PERFORMANCE EVALUATION FOR SPORTS TEAM SELECTION USING DATA MINING TECHNIQUES

Isha Sachin Thakare[1]
Sayali Rajesh Suyal[2]
Kirti Yogesh Pandav [3]

Abstract

Indian tradition emphasizes a lot on sports as well as physical fitness and nowadays sports have gained a vital importance and prestigious recognition in India. Making an international mark in the field of sports also has become influence on national morale. This leads towards the gathering of best sport talent through efficient team selection. Performance in Team Sports is carried out through a long term and methodical training process so as to improve the skills of players in a team. In order to meet the increasing competitive demands of the changing times, the coaches and sports team managers can be aided by various data mining techniques. In this paper, we are exploring data mining techniques like Classification, Association Rules to evaluate the performance of players for team selection.

Key Terms — Performance Evaluation, Data Mining, Classification, Association Rules

Introduction

Awareness of physical fitness is growing rapidly all over the world. Sports provide the comprehensive platform fulfilling the need for physical fitness. Today’s generation is incredibly enthusiastic about the sports and hence the sports have gained wide attention. Financial involvement in the sports has made it more methodical, improvising the performance to the great extent. The conventional training provided to the elite sportsmen is not sufficient as the sports competitions have become professional in recent years. Emergent attention and new techniques evolved in sports are generating incredibly exceeding the sports data. The overall aim of this paper is to gather data of the sportsmen and explore the relationships between the criteria for selecting the preeminent players using Data Mining techniques.

For team selection, among various team events, we have selected Handball, a second fastest game after Ice Hockey. In this paper, we are using aforementioned data from Changu Kana Thakur Arts, Commerce and Science College, New Panvel, affiliated to University of Mumbai. A dynamic approach to analyze this data to evaluate the performance of the player is Data Mining. Data Mining is the knowledge discovery process of extracting and analyzing the relevant data from huge database, as to explore

[1] Assistant Professor, CKT College, Panvel.
[2] Assistant Professor, CKT College, Panvel.
[3] Assistant Professor, CKT College, Panvel.

patterns or relationships. It can be proved as a new powerful technology with great potential to help the coaches and sports team managers for selecting the splendid players.

Data Mining techniques are categorized into Predictive and Descriptive models. A Predictive model makes a prediction about values of data using known results found from different data. Predictive model include Classification, Regression, Time Series Analysis and Prediction. A Descriptive model identifies patterns and relationships in data by examining the existing properties of data. Clustering, Summarization, Association Rules and Sequence Discovery are viewed as Descriptive models [1]. Among all the above techniques of Data Mining, we have used the Classification and Association Rules for analyzing the data. Classification is a Data Mining technique used to predict group membership for data instances.

Association Rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an Association Rule would be "If a customer buys a mobile, he is 80% likely to buy its screen-guard also." It is used to predict whether the performance of the player is excellent, average or poor depending on different sports criteria.

Data collection and pre-processing

In our research paper, data is acquired from Changu Kana Thakur Arts, Commerce and Science College, New Panvel, affiliated to University of Mumbai. Before applying the Data Mining techniques the collected raw data needs to be pre-processed into an expedient format.

On the basis of collected data, some attributes are considered to evaluate player's performance for final team selection. The Attributes used for classifying the player's performance are mentioned in Table I.

Attributes	**Description**	**Values**
Age	The running age of player	Between 18 years to 28 years
Running	The physical running capacity in minutes	Minimum 12 minutes
Experience	Player's experience of playing the game calculated in years	Value in years
Achievement	Player's achievements at different levels	University, State or National level

TABLE II. SAMPLE OF PLAYER DATA COLLECTION

Name	Age	Running	Experience	Achievement
Ajay Pawan Kanwar	19	32	6	State
Krunal Jitendra Jagtap	20	31	7	State
Randhir Rajendra Singh	18	35	1	University
Nilesh Rajaram Khairnar	18	35	6	State
Akshay Shankar Shetty	20	25	7	National
Prashant Daji Khose	20	25	1	University
Girish Ankush Mohite	21	30	8	National
Jayesh Rohidas Shinde	20	35	8	State
Pramod Mallikarjun Kumbhar	26	30	8	State
Mangesh Bajirao Pandav	22	32	6	University
Arun Manik Ghemud	27	30	7	State
Pankaj Dattatray Mhatre	18	35	6	University
Awdesh Kumar Morya	21	35	2	University
Dikshant Gopal Kadu	19	32	7	University
Shakti Dilip Thakur	21	20	2	State
Akash Gorakhnath Jadhav	19	15	2	State
Anthoney George Daniel	21	35	11	National
Nilesh Hanumant Patil	21	30	1	University
Vishal Udhaysingh Phadtare	21	25	3	State
Chandraroshan Sabhajit Bari	20	34	7	University
Sushant Jhawle	25	25	5	University
Yogesh Ananda Tawar	25	30	6	State
Satish Ragupati Mokal	27	20	2	State
Vikram More	22	25	6	University

Research methodology

Association Rules: Association is a Data Mining function that discovers the probability of the co-occurrence of items in a collection. The relationships between co-occurring items are expressed as Association Rules.

