

Study and Analysis of Data Mining Algorithms for Identifying the Students' for Psychology Motivation

S. Peerbasha¹ and M. Mohamed Surputheen²

¹Research Scholar, ²Associate Professor

^{1&2}Department of Computer Science, Jamal Mohamed College, Tiruchirappalli, Tamil Nadu, India
E-Mail: bashapeer2003@gmail.com

Abstract - The development of many educational institutions is based on the performance of students learning and understanding capabilities. Here, we analyzed their academic profile with their grades and various cumulative attributes. The academic performance in learning their subjects could be improved by motivational approach. The analysis of student performance is carried out through knowledge-based data mining process. But, the problem is arrived by a probability of information prediction accuracy from student data set which is not accurate. Here, we propose a novel machine learning algorithm based on subspace clustering and multi-perspective classification techniques to identify psychological motivation required students. Also, the extraction of relational patterns to form enhanced clustering classes is done. This discovers the innovative relations between students and their educational performance in the various attributes using surf scale nested clustering approach based on an intelligent predicting system from soft computing processing tasks. This improves the data prediction rate by considering the time factor analysis and complexity to design and develop an efficient clustering algorithm which maximizes the clustering and classification accuracy for improving academic performance.

Keywords: Knowledge Mining, Prediction, Cluster, Educational Data Mining, Classification

I. INTRODUCTION

Educational data mining is a significant role for prediction of information and relational data that are useful for student's performance analysis. Data mining approaches will classically analyses the importance of the information process in the various fields of educational factors like student activities, competitions, talents, etc. Machine learning techniques carries out the performance analysis by understanding complementary role which is nothing but semantic relations in educational data mining. Cluster analysis method is one of the leading analytical methods in data mining. It will influence the clustering results directly. Intra-Micro subspace clustering algorithm avoids computing the distance of each data object to the cluster centers again and again and save the execution time. Formally, given a set of dimensional points and a function that gives the distance between two points for weight age analysis required to compute cluster centers, such that the clustering points falling in the same clusters are similar and positions that are in the different group are dissimilar.

The supervised learning approach formed by Classification of data in which students are grouped into identified classes. Classification rules may be determined from a part of data known as training data. It mainly focuses on analyzing and understanding students' educational data that indicates their educational performance, and generates specific rules, classifications, and predictions to help students in their future educational performance.

Data analysis can be done using classification and prediction, and it helps in dividing the data into classes by which we can easily predict the future trends.

1. Classification helps in putting labels on the data and in prediction helps in forecasting the value of data.
2. There are various techniques of classification compared with our novelty like decision tree classifier, rule-based classifiers, Bayesian classifiers, support vector machine, k-n-n classifier are well classified with the trained dataset with optimization

A. Problematic Consideration

Mostly Relational measures are based on the nearest representation of the identified dataset. There is inconsistency in the clustering accuracy, and many rules produced by decision algorithms lead to timing consumption. Partial classification algorithms are the fastest ones, but many make to construct the classification system due to make noise such methods in building the rules. To identify the significant variables that affects and influences the performance of undergraduate students.

1. High dimensional datasets process
2. Relational analysis –inconsistent attribute
3. Frequent dataset – Clair dataset
4. Data Prediction – dissimilarity observation
5. Data redundancy – Terminating complex data
6. Midpoint missing values carry cluster reviews to similarity analyses.

The high dimensional categorical dataset is challenging to cluster therefore a fully automated algorithm has been proposed. Academic performance of the student is not a result of only one deciding factor besides it heavily hinges

on various factors like personal, socio-economic, psychological and other environmental variables.

3. To improve the clustering accuracy, design an intra-micro subspace clustering to optimize the clustering weight age resultant accuracy.

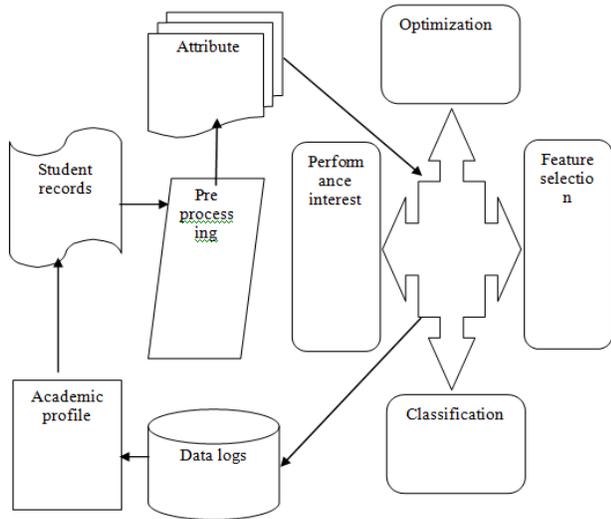


Fig.1 Prediction of Performance of student

The accuracy level is low due to process the higher dimensionality datasets. Due to lacks of redundant problems datasets doesn't give resultant of real-world decisions.

