

A Study on Prediction of Student Academic Performance based on Expert Systems

¹Kandula Neha , ²Dr S. Jahangeer Sidiq

¹Research Scholar, CSE
Lovely Professional University
Punjab, India
neha09kandula@gmail.com

²Assistant Professor, CSE
Lovely Professional University
Punjab, India
jahangir.25453@lpu.co.in

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 16 April 2021

ABSTRACT: In recent years , research evolution in domain of education focus on analytics and which provides insights on students academic performance. The tremendous growth of instructional institutions' electronic information provides the chance to extract info which will be wont to predict students' overall success, predict students' dropout rate, appraise the performance of academics and instructors, improve the learning material in step with students' desires, and far a lot of. This paper aims to review the latest trends in predicting students' performance in educational activity. we explain the measures of determinant educational success and highlight the strengths and weaknesses of the foremost common data processing (DM) tools and strategies used today. Moreover, we offer a fashionable literature review of the EDM work that has been revealed during the past years with target the prediction of educational performance in educational activity. we tend to analyze the foremost normally used options and strategies in predicting educational accomplishment, and highlight the advantages of the principally used DM tools in EDM. The results of this paper might assist researchers who are working to hold out EDM solutions within the domain of education as we tend to highlight the type of options that the previous researches found to possess important impact on the prediction, likewise because the edges and downsides of the DM strategies and tools used for predicting educational outcomes.

Keywords—Educational Data Mining, Prediction, Analytics, Machine Learning, Deep Learning

I INTRODUCTION

The huge quantity of digital content on the market in several areas has actuated analysis in, and also the development of, totally different disciplines that build it easier to look, organize and analyze this content. data processing and machine learning are disciplines that have emerged to research this info in an automatic manner,[16] by finding patterns and relationships in information, and that they have conjointly contributed to resolution advanced issues. During the past decades, the most important innovations in educational systems are related to the introduction of new technologies [9].

Educational data processing (EDM) may be a field that exploits statistical, machine-learning, and data-mining (DM) algorithms over the various forms of instructional information. Its main objective is to investigate these forms of information so as to resolve educational analysis problems[24].EDM seeks to use these knowledge repositories to raised perceive learners and learning, and to develop process approaches that mix knowledge and theory to rework follow to profit learners. EDM has emerged as a look space in recent years for researchers everywhere the planet from totally different and connected research areas, that square measure as follows[3]

- 1) Offline education attempt to transmit information and skills based on face-to-face contact and conjointly study psychologically on however humans learn. psychometric and applied mathematics techniques are applied to knowledge, like student's performance, curriculum, etc., that was gathered in classroom environments.
- 2) E-learning and learning management system (LMS). E-learning provides on-line instruction, and LMS conjointly provides communication, collaboration, administration, and reporting tools. net mining (WM) techniques are applied to student's knowledge keep by these systems in log files and databases.[20]
- 3) Intelligent tutoring system (ITS) and adjustive instructional hypermedia system (AEHS) square measure another to the just put- it-on-the-web approach, making an attempt to adapt teaching to the needs of every

specific student. DM has been applied to knowledge picked up by these systems, like log files, user models, etc.

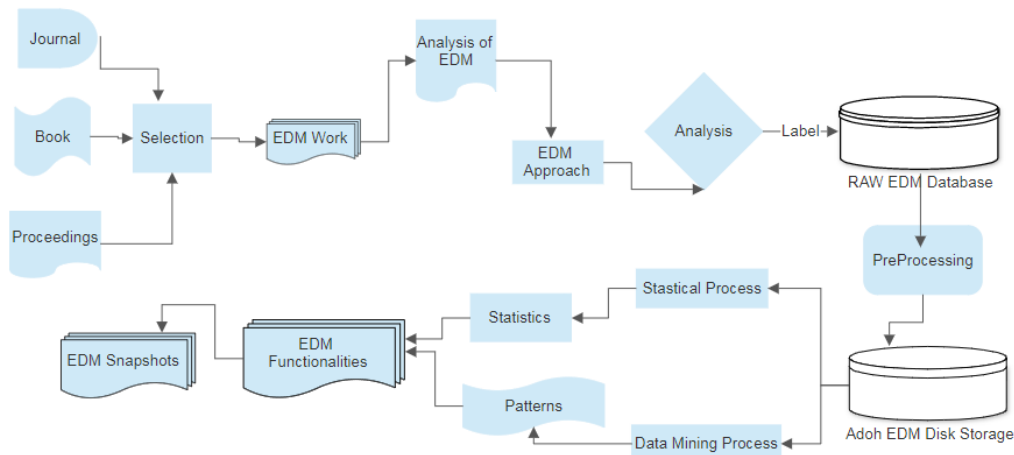


Fig:1 Workflow of DM Approach performed to analyse, classify, represent and mine data of EDM related works[15]

II FEASIBLE STUDY OF DIFFERENT APPROACHES FOR STUDENT PREDICTION

In the previous few years there has been a proliferation of research within the EDM field DL architectures. This article presents a review on the literature of DL techniques applied to EDM.[14]

The primary contributions of this text square measure as follows:

- (i) Summarize the most EDM tasks and classify the existing works that have applied DL on every of those tasks.
- (ii) establish the tasks that have gained major attention and those that square measure still undiscovered.
- (iii) Describe and categorise the most public and personal datasets utilized to coach and check DL models in EDM tasks.
- (iv) Introduce key DL ideas and technologies, describing the techniques and configurations most generally used in EDM and its specific tasks.
- (v) Discuss future directions for analysis in DL applied to EDM based on the knowledge gathered during this study[1]

