International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

A Novel Educational Data Mining Model using Classification Algorithm for evaluating Students' E-learning Performance

S. Arumugam[1*], A. Kovalan[2], A.E. Narayanan[3]

[1]Dept. of Computer Science and Applications, Periyar Maniammai Institute of Science and Technology, Thanjavur, India
[2]Dept. of Computer Science, VSB College of Arts and Science, VSB Group of Institutions, Karur, India
[3]Dept. of Computer Science Engineering, Periyar Maniammai Institute of Science and Technology, Thanjavur, India

Corresponding Author: arumugammca@pmu.edu

DOI: https://doi.org/10.26438/ijcse/v7i5.616624 | Available online at: www.ijcseonline.org

Accepted: 23/May/2019, Published: 31/May/2019

***Abstract* -** It is possible to assess the learning behavior in online systems or computed based learning backgrounds using data mining techniques on 'e-learning session activity log data' from different learning sessions. This information is very useful to improve the e-learning system better. There is a possibility to identify learners' performance well before the conduction of an examination. The objective of the research is to find weather it is possible to apply Data Mining techniques on this transformed dataset and to predict some information. The educational dataset is used for analyze and also improve any e-Learning models. This research work proposes an Educational Data Mining (EDM) model which provides good performance with precision, recall and f-score. It shows the predictability of students' grades by mining the e-learning session log data.

Keywords: E-learning, Learning Analytics(LA), Technology Enhanced Learning(TEL), Educational Data Mining(EDM).

I. INTRODUCTION

E-Learning is a type of learning to describe using computer, web or electronics based learning and teaching. This learning is done inside the class room as well as away from the class room.

Data Mining (DM) is the process of sorting through large data sets to identify patterns and establish relationships to solve problems through data analysis.

Learning Analytics (LA) helps us to measure, gather, study and result of data about learners and their backgrounds. It is used to understand the effectiveness of learning sources and also the environments. In the last decade, LA and EDM elevated a lot of awareness in this research area [1].

To obtain valuable knowledge from a process is known as Process Mining (PM). From event logs process oriented knowledge would be extracted by applying some PM techniques like process model discovery and conformance checking[1]. Having focus on DM methods adopted in educational circumstance, applying the common methods and applications on the collected data is an important part of the LA and EDM evaluation [2].

Educational Data Mining (EDM)

The authors of paper [1] captured students' time series of activities during six e-learning sessions of laboratory sessions and analyzed its Cyclomatic Complexity and found some correlation between Cyclomatic Complexity of the sessions and Final Grades. In [3], the authors tried to found the relationship between the cyclomatic complexity numbers of the predictable programming problems answered by the students during the assessment period of a programming course with the grades they have obtained. It was proposed as a new data transformation method to transform this time series data in to simple, numerical dataset [4]

EDM is a technique to extract knowledge educational dataset. The information recorded by the student, teacher and the system are analyzed. The main aim is to find prototype of the system. Another objective is to discover the students' learning behavior patterns. Further, the authors say that EDM is the process of converting raw data to meaningful information in the educational system. These information help students, teachers, and educational designers.

The Techniques Used for EDM

Ryan Baker categorized the types of EDM tasks as follows: [5]

➢ Prediction
- Classification
- Regression

- Density estimation
- ➢ Relationship mining
- Association rule mining
- Correlation mining
- Sequential pattern mining
- Causal data mining
- ➢ Distillation of data for human judgment
- ➢ Clustering

The available dataset to be preprocessed to apply data mining techniques. The purpose is to retrieve the knowledge from the dataset. Statistical analysis like correlation, regression can also be applied to mine the knowledge to predict the result. The complex data mining techniques like classification, clustering, association rule mining, pattern mining, text mining etc., can also be implemented to foresee the result of the students.

In [6], the authors highlighted that several studies had been demonstrated to apply the DM techniques that could fruitfully be implemented into E-learning atmospheres.

In [7], the authors present how the data mining technique is used to find the mistakes that they do while answering the problems.

In [8], the authors survey the role of association rule mining and its application in EDM Learning Management Systems (LMS). The drawbacks of knowledge discovery process are studied and also some possible way out to overcome those drawbacks are mentioned.