B. Objective Consideration

1. To implement a different clustering and classification technique for relational analysis in high dimensional datasets.
2. To intent, a surf scale nested clustering (SSNc) algorithm to measure the relative closeness of data points.

II. DISCUSSION

To identify the students' traits, the manual extraction of student's records would be a huge task. Considering a higher number of parameters will help to predict more detailed information about the student. Evaluating the weakness and strengths of the individual student and analysis on effort given by the teacher to an individual student improves the performance of the educational institution.

Instead of considering only academic parameters, we could also consider personal characteristics, behavior, family history, records to predict the performance of a student [8]. The whole procedure can be pictorially depicted in fig. 2.

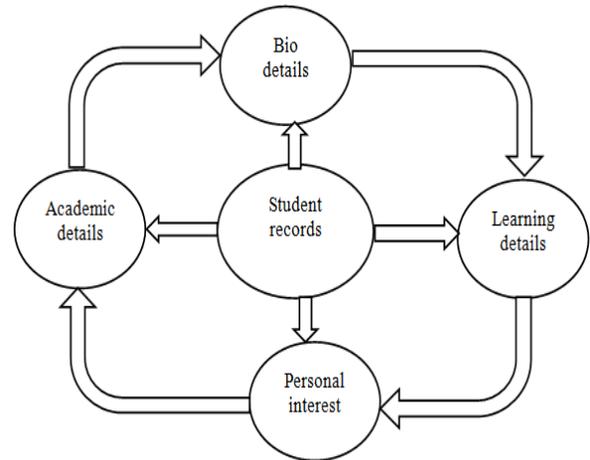


Fig. 2 Attribute consideration in student records

TABLE I STUDENT ATTRIBUTED IN EDUCATIONAL DATASET

Attribute	Description	Possible Values
GENDER	Student's gender	{Male, Female}
NAT	Nationality	{Indian, Gulf, Arab, Non- Arab}
FLANG	First Language	{Arabic, English, Hindi, Urdu, Other}
TEACHLANG	Teaching language S	{English, Arabic}
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%)}
STATUS	Student status depending on his/her earned credit hours	{Freshman (< 32), Sophomore (33 - 64), Junior (65 - 96), Senior (> 96)}
LOC	Living Location	{Ajman, Sharjah, Dubai, Abu Dhabi, Al-Ain, UAQ, RAK, Fujairah, University Hostel}
SPON	Sponsorship	{Yes, No}

TABLE II RELATIONAL ATTRIBUTES CONSIDERATION

Attribute	Range
GPA	Very Good (81), Good (68), Pass (61), Excellent (60)
GENDER	Female (174), Male (96)
STATUS	Freshman (109), Sophomore (62), Junior (53), Senior (37)
NATCAT	Arab (180), Other (34), Gulf (29), Local (23), Non-Arab (4)
FLANG	Arabic (233), Other (18), Hindi-Urdu (16), English (3)
TEACHLANG	English (248), Arabic (20)
LOC	Ajman (123), Sharjah (90), Dubai (18), University Hostel (13), RAK (11), UAQ (10), Abu Dhabi (3), Fujairah (1), Al-Ain (1)
TRANSPORT	Car (175), University Bus (54), Walking (21), Public Transport (20)
HSP	Excellent (100), Very Good (High) (63), Very Good (50), Good (High) (33), Good (19), Pass (High) (4), Pass (1)
PWIU	No (262), Yes (8)
DISCOUNT	No (186), Yes (84)
SPON	No (210), Yes (60)
FRIENDS	Average (81), High (75), Medium (67), Above Medium (27), One (13), None (7)
WEEKHOURS	Average (122), Very limited (57), Medium (40), High (21), Very High (16), None (14)

Our initiated proposal objective based on improves the clustering accuracy for considering problematic issues. To improve the clustering performance, we present an efficient surf scale nested clustering algorithm.

In existing approaches, the method estimates the cluster similarity in an incremental manner by increasing the number of dimensions or features in estimating the cluster similarity between data points.

The surf scale nested clustering (SSNC) method avoids computing the distance of each data object to the cluster centers again and again and save the running time.

Surf scale nested clustering structure to predict data in every iteration, which is to be used in the next iteration for semantic relation .This relevantly measures the semantic similarity closeness of cluster group.

