IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 23, Issue 1, Ser. I (Jan. – Feb. 2021), PP 44-48
www.iosrjournals.org

A Comparative Study on use of Data Mining Algorithm in EDM Development

Joydip Kumar Sarmah[1], Dr. Siddhartha Baruah[2]
[1](Department of Computer Applications, Jorhat Engineering College, Jorhat, Assam, India)
[2](Department of Computer Applications, Jorhat Engineering College, Jorhat, Assam, India)

Abstract: *EDM is emerging field for exploring educational dataset. Student's performance, Enrollment capabilities and development of EDM application is highly used today while analyzing the student data. EMD id specialized with the capabilities of Visualization, Recommendation, Performance probability, Enrollment outcomes etc. Basic five data mining algorithm is compared – Decision Tree, Naïve Bayes, K Nearest neighbor, Neural network and Support Vector Machine (SVM). The accuracy of various authors work has been compare. The paper tried to combine a comparative study with respect to EDM functioning and its development with the help of core Data Mining Techniques.*

Date of Submission: 18-01-2021 Date of Acceptance: 03-02-2021

I. Introduction

Educational Data Mining (EDM) is the fields where various techniques and tools are used within Data Mining domain for development manage analysis and extract information from the historical or present data of student. EDM is one of the emerging and exploring topics of Data Mining field. Educational data comes from large scale extraction and EDM plays vital role in generating information regarding student data. EDM also carry all the Educational Statistical data for predicting the performance of students in particular institution. The author Rayen S. Baker and Kalina Yacef **[1]** identified some goals of EDM.

1. EDM will help to predict student's future learning behavior.
2. EDM help to develop new model or improve existing model using latest data mining aspects.
3. EDM also highlight the present educational system which suggests to be improved and better.
4. New scientific knowledge and improved software will also be studied by EDM perspective.

It is always challenging for Institute to admit best student as per academic subjects. Today's perspective regarding education in most institutes to provide satisfactory performance and opportunity globally to establish. In educational institute academic performance and student's placement is one of the main dependencies to prove its existence. EDM and Data mining interrelate to provide guidance in improving the decision and knowledge for best performance.

Data Mining is useful in the fact that it will provide satisfactory decision or knowledge by providing the dataset of students. Data Mining deals with system where any large dataset can be proceed for knowledge. The historical data of students can help to predict the best knowledge where forecast of student performance can be determined. The knowledge extraction process of Data Mining always tries to improve student's behaviors, success rate of an institute, course publicity etc. An improved educational institute is liable to provide quality of education, student's intake capacity, best career counseling etc. The improvement of traditional decision making system with data mining approach makes an institute best in globally competition.

The main objective of the paper is to generate the use of data mining approaches and techniques in the field of Education Data Mining and to review paper in large basis to generate an idea for establishing data mining techniques while improving the EDM. The paper also put forward some essential information regarding improvement of student's performance predictions. Data Mining has many approaches depending on the use of dataset and its improvement.

Data mining techniques have mainly three key approaches – Classification, Clustering and Regression. Classifications Techniques fall under supervised learning allow defining the class variable where as unsupervised learning which is Clustering don't allow such definition of class variable. Class variable is the predictive outcome of some conditions depending upon the question we are asking or finding knowledge. Regression Techniques allow variables to predict the definite outcomes. Data mining is always useful in generating patterns of information. Data Mining is also capable of handling statistical data either historical based or predictive based the extraction of knowledge is always concluded. The paper tried to study those

techniques and associated literature which develop the EDM and help to conclude outcomes while handling the educational data set.

II. Objective

Objective is always set to proper define the goal of the paper, so here are some objective while developing the paper.

1. The objective of the paper is to review paper to understand the use of techniques in EDM
2. The paper also put forward some research on basis of such data mining techniques.
3. The paper also focused on discussing some statistical tools use by such techniques while predicting outcomes.
4. The paper also tried to set relationship of data extraction and techniques involvement.

III. Literature Review

Literature review is conducted in specialization of three main fields with respect to use of EDM while extracting or gaining knowledge. Various author put forward many research work related to development of EDM which is being categorized into the following fields.