In [9] the authors show how to provide effective e-learning system using data obsessed approaches. This was presented using a case study of an e-learning project

In[1], the authors proposed a Learning Analytics approach for understanding the learning behavior of students. They have taken the interaction Technology Enhanced Learning(TEL) tools. This work shows that from communication data of the students, the learning process can be found. They have done their study with the data set of I year Computer Engineering students laboratory classes. They made use of PM methods to investigate and compare the learning procedures of the students. The authors measured the understandability of the student process models through a complexity metric. Finally it was matched up to the academic performance of grouped students. They concluded that the complexity with the result is positive correlation and with the course it is a negative correlation.
The results of that study [1] have been summarized into four main conclusions:

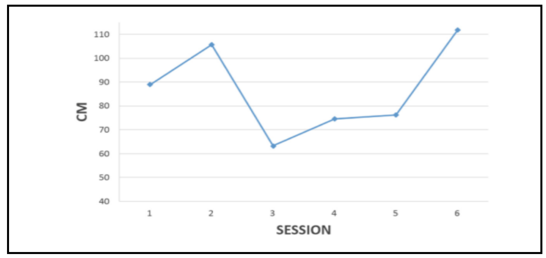
➢ The average Cyclomatic Complexity of McCabe (CM) of average of students intermediate grades is having negative correlation. Hence, CM helps to predict the difficulty of a session for students.

➢ The average CM of high-graded students is calculated. It is higher than the one of low-graded students. Hence, CM explains the interactive behavior of students.

➢ CM elucidates the resemblance of student behaviors at granularity level.

➢ CM can give indication to find the interactive and cognitive behavior of students which helps to improve the educational model.

The following two graphs from that previous study clearly explain one important nature that is if complexity of the session is high then the average final grades will be low.



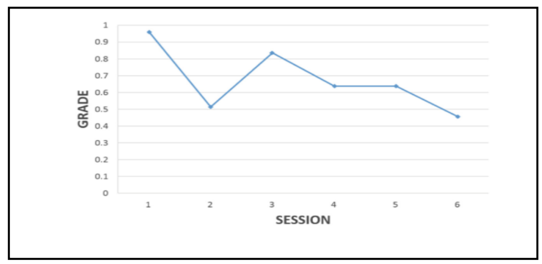


Figure 1. AverageofCyclomaticComplexity(CM)persession(top),averageofintermediategradespersession(bottom) – Results from [1]

The results explained by CM in the previous research [1] are somewhat incomplete. It reveals only the following facts:
1) Difficulty of a e-learning session is directly proportional to complexity of the session
(In terms of CM),
2) Grades are inversely proportional to complexity of the session (in terms of CM).
But, there are possibilities to discover much unknown knowledge from this e-learning session log data if it is applied by using suitable data mining techniques. This is the main problem which is addressed in this work.

In previous works, the complexity of course, curriculum design have been analyzed. Prediction of the students performance from their activities is quit reasonable and interesting, that too from the time series data.

In this paper contains introduction in section I, In section II, the algorithm used to find the Educational Data Mining task. Section III explains the proposed model to implement the algorithm. The implementation and evaluation techniques are well explained in section IV. In section V, the results discussed and interpreted. Section VI concludes the findings and discusses the scope of future works.

II. ALGORITHMS USED IN EDM TASK

Cyclomatic Complexity

Cyclomatic complexity is software metric, used to indicate and measure the complexity of a program. It was developed by Thomas J. McCabe, Sr. in 1976 [wiki] and commonly referred as Cyclomatic Complexity of McCabe (CM). Apart from applying it on a computer software code, we can apply it on any such complex systems.

In [1], Cyclomatic Complexity is used as an Indicator of Student Behavior. To compute CM it is given G = (V, E) as a control-flow graph.

E = the number of independent paths
V = the number of activities
The Cyclomatic metric can be calculated using:

$CM = E - N + 2P$, where

E = the number of edges of the graph.
N = the number of nodes of the graph.
P = the number of connected components.

J48 Decision Tree Classifier

To classify the data the J48 Decision tree classifier is used. This algorithm creates a decision tree using the attribute values in the dataset. It discriminates the instances clearly when it runs into the given dataset. This feature helps us to have the highest information gain. In the possible values of this feature, which does not have ambiguity, that is, for which the data occurrence of the same group have the same value of the final variable. It terminates that branch and assigns it to the target value.

The authors of [10] explained various possibilities of the Data Mining algorithms and their uses in the analyses of the education data. They are

1. Classification algorithms – To predict discrete variables
2. Regression algorithms – To predict continuous variables
3. Segmentation algorithms – To divide data into groups, or to apply clusters properties.
4. Association algorithms – To find correlations between the attributes in a dataset.
5. Sequence analysis – To summarize sequences or episodes

Students' performance is not required to depend only on their academic performance, instead some other factors that too have equal and more influences. It was explained in [11], by applying the data mining techniques and finding the results would help the Institution and the students for their betterment.