A. Intra -Micro Subspace Clustering Technique (IMSc)

The resultant classes are from surf space nested Clustering technique is used to place data elements into micro subspaces clusters in neural network classification for related groups without advance knowledge of the group description.

The intra -micro subspace clustering technique (IMSc) belongs to an unsupervised learning and it is used to discover a new set of categories for.

Proposed clustering groups the data instances in to micro subsets in such a manner that higher weight age similar instances are grouped together while different instances

belong to different groups for student performance prediction.

Such the classifications are highly improved by multi-perspective subsequent neural networks (MPSNc) for improving the performance from student datasets.

III. COMPARATIVE ANALYSIS

Educational data mining field concentrate on prediction more often as compare to generate exact results for future purpose.

Statistics are generated based on all classification algorithms and comparison of all five classifiers is also done in order to predict the accuracy and to find the best performing classification algorithm among all [1].

Currently in Malaysia, the lack of existing system to analyze and monitor the student progress and performance is not being addressed. [2].

In each area of business, there is an increasing urge to extract the knowledge from data in a timely manner in order to be able to make shifts that ensure a competitive advantage.

Thus, the knowledge of methods and techniques of big data processing is currently a tool of a modern economist. [3]. The following are the techniques prescribed by the various authors and the limitations are as follows.

TABLE III TECHNIQUES USED BY VARIOUS AUTHORS

Title	Author(s)	Year / Journal	Technique	Debate/limitations
Classification and prediction based data mining algorithms to predict slow learners in the education sector	ParneetKaura, Manpreet Singh,	2015/ICRTC	Multilayer Perception, Naïve Bayes, SMO, J48 and REPTree using WEKA an Open source tool	Some factors that are considered to influence the performance of the dataset identified.
A Review on Predicting Student's Performance using Data Mining Techniques	Amirah Mohamed Shahira,*, WahidahHusaina	2015/ICRTC	Decision Tree, Neural network, Naive Bayes	Decision Tree and Neural Network are the two methods highly used by the researchers for predicting students performance not supportive to large-scale datasets
A survey on novelty detection using level set Methods with Adaptive Boundaries	Xuemei Ding, Yuhua Li, A. Belatreche and Liam P. Maguire,	2013/IEEE	Focuses on the Identification of shapes or patterns, space and learning along With density taxonomy models.	Mining is profoundly affected by the absence. Unknown data extraction process be non related. The considerable dataset that has been left during the classification process
Machine learning techniques in the education process of students of economics	J. Bucko*, L. Kakalejčik*	Mipro 2017	Machine learning Techniques, fundamental assumptions of decision support systems	The difficulty of the course content. Irrelevant predictions of the relations.
Application of data mining classification techniques On soil data using R.	Nikhita Awasthi, Abhay Bansal	2017/ijaec	Artificial neural network and Support vector machine.	Mining analyzed and explored to discover some unknown facts, patterns, classifications, and correlations.
Educational Mining: A Comparative Study of Classification Algorithms Using WEKA	Biriyani Rajeshinigo 1, J. Patricia Annie Jebamalar	2017/IJIRCCE	Decision trees, Naive Bayes classifier, random forest, Multilayer Perceptron and Support vector machine	Be giving lower accuracy among the algorithms compared. The C4.5 algorithm provided lower accuracy because of the use of continuous data as the attribute values
Clustering Algorithms Applied in Educational Data Mining	Ashish Dutt, Saeed Aghabozrgi	2017/IJEE	Expectation Simple k-Means and X-Means	Sets to reveal the beforehand concealed data to important. ata that could be utilized for both vital and also learning picks up.
A Survey on Data Mining and Its Current Research Directions	Er. Anita Devi, Er. Jasjeet Kaur	2017/IJARC	Association, correlation, clustering and neural network. It is also conducting a formal review of the application of data mining such as the education sector	Provide positive and appropriate results thought which more and more appropriate and valuable information can be found out.

Every second peta bytes of data are getting generated by social media, commercial websites, surveys, telecommunications, etc. and it needs to be analyzed and explored to discover some unknown facts, patterns, classifications and correlations [4]. This provokes data classifiers as regulated learning means with the capability of creating intelligible output [5], generally utilized as a part of data mining to think about the data and produce the tree, and it decides that will be used to define forecasts.

The other calculation utilized, for gathering students is K-means clustering which doles out an arrangement of perceptions into subsets [6]. In the field of scholastics, data

mining can be exceptionally helpful in finding valuable data which can be utilized for profiling students in light of their scholarly record. Use of Apriori calculation for student profiling which is one of the prominent methodologies for mining affiliations [7], i.e., finding co-relations among the set of things.

