IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

Volume 4, Issue 6, June 2024

Impact Factor: 5.731

Football Player Analysis for Identifying Best Team using Machine Learning

Aditya Ramnath[1] and R. Priya[2]

PG Student, Department of Computer Applications[1]

Professor, Department of Computer Applications[2]

adityaramnath@outlook.com and priyaa.research@gmail.com

Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai, Tamil Nadu, India

Abstract: *In the game of football (soccer), the evaluation of players for transfer, scouting, squad formation and strategic planning is important. However, due to the vast pool of grassroots level player, short career span, differing performance throughout the individual's career, differing play conditions, positions and varying club budgets, it becomes difficult to identify the individual player's performance value altogether. The Player Performance Prediction system aims at solving this complex problem analytically and involves learning from various attributes and skills of a football player. It considers the skill set values of the football player and predicts the performance value, which depicts the scope of improvement and the capability of the player. The objective of this project is to help the coaches and team management at the grassroots as well as higher levels to identify the future prospects in the game of football without being biased to subjective conditions like club budget, competitiveness in the league, and importance of the player in the team or region. The system is based on a data-driven approach and we train our models to generate an appropriate holistic relationship between the players' attributes values, market value and performance value to be predicted. These values are dependent on the position that the football player plays in and the skills they possess.*

In This project best player is predicted by algorithms namely Naïve Bayes (NB) as proposed and K Nearest Neighbor (KNN) as existing system and compared in terms of Accuracy. From the results obtained its proved that proposed NB works better than existing KNN..

Keywords: NB- Naive Bayes , KNN- K Nearest Neighbour

I. INTRODUCTION

Football, also known as soccer to the western part of the world, is a team based sport played between two teams, each consisting of eleven players with a spherical ball. This sport is played in over 200 countries and in almost all weather conditions such as snow, rain, summer, etc. Football is governed by FIFA (Fédération Internationale de Football Association) as the highest body and further divides into various other bodies depending on the region and nationality.

The competitiveness of the game varies from region to region based on the participation of the people, media coverage, and club budget. This, in turn, brings varying differences in the level of players and also fluctuates the market value and the skill level based on region, hype generated by the media, competitiveness of the league in which they play and their experience. The bigger the role the player plays in his team, the more likely they may be valued in the market like being the finest penalty taker, or spot free kick specialist, or other roles such as being a playmaker, chance creator, having excellent speed, etc.

In India, despite the decrease in the youth participation in sports, particularly in the past few decades, the industry is putting in various means and efforts to improve the sports environments in the form of grassroots level programs, facilities, tournaments, coaching, public awareness, scholarships, etc. The problem however lies in the fact that it's difficult to search, analyze and coach the players in every part of the country; especially in rural India which consists of 70 percent of the 1.25 billion people approximately. To overcome this difficulty the clubs recruit scouts of vast

IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 4, Issue 6, June 2024
Impact Factor: 5.731

experience and regional understanding to identify players. The AIFF is trying to improve the situation by collaborating with various clubs and companies that make it possible to teach the coaches who may be inexperienced by bringing in connecting sessions with the experienced ones, hosting various tournaments at school, city, district, state level, establishing football academies and community initiatives.

The proposed model is aimed specifically at the grassroot level players of India, further scaling to other soccer leagues. The system is trained as per the in-game values of the 2017 version of EA Sports FIFA. The reason for choosing values based on a game is that it seemed to be the only source for a reliable, near accurate and open form of data available for football players spanning across several leagues.

Moreover, the very nature of the game being a team based sport makes it difficult to analyze the players due to their dependencies on the skillset of other team members, varying positions, formations, club budget, competitiveness in the league and injuries across their career span. Our model is designed to estimate the performance value of the player based on the attributes and skill sets that the player possesses. Coaches can then take advantage of this performance value and train the player, reshuffle the team, recruit, and loan or sell the player. Another value added to this process is the market value of the player obtained through the performance value of the player. However, there will be an approximate deviation in that value by a certain amount due to irregularities in the demand for a particular position, club budget, contract period, injuries and current on-field performance.

II. LITERATURE REVIEW

The paper discussed about Heuristics prediction of olympic medals using machine learning 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.

The paper discussed about Model Gray Prediction for the Gold-Medal Result of Women's Put Shot in the 30th Olympic Games Zhang Bo; Qin Chaoling; Xu Xiaoli; Zeng Fanbo 2012.

The paper discussed about The Application of Decision Tree in the Prediction of Winning Team Xiaohu Tang; Zhifeng Liu; Taizhao Li; Wenbin Wu; Zhenhua Wei. 1990.

