A systematic review of knowledge mining based clustering and classification for predicting sports interest from student performance

A. Basheer Ahamed[1*], M. Mohamed Surputheen[2]
[1.] Assistant Professor, Department of Information Technology,
[2.] Associate Professor, Department of Computer Science,
Jamal Mohamed College (Autonomous), Tiruchirapalli – 620 020. Tamilnadu, India.

***Abstract*:** Data mining is an incredible part to inexpensive the information investigation in different fields to predict the execution and arrange the information from sports dataset. The information mining part affects the data prediction to bargain the factual information for classification. The data prediction model for student information not to recognize the activities carried toward becoming to distinguish the sports players. Information mining and Machine learning (ML) is one of the keen strategies that have indicated likely outcomes in the areas of classification and prediction. One of the extending zones requiring great predictive precision is sports prediction, because of the vast financial sums associated with wagering. Moreover, sports supervisors and sponsors are taking a stab at classification models so they can comprehend and detail systems expected to the student for matches. These models depend on various elements engaged with the diversions, for example, the sports interest of chronicled events, player execution markers, and resistance data. To propose new clustering and classification based on multi-objective to predict the sports performance from students that examination gives a first investigation of the sports mining in ML, concentrating on the use of information mining to sport comes about prediction. In doing as such, to distinguish the learning strategies used, information sources, proper methods for show assessment, and particular difficulties of predicting sport comes about. This at that point drives us to propose a novel sports prediction structure through which ML can be utilized as a learning technique. Our examination will ideally be educational and of utilization to those performing future research in this sports application zone.

Keywords: data mining, prediction, classification, interest score, sports mining.

1. INTRODUCTION

Data mining technique has attracted attention in the information industry and society as a whole, because of the significant amount of data and the imminent need to transform that data into useful information and knowledge. Recently conducted studies with successfully demarcated results using this technique, to estimate several parameters in a variety of domains. However, the practical use of data in some areas is still developing, as is the case with sports, which has shown moderate growth. In the sports mining activity that is present a systematic review of the literature about research involving sports data mining

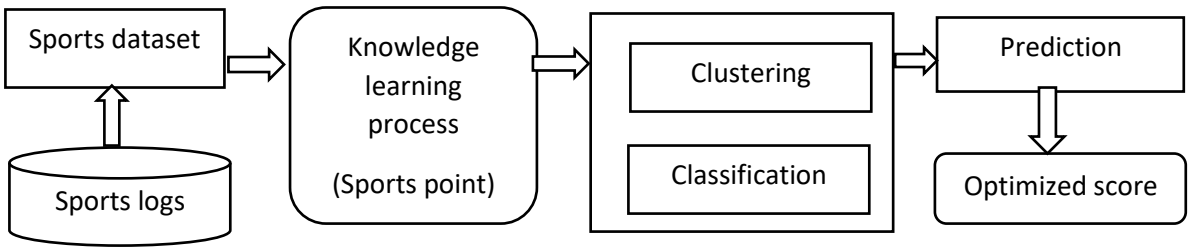

Figure 1 process of sports mining

In this work, we design and construct a sports information analytics framework for sports behavior mining and personalized sports services using knowledge learning process as shown in figure 1. We analyzed users' exercise data that were collected in Sports platform to discover users' periodic sports patterns and the trend of performance prediction, and those analytical result can be applied for various sports applications including those listed below:

Talent is a term used in any field to point out the outstanding skills or performance of individuals. The non-universal status is based on the student interest. In general terms, according to the talent has several properties which can be genetically transmitted and partially specific interest on sports by stage, there are particular predictors used to foresee which individuals have the best chances for success in the extraction better opportunities for student support in sports depend on the field in question. Here, the study focuses on sports and more specifically, soccer. In sports, the predictors for identifying talent are not necessarily precisely defined.

The table formulated the domain attribute value to process the sports interest

Attribute	Description	Possible values
GENDER	Student gender	{Male, Female}
NATCAT	Nationality category	{Local, Gulf, Arab, Non-Arab}
FLANG	First Language	{Arabic, English, Hindu-Urdu, Other}
HSP	High School Percentage	{Excellent (90% to 100%), Very Good (High) (85% to 89.9%), Very Good (80% to 84.9%), Good (High) (75% to 79.9%), Good (70% to 74.9%), Pass (High) (65% to 69.9%), Pass (60% to 64.9%)}
Curriculum sort, drawing, games	certificate	{Freshman (< 32), Sophomore (33 - 64), Junior (65 - 96), Senior (> 96)

The following are the consideration in sport mining knowledge mining process whether the techniques have enhance the features for prediction result

1.1 Clustering and classification

The approach introduced in this research is motivated by the importance of the talent identification stage from domain attribute values of student and sports learners as well as the debate whether science methods can help when it comes to identifying talent in sports are grouped. There are a lot of talented students classified who might not have other ways of demonstrating their expertise except through social media. A sports community in an educational institution can also leverage the benefits by exploiting the information available on student achievements from records.