Given a set of items I = {I1, I2,……., Im} and a database of transactions D = {t1, t2,….,tn} where ti = {Ii1, Ii2, ……., Iik} and Iij Є I , an association rule is

an implication of the form X ⇒ Y where X, Y ⊂ I are sets of items called itemsets and X ∩ Y = ф [1] . The strength of the Association Rule can be measured in terms of its support and confidence.

The support(s) for an association rule X ⇒ Y is the percentage of transaction in the database that contains X ∪ Y.

The confidence (α) for an association rule X ⇒ Y is the ratio of the number of transactions that contains X ∪ Y to the number of transactions that contain X [3].

The association between various data items can be found out by mining Multilevel Association Rules, Multidimensional Association Rules and/or Quantitative Association Rules. Multilevel Association Rules involve concepts at different levels of abstraction. Multidimensional Association Rules involve more than one dimension or predicate. Quantitative Association Rules involve numeric attributes that have implicit ordering among values. Each Quantitative Association Rule has quantitative attributes on left – hand side of the rule and one categorical attribute on right-hand side of the rule [2].

Aquan1 ^ Aquan2 → Acat

We have mined Quantitative Association Rules from the pre-processed data as mentioned in Table III.

TABLE III. ASSOCIATION RULES

Association Rule	**Support (%)**	**Confidence (%)**
age (P1,18..23) ^ running (P1,21..30) ^ experience(P1,0..5) ^ achievement (P1,”University”) -> Player_performance (P1,"Average ")	8	14.28
age (P2,24..28) ^ running (P2,21..30) ^ experience(P2,0..5) ^ achievement (P2,”University”) -> Player_performance (P2,"Average ")	4	7.14
age (P3,18..23) ^ running (P3,21..30) ^ experience(P3,>5) ^ achievement (P3,”University”) -> Player_performance (P3,"Average ")	4	7.14
age (P4,18..23) ^ running (P4,>30) ^ experience(P4,0..5) ^ achievement (P4,”University”) -> Player_performance (P4," Average ")	4	7.14
age (P5,18..23) ^ running (P5,>30) ^ experience(P5,>5) ^ achievement (P5,”University”) -> Player_performance (P5," Excellent ")	16	44.44
age (P6,18..23) ^ running (P6,12..20) ^ experience(P6,0..5) ^ achievement (P6,”State”) -> Player_performance (P6," Average ")	8	14.28
age (P7,24..28) ^ running (P7,12..20) ^ experience(P7,0..5) ^ achievement (P7,”State”) -> Player_performance (P7," Average ")	4	7.14
age (P8,18..23) ^ running (P8,21..30) ^ experience(P8,0..5) ^ achievement (P8,”State”) -> Player_performance (P8,"Average ")	4	7.14
age (P9,24..28) ^ running (P9,21..30) ^ experience(P9,>5) ^ achievement (P9,”State”) -> Player_performance (P9,"Average ")	12	21.42
age (P10,18..23) ^ running (P10,>30) ^ experience(P10,>5) ^ achievement (P10,”State”) -> Player_performance (P10," Excellent ")	16	44.44
age (P11,18..23) ^ running (P11,21..30) ^ experience(P11,>5) ^ achievement (P11,”National”) -> Player_performance (P11,"Average ")	8	14.28
age (P12,18..23) ^ running (P12,>30) ^ experience(P12,>5) ^ achievement (P12,”National”) -> Player_performance(P12," Excellent ")	4	11.11

Classification

Classification is an analytical task where the classifier is constructed to predict the categorical labels i.e. Classes.

Given a database D={t1,t2,……,tn} of tuples (items, records) and a set of classes C={c1,…..,cm}, the classification problem is to define a mapping f : D → C where each ti is assigned to one class. A class, Cj, contains precisely those tuples mapped to it; i.e. $Cj=\{ti \mid f(ti)=Cj,\ 1 \leq i \leq n, \text{ and } ti \in D\}$[1].

Classification is a process that works in two step in which data are simplified for illustrative purposes. The first step consist of the learning step (or training phase) in which the classifier is built describing a predetermined set of data classes or concepts. A classification algorithm builds the classifier by analyzing or “learning from” a training set made up of database tuples and their associated class labels. A tuple, X, is represented by an n-dimensional attribute vector X=(x1,x2,…,xn), depicting n measurements made on the tuple from n database attributes, respectively, A1,A2,…,An. Each tuple, X, is assumed to belong to a predefined class.

The learning step expands with use of model for classification which is the second step. First, the predictive accuracy of the classifier is estimated. A test set is used, made up of test tuples and their associated class labels. They are independent of the training tuples, meaning that they were not used to construct the classifier. The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier. The associated class label of each test tuple is compared with the learned classifier’s class prediction for that tuple. If the accuracy of the classifier is considered acceptable, the classifier can be used to classify future data tuples for which the class label is not known [2].