Bharadwaj and Pal concluded in [12], classification techniques could be applied on the dataset to envisage the result. They applied decision tree algorithm to predict the performance of the students. To apply this method they have taken the students information like attendance, assignment, test and seminar marks from the dataset.

It was explained that the prediction of performance of the students are influenced by the factors or attributes. In [13], the authors attempted to examine the attributes of the other students which could affect the prediction process. They suggested that this would help to provide better accuracy on the results.

In [14], the authors explained the role of WEKA tool. It was considered as a landmark system in data mining and machine learning. Using this tool the source code can be freely accessed to develop and facilitated the creation of projects. In that paper, WEKA tool was used to analyze the classification algorithms and predicts the students' performance.

Many classification algorithms help us to predict student's performance [15]. It is tabulated and compared various algorithms to expect the result of the students.

To extract the knowledge from the raw data, the machine learning techniques and the data mining concepts help us lot. This result provides us interesting possibilities from the education domain [16]. In this paper, the authors used ID3 Classification algorithm to do the analysis.

Educational domain data cannot be taken for analysis directly. Instead, it should be converted as to apply the data mining algorithms. EDM techniques help to get better in e-learning design.

In [17], authors tried to implement an algorithm to classify huge log data. And resulted that feature based auto recommendation algorithm is suitable.

In [18], K-Nearest Neighbor algorithm used to classify student databases to predict their salary from two different databases.

III. THE PROPOSED MODEL OF EDM SYSTEM

Educational Process Mining (EPM), Dataset of UCI

This is a Learning Analytics Data Set which has been offered to University of California, Irvine (UCI). It is a machine learning repository by the people of Non-Linear Complex Systems Laboratory, Italy and Department of Industrial Design, Eindhoven University of Technology, Netherlands. It contains non-liner time series data. The attributes of the dataset are selected and converted into a desirable design that will be suitable for complex Process Mining tasks. But, applying a simple clustering and classification algorithm require processing of the data further.

The dataset is taken from a group which contains 115 Engineering students of I year from University of Genoa. DEEDS (Digital Electronics Education and Design Suite) is a e-learning tool in digital electronics . The course content is available to the students, and they were asked to solve various problems with various levels of complexity.

In the previous work [4], the data set contains the time series dataset of students activities during six sessions of laboratory of digital electronics course. It contains the students' data of each session. It has around 100 CSV files, which contain students log of the students laboratory sessions. This has 13 attributes. files each dedicated to a specific student log during that session. Each file contains 13 features. The authors measured the average performance using J48 classification method by measuring of Precision, Recall and F-Score. This study finds out the poor grade and high grade students from their e-learning behavior.

Proposed Model for EDM

The following diagram explains the proposed model of the educational data mining system. The UCI EPM data of activity logs is in the form of text file and students grades are in excel files. In this following model, the text and excel format data is converted in to a simple multidimensional text data and each time series data of 6 sessions will be converted in to numeric data.