1.2 Prediction and classification

The main issue is that there are too many predictions and classification methods don't provide exposure to give the correct result. The presented study was done in prediction and classification based on multi fact cluster evaluation to predict the best performance to motivate sports person. The research

objective is to identify attribute by quick case information feature identification from student achievements to evaluate average weightage model. The further knowledge processing based on natural language processing (NLP) is used for the methodological basis in clustering and classification. Finally, it is classification results re to predict the sports interested with high accuracy if this can be done automatically through our proposed method.

Periodic Pattern Mining

For the user exercise records, there are lots of attributes such as exercise time, exercise type, duration of exercise et al. Therefore, we focus on discovering multi-attribute periodic patterns of each user, e.g., how often this user bikes for an hour on Monday afternoon and how often this user runs for thirty minutes on Sunday. Hence, it is a kind of periodic pattern mining problem. Intermittent pattern mining is the task of discovering patterns that periodically appear in data. Detecting these periodic patterns is helpful to understand data and make strategic decisions. Therefore, we try to find regular trends in every user exercise records.

Objective and consideration of this research

- In this research, the primary goal is to predict the performance of the student in achieving sports legal actives for the student as players in the future.
- To find the complex predicted with the data variables that may impact how the student be a player will perform, and determine what their student sports statistics will be for a future.
- To design an efficient extraction pattern model to improve the sports learning. Which the prediction information be well supported by sports learning to identify the future players.
- To design an efficient extraction pattern model to improve the sports learning. Which the prediction information be well supported by sports learning to identify the future players.

2. DISCUSSION

The present reviews previous research on data mining systems to predict sports results and evaluates the advantages and disadvantages of each system Data mining can be used to solve many problems today. Available datasets such as learning educational insights after some time can be information mined to acquire exact predictions of how the data will look like later on. Our concern is to utilize a lot of student sports datasets over the order to determine the performance of students is developed as players in the future. This way, coaches, players and educational organization can adequately prepare for the future and make the best decisions for their teams. Also, students are shine in fantasy games can optimize their talents so that they have the best chance to win. The various tremendous approach that does not well proceed to shine the information mining described by multiple authors reviews are s follows,

2.1.1 Frequent data Analyze and extraction

Clustering and classification technique is used to analyze and extract the useful information from the data. The knowledge of student data is stored on dataset on a centralized resource for classification schemes. It helps to analyses, categories and to find the relationships among the data using various algorithms such as classification, clustering, regression, association rule mining and anomaly detections. Classification is used for prediction.

In this research to use novel statistical clustering and Decision classification algorithm to predict the performance of a sports player.

2.1.2 Statistical clustering

In statistical clustering, the framework is prepared with the set of information, and this information comprises numerous records each having various properties. The data used to develop the structure is known as the preparation information. Each file will have a class mark. The goal is to examine the info information to build up the model for each class utilizing the traits. Though this model is used to test the knowledge for which they little class marks are obscure to public cluster groups.

2.1.3 Decision classification in point of class labels

Decision classification begins with the root node and branches out into internal node or leaf node by students interest rate. Each interest rate represents data attributes that fall within the range of values. Data points travel through the internal nodes to reach the final node which is known as a leaf node. The decision that identifies the match case rule of pointed data that the classifiers are simple, inexpensive and can build the models quickly even though the training set is huge with different classifiers.

3. COMPARATIVE STUDY

The following are the case study reviewed by different authors by the line of techniques flows with debate and its limitations. The most current predicting models use academic credit worth information from the curricula, ignoring extracurricular activities, while there is evidence from other research fields that some activities like sports can be related to academic performance