Here the classifier is constructed to predict the categorical labels among “Selected” and “Rejected”. The corresponding class label is provided to each training tuple. The Classification rules obtained at the end of the training phase are as mentioned in Table IV.

TABLE IV. CLASSIFICATION RULES

IF Player_performance = Poor AND Physical_fitness = Good AND Family_support = Yes THEN Player =Rejected
IF Player_performance = Poor AND Physical_fitness = Good AND Family _support = No THEN Player =Rejected
IF Player_performance = Poor AND Physical_fitness = Poor AND Family_support = Yes THEN Player =Rejected
IF Player_performance = Poor AND Physical_fitness = Poor AND Family_support = No THEN Player = Rejected
IF Player_performance = Average AND Physical_fitness = Good AND Family_support = Yes THEN Player = Selected
IF Player_performance = Average AND Physical_fitness = Poor AND Family_support = Yes THEN Player = Rejected
IF Player_performance = Average AND Physical_fitness = Good AND Family_support = No THEN Player = Rejected
IF Player_performance = Average AND Physical_fitness = Poor AND Family_support = No THEN Player = Rejected
IF Player_performance = Excellent AND Physical_fitness = Good AND Family_support = Yes THEN Player = Selected

IF Player_performance = Excellent AND Physical_fitness = Good AND Family_support = No THEN Player = Selected
IF Player_performance = Excellent AND Physical_fitness = Poor AND Family_support = Yes THEN Player = Rejected
IF Player_performance = Excellent AND Physical_fitness = Poor AND Family_support = No THEN Player = Rejected

The Classification rules mentioned in Table IV predict Player's selection for the final team. The outcome of the Classification process is the set of Classification rules which predict the inclusion of any player in the team. The rejected players can approach for the better training to enhance their performance.

Conclusion

This paper explores the prospective efficiency of Data Mining techniques for enhancing the team selection process. The Association rules are used to find the player's performance and through the Classification rules we have selected best players for the final team. By applying above mentioned techniques on the player's data, the selected players forming a team are listed in Table V.

TABLE V. FINAL SELECTED TEAM

Name	**Age**	**Running**	**Experience**	**Achievement**
Ajay Pawan Kanwar	19	32	6	State
Krunal Jitendra Jagtap	20	31	7	State
Nilesh Rajaram Khairnar	18	35	6	State
Akshay Shankar Shetty	20	25	7	National
Prashant Daji Khose	20	25	1	University
Jayesh Rohidas Shinde	20	35	8	State
Pramod Mallikarjun Kumbhar	26	30	8	State
Mangesh Bajirao Pandav	22	32	6	University
Arun Manik Ghemud	27	30	7	State
Pankaj Dattatray Mhatre	18	35	6	University
Dikshant Gopal Kadu	19	32	7	University
Shakti Dilip Thakur	21	20	2	State
Akash Gorakhnath Jadhav	19	15	2	State
Anthoney George Daniel	21	35	11	National
Nilesh Hanumant Patil	21	30	1	University
Chandraroshan Sabhajit Bari	20	34	7	University
Yogesh Ananda Tawar	25	30	6	State

We are further on working on the analysis of the efficient team selection for the players belonging to different localities (Urban and Rural) and different socio-economic status using some more predictive and descriptive Data Mining techniques.

References

[1] Margaret H. Dunham, “Data Mining: Introductory and advanced Topics”, Pearson 2013.

[2] Jiawei Han, Micheline Kamber, “Data Mining Concepts and Techniques”, Elsevier 2009.

[3] Vipin Kumar, Pang-Ning Tan, Michael Steinbach, “Introduction to Data Mining”, Addison-Wesley, 2005. ISBN : 0321321367

[4] Rajan chattamvelli, “Data Mining Methods” Narosa 2009

[5] David Cheung, Graham Williams, Qing Li, “Advances in Knowledge Discovery and Data Mining”, PAKDD 2001

[6] Bing Liu, Wynne Hsu, Yiming Ma, “Integrating Classification and Association Rule Mining”, KDD-98 Proceedings

[7] Carson K. Leung, Kyle W. Joseph, “Sports data mining: predicting results for the college football games”, Procedia Computer Science 35 (2014) 710 – 719

[8] Chenjie Cao, “Sports Data Mining Technology Used in Basketball Outcome Prediction”.

[9] Neelamadhab Padhy, Dr. Pragnyaban Mishra, Rasmita Panigrahi, “The Survey of Data Mining Applications And Feature Scope”, (IJCSEIT), Vol.2, No.3,page no-43 June 2012 ,DOI : 10.5121/ijcseit.2012.2303

[10] Ms. D. P. Vaidya, Dr. Sanjay U. Makh, “Data Mining System Architecture for Training Plan Selection for the Swimmers”, International Journal of Engineering and Science ISSN: 2278-4721, Vol. 1, Issue 2 (Sept 2012), PP 35-41

[11] Smita, Priti Sharma, “Use of Data Mining in Various Field: A Survey Paper”, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 3, Ver. V (May-Jun. 2014), PP 18-21

[12] Kalyani M Raval, “Data Mining Techniques”, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012, ISSN: 2277 128X