1. Student's Performance Prediction
2. Application development for Effective and efficient task flow for managing data.
3. Influence of student's enrollment

Student's Performance Prediction

Student's Prediction is one of the important tasks as the performance and quality of any institute is directly depending upon the success of student. In data mining field there are many methods how we can conduct the performance of students like according to Usamah et al. (2013) **[2]** student's performance can be obtained by learning assessment and co-curriculum. Some author put forward important attributes like according to Amirah Mohamed Shahiri et al. (2015) [3] CGPA performance is important factor to predict what the student will perform. The author also stated that CGPA relate to tangible value which helps many researchers to implement data mining methods for prediction. It has been highly notices in research paper that the prediction based on CGPA is mostly used because ultimately it help to the state the future survival status of any students. Beside the CGPA some author put forward the demographic knowledge like gender, age, family background and disability to state the performance of the students. Demographic attributes gender is highly influential with predicting the students performance as according to the author Meit at al. (2007) [4] found that female student have more positive learning style compare to the male students. The author also proved that the learning strategy of female student is better than male student and it highly influence the performance of the students. Next important attributes that can be considered for the prediction of performance of student is the psychometric performance. Psychometric performance is identified as the interest of students, their behavior, engage time etc. According to M. Mayilvaganan et al. [5] it is very important for lecturer to evaluate student based on their personal interest and behavior. The student performance is highly depend upon the above three attributes. Sometime the background of student becomes much influential to score good academic result. The researcher always tries to put prediction depending upon the above mention attributes while predicting the performance.

Application Development for Effective and Efficient task flow for managing data.

Application development is one of the major tasks for managing data especially if the data is majorly very high in amount. Student data sometime can be available in both past statistics as well as present prediction. Application help to organize and provide visualization output of any knowledge extraction information. The author Cristobal Romero and Sebastian Ventura [6] suggested some development of applications which are listed below.

1. Analysis and Visualization of Data – Application based on EDM culture always provide importance in visualization task as it helps the non programming people to interact with the system. Visualization provides easy understanding of analysis of student's data.
2. Providing Feedback for Support Instructors – The faculty always needs improvement in academic path so EDM application also helps to collect feedback from the student.
3. Recommendation for the Student – Student Recommendation task is very important as it help them to choose the subject they are suppose to score best. Recommendation application basically run by implementing Data Mining techniques like Neural Network or Classification or Regression to provide best appropriate analysis for student.
4. Students Modeling – It is the section where cognitive model for student will be developed based on their skills and declarative knowledge.

5. Detecting Undesirable student's behavior – this application is developed keep in knowledge to detect the abnormality or unusual possession of student's behavior. This application tries to provide solution to the student to improve or develop in better way.
6. Grouping Student – This EDM application development help to combine student based on their personality, likings, nature of study etc so that the faculty or instructor can build personalized learning system.
7. Social Network Analysis – This EDM application development help to analysis the network of relationship the student maintain socially, the analysis help to judge the nature of student and their personality.
8. Constructing Courseware – This application generally help to build a structure of course based on the student capacity, the instructor or faculty can customize the course based on student's way of learning.
9. Planning and Scheduling – The application development for planning and scheduling provide analysis based on future courses, course scheduling, resources allocation, helping in admission and counseling process etc.

Influence of Student's Enrollment

Enrollment of student in Science Technology Engineering and Mathematics (STEM) courses s one of the major task for the institution for best enrollment. Data mining algorithm are mostly used to judge the student's capacity on which the stream will be given. Many universities try to develop or built their own STEM analysis application for enrolling student and the success rate is high in respect to the result of outcomes. Author Fong et al. [7] used back propagation techniques to develop application based on enrollment task for student. Author Stephen Kahara et al. [8] used the enrollment of student in respect to the STEM courses where the Cross Industry Standard Process for Data Mining framework is used to development application to provide best enrollment details of students.

IV. Attributes Selection in Association of Data Mining Algorithm

The author Amirah Mohamed Shahiri et al. (2015) [3] highlighted some important attributes and their selection based upon the data mining algorithms. Those are listed below:-

Attributes	Methods
1. Internal Assessment, CGPA	Support Vector Machine (SVM)
2. Internal Assessment, CGPA, Extra Curricular Activity	Naïve bayes, K Nearest neighbor, Support Vector Machine
3. CGPA, Student Demographic, High School Background, Scholarship, Social network interaction	Naïve Bayes
4. Internal Assessments, External Assessment, Student Demographic, High School Background	Neural Network
5. External Assessment, Student Demographic, High School Background	Neural Network, Naïve Bayes
6. Internal Assessment	Decision Tree, Neural Network, K Nearest Neighbor
7. CGPA, Student Demographic, High School Background, Scholarship, Social network Interaction, Internal Assessment, CGPA, Extra Curricular Activities, External Assessment, Psychometric Factors, Soft Skills	Decision Tree

V. Result Accuracy and Algorithms

Methods	Attributes	Results	Authors
1. Decision Tree	Internal Assessments	76%	Romero et al. (2008) [9]
	Psychometric Factors	65%	Gray et al. (2014)[10]
	External Assessment	85%	Bunkar et al. (2012)[11]
	CGPA	91%	Jishan et al. (2015)[12]
	CGPA, Student Demographic, High School Background, Scholarship, Social Network Interaction	73%	Osmanbegovic and Suljic (2008) [14]
	Internal Assessment, CGPA, Extra- Curricular Activities	66%	Mayilvaganan and Kapalnadevi (2014) [5]
	Student Demographic, High School Background	65%	Ramesh et al. (2013) [13]
	Internal assessment, Student	90%	Elakia et al. (2014)[15]

	Demographic, Extra-curricular activities		
	External assessment, CGPA, Student Demographic, Extra-curricular activities	90%	Natek and Zwilling (2014)[16]
	Psychometric factors, Extra-curricular activities, soft skills	88%	Mishra et al. (2014) [17]
2. Neural Network	Internal assessments	81%	Wang and Mitrovic (2002) [18]
	Psychometric factors	69%	Gray et al. (2014)[10]
	External assessment	97%	Arsad et al. (2013)[19]
	CGPA	75%	Jishan et al. (2015)[12]
	CGPA, Student Demographic, High school background,Scholarship, Social network interaction	71%	Osmanbegovic and Suljic (2008) [14]
	Student Demographic, High school background	72%	Ramesh et al. (2013) [13]
	External assessment, Student Demographic, High school background	74%	Oladokun et al. (2008)[20]
	Internal assessments, External assessment	98%	Anupama and Vijayalakshmi (2012) [21]
3. Naïve Bayes	CGPA, Student Demographic, High school background, Scholarship, Social network interaction	76%	Osmanbegovic and Suljic (2008) [14]
	Student Demographic, High school background	50%	Ramesh et al. (2013) [13]
	CGPA	75%	Jishan et al. (2015)[12]
	Internal assessment, CGPA, Extra-curricular activities	73%	Mayilvaganan and Kapalnadevi (2014) [5]
4. K Nearest Neighbor	Psychometric factors	69%	Gray et al. (2014) [10]
	Internal assessment, CGPA, Extra-curricular activities	83%	Mayilvaganan and Kapalnadevi (2014) [5]
	Internal assessment	82%	Bigdoli et al. (2003)[22]
5. Support Vector Machine	Psychometric factors	83%	Sembiring et al. (2011) [23]
	Internal assessment, CGPA, Extra-curricular activities	80%	Mayilvaganan and Kapalnadevi (2014) [5]
	Internal assessment, CGPA	80%	Hamalainen et al. (2006) [24]

VI. Conclusion

EDM techniques are most important while extracting knowledge for development of student. The paper is highly motivated upon the various author's work which lead and executed best implementation of EDM either by application or knowledge purpose. The large scale of student data is handled and manipulated by EDM applications. The paper tried to compare and put forward the work of various author and their recommendation for future work is highly appreciable. The five Data Mining techniques are considered to be used most in student analysis those are SVM, K nearest Neighbor, Naïve Bayes, Neural Network and Decision Tree. In future the study will be conducted using the real data set for predicting and analyzing the future probability of student's either enrollment or performance betterment.

References

[1]. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (AdultTreatment Panel III) Third report of the national cholesterol education.

[2]. U. bin Mat, N. Buniyamin, P. M. Arsad, R. Kassim, An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, in: Engineering Education (ICEED), 2013 IEEE 5th Conference on, IEEE, 2013, pp. 126–130.

[3]. Amirah Mohamed Shahiria, Wahidah Husaina , Nur'aini Abdul Rashida , A Review on Predicting Student's Performance using Data Mining Techniques , The Third Information Systems International Conference , Procedia Computer Science 72 (2015) 414 – 422

[4]. S. S. Meit, N. J. Borges, B. A. Cubic, H. R. Seibel, Personality differences in incoming male and female medical students., Online Submission.

[5]. M. Mayilvaganan, D. Kalpanadevi, Comparison of classification techniques for predicting the performance of students academic environment, in: Communication and Network Technologies (ICCNT), 2014 International Conference on, IEEE, 2014, pp. 113–118.

[6]. Cristobal Romero, ebastian Ventura , Educational Data Mining: A Review of the State of the Art, IEEE Transactions on Systems Man and Cybernetics Part C 40(6):601-618.

[7]. Fong, S., Yain-Whar, S., Robert, P., & Aghai, B. (2009). Applying a Hybrid Model of Neural Network and Decision Tree Classifier for Predicting University Admission. IEEE.

[8]. Stephen Kahara Wanjau, George Okeyo, Richard Rimiru, Data Mining Model for Predicting Student Enrolment in STEM Courses in Higher Education Institutions, International Journal of Computer Applications Technology and Research Volume 5–Issue 11, 698-704, 2016, ISSN:-2319–8656.

[9]. C. Romero, S. Ventura, P. G. Espejo, C. Hervas, ´ Data mining algorithms to classify students, in: Educational Data Mining 2008, 2008.

[10]. G. Gray, C. McGuinness, P. Owende, An application of classification models to predict learner progression in tertiary education, in: Advance Computing Conference (IACC), 2014 IEEE International, IEEE, 2014, pp. 549–554.

[11]. K. Bunkar, U. K. Singh, B. Pandya, R. Bunkar, Data mining: Prediction for performance improvement of graduate students using classification, in: Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on, IEEE, 2012, pp. 1–5.

[12]. S. T. Jishan, R. I. Rashu, N. Haque, R. M. Rahman, Improving accuracy of students final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique, Decision Analytics 2 (1) (2015) 1–25.

[13]. V. Ramesh, P. Parkavi, K. Ramar, Predicting student performance: a statistical and data mining approach, International Journal of Computer Applications 63 (8) (2013) 35–39.

[14]. E. Osmanbegovic, M. Sulji ´ c, Data mining approach for predicting student performance, Economic Review 10 (1).

[15]. G. Elakia, N. J. Aarthi, Application of data mining in educational database for predicting behavioural patterns of the students, Elakia et al,/(IJCSIT) International Journal of Computer Science and Information Technologies 5 (3) (2014) 4649–4652.

[16]. S. Natek, M. Zwilling, Student data mining solution–knowledge management system related to higher education institutions, Expert systems with applications 41 (14) (2014) 6400–6407.

[17]. T. Mishra, D. Kumar, S. Gupta, Mining students' data for prediction performance, in: Proceedings of the 2014 Fourth International Conference on Advanced Computing & Communication Technologies, ACCT '14, IEEE Computer Society, Washington, DC, USA, 2014, pp. 255–262. doi:10.1109/ACCT.2014.105.

[18]. T. Wang, A. Mitrovic, Using neural networks to predict student's performance, in: Computers in Education, 2002. Proceedings. International Conference on, IEEE, 2002, pp. 969–973.

[19]. P. M. Arsad, N. Buniyamin, J.-l. A. Manan, A neural network students' performance prediction model (nnsppm), in: Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on, IEEE, 2013, pp. 1–5.

[20]. V. Oladokun, A. Adebanjo, O. Charles-Owaba, Predicting students academic performance using artificial neural network: A case study of an engineering course, The Pacific Journal of Science and Technology 9 (1) (2008) 72–79.

[21]. D. M. S.Anupama Kumar, Appraising the significance of self regulated learning in higher education using neural networks, International Journal of Engineering Research and Development Volume 1 (Issue 1) (2012) 09–15.

[22]. B. M. Bidgoli, D. Kashy, G. Kortemeyer, W. Punch, Predicting student performance: An application of data mining methods with the educational web-based system lon-capa, in: Proceedings of ASEE/IEEE frontiers in education conference, 2003.

[23]. S. Sembiring, M. Zarlis, D. Hartama, S. Ramliana, E. Wani, Prediction of student academic performance by an application of data mining techniques, in: International Conference on Management and Artificial Intelligence IPEDR, Vol. 6, 2011, pp. 110–114.

[24]. W. Ham¨ al¨ ainen, ¨ M. Vinni, Comparison of machine learning methods for intelligent tutoring systems, in: Intelligent Tutoring Systems, Springer, 2006, pp. 525–534.

Joydip Kumar Sarmah. "A Comparative Study on use of Data Mining Algorithm in EDM Development." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 23(1), 2021, pp. 44-48.