To identify learning troubles right on time, to assist students with particular learning incapacities, to enhance the administration of learning objects [8]. The discoveries from this examination add to the educational network's comprehension of students' commitment and execution in MOOCs [12] and furthermore furnish the more extensive

learning investigation network with proposals. Consider fuzzy c-means (FCM) as a calculated and algorithmic setting to manage the issue of abnormality detection [11]. Utilizing a sliding window, the time arrangement is isolated into various subsequences, and the accessible spatiotemporal structure inside each time window is found utilizing the FCM strategy

Different endeavors per question-related adversely with student achievement [13]. A different direct relapse show and different rules are displayed for teachers'. The pretest-posttest instrument uncovers measurably noteworthy picks up in information for each course area and all areas consolidated [14]. The outcomes show that pretests can set up students' earlier information toward the start of the semester, while posttests measure learning toward the finish of the course. The procedure of extraction of complex and new data from the large dataset that has been left amid the classification procedure is known as oddity discovery.

IV. CONCLUSION

Predicting student's performance is mostly used to help the educators and learners improving the student performance. This research reviewed previous studies on predicting students' performance with various analytical methods. By calculating surf scale based clustering and Intra -micro subspace clustering to reduce the leads consumption services to improve the clustering and classification accuracy to produce the best results for learning and teaching process. Finally our proposed method intra micro subspace–multi-perspective neural classifier produces higher prediction accuracy which is required for motivation. While for prediction techniques, the classification method is frequently used in educational data mining area for psychology improvement. In conclusion, the intra micro subspace clustering with multi-perspective classifier predicts the optimized result that has motivated us to carry out further research to be applied in our environment. It will help the educational system to monitor the students' performance in a systematic way and supportive o real word entities.

REFERENCES

- [1] Mandeep Kaur and Vimal Dev, "A Review on Performance Prediction of Students using Data mining", *Journal of Advanced Research in Information Technology, Systems & Management*, Vol. 2, No 3&4, 2017.
- [2] Amirah Mohamed Shahiri, Wahidah Husain and Nur'aini Abdul Rashid, "A Review on Predicting Student's Performance using Data mining techniques", *Procedia Computer Science*, Vol. 72, pp. 414 – 422, 2015.
- [3] J. Bucko and L. Kakalejcl, "Machine learning techniques in the education process of students of economics", *IEEE, MIPRO 2017*, pp 22-26, 2017.
- [4] Nikhita Awasthi and Abhay Bansal, "Application of Data mining classification techniques on soil data using R", *International Journal of Advances in Electronics and Computer Science*, Vol. 4, No. 1, Jan. 2017.
- [5] D. Rajeshinigo and J. Patricia Annie Jebamalar, "Educational mining: A comparative study of classification algorithms using WEKA", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 5, No. 3, March 2017.
- [6] Dutt, S. Aghabozrgi, M.A. Binti Ismail and H. Mahroean, "Clustering algorithms applied in educational data mining", *International Journal of Information and Electronics Engineering*, Vol. 5, No. 2, March 2015.
- [7] Er. Anita Devi and Er. Jasjeet Kaur, "A Survey on Data mining and its current research directions", *International Journal of Advanced Research in Computer Science*, Vol. 8, No. 4, May 2017 (Special Issue).
- [8] Ivon Arroyo and Beverly Park Woolf, "Inferring learning and attitudes from a Bayesian Network of log file data", *Amsterdam: IOS Press*, pp. 33-40, 2005.
- [9] Cristobal Romero and Sebastian Ventura, "Educational Data Mining: A Review of the State-of-the Art", *IEEE transactions on Systems, Man and Cybernetics- Part c: Applications and Reviews*, Vol. 40, 2010.
- [10] V. Ramesh, P. Parkavi and K. Ramar, "Predicting Student Performance: A Statistical and Data Mining Approach", *International Journal of Computer Applications*, Vol. 63, No. 8, Feb. 2013.
- [11] Hesam Izakian and Witold Pedrycz, "Anomaly Detection in Time Series Data using a Fuzzy c means clustering", *IEEE Xplore*, 26 Sept. 2013.
- [12] Coffrin, L. Corrin, P. De Barba and G. Kennedy, "Visualizing patterns of student engagement and performance in MOOCs", in *Proc of the learning analytics and knowledge, ACM*, pp. 83-92, 2014.
- [13] Hesamizakian and Witoldpedrycz, "Anomaly detection and characterization in spatial time series data: a cluster-centric approach", *IEEE transactions on fuzzy systems*, Vol. 22, No. 6, December 2014.
- [14] R. Bowman, O. Gulacar, and D. B. King, "Predicting student success via online homework usage", *Journal of Learning Design*, Vol. 7, No. 2, pp. 47-61, 2014.