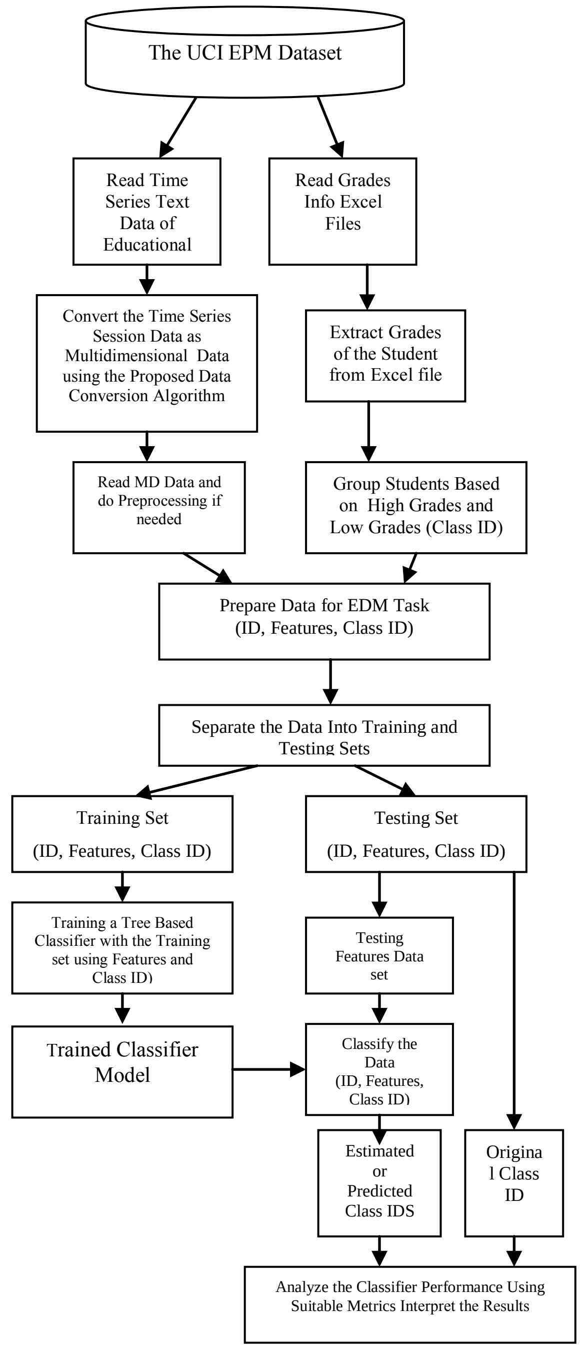


Figure 2. The proposed Model of the EDM System

The Figure 2 explains that, two set of data were taken from the UCI EPM dataset, time series data of the sessions and grades. The time series data is converted into Multidimensional Data(MD) and the students are grouped into high grade students and low grade students from the grade excel file and create an Education Data Mining(EDM) model. Now, it is separated as training set and testing set. Then, classify the data, analyze the performance and interpret the results.

IV. THE IMPLEMENTATION AND EVALUATION

The time series data conversation system and EDM system were developed using Matlab version 7.4.0 (R2007a), implemented with the help of WEKA tool and classified using J48 classifier and Random Tree classifier.

The Metrics and Validation Method Used for Performance Evaluation

In order to measure the performance of an EDM system, suitable metrics are required. In which, precision is used as a main metric. Classifier performance depends on the characteristics of the data to be classified. Performance of the proposed EDM system is measured using the following metricsPrecision, Recall, F_Score, True Positive Rate(TPR) and False Positive Rate(PR). Further, Multiple Significant tests were made using K-Fold cross validation.

Confusion Matrix

The performance of a classifier is broken down into a *confusion matrix*. A confusion matrix explains the type of classification mistakes created by a classifier. The following table shows the format of a typical confusion matrix.

Table 1: A confusion matrix.

Predicted Class		
Positives	**Negatives**	**Actual Class**
TP	FN	**Positives**
FP	TN	**Negatives**

The breakdown of a confusion matrix is:

- True Positives –TP is the number of positive examples correctly classified
- False Negatives -FN is the number of positive examples misclassified as negative
- False Positives –FP is the number of negative examples misclassified as positive
- True Negatives –TN is the number of negative examples correctly classified

The Metrics

To measure the performance of a student precision, recall, f-score are taken into consideration,

Precision

It is used to measure of classifiers exactness. A low precision indicates the presence of large number of false positives in the classification results. Precision is also known as Positive Predictive Value. This can be calculated using :

$$\text{Precision} = \frac{\text{True Positive(TP)}}{\text{True Positive(TP)} + \text{False Positive(FP)}}$$

High precision point outs that an algorithm returned significantly good relevant results.

Recall

It calculates the proportion of actual positives which are correctly identified as such.

$$\text{Recall} = \frac{\text{True Positive(TP)}}{\text{True Positive(TP)} + \text{False Negative(FN)}}$$

It is used to measure the classifiers completeness. High recall reveals that an algorithm returned the best of the appropriate results. A low recall informs that many False Negatives.

F_Score

F-Score helps to find the harmonic mean of precision and recall. It is calculated as follows:

$$\text{F\_Score} = 2 \times \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

True Positive Rate (TPR)

Recall is referred as Sensitivity or True Positive Rate.

$$\text{TPR} = \text{Recall} = \frac{\text{True Positive(TP)}}{\text{True Positive(TP)} + \text{False Negative(FN)}}$$

False Positive Rate(FPR)

The False Positive Rate is calculated using the following relation:

$$\text{FPR} = \frac{\text{False Positive(FP)}}{\text{False Positivee(FP)} + \text{True Negative(TN)}}$$

The K-fold Cross-Validation

In this work, K-Fold cross validation is selected as the main metric for evaluating the performance of the classifier and to interpret the results with improved significance. The data are randomly separated into K mutually exclusive subsets or folds $d_1, d_2, \ldots, d_k$, each approximately equal in size. The training and testing is performed k times. In the first iteration, subsets $d_2, \ldots, d_k$ all serve as the training set to obtain a first model, it is tested on d_1; the second iteration is trained in subsets $d_1, d_3, \ldots, d_k$ and tested on d_2; and so no. The advantage of this method is the accuracy which will not get much affected due to random division of data. Every data record in the dataset will present in a test set exactly once, and will present in a training set for k-1 times so that we will get much accurate measure. The variance of the resulting estimate is getting reduced while k is increased[19].

V. THE RESULTS AND DISCUSSION

In order to perform comparative study, the CM values from[1] are used. To make it fit with Precision, Recall and F-Score graphs, scaled the original values of CM[1] with a scaling factor of 1/50.

Scaled Value of CM[1]
CM=CM[1]/50

Table 2 – The Cyclomatic Complexity of Sessions [1]

	Session 2	Session 3	Session 4	Session 5	Session 6
CM[1]	105	63	74	76	112
CM	2.1	1.26	1.48	1.52	2.24

Table 2 explains that, the Cyclometic Complexity of the sessions 2 to 6 in 1:50 scale.

Table 3 - The Performance on the Combined All Session Data

Classifier	Grades	Precision	Recall	F-Measure
J48	Low	0.692	0.623	0.656
	High	0.661	0.726	0.692
	Average	0.676	0.675	0.674
Random_ Tree	Low	0.604	0.595	0.600
	High	0.606	0.614	0.610
	Average	0.605	0.605	0.605

Table 3, compares the J48 classifier and Random Tree classifier with precision, recall and f-measure at three levels namely low, average and high grade students for combined all session data.

Table 4 - The Performance On Individual Session Data

Classifier	Grades	Precision on Different Session Data				
		2	3	4	5	6
J48	Low	0.701	0.547	0	0	0.905
	High	0.4	0	0.708	0.879	0
	Average	0.609	0.302	0.508	0.773	0.819
Random Tree	Low	0.627	0.7	0.313	0.067	0.897
	High	0	0.574	0.731	0.868	0
	Average	0.436	0.644	0.613	0.772	0.812

Table 4 clarifies that, the comparison of J48 and Random Tree classification for the individual sessions of three categories of the students, that is low, average and high grade students.

The results of the previous study [1] reveals the below mentioned facts:

1. Difficulty of a e-learning session is directly proportional to complexity of the session (in terms of CM),
2. Grades are inversely proportional to complexity of the session (in terms of CM).

In addition to that, this proposed EDM model revealed some interesting facts. The following charts show the outcomesof the proposed EDM system with the data transformed into time series data. Since the UCI EPM data set doesn't contain the grades of first e-learning session , this analysis considered data between session-2 and Session-6.

Results based on Session-wise Data
Interpretation of Results in term of Precision
The following graphs show the predictability of low grades.

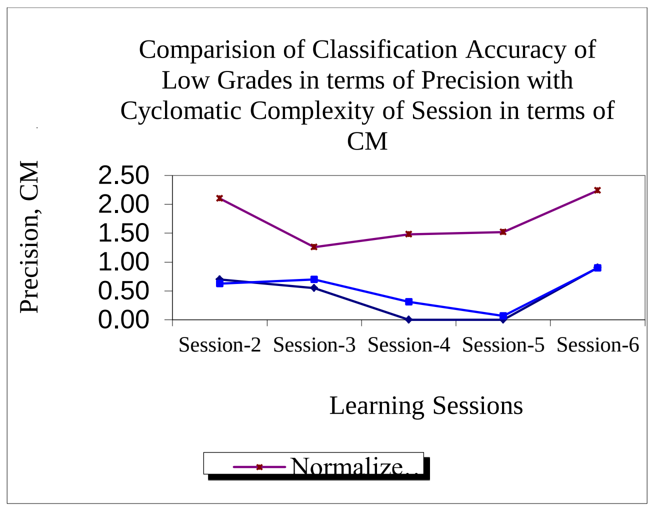


Figure 3: The Predictability of Low Grades

Predictability of potential low grade students (identifying poor performing students) is directly proportional to complexity of the e-learning session. It means, if the session complexity is very low, then not able to identify the poor individual students based on their behavior in e-learning session. This shows that a complex e-learning earning session can really make it possible to identify individual poor performing students (by mining the activity logs).

The following graph shows that the predictability of High grades.

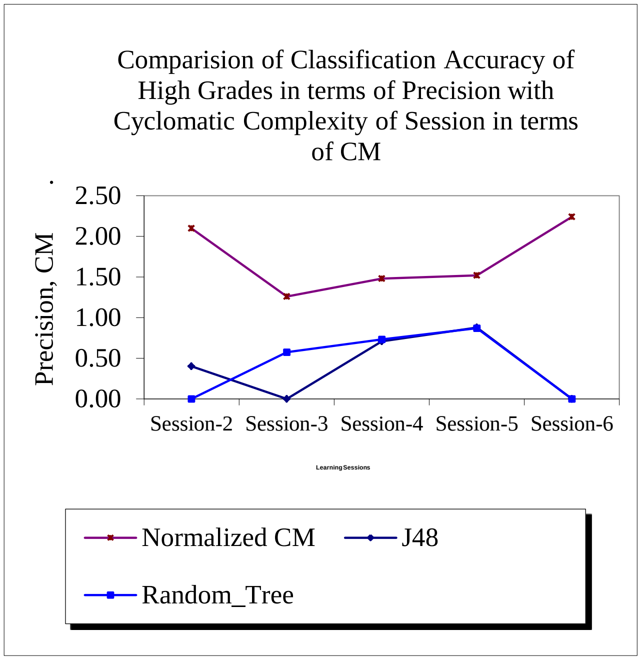


Figure 4: The Predictability of High Grades

Predictability of potential high grade students (identifying good performing students) is inversely proportional to complexity of the session. It means, if the session complexity is very high, then it is not able to identify the good performing individual students based on their behavior in an e-learning session. It implies that, only a session with optimum complexity can really make it possible to identify individual good performing students (by mining the activity logs).

The following graph shows that the predictability of grades (combined) is slightly complex by an e-learning session measured with CM. It also signifies that, lower level of prediction of the students' performance can be done by mining the activity logs.

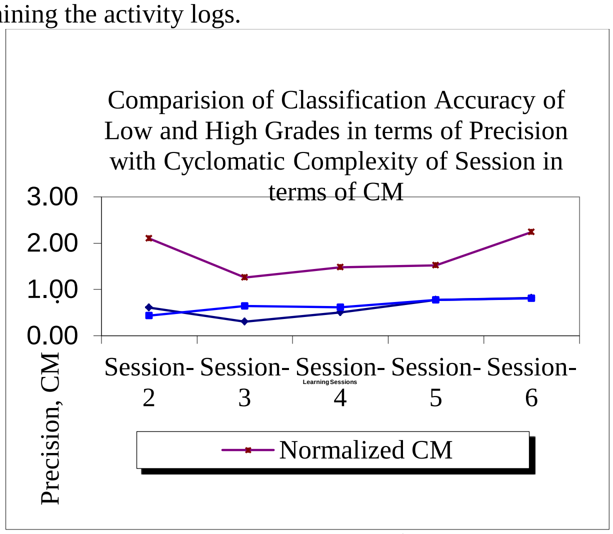


Figure 5: The Predictability of Grades

Results based on All Sessions Data
Predictability of Grades

Precision, Recall, F-Score, TP_Rate and FP_Rate charts show the predictability of the grades of the proposed system

The two tree based models present acceptable level of performance. The predictability of the J48 classifier based model is better than Random Tree classifier model.

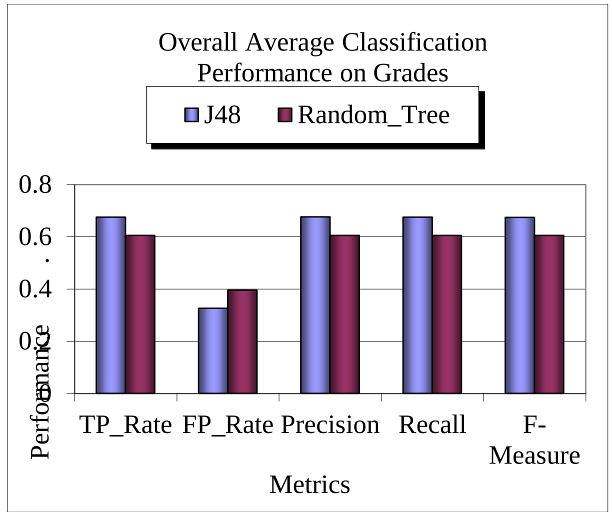


Figure 6: The Predictability of Grades

The following chart shows the predictability of lower grade Students. As shown in the following chart, the performance of the J48 classifier based model was better than Random Tree classifier model.

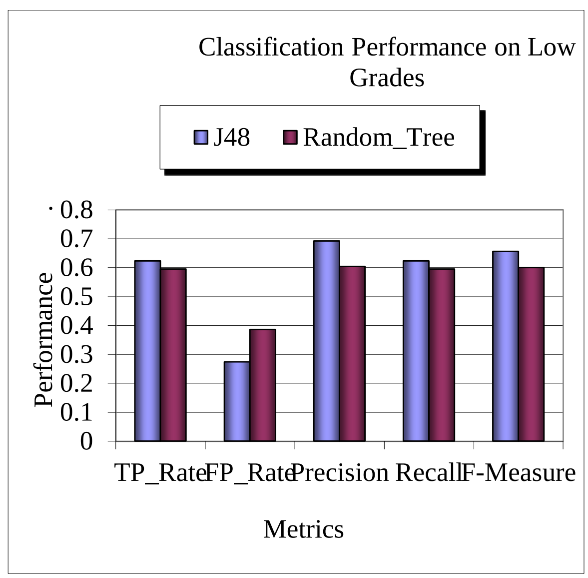


Figure 7: The Predictability on Low Grades

The following chart shows the predictability of higher grade Students. As shown in the following chart, the performance of the RBF Neural Network based model was better than all other compared models. J48 based model provides next better performance with respect to recall.

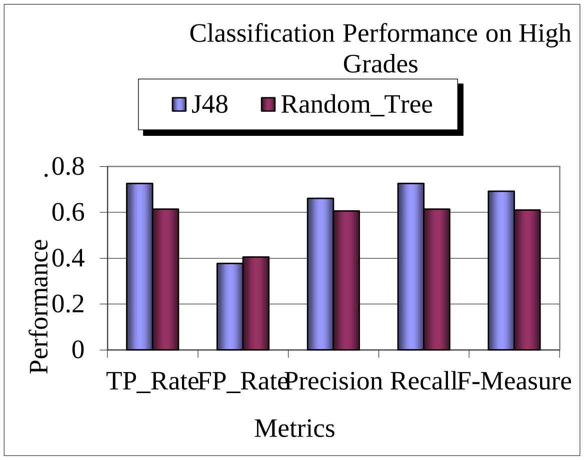


Figure 8: The Predictability on High Grades

In the overall session data, J48 provides to be providing good classification performance. But J48 failed to classify individual session data in some cases. In the figure 4 and 5, shows that for some cases, J48 is not able to classify High Grades as well as Low Grades and provided very low precision. Similarly, the precision in the case of Random Tree on some individual session data was zero. It reveals that it is not at all able to classify that particular class in that session data.

The results of this work can be summarized into the following findings:

1. Predictability of potential low grade students (identifying poor performing students) by mining the e-learning activity logs is directly proportional to complexity of the e-learning session.
2. Predictability of potential high grade students by mining the e-learning activity logs (identifying good performing students) is inversely proportional to complexity of the session.
3. Simple Tree based classifiers such as J48 is not able to predict grades by learning the models from small individual e-learning activity logs. Further study is recommended in this regard.

VI. Conclusion

A classification based EDM system is developed and implemented to mine the dataset. It makes use of two tree based classification algorithms to mine some unknown knowledge from the data using the classification results.

As mentioned in previous section, the results of that previous study [1] only revealed some facts by correlating Cyclomatic Complexity and Grades. It is also found that such correlation exists between the metrics (Precision, Recall, F-Score etc.,) and CM. In addition to that, the result shows that the outlines the possibility of estimating individual students' performance only based on their browsing/learning behavior during the e-learning session using the activity log of the e-learning session.

Further, it is learned that the J48 based classifier model could able toreach between 0.6 and 0.7 accuracy in terms of Precision with the help mining activity on whole data set. The results with individual session data were not that much significant when J48 is used. The study finds that the system will perform well if there will be sufficient data for training the classification model and a good subspace clustering algorithm. So, the outcome will be effective as mining activity is applied on considerably big dataset and suitable clustering algorithm.

In future works, an EDM system could be improved further by using neural network based classification model. This model will address the issues mentioned earlier and improve the performance of the of the EDM system in terms of predictability of student grades.

REFERENCES

[1]. M. Vahdat, L. Oneto, D. Anguita, M. Funk, M. Rauterberg.: "A learning analytics approach to correlate the academic achievements of students with interaction data from an educational simulator". In: G. Conole et al. (eds.): EC-TEL 2015, LNCS 9307, pp. 352-366. Springer, DOI: 10.1007/978-3-319-24258-3 26,2015

[2]. M. Vahdat, A. Ghio, L. Oneto, D. Anguita, M. Funk, M. Rauterberg, "Advances in learning analytics and educational data mining, in: European Symposium on Artificial Neural Networks", Computational Intelligence and Machine Learning, 2015.

[3]. Mohamed, N., Sulaiman, R.F., Endut, W.R.: "The use of cyclomatic complexity metrics in programming performance's assessment". ProcediaSoc. Beha. Sci. 90, 497-503, 2013

[4]. S. Arumugam, A. Kovalan, A.E.Narayanan " A Study of Easy Educational Data Mining for E-Learning Log Data from complex and large Dataset", IJIET, Volume 11 Issue 1, 39-47, August 2018.

[5]. R. Baker ,"Data Mining for Education". In McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevier, 2010

[6]. Jay Gholap,"Performance Tuning Of J48 Algorithm For Prediction Of Soil Fertilityby" Asian Journal of Computer Science and Information Technology,Vol 2,No. 8, 2012

[7]. Merceron and Yacef, "Revisiting interestingness of strong symmetric association rules in educational data", International Workshop on Applying Data Mining in e-Learning (ADML'07), Second European Conference on Technology Enhanced Learning,2007

[8]. García et al ,"Drawbacks and solutions of applying association rule mining in learning management systems", International Workshop on Applying Data Mining in e-Learning (ADML'07), Second European Conference on Technology Enhanced Learning, 2007

[9]. Bravo et al.,"A Problem`-Oriented Method for Supporting AEH Authors through Data Mining" International Workshop on

Applying Data Mining in e-Learning (ADML'07), Second European Conference on Technology Enhanced Learning, 2007
[10]. ShafiqAslam and Imran Ashraf , "Data Mining Algorithms and their applications in Education Data Mining", International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 7, July 2014
[11]. Amjad Abu Saa, "Educational Data Mining & Students' Performance Prediction", International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5, 2016
[12]. Brijesh Kumar Baradwaj, Saurabh Pal , "Mining Educational Data to Analyze Students Performance" International Journal of Advanced Computer Science and Applications, Vol. 2, No. 6, 2011
[13]. Mashael Al luhaybi, Allan Tucker and Leila Yousefi ,"The Prediction Of Student Failure Using Classification Methods: A Case study", IPPR, SOENG, DaMi, CSIT, AIS, CSE, CSIP, CCNET - 2018 pp. 79–90, 2018. © CS & IT-CSCP 2018
[14]. DriyaniRajeshinigo, 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, Issue 3, March 2017
[15]. Shimaa Abd Elkader AbdElaal, "E-learning using data mining", Chinese-Egyptian Research Journal
[16]. ChandiniLulla, YashAgarwal, SnehalKankariya, PrateekSakaray, PankajaAlappanavar ,"Student Academic Performance Prediction using Machine Learning and Data Mining Techniques",, International Journal of Computer Science and Mobile Computing, Vol.6 Issue.5, pg. 301-307, May- 2017
[17]. N.Sujatha , K. Prakash, "An Efficient and Scalable Auto Recommender System Based on Users Behavior", International Journal of Scientific Research in Computer Science Engineering, Vol.6, Issue.6, pp.35-40, December 2018
[18]. Himanshi , Komal Kumar Bhatia, "Prediction Model for Under-Graduating Student's Salary Using Data Mining Techniques", International Journal of Scientific Research in Network Security and Communication, Volume-6, Issue-2, April 2018
[19]. Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", 2nd Edition, Morgan Kaufmann Publisher, ELSEVIER, 2006

AUTHORS PROFILE

S. Arumugam. received his M.C.A., and M.Phil. in computer Science from Bharathidasan University. Now working as Assistant Professor at the Department of Computer Science and Application, PeriyarManiammai Institute of Science and Technology, Vallam, Thanjavur, Tamilnadu India. He authored several books/course materials for postgraduate students of PMIST and published papers in national and international Journals. He is a Life Member of Indian Society for Technical Education (ISTE) and Computer Society of India(CSI). He organized several workshops on different topics. Currently he is pursuing Ph.D. in Computer Science a PMIST, Tamilnadu, India .

A. Kovalan received his MCA, and PhD in Computer Science from Bharathiyar University. He has presented more than 20 papers at National and International Conferences. He has published more than 20 papers at National and International Journals. He has contributed in the development of Software on Online Entrance Examination system(PMU EES v1.0) at PMIST. His area of interest includes Image Processing, Signal Processing, Datamining. He is providing research guidance for Ph.D scholars from different areas of research and has presented several invited talks in this areas of research.

A. E. Narayanan received his M.Tech (Computer and Information Technology) from Manonmaniam Sundaranar University and Ph.D in Computer Science Engineering from Periyar Maniammai University. He has presented more than 10 papers at National and International Conferences.
He has published at National and International Journals. He organizes various National Conferences such as RITIDS in association with Ministry of Earth Sciences, Govt. of India New Delhi. He also conducted various National Level workshops in collaboration with IIT Bombay, ISTE and MHRD and guided various M.Tech students for their project and research work. His areas of research are Digital Image Processing and Fault Tolerant Systems. He is the Member of ISTE, CSI and IEI. He is providing research guidance for Ph.D scholars from different areas of research and has presented several invited talks in this areas of research.