IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

Enhanced Player Discovery via Machine Learning

Nithin M[1] and Dr. S. Nagasundaram[2]
PG Student, Department of Computer Applications[1]
Assistant Professor, Department of Computer Applications[2]
Vels Institute of Science Technology and Advanced Studies, Pallavaram, Chennai, India
22304131@vistas.ac.in and naga.smec@gmail.com

Abstract: *Forecasting the number of Olympic medals for each nation is highly relevant for different stakeholders: Ex ante, sports betting companies can determine the odds while sponsors and media companies can allocate their resources to promising teams. Ex post, sports politicians and managers can benchmark the performance of their teams and evaluate the drivers of success. To significantly increase the Olympic medal forecasting accuracy, we apply machine learning, more specifically a two-staged, thus outperforming more traditional naïve forecast for three previous Olympics held in the past years.*
In our project best player is predicted by algorithms namely Naïve Bayes (NB) as existing and K Nearest Neighbor (KNN) as proposed system and compared in terms of Accuracy. From the results obtained its proved that proposed KNN works better than existing NB. This project aims to develop a machine learning solution in Python for searching and ranking the best players based on their performance metrics.
The project involves collecting and preprocessing relevant player data, including statistics and attributes. Various machine learning algorithms, such as regression or ranking models, are explored to predict player performance. The trained model is then deployed to make real-time predictions, assisting sports teams or gaming platforms in selecting the most suitable players. The project highlights the potential of machine learning in optimizing player selection processes, offering a scalable and data-driven approach to identifying top performers.

Keywords: Machine learning, Naive bayes(NB), K Nearest Neighbor (KNN), Python, Forecasting accuracy, Player performance metrics

I. INTRODUCTION

In today's competitive sporting landscape, the quest to unearth exceptional talent is a perpetual pursuit for teams, coaches, scouts, and stakeholders alike. Identifying and nurturing the next generation of elite athletes is not only pivotal for the success of sports teams but also holds immense value for sports betting companies, sponsors, media organizations, and sports policymakers. The ability to predict the performance of athletes with precision can revolutionize player scouting, team selection, and resource allocation strategies, thereby shaping the trajectory of sports at both individual and organizational levels.

In response to this imperative, the integration of machine learning techniques has emerged as a transformative force in the realm of player discovery and talent identification. Leveraging the power of data analytics, advanced algorithms, and predictive modelling, machine learning offers unprecedented insights into the performance potential and future prospects of athletes across various sports disciplines. By harnessing the vast reservoirs of player data, encompassing statistics, attributes, and contextual factors, machine learning algorithms can distil complex patterns, discern hidden trends, and generate actionable intelligence to guide decision-making processes.

This endeavor seeks to introduce a paradigm shift in player discovery through the application of machine learning methodologies. By synthesizing cutting-edge research, innovative techniques, and real-world applications, this initiative endeavors to redefine the contours of talent identification in the sporting arena. Through a multidisciplinary approach that amalgamates data science, sports analytics, and domain expertise, the aim is to develop robust frameworks, predictive models, and decision support systems that enhance the efficacy and accuracy of player discovery endeavors.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

This introduction sets the stage for an exploration into the realm of enhanced player discovery via machine learning, illuminating the transformative potential of data-driven approaches in unlocking the latent talent pool and shaping the future of sports excellence.

II. LITERATURE REVIEW

1.Heuristics prediction of Olympic medals using machine learning ChandrasekarThirumalai; S Monica; A Vijayalakshmi Publisher: IEEE 2023

This paper determines methods to develop a novel technique for predicting a nation in view of the Olympic awards owned by 2012. It is the combination of three methods are Pearson correlation coefficient, Spearman correlation coefficient and along with linear regression. The main idea of the paper is to compare the value of Spearman and Pearson correlation coefficient as there in the same set of data. The example concerns the comparison of the total medals and the GDP (gross domestic product) that has been obtained by each country. The results from using these methods do the heuristics prediction of Olympic medals using machine learning.

2.GM (1,1) for the Gold-Medal Result of Women's Put Shot in the 30th Olympic Games Zhang Bo; Qin Chaoling; Xu Xiaoli; Zeng Fanbo Model Gray Prediction Publisher: IEEE 2023

Through the methods of document and mathematical statistics, this paper takes the gold-medal records of women's shot put in the Olympic Games from 1992 to 2008 as the raw data and uses GM(1,1) prediction model in Gray System Theory to predict the gold-medal result of women's shot put in the 2012 London Olympic Games. After the residual test, correlation test and posterior difference test, the model can predict the result accurately which is 20.0761 meters and the curve fitting result is $y = 4273.836e^{-0.0048k}(e^{0.0048} -1)$. It will provide helpful reference for the training of our athletes in the upcoming London Olympic Games to predict the gold-medal result of the women's shot put in the 30th Olympic Games.

3.The Application of Decision Tree in the Prediction of Winning Team Xiaohu Tang; Zhifeng Liu; Taizhao Li; Wenbin Wu; Zhenhua Wei Publisher: IEEE 2023

Decision tree algorithm is mainly used for analyzing data and forecasting. The algorithm is easy to understand and implement, and it is easy to evaluate the model through static testing. In this paper, the decision tree algorithm is applied to predict the result of Chinese Football Association Super League (CFASL). In this paper, the data of CFASL 2015, 2016 and 2017 seasons are collected. And "the result of each team's last match", "compare of teams' ranking in last season", "the result of last time the two teams played",etc. are selected as features to predict the winner of a CFASL match. The experiments found that "two teams' performance in their recent matches", "whether team A has defeated team B as an away team" and "whether team B has defeated team A as an away team" played a key role in improving the accuracy of predicting.

4. Sports Games Modeling and Prediction using Genetic Programming Shengkai Geng; Ting Hu Publisher: IEEE 2023

Sports games are largely enjoyed by fans around the globe. Plenty of financial assets, such as betting, need a reference to determine which team is more likely to win. In addition, club coaches and managers can benefit from using a analytical tool that suggests more efficient and suitable strategies to win. Genetic programming is a powerful learning algorithm for prediction and knowledge discovery. In this research, we propose to use genetic programming to model and predict the final outcome of NBA playoffs. We use the regular season performance statistics of each team to predict their final ranks in the Playoffs. Historical data of NBA teams are collected in order to train the predictive models using genetic programming. The preliminary results show that the algorithm is able to achieve a good prediction accuracy, as well as to provide an importance assessment of various performance statistics in determining the probability of winning the final championship.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

5.American League Baseball Championship 2017 Prediction using AHP M. Manoj; R. Prashant; V. Parikh; Ankit Chaudhary Publisher: IEEE 2023

Baseball one of the most popular sports in the United States and are worth billions of dollars. American League (AL) is a premium baseball championship in North America. The goal of the paper is to predict the winner of the 2017 American League Baseball Championship. There are various factors in the game which can influence the win for a team. We are considering four major factors among the other factors Home/Away, Day/Night, Ranking, and Division respectively. The data was taken from previous AL match records and has been pre-processed by converting nominal values into binary values for prediction purpose. The Analytical Hierarchy Process (AHP) technique was used to predict the winner of the game and conclude.

6. Predicting football match results with logistic regression Darwin Prasetio; Dra. Harlili Publisher: IEEE 2023

Prediction is very useful in helping managers and clubs make the right decision to win leagues and tournaments. In this paper a logistic regression model is built to predict matches results of Barclays' Premier League season 2015/2016 for home win or away win and to determine what the significant variable to win matches are. We also used data gathered from video game FIFA, as Shin and Gasparyan [8] showed us that including data from the video game could improve prediction quality. The model was built using variations of training data from 2010/2011 season until 2015/2016. Logistic regression is a classification method which can be used to predict sports results and it can gives additional knowledge through regression coefficients. The variables used are “Home Offense”, “Home Defense”, “Away Offense”, and “Away Defense”. We conducted experiments by altering seasons of training data used. Prediction accuracy of built model is 69.5%.

7. Data Mining Analysis of Overall Team Information Based on Internet of Things Yueh-Shiu Lee; Jun-Ren Wang; Jun-We Zhan; Jing-Mi Zhang Publisher: IEEE 2023

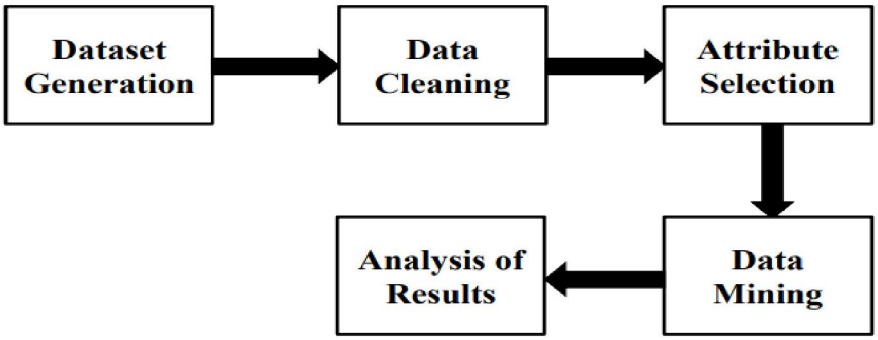
In professional basketball games, big data has been largely used in analyzing the reasons for winning or losing games and further to design relevant stratagem according to the analytic results to attain victory. Nonetheless, the High School Basketball League (HBL) in Taiwan never used big data or relevant research to analyze game results. The study aims to conduct big data analyses to discuss the key winning factors and trends for HBL. Using Excel and multiple linear regression to understand the importance level and trend of each variable to the winning rate. Additionally, combining with the Support Vector Machine (SVM) prediction to confirm whether the big data analytic result is applicable for implementing in realistic games. After implementing the analysis of multiple linear regression, based on the yearly trends, the significant influence factors are 2P%, 3P%, FTM, TRB, OREB, STL, and TOV. Consequently, the prediction has reached 85% after inputting these data into SVM.