Title	Author's	Year / Journal	Debate/limitations
Monitoring Student Performance Using Data Clustering and Predictive Modelling	Alana M. de Morais and Joseana M. F. R. Araújo, Evandro B. Costa	2014/IEEE	analyzes the usage of Open Learning Data to characterize behavioral profiles of answers using multivariate analysis techniques
Extracurricular activities as dropout prediction factors in higher education using decision trees	Tomas Hasbun, Alexandra Araya, and Jorge Villalon1	2016/IEEE	extracurricular activities to predict dropout in students/use academic credit worth information from the curricula, ignoring extracurricular Activities
Clustering Analysis of Sports Performance based on Ant Colony Algorithm	Wang Jian1, Hong Zhi-hua2, Zhou Zhi-yong3	2016/IEEE	Investigates the principle of the k-means clustering algorithm. Because it is easy to converge into local minimum and is also sensitive to noise, isolated point data have a great impact on the average value
M5P Model Tree in Predicting Student Performance: A Case Study	S Chaitanya Kumar, E Deepak Chowdary, Venkatramaphanikumar	2016/IEEE	decision tree induction approach called Multivariate Regression prediction model M5P has been used for predicting performance as online-learning skills, problem-solving efficiency, time management, sports participation.

Evaluating the Effect of Program Visualization on Student Motivation	J. Ángel Velázquez-Iturbide	2017/IEEE	Motivation is probably the most often-cited effect on students of using visualizations/objective data and calibrated questionnaires, of the impact of viewing on student motivation. focuses on useful mining patterns and discovering helpful knowledge from the educational information
Educational Data Mining & Students' Performance Prediction	Amjad Abu Saa	2016/IJACSA	Motivation is probably the most often-cited effect on students of using visualizations/objective data and calibrated questionnaires, of the impact of display on student motivation.

3.1.1 Linear Classification mining

Our structure is novel and remarkable in three diverse ways. Immediately, the framework contains two essential parts: an oversee based reasoner and a Bayesian framework component[1]. This is a compound one as in two one of good systems facilitates in predicting the delayed consequences of sports matches.

The purpose of classification is to predict a goal variable (class) by building a classification show in light of an arrangement dataset [2], and a short time later utilizing that model to predict the estimation of the class of test data.

The objective is to lift or point of confinement this limit subject to prerequisites that are moreover immediate. In this model, weights for each one of the bits of knowledge were settled and a while later used to figure a numerical score for each player [3, 4]. This work shows an audit of continuous headways in diversion and unwinding with the compliment on advancement and computational frameworks which were assembled in a practical sports and social stage, to locate customers' irregular sports outlines and the example of heart rate change in the midst of exercise[5]. Since the dataset isn't merely extraordinarily tremendous yet also winding up quickly.

3.1.2 Clustering and social mining prediction

Predicted practices of students on each gathering, which were described in the last progress. In our examination, it describes that the prediction will use a backslide approach [6], and the clustering will execute K-suggests figuring. The results displayed prescribe that including extracurricular activities are useful to watch specific practices that are all in all related with the withdrawal ponder related to dropout [7, 8]. to converge into neighborhood minimum and is furthermore tricky to racket, detached point data incredibly influence the typical regard, an upgraded clustering computation in perspective of underground creepy crawly state change had short-lived information[9,10]. The standard data is gotten by demonstrating events as the transient progressions using a lexicon of non-unclear transient illustrations.

Motivation is likely the as often as possible referred to the effect on students of using observations. However, to the makers' data, the affirmation merely is described, with, most ideal situation, surveys and considers being used to collect tendencies or estimations [12].

To demonstrate that nature-moved figurings are similarly profitable inside the space of diversion [13], mainly to acquire secured and effective getting ready designs concentrating on various parts of execution. Educational Data Mining (EDM) is the field of think stressed over enlightening mining data to find entrancing cases and learning in informative affiliations [14]. This examination is correspondingly emphasized over this subject, especially, the students' execution. Most cases he proposed positive Euler segment to bit k-suggests and spectral clustering with a specific end goal to make Euler k-means and Euler powerful clustering, independently.

Distinctive target limits are used to evaluate the suitability of the bundle by taking a gander at some regular property minimization of the clusters, expel measures [18], amass symmetry and thickness [20]. Precision diverged from individual classifiers and besides heterogeneous models show best results over homogeneous models or buts and benchmark educational accumulations.

This review points to centers around various issues existed in the streams for sports preparing, talks about the application guideline of the information mining innovation in sports mining

4. Conclusion

This work presents a big data analytical framework for sports behavior analysis and personalized talents. The frame can effectively and efficiently analyze users' exercise data to discover users' periodic sports patterns and the trend of heart rate change during exercise. The analytical result can serve as the important core for personalized sports applications. Moreover, we also group the individual result to find the clustering result, and those analytical result can be applied for advanced sports prediction application Finally, the improved algorithm is applied to mining the evaluation of student performance, including the comprehensive evaluation and single subject level evaluation. By comprehensive evaluation, it's easy to find out the difference between students of the different type, which used to support teaching, and it is valuable for teachers and students to improve their work. And for the other one evaluation, the clustering can relieve the insufficient of traditional grades which lack the flexibility to change with the change of external influences

Author Contributions:

Mr. ABA carried out the preparation of the survey and executes the characterization studies and contributed to the main text of the manuscript. Dr. MMS checked the information and flow of the text to maintain a better readability. Further this research work is not funded by any agency.

Compliance with ethical standards:

Conflict of interest: The authors declared that they have no conflict of interest.

References

1. Byongho Min, Jinhyuck Kim, Chonghyoun Choe, Hyeonsang Eon, Robert Ian (Bob) McKay, "A Compound Framework for Sports Prediction: The Case Study of Football," Knowledge-Based Systems, vol. 21, issue 7, October 2008, pp. 551-562, Elsevier Science Publishers B. V. Amsterdam, The Netherlands
2. Michael Baulch, "Using Machine Learning to Predict the Results of Sporting Matches," Department of Computer Science and Electrical Engineering, University of Queensland
3. Lloyd Smith, Bret Lipscomb, Adam Simkins, "Data Mining in Sports: Predicting Cy Young Award Winners," Journal of Computing Sciences in Colleges, vol. 22, issue 4, April 2007, pp. 115-121, Consortium for Computing Sciences in Colleges

4. Baca, P. Dabnichki, M. Heller, and P. Kornfeind. Ubiquitous computing in sports: A review and analysis. Journal of Sports Sciences, 27(12):1335–1346, 2009
5. Vincent S. Tseng; Chih-Hsin Chou; Kai-Qi Yang; Jerry C. C. Tseng"A Big Data Analytical Framework for Sports Behavior Mining and Personalized Health Services"2017 Conference on Technologies and Applications of Artificial Intelligence (TAAI) Pp178 - 183: 2017
6. Alana M. de Morais and Joseana M. F. R. Araújo, Evandro B. Costa," Monitoring Student Performance Using Data Clustering and Predictive Modelling " IEEE,2014.
7. Y. Li and Y. Zhang. App. of data mining tech in sports training. In Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on, pages 954–958. IEEE, 2012.
8. Tomas Hasbun, Alexandra Araya and Jorge Villalon "Extracurricular activities as dropout prediction factors in higher education using decision trees" IEEE 2016.
9. Jian1, Hong Zhi-hua2, Zhou Zhi-yong3 "Clustering Analysis of Sports Performance based on Ant Colony Algorithm" IEEE 2016.
10. M.-S. Dao and N. Babaguchi. Sports event detection using temporal patterns mining and web-casting text. In Proc. of the 1st ACM workshop on Analysis and retrieval of events/actions and workflows in video streams, pages 33–40. ACM, 2008.
11. S Chaitanya Kumar, E Deepak Chowdary, Venkatramaphanikumar "M5P Model Tree in Predicting Student Performance: A Case Study" IEEE,2016
12. J. Ángel Velázquez-Iturbide "Evaluating the Effect of Program Visualization on Student Motivation" IEEE,2017.
13. I. Fister, K. Ljubicˇ, P. N. Suganthan, and M. Perc. Computational intelligence in sports: challenges and opportunities within a new research domain. Applied Mathematics and Computation, 262:178–186, 2015.
14. Amjad Abu Saa" Educational Data Mining & Students' Performance Prediction " IJACSA,2016.
15. Methods Saima Sayyed, Rugved Deolekar, A Survey on Novelty Detection using Level Set ICICCT,.2017.
16. Jian-Sheng Wu, Wei-Shi Zheng, Euler Clustering on Large-scale Dataset IEEE 2017.
17. Golda George, Latha Parthiban,Multi-objective hybridized firefly algorithm with group search optimization for data clustering IEEE, 2016.
18. Ahmed, M., Afzal, H., Siddiqi, I., Khan, B.: Mcs: Multiple classifier systems to predict the churners in the telecom industry .2017.
19. Ahmed, M., Rasool, A.G., Afzal, H., Siddiqi, I.: Improving handwriting based gender classification using ensemble classifiers. Expert Systems with Applications 85, 158–168 (2017)
20. Singh, Karanjit, and Shuchita Upadhyaya. "Outlier detection: applications and techniques." Int. Journal of Comp. Science Issues, Vol. 9.1, pp. 307-
21. Vaishali, Vaishali. "Credit Card Fraud Detection using Clustering Approach. "International Journal of Computer Applications Vol. 98.3, pp. 29-32, 2014.