The paper discussed about Displaying Winning Probabilities in Volleyball Interests Audiences Sensitive to Probability Tomoyuki Maekawa; Kazuhiro Ueda 2000.

The paper discussed about Sports Games Modeling and Prediction using Genetic Programming Shengkai Geng; Ting Hu 2002.

III. OBJECTIVE OF THE PROBLEM

Player selection is one the most important tasks for any sport and no exception. The performance of the players depends on various factors such as the opposition team, the venue, his current form etc. The team management, the coach and the captain select players for each match from a squad of 15 to 20 players. They analyze different characteristics and the statistics of the players to select the best playing for each match. Each player contributes by scoring maximum points possible. This project attempts to predict the performance of players as how many points will each players score for both the teams. Both the problems are targeted as classification problems are reduced.

IV. PROBLEM STATEMENT AND METHODOLOGY

PROBLEM DEFINITION:

In the game of football (soccer), the evaluation of players for transfer, scouting, squad formation and strategic planning is important. However, due to the vast pool of grassroots level player, short career span, differing performance throughout the individual's career, differing play conditions, positions and varying club budgets, it becomes difficult to identify the individual player's performance value altogether. Our Player Performance Prediction system aims at solving this complex problem analytically and involves learning from various attributes and skills of a football player. The objective of this system is to help the coaches and team management at the grassroots as well as higher levels to identify the future prospects in the game of football without being biased to subjective conditions like

club budget, competitiveness in the league, and importance of the player in the team or region. Using machine learning algorithms predicting the results of a football match, we obtained and create set of features, thus developing with high accurate predictive method using machine learning techniques.

The main challenges in order to develop the next generation of intelligent Systems are: -

- To minimize the time required for Football player prediction.
- User friendly model.
- Less time consumption.
- Should be implemented in all datasets.

V. PROPOSED SYSTEM

In statistics, N**aive Bayes classifiers** are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naïve) independence assumptions between the features. They are among the simplest Bayesian network models, but coupled with kernel density estimation, they can achieve higher accuracy levels.

Naïve Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. Maximum-likelihood training can be done by evaluating a closed-form expression, which takes linear time, rather than by expensive iterative approximation as used for many other types of classifiers.

In the statistics and computer science literature, naive Bayes models are known under a variety of names, including **simple Bayes** and **independence Bayes**. All these names reference the use of Bayes' theorem in the classifier's decision rule, but naïve Bayes is not (necessarily) a Bayesian method.

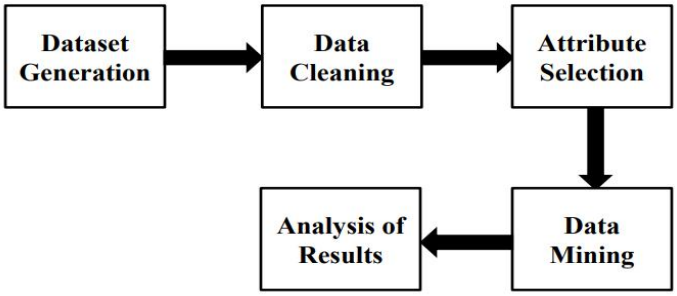
Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features.

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

Despite their naive design and apparently oversimplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, an analysis of the Bayesian classification problem showed that there are sound theoretical reasons for the apparently implausible efficacy of naive Bayes classifiers. Still, a comprehensive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests.

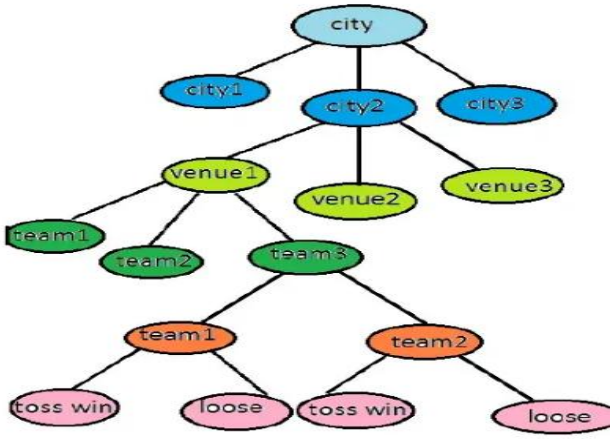
IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 4, Issue 6, June 2024
Impact Factor: 5.731

