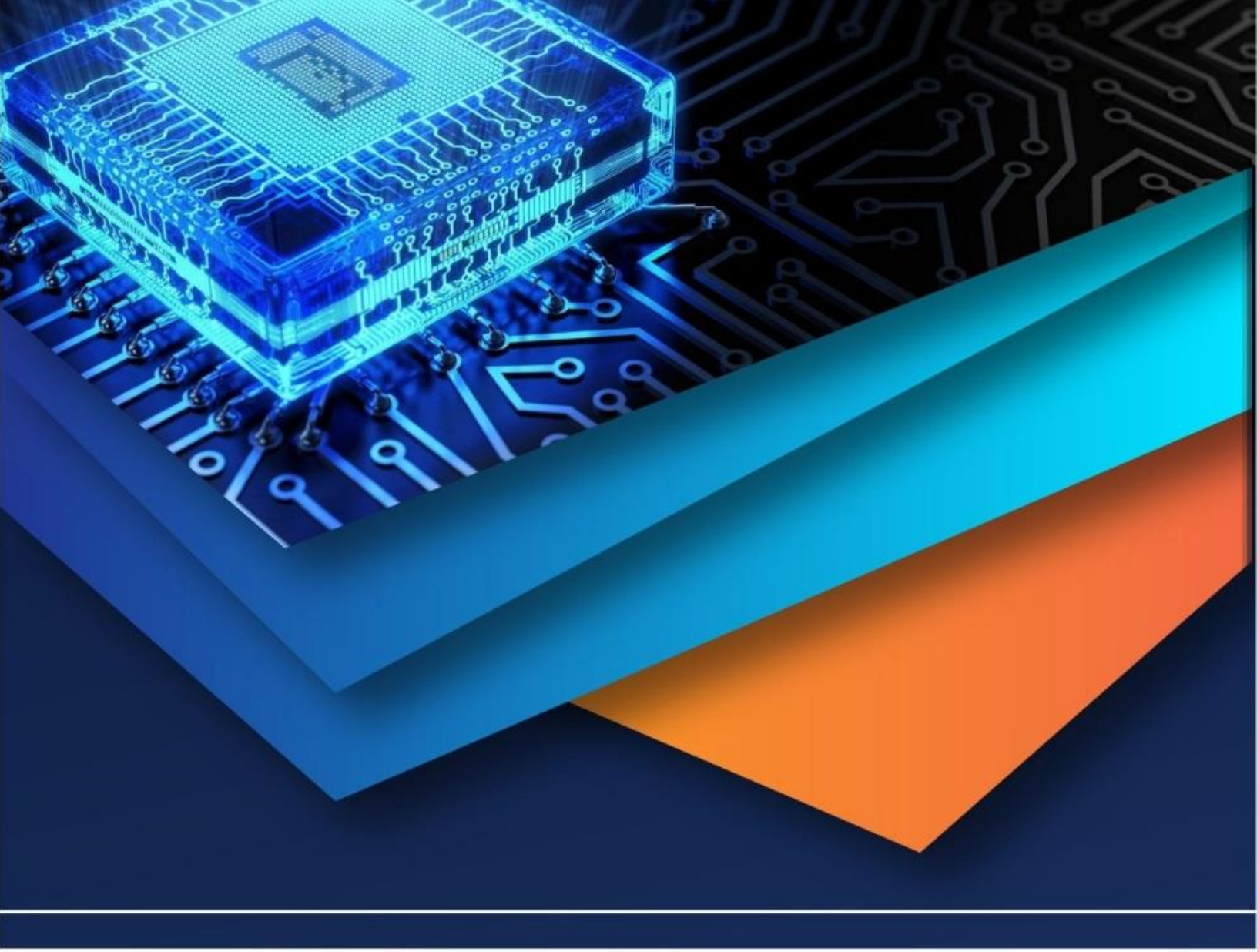
Figure 9: Performance metrics for one position

VII. CONCLUSION

Choosing suitable position is critical for player in any game, but it's slightly more difficult in football due to large number of positions. Likely, they will simply accept theposition that is suggested. If they do this, they may or may not be successful. To prevent this confusion, we will provide idea of their position based on their skills.A coach or player must find a player who is qualified for a specific position, but doing so might be difficult for a number of reasons. So, if we predict it beforehand it will be easier for practice and performance. This project will provide the best suitable positions for the players with the highest accuracy.
There is room for further improvement. Larger dataset would also help in achieving better results.

REFERENCES

[1] Chazan-Pantzalis, Victor &Tjortjis, Christos. (2020). Sports Analytics for Football League Table and Player Performance Prediction.
[2] Fernández, Javier & Medina, Daniel & Gómez Díaz, Antonio & Arias, Marta &Gavaldà, Ricard. (2016). From Training to Match Performance: A Predictive and Explanatory Study on Novel Tracking Data. 136-143. 10.1109/ICDMW.2016.0027.
[3] Verstraete, Kenneth &Decroos, Tom &Coussement, Bruno &Vannieuwenhoven, Nick & Davis, Jesse. (2020). Analyzing Soccer Players' Skill Ratings Over Time Using Tensor-Based Methods. 225-234. 10.1007/978-3-030-43887-6_17
[4] Westin, Lena. (2008). Missing data and the preprocessing perceptron.
[5] Luaces, Oscar &Díez, Jorge &Barranquero, Jose & del Coz, Juan &Bahamonde, Antonio. (2012). Binary relevance efficacy for multilabel classification. Progress in Artificial Intelligence. 1. 10.1007/s13748- 012-0030-x.
[6] Ali, Jehad & Khan, Rehanullah& Ahmad, Nasir & Maqsood, Imran. (2012). Random Forests and Decision Trees. International Journal of Computer ScienceIssues(IJCSI)9
[7] Machine Learning for Position Detection in Football by Martin Frey, Elvis Murina, Janick Rohrbach, Manuel Walser, Patrick Haas, Marcel Dettling



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY