Learning Variant Multi perspective Machine Learning clustering algorithm for predicting student interest in sports

A. Basheer Ahamed[1], Dr. M. Mohamed Surputheen[2,] Dr.M. Rajakumar[3]
[1]*Research Scholar, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous) (Affiliated to Bharathidasan University), Tiruchirappalli, Tamil Nadu, India.*
[2,3]*Associate Professor, PG and Research Department of Computer Science, Jamal Mohamed College (Autonomous) (Affiliated to Bharathidasan University), Tiruchirappalli, Tamil Nadu, India.*

***Abstract*—Learning the interest of students in sports turns out to be more significant assignment towards the improvement of sports around worldwide principles. Quantities of methods have been prescribed before to foresee the students' advantage on sports. The interest on sports has been anticipated by time spent and recurrence of playing in prior draws near, yet endures to deliver higher outcomes on forecast and the sports distribution analysis model identifies the list of sports interests of various students by analyzing the traces of student activities in academic and social environments. To work on these presentations, an effective Learning Style based Sports prediction model (LSPM) is suggested in this article. Learning type Sports prediction model identifies the list of sports interests of various students by analyzing the traces of student activities learning style in academic and social environments. The logs of sports activities are split into number of time window and for each of them, the method estimates the frequency of sports activity on each interest based on learning types. According to these measures the method performs learning type analysis and measure sports factors. Based on these values, the method performs interest prediction. The LSPM (Learning Type Sports Prediction) achieve higher interest prediction performance with less false ratio and it provides more accuracy.**

***Index Terms*—Learning Type Prediction, Sports Prediction, Learning types, ML, sport factors**

I. INTRODUCTION

The instructive organizations are the primary source in recognizing the sports gifts in any country. The interests of the students are recognized around various sports, with the goal that the sports advancement of any nation can be performed. Any Educational organization would have number of students and every one of them would have number of interest and they include in playing various spots all through their scholar. Recognizing the student with more interest on the particular sports would be useful in distinguishing the specific understudy towards sports advancement. Enormous population countries like India, the talents are covered up all over and mining the sports abilities accessible in various understudies towards sports development is essential.The students are including in various exercises all through their scholastic profession. They would talk about various sports and by simply talking about a sport it can't be concluded that the student is intrigued on unambiguous games. To distinguish a games interest on a student, it is important to consider, the quantity of plays, time spent and the quantity of meetings the client partook in the game, etc. By monitoring their communication being exchanged, the interest of the users can be found efficiently. However, the sports mining can be handled by considering the social data in efficient way in predicting user interest but it lacks in accuracy. The LSPM analysis model consider different factors like the plays, tags, chats, learning style like visualization, Auditory, Kinesthetic, and read-write types include in prediction different sports interest.

II. LITERATURE REVIEW

Some of the mining techniques are analyzed in this section.The problem on sports mining has been

approached with several techniques in history. Few techniques among them are analyzed in this section.

The application of data mining techniques towards school failure detection is analyzed in [1], where the logs of different students around 600 are obtained from school placed in spain. The logs are clustered and decision tree techniques is implemented for classification. The web orient systems are used for analysis and the author present a survey on such systems around the quality in e-learning using various techniques. Similarly, the performances of different educational sectors are examined based on the group of users and environment [3].

Decision tree-based behavior analysis model (DTree-BAM) is presented in [4], which examines the student peception and index them in the tree which has been used for classification. The secondary student's performance is examined using different techjniques namely random forest, SupportVectorMachine, and NeuralNetwork [5]. The failure of students w.r.t to acadamic are examinedd with various data mining techniques in [6], and the students are classified around the state using them.

The interest prediction of students around the subjects is handled with trained mining model in [7], which points the student interest on different subjects to improve the performance of teaching model.

The success prediction of students is handled with regression tree [8], which identifies the success rate of students. Similarly, the performances of students are examined with more techniques in [9].

The Euler K means approach is used for grading of students according to Stiegel manifold scheme. The method produces high quality clustering and grading performance. The firefly algorithm is used for student rating and the multi feature firefly algorithm ((MFF-GSO) uses Group Search Optimizer (GSO) towards clustering.

The performance prediction of telecom industry is approached with efficient classifier in [12], which uses UCI data set. A predictive modeling is discussed towards measurement of student performance [13], which group the students based on the answers given. The K means algorithm is used for clustering and behavior analysis is performed to measure performance.

The extracurricular activities are used in the measurement of student performance which uses decision trees [14]. The method analyzes the performance in implicit manner according to the previous records. An Ant colony orient performance analysis scheme in sports is presented in [15], which generates similarity rules and use them to measure the performance. The records are grouped using K means and the ACO-Kmeans approach predicts the user interest and measure performance.

A multivariant regression prediction model M5P is presented in [16], which works on the basis of decision tree towards student performance analysis.

In [17] Basheer Ahamed suggested a Time Variant Multi Perspective Hierarchical Clustering Algorithm for Predicting Student Interest in Sports Mining. To enhance performance on student sports interest, the time stamp-based clustering examination done here.

In [18] Basheer Ahamed suggested a Sports Distribution Analysis Model (SDAM) as a means of enhancing performance. Using traces of student actions in academic and social domains, the sports distribution analysis model may identify the sports interests of various students. The logs of sports activities are divided into a number of time windows, and the approach estimates the frequency of sports activity for each interest for each of those time windows. The approach performs distribution analysis and measures the elements affecting sports distribution based on these data. The approach predicts interest based on these numbers. There is a lower false positive rate with the SDAM.

The above-mentioned work shows some poor in accuracy and false ratio. This LSPM model is developed to show higher accuracy and less false ratio.

A. Learning Style based Sports prediction model

Learning type Sports prediction model uses the records of various academic years which contain data connected with the student actions on various sports. Each record of the log contains information which include, sports played, sports video watched (visual), time spent on discussing about particular sports (auditory), time spent on learning the sports by well-wisher (kinesthetic), tagged in social media, argument on sports and so on. All these information is extracted by preprocessing the logs and for each of the sport, the method estimates the frequency support.

B. Feature Extraction

The sports hint of the instructive area has been recovered in this part. To start with, the whole follow

has been navigated and from each follow, the technique separates rundown of elements and distinguishes the interesting highlights. Everything the hints of sports log is confirmed for their presence on aspect and worth. The records with missing elements are disposed of from the set. As per the highlights distinguished, the technique divides the whole logs into number of gatherings as indicated by the quantity of time stamp. Further, from each time stamp groups, the technique separates the committed highlights and given for various examination.

C. Pseudo Code of Feature Extraction:

Given: Sports Log Trace SLT

Obtain: [Feature Set Fs, Interest part IP]

D. Start

Read sports Log trace SLT.

The sports log trace would be having the logs generated in different time stamp. It is necessary to monitor the list of time stamps as TSLO (Time Stamp List Order).

$$\text{TSLO} = \int_{i=1}^{\text{size(SLT)}} \sum(\text{TimeStamp} \in \text{Ts(i)}) \cap \text{TSLO}$$

Presently the sports logs ought to be gathered by the time stamps recognized. As per this, the rundown of sports group bunches is produced.

$$\text{SGP} = \int_{i=1}^{\text{size(TSLO)}} \text{Generate Sports Group Sgp(i)}$$

For each sports group, the logs are indexed according to the time stamp.

$$\int_{i=1}^{\text{size(SLT)}} \text{Sgp(k)} \leftarrow \text{SLT(i)? k} == \text{SLT(i).Time}$$

Presently, the logs are gathered by various sports bunch which are parted by various time stamp produced.

Presently, the logs are assembled by various sports bunch which are parted by various time stamp produced. Presently from each log accessible in each sports bunch, the technique extricates various elements like number of times played, all out time spent on play, number of talks, number of videos watched, number of offers, number of arguments with friend's time spent on visit, etc. This large number of highlights are separated and produced as element vector for each time stamp sports bunch. Created highlight vector are utilized to perform examination on various elements.

For each time stamp sports group sgp

Identify list of interests in sports iis = $\int_{i=1}^{\text{size(sgp)}} \sum \text{sgp(i).Interest} \ni \text{IIS}$

For each interest iis

Compute number of plays Ntalk

$$\text{Ntalk} = \int_{i=1}^{\text{size(sgp)}} \sum \text{sgp(i).Interest} == \text{iis \&\& } Sgp(\text{i}).\text{type} = \text{play}$$

Compute time spent on play Pvideo(pv).

$$\text{Pv=pv+}\int_{i=1}^{\text{size(sg)}} \sum \text{sg(i).Interest} = = \text{iis \&\& } Sgp(\text{i}).\text{type} = \text{play \&\& } Sgp(\text{i}).\text{Time}$$

Compute number of visits Nvt

$$\text{Nvt} = \int_{i=1}^{\text{size(sgp)}} \sum \text{sgp(i).Interest} == \text{iis \&\& } Sgp(\text{i}).\text{type} = \text{visit}$$

Compute time spent on visit as NVT.

$$\text{Vtime=vtime+}\int_{i=1}^{\text{size(sgp)}} \sum \text{sgp(i).Interest} == \text{iis \&\& } Sgp(\text{i}).\text{type} = \text{visit \&\& } Sgp(\text{i}).\text{Time}$$

Compute number of chats cn = $\int_{i=1}^{\text{size(sgp)}} \sum \text{sgp(i).Interest} == \text{iis \&\& } Sgp(\text{i}).\text{type} = \text{chat}$

Compute number of shares sn = $\int_{i=1}^{\text{size(sgp)}} \sum \text{sgp(i).Interest} == \text{iis \&\& } Sgp(\text{i}).\text{type} = \text{share}$

Feature Vector Fv = [Ntalk, Pv, Nvt, Vtime, cn,sn]

Add to feature set.

End

End

Stop

The feature extraction algorithm represents the way how the feature set is generated towards analysis of sports interest to predict the sports interest.

The component extraction calculation addresses the way how that the list of capabilities is created towards analysis of sports revenue to anticipate the sports revenue.

E. Sports Frequency Analysis:

The sport recurrence analysis is the most common way of breaking down the sports interest as per the recurrence of event of the sports in various ways like play,watch, offer and agumentt. The specific interest would be played, shared, chat, and watching videos in different time stamp. However, the frequency of the activity belongs to the specific sports vary on different time stamp. By considering this, analyzing the frequency of sport would help to predict the sports interest. It is performed by computing the

frequency measure based on the different features extracted in the feature extraction stage.

The frequency of playing the sports is measured as follows:

$$FreqP = \sum_{i=1}^{size(Sgp)} Sgp(Fv(Ntalk)) \Big/ size(sg)$$

The visit frequency is measured as follows:

$$FreqV = \sum_{i=1}^{size(Sg)} Sg(Fv(NVis)) \Big/ size(sg)$$

The social frequency is measured as follows:

$$FreqS = \frac{\sum_{i=1}^{size(Sgp)} Sg(i).Nc + Sgp(i).Ntalk + Sgp(i).Ns}{size(sgp)}$$

$$\text{Frequency Interst Support FIS} = \frac{FreqP \times FreqV \times FreqS}{3} \times \frac{\sum_{i=1}^{size(Sgp)} Sgp(i).Ptime + Sgp(i).vtime}{size(sgp)}$$

The estimated interest frequency support measure is used in the prediction of user sports interest to perform sports mining.

The sports distribution analysis is the process of analyzing the distribution of the sports interest in different time stamp. The method maintains the sports logs of different time stamp, the method extracts the features of the log and measure the frequency support of various sports interest according to the frequency analysis. Based on the frequency analysis and the frequency threshold used by the method, the method analyzes the distribution of the sports in different time stamp. To perform the analysis, the method computes the interest distribution support measure which is computed according to the frequency support and threshold. Based on the result of sports analysis, the method performs interest prediction.

Pseudo Code of Learning Style based Sports Distribution Analysis:

Given: Sports Log Trace SLT, Student ST

Output: IDSs

Start

Read SLT, ST.

[Feature Set Fs, Interest Set IS] = Feature Extraction [SLT, ST]

For each interest I

IFS = Frequency Analysis (Interest,Fs)

Compute Learning Interest Distribution Support IDS.

$$LIDS = \frac{\sum_{i=1}^{size(Sgp)} I(Sgp(i)).IFS > Th}{size(Sg)}$$

Add to LIDS set LIDSs = ∑ Ids ∪ LIDS

End

Stop

Learning Style based Sports Prediction Analysis algorithm explains how the learning interest distribution analysis is performed. It has been measured according to the frequency and the distribution of the support in different time window records. Assessed sports appropriation support measure towards interest forecast.

F. Interest Prediction

The interest forecast on sports towards sports mining is performed here. The strategy peruses the sports follow created and kept up with by the timestamp. From the follow, the technique produces the sports bunch in light of different time stamp records created by past activities Extracted features are used to perform sports frequency analysis and learning Style based distribution analysis.

III. RESULTS AND DISCUSSION

Learning Style based Sports prediction model approach is implemented in java. The method focused on different factors for analysis. Similarly, the same is applied to different techniques.

Parameter	Value
No of Sports	5
Number of Sub class	3
Number of Students	5000
Number of Time window logs	2 years

Table 1: Evaluation Details

The key and value used for the performance evaluation of LSPM model is presented in Table 1. The students are classified in 3 levels and one year data has been collected. In total, the logs of 2000 students are considered for evaluation

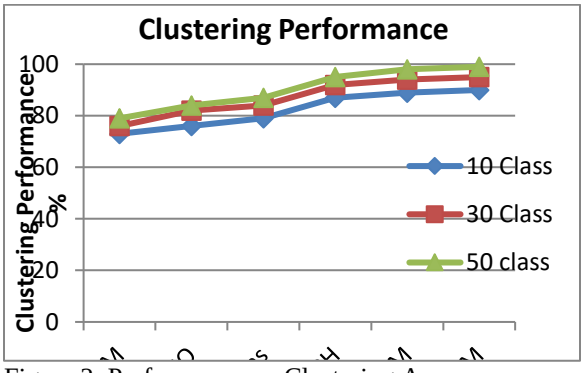

Figure 2: Performance on Clustering Accuracy

The accuracy of clustering generated by LSPM algorithm is measured and presented in Figure 2, where the LSPM approach indicates the higher accuracy achievement

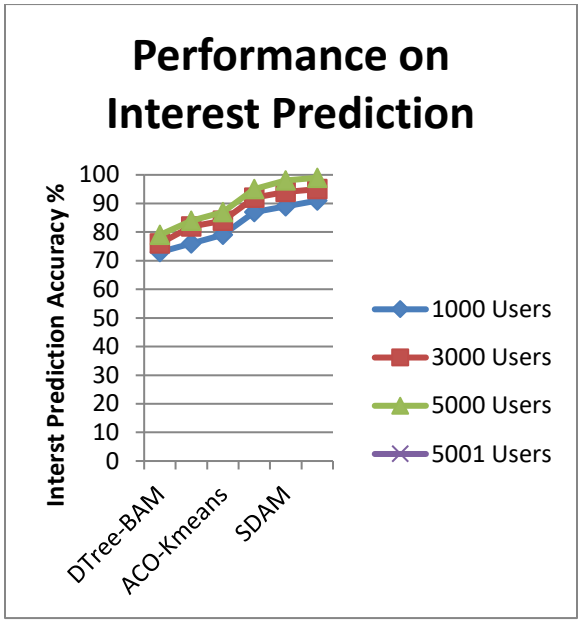

Figure 3: Interest Prediction Accuracy

The accuracy on interest prediction is measured for LSPM and plotted in Figure 3, where LSPM has marked the higher accuracy in interest prediction.

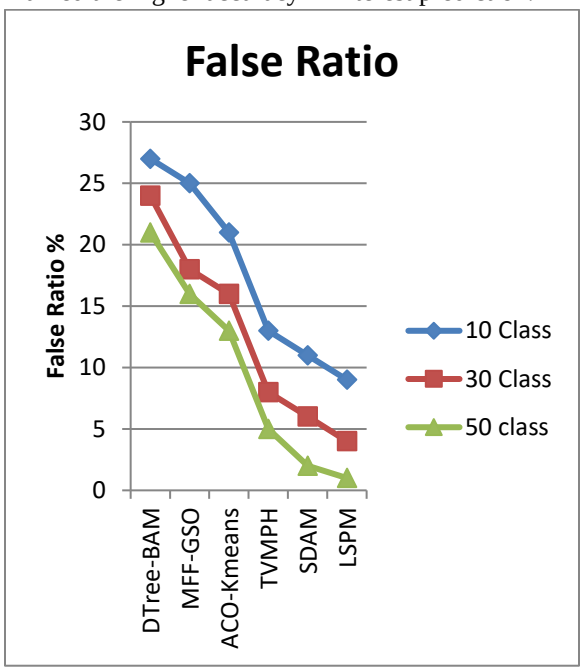

Figure 4: false ratio in interest prediction

The ratio of false interest prediction is measured for LSPM approach and plotted in Figure 4, which produce the lower rate of LSPM

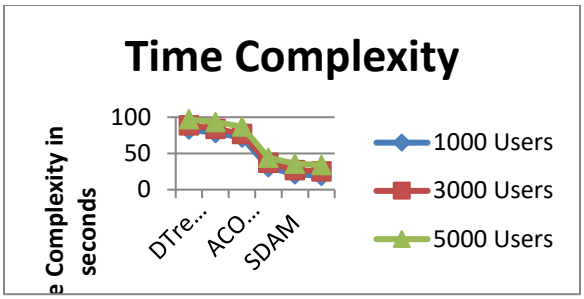

Figure 5: Time complexity

The value of time taken by the SDAM in interest prediction is measured and plotted in Figure 5, which denote the LSTM has taken less time.

IV. CONCLUSION

This research work focused on a Learning Style based Sports Prediction (LSPM) explains towards interest prediction of students and sports mining. The sports logs of various students generated in different time are obtained and grouped according to the time stamps identified. The clustered results are used to extract the features and perform Sports frequency and distribution analysis on various interests. According to the result of interest distribution analysis, the method performs interest prediction. Further, sports frequency analysis and interest distribution analysis are performed. The IDS measure is used to predict the student interest. The method produces higher performance in clustering and interest prediction with less false ratio.

REFERENCES

[1] Carlos Marquez-Vera, et. Al, "Predicting school failure and dropout by using data mining techniques". IEEE Journal of Latin-American Learning Technologies, Vol. 8, No. 1, February 2013.

[2] C. Romero and S. Ventura, "Educational data mining: A survey from 1995 to 2005," Expert Systems. Appl., vol. 33, no. 1, pp. 135-146, 2007.

[3] C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Trans.Syst., Man, Cybern. C, Appl. Rev., vol. 40, no. 6, pp. 601-618, Nov. 2010.

[4] V.P. Bresfelean, "Analysis and Predictions on Students' Behaviour using Decision trees in

WEKA Environment", Proceedings of the ITI 2007 29th Int Conf. on Information Technology Interfaces, June 25-28, 2007.

[5] P.Cortez and A.Silva, "Using Data Mining To Predict Secondary School Student Performance", In EUROSIS, A. Brito and J. Teixeria (Eds.), 2008, pp.5-12.

[6] P. Bresfelean , et. Al "Determining Students' Academic Failure Profile Founded on Data Mining Methods", Proceedings of the ITI 2008 30th International Conference on Information Technology Interfaces, June 23-26 2008.

[7] B.K. Baradwaj, S. Pal, "Mining Educational Data to Analyze Students' Performance", (IJACSA) International Journal of Advanced Computer and Application, Vol. 2, No. 6, 2011.

[8] Z.J.Kovacic(2010),"Early Prediction of Students Success: Mining Student Enrolment Data", paper presented at proceedings of Informing Science & IT Education Conference(InSITE), casino Italia,June,19-24-2010

[9] Saima Sayyed, Rugved Deolekar, A Survey on Novelty Detection using Level Set Methods ICICCT,.2017.

[10] Jian-Sheng Wu, Wei-Shi Zheng Euler Clustering on Large-scale Dataset IEEE 2017.

[11] Latha Parthiban, Multi objective hybridized firefly algorithm with group search optimization for data clustering Golda George, IEEE ,2016.

[12] Ahmed, M., et. Al Mcs: Multiple classifier system to predict the churners in the telecom industry .2017.

[13] Alana M. de Morais, et. Al " Monitoring Student Performance Using Data Clustering and Predictive Modelling " IEEE,2014.

[14] Tomas Hasbun, et. Al "Extracurricular activities as dropout prediction factors in higher education using decision trees"IEEE 2016.

[15] Jian, et. Al "Clustering Analysis of Sports Performance based on Ant Colony Algorithm"IEEE 2016.

[16] S Chaitanya Kumar, et. Al "M5P Model Tree in Predicting Student Performance: A Case Study" IEEE,2016

[17] Ahamed, A.B. and Surputheen, M.M., Time Variant Multi Perspective Hierarchical Clustering Algorithm for Predicting Student Interest in Sports Mining. International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-4, PP-7313-7317, November 2019.

[18] A Basheer Ahamed, M. Mohamed Surputheen. Real Time Sports Distribution Analysis Model for Improved Student Interest Prediction in Sports Mining. international Journal of Advanced Science and Technology Vol. 29, No. 5s, (2020), pp. 513-522.

[19] Jayasree, R., & Selvakumari, S. (2023). Design of a Prediction Model to Predict Students' Performance Using Educational Data Mining and Machine Learning. This study introduces a model utilizing Back Propagation Neural Networks (BPNN) and compares its accuracy against Support Vector Machines and Decision Trees, showing enhanced predictive performance in academic settings (Eng. Proc., 2023).

[20] MLSA 2023 Proceedings on Machine Learning for Sports Analytics. Published in 2023, this collection explores various machine learning applications in sports, including predictive analysis for student engagement in sports activities.