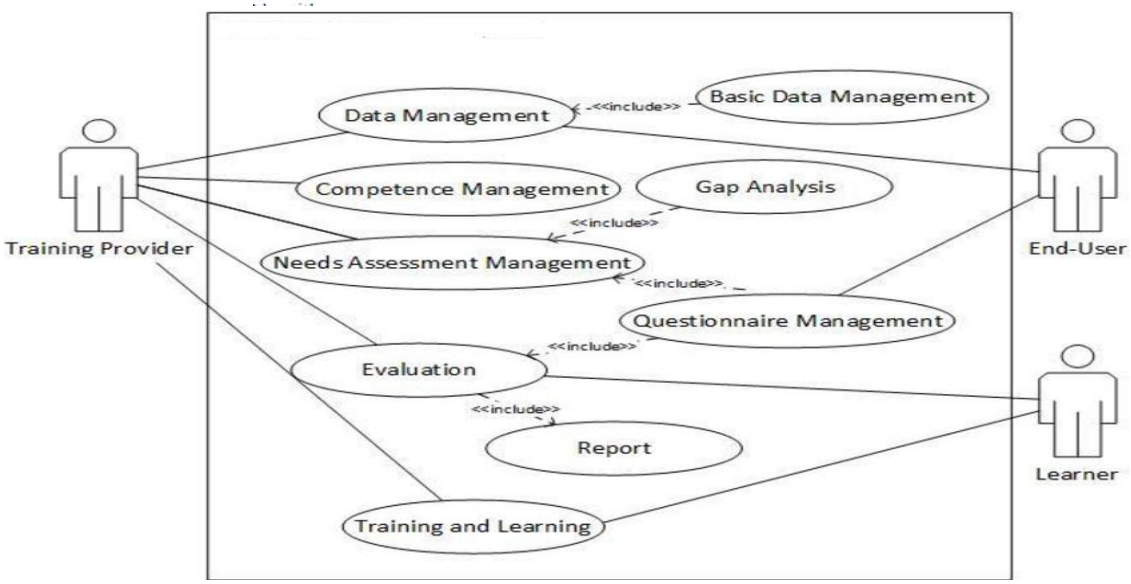
VI. SYSTEM ARCHITECTURE



UML DIAGRAMS:

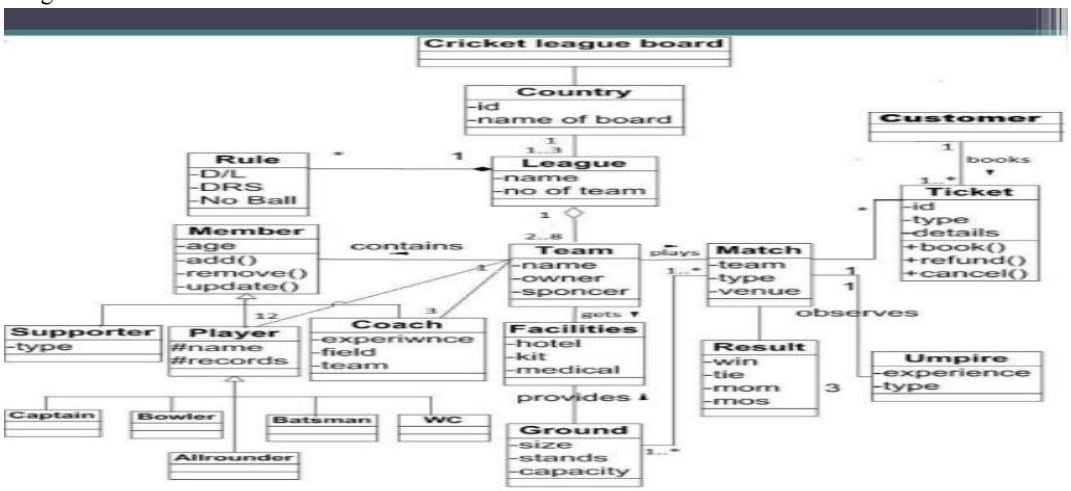
Use case Diagram:



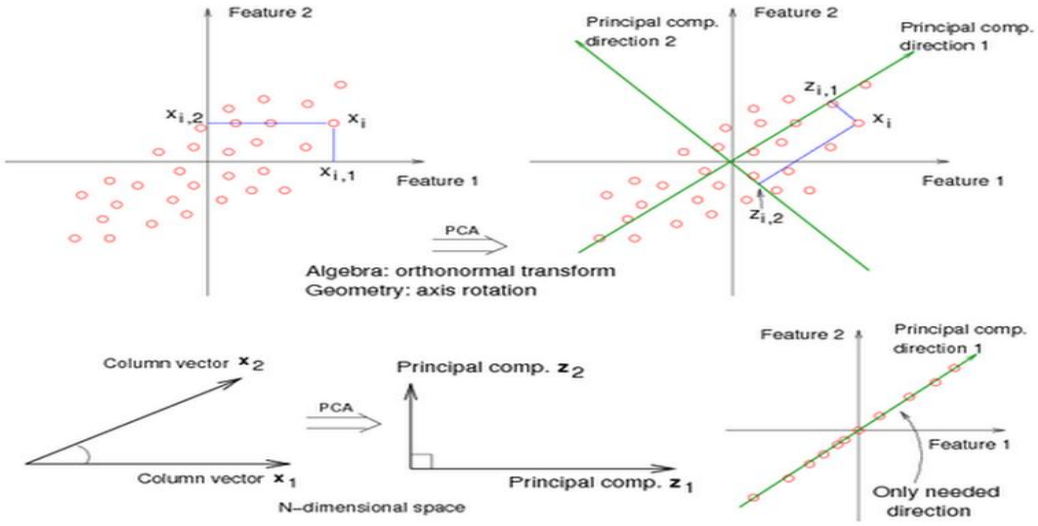


IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 4, Issue 6, June 2024
Impact Factor: 5.731

Class Diagram:

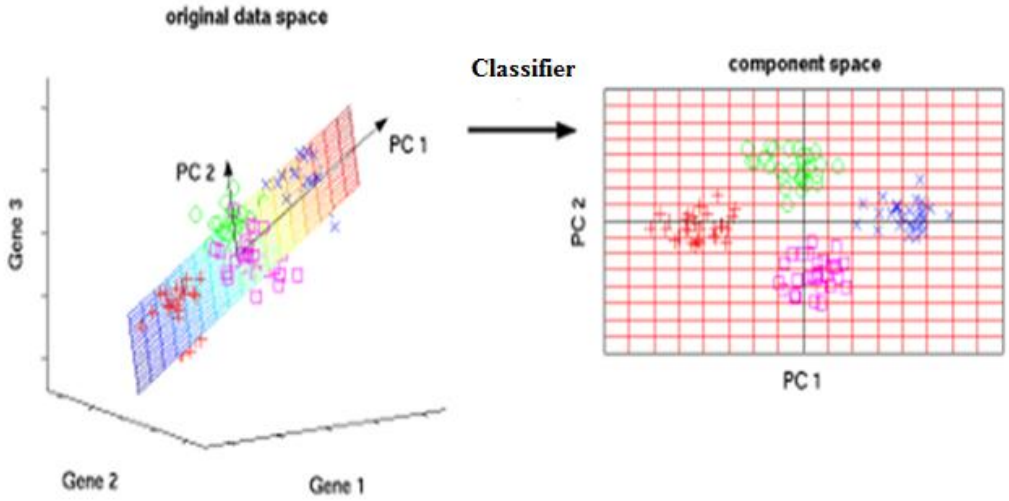


Deployment Diagram:



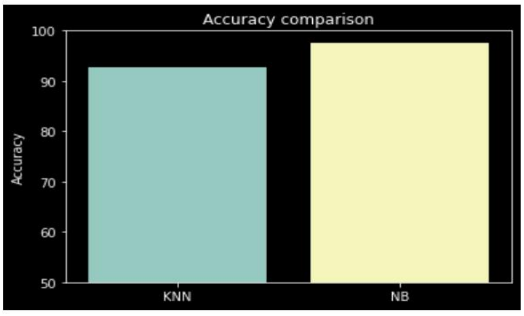
IJETIR
ISSN (Online) 2583-0554
International Journal of Emerging Technologies and Innovative Research (IJETIR)
Volume 4, Issue 6, June 2024
Impact Factor: 5.731

Component Diagram:



OUTPUT:

	Agorithms	accuracy
0	KNN	92.82
1	NB	97.61



	Unnamed: 0	FTR	HTP	ATP	HM1	HM2	HM3	AM1	AM2	AM3	HTGD	ATGD	DiffFormPts
0	0	NH	0.0	0.0	M	M	M	M	M	M	0.0	0.0	0.0
1	1	H	0.0	0.0	M	M	M	M	M	M	0.0	0.0	0.0
2	2	H	0.0	0.0	M	M	M	M	M	M	0.0	0.0	0.0
3	3	NH	0.0	0.0	M	M	M	M	M	M	0.0	0.0	0.0
4	4	NH	0.0	0.0	M	M	M	M	M	M	0.0	0.0	0.0
5	5	NH	0.0	0.0	M	M	M	M	M	M	0.0	0.0	0.0
6	6	NH	0.0	0.0	M	M	M	M	M	M	0.0	0.0	0.0
7	7	NH	0.0	0.0	M	M	M	M	M	M	0.0	0.0	0.0
8	8	H	0.0	0.0	M	M	M	M	M	M	0.0	0.0	0.0
9	9	H	0.0	0.0	M	M	M	M	M	M	0.0	0.0	0.0