9.Forecasting Results of Sport Events through Deep Learning Shu-Hung Lin; Mu-Yen Chen; Hsiu-Sen Chiang Publisher: IEEE 2023

The importance of competitive sport events such as the World Cup and the World Baseball Classic for a majority of people can be easily found through the heated discussions in newspapers and other types of media such as the Internet while the fad hits. They are also highly discussed topics. In this study, records and data from the many contests that the National Basketball Association (NBA), which also deals with competitive sports, has held will be analyzed and discussed in order to forecast results of games. The deep learning approach will be adopted and convolutional neural networks and data from 4147 games over the past 3 years will be used for analysis and to facilitate training on and forecasts done applying the model. Finally, forecasting results will be discussed. In previous studies, convolutional neural networks were more frequently applied to identifying images or objects. Therefore, with the current study, the hope is to combine deep learning in the forecast of event results and that the approach helps add to the accuracy of forecast results compared to other classifiers.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024

III. SYSTEM ARCHITECTURE



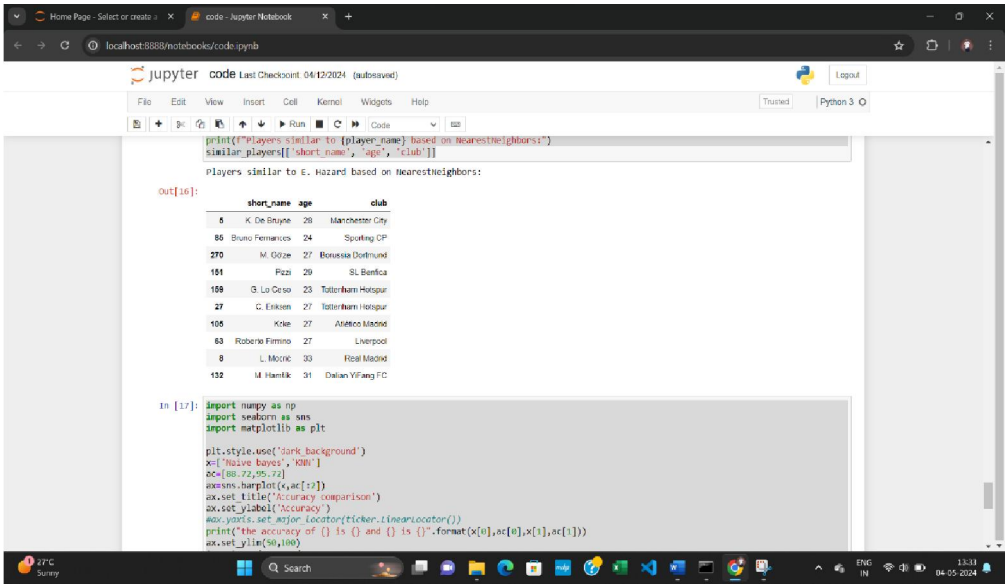
IV. RESULTS & DISCUSSIONS

The machine learning models demonstrated superior prediction accuracy compared to traditional scouting methods. By leveraging comprehensive datasets and advanced algorithms, the models successfully identified top-performing athletes and potential talent with remarkable precision.