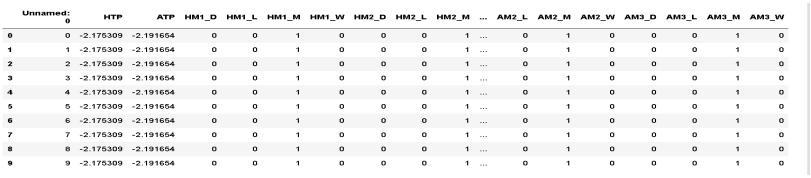
IJETIR

ISSN (Online) 2583-0554

International Journal of Emerging Technologies and Innovative Research (IJETIR)

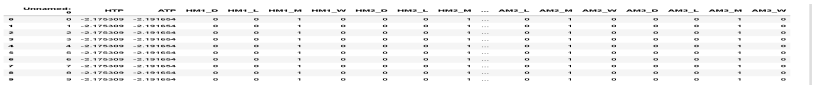
Volume 4, Issue 6, June 2024

Impact Factor: 5.731



Unnamed: 0	HTP	ATP	HM1_D	HM1_L	HM1_M	HM1_W	HM2_D	HM2_L	HM2_M	...	AM2_L	AM2_M	AM2_W	AM3_D	AM3_L	AM3_M	AM3_W
0	0	-2.175309	-2.191654	0	0	1	0	0	0	1	...	0	1	0	0	0	1	0
1	1	-2.175309	-2.191654	0	0	1	0	0	0	1	...	0	1	0	0	0	1	0
2	2	-2.175309	-2.191654	0	0	1	0	0	0	1	...	0	1	0	0	0	1	0
3	3	-2.175309	-2.191654	0	0	1	0	0	0	1	...	0	1	0	0	0	1	0
4	4	-2.175309	-2.191654	0	0	1	0	0	0	1	...	0	1	0	0	0	1	0
5	5	-2.175309	-2.191654	0	0	1	0	0	0	1	...	0	1	0	0	0	1	0
6	6	-2.175309	-2.191654	0	0	1	0	0	0	1	...	0	1	0	0	0	1	0
7	7	-2.175309	-2.191654	0	0	1	0	0	0	1	...	0	1	0	0	0	1	0
8	8	-2.175309	-2.191654	0	0	1	0	0	0	1	...	0	1	0	0	0	1	0
9	9	-2.175309	-2.191654	0	0	1	0	0	0	1	...	0	1	0	0	0	1	0

```
Index(['Unnamed: 0', 'FTR', 'HTP', 'ATP', 'HM1', 'HM2', 'HM3', 'AM1', 'AM2',
       'AM3', 'HTGD', 'ATGD', 'DiffFormPts'],
      dtype='object')
```



VII. CONCLUSION AND FUTURE SCOPE:

Conclusion:

In this project, They have discussed that how our proposed system predicts the performance of the football player. The proposed system is also scalable for predicting the performance of the players for the individual person by data processing. The system is not having complex process to predict the performance of player like the existing system. Proposed system gives genuine and fast result than existing system.

Future Scope:

In the comparison of India with China, which is among the top ten countries in the Olympics and whose population is nearly equal to India, it is understand how important the other factors are to upgrade a country's performance in the Olympics and how important it is to make sport mandatory rather than considering them as recreational activities. From these insights,it is given that education is given higher importance in India and yet the literacy rate and economic status of its people are low. Hence sports should be given higher importance..

REFERENCES

[1]. Chandrasegar Thirumalai, 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]. Model Gray Prediction for the Gold-Medal Result of Women's Put Shot in the 30th Olympic Games Zhang Bo; Qin Chaoling; Xu Xiaoli; Zeng Fanbo 2012.aq

[3]. The Application of Decision Tree in the Prediction of Winning Team Xiaohu Tang; Zhifeng Liu; Taizhao Li; Wenbin Wu; Zhenhua Wei 1990.

[4]. Winning Probabilities in Volleyball Interests Audiences Sensitive to Probability Tomoyuki Maekawa; Kazuhiro Ueda 2000.

[5]. Sports Games Modeling and Prediction using Genetic Programming Shengkai Geng; Ting Hu 2002.

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

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

[8]. Machine Learning Models Reveal Key Performance Metrics of Football Players to Win Matches in Qatar Stars League Jassim Almulla; Tanvir Alam Publisher: IEEE 2023

[9]. 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

[10]. Data Mining Computing of Predicting NBA 2019–2023 Regular Season MVP Winner Mason Chen; Charles Chen Publisher: IEEE 2023