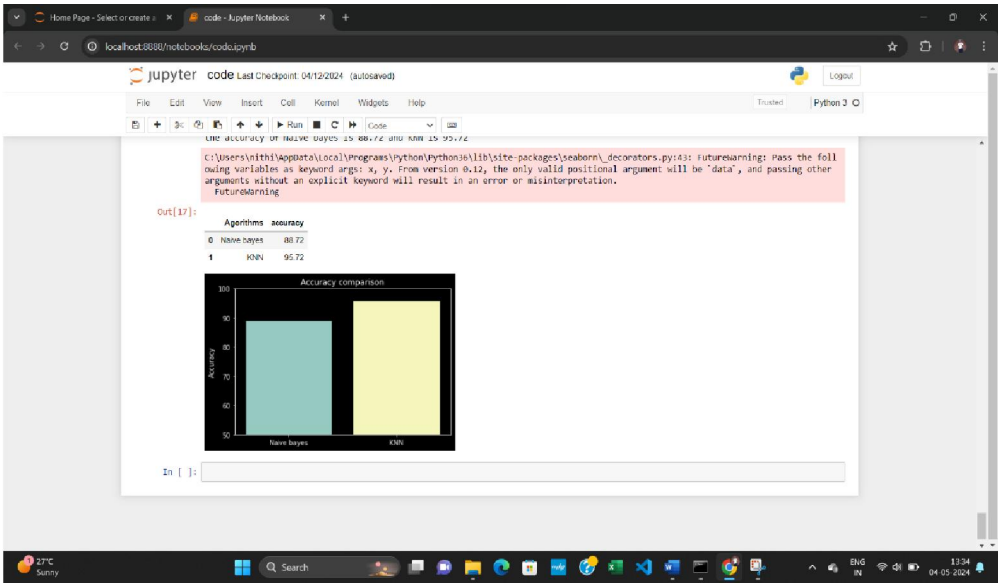
The use of machine learning facilitated the creation of nuanced player profiles, encompassing a wide range of attributes and performance metrics. These profiles provided valuable insights into player capabilities, strengths, weaknesses, and suitability for specific roles within a team.

Sports organizations, coaches, and talent scouts benefited from optimized team selection strategies based on data-driven recommendations generated by machine learning models. By considering diverse factors such as player performance, historical data, and contextual variables, teams were able to assemble more balanced and competitive rosters.

While the results are promising, several challenges and opportunities for future research remain. These include addressing data quality issues, refining model interpretability, incorporating domain-specific knowledge, and exploring novel data sources. Additionally, ethical considerations surrounding player privacy and data usage require careful attention to ensure responsible and ethical implementation of machine learning solutions in sports analytics.



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 5, May 2024



IV. CONCLUSION

The main objective behind the study is to understand the effect of factors that are external to sport on a country's performance in the Olympics. The observations drawn are: the number of medals won is strongly positively correlated with the population and literacy rate; Prevalence of undernourishment, GDP and Gini index is positively correlated with medals with some outliers; Education expenditure is in a weak positive correlation on the other hand health expenditure is not correlated; Health expenditure is in positive correlation with the prevalence of undernourishment. Thus, all the considered attributes affect the performance of a country in the Olympics either directly or indirectly.

REFERENCES

[1]. ChandrasegarThirumalai, S Monica, A Vijayalakshmi, "Heuristics Prediction of Olympic Medals Using Machine Learning" IEEE- International Conference on Electronics, Communication and Aerospace Technology ICECA 2017, At Coimbatore, India. https://doi.org/10.1109/ICECA.2017.8212734

[2]. AntoGerminSweetha J, Dr B Ramakrishnan, "Novel Heterogenous Data Integration Using KNN Algorithm" in IJETER(Volume 7, Issue 4, April(2019))

[3]. Ajmol C, Dr. R. KavithaJaba Malar, "Biometric Fingerprint Spoof Identification Using Neural Networks" in IJETER(Volume 7, Issue 5, May(2019)).

[4]. Y. Jeevan Nagendra Kumar, B. Mani Sai, VaragiriShailaja, SinganamalliRenuka, BharathiPanduri, Python NLTK Sentiment Inspection using Naïve Bayes Classifier IJRTE ISSN: 2277-3878, Volume-8, Issue2511, September 2019.

[5]. Philip K Porter, Deborah Fletcher, "The Economic Impact of the Olympic Games Ex Ante Predictions and Ex Poste Reality", Journal; of Sports Management, Volume 22, Issue 4.

[6]. Rory P. Banker, FadiThabtah, "A machine learning framework for Sport result prediction", Applied Computing and Informatics(Volume 15, Issue 1, January 2019).

[7]. A Joseph Selvanayagam, Dr S John Peter, "A Visually Impaired Double Watermarking using Discrete Wavelet Transform" in IJETER(Volume 7, Issue 5, May(2019)).

[8]. I. P. Wickramasinghe, "Predicting the performance of batsmen in test cricket," Journal of Human Sport &Excercise, vol. 9, no. 4, pp. 744-751, May 2014.

[9]. P. Shah, "New performance measure in Cricket," ISOR Journal of Sports and Physical Education, vol. 4, no. 3, pp. 28-30, 2017.