Table 1: Authors and their Study of Approaches

Author	Study
Amin Zollanvari	Predicting students GPA based on self-regulatory learning behaviours [2]
Jie Xu, KyeongHo Moon	In this article type of prediction is predicting graduate student performance, the future performance of the student and student performance in completing graduate based on their continuing education files is necessary for viable doing vital crucial pedagogical mediation to guarantee learners' on-time and fair graduation.[11]
Simon Fong, Yain-Whar Si	forecasting similarities between classes of universities and secondary students based on constructed model RSAU (recommender system for admission to university) to get admission from secondary school to university via the recommended system, input fields=19, data set=2400 students.[19]
Shaojie Qu, Kan	Predicting achievement of student in smart campus (online class)[17]
V.O. Oladokun, A.T. Adebajo	The main objective of this article is predicting the likely performance of the student in the university based on given data at the time of admission [25]
Suhas S Athani Sharath Kodli	Predicting student behaviour and academic performance based on some factors [21]

III FEW METHODS OF DATA MINING WITH PRO'S & CON'S

Decision trees (Clark 2013; Kaushal and Shukla 2014; Yu-Wei and David. 2015) : A classification methodology during which every internal node is a “test” on a feature, every branch function the result of the check, and every leaf node correspond as a category label (decision taken when computing all

features)[7,12,26]. It is Easy to grasp and Can handle missing values and Could suffer from over fitting with Less accuracy with continuous variables

Support vector machine (SVM)(Clark 2013; Harrington 2011; Tomar and Agarwal 2013) : A supervised learning model with associated learning algorithms that analyze knowledge being employed for classification and multivariate analysis by applying the statistics of support vectors to categorize untagged knowledge.[7,23,10]. It has High accuracy and Can handle totally different knowledge sorts with Effective in high dimensional area. It also has Black box training processes which will take time with High algorithmic quality.

Naïve Bayes(Harrington 2011; Yu-Wei and David 2015) : A classification methodology that assumes that the prognostic options area unit not absolutely freelance which there aren't any hidden options that would have an effect on the method of prediction[23,26]. It is Simple to use and Can trot out missing and wheezy knowledge with Black box and assumes that each one options area unit freelance and equally vital.

Neural Networks(Clark 2013; Kaushal and Shukla 2014;Tomar and Agarwal 2013) : A method that learns to perform tasks by considering examples, usually while not being programmed with any task-specific rules[7,12,10]. It has High accuracy and Can handle missing and wheezy knowledge with Black box and difficult in managing huge knowledge with High quality

Logistic regression(Geng 2006; Yu-Wei and David 2015) : A applied math model that's usually taken to use to a binary variable. additionally formally, a supply model is one wherever the log-odds of the likelihood of an incident may be a linear combination of freelance or predictor variables[8,26]. It is Easy to grasp and provides likelihood outcome which does not handle missing values well and could suffer from over-fitting.

K-nearest neighbour(Clark 2013; Yu-Wei and David 2015) : A classification and regression methodology that stores obtainable cases and classifies new cases supported a similarity live. A case is classed by a majority vote of its neighbours, with the case being allotted to the category commonest amongst its K nearest neighbours measured by a distance operate.[7,26]. It is Nonparametric and easy to grasp the output. The Robustness to wheezy coaching knowledge and is difficult in handling mixed knowledge kind Assumes that each one options area unit equally important Sensitive to outliers.

Rule induction(An et al. 1997; Domingos 1995) : An area unit of machine learning wherever formal rules (If-then) are extracted supported applied math significance from a group of knowledge observations Low computational-space price will create effective use of applied math measures to combat noise[4]. It has Slow coaching time and has hassle recognizing exceptions or little with low-frequency sections of the area

IV AUTHOR(S) PREDICTION PREDICTOR FEATURES DM METHOD(S) DM TOOL RESULTS

S.No	Author & Title	Demographics	Pre - Enrolment Fetures	Post - Enrolment Fetures	Tool Result
1	Nghe et al. (2007) Predict students' GPA at the end of the first year of their Master program using three models at the Asian Institute of Technology (AIT), Thailand [13]	Gender, marital status, income, and age	Academic institute, previous GPA, English proficiency, and TOEFL score		Decision-trees (J48) Bayesian-tree Weka J48 produced better accuracy a) (91.98%) for 2 classes (pass/fail), b) (67.74%) for 3 classes (Fail/Good/Very Good) c) (63.25%) for 4 classes (Fail/Fair/Good/ Very Good)
2	Nghe et al. (2007) Predict students' GPA at the end of the third year using three models in CanTho University (CTU),Vietnam[13]	Gender, age, family, job, and religion	English skill, entry marks range, the field of study, and faculty	second-year GPA	Decision-trees (J48) Bayesian-tree Weka J48 produced better accuracy a) (92.86%) for 2 classes (pass/fail), b) (84.18%) for 3 classes (Fail/Good/Very Good)

					c) (66.69%) for 4classes (Fail/Fair/Good/Very Good)
3	Yadav et al. (2011) Predict computer master students' performance at VBS Purvanchal University in Jaunpur, India.[27]			attendance, test grade, seminar grade, assignment grade, and lab work	Decision trees (ID3, CART, and C4.5) Weka CART produced the best accuracy (56.25%) followed by ID3 (52.08%) then C4.5 (45.83%)
4	Sembiring et al. (2011) Predict final grade of students from the faculty of computer systems and software engineering at the University of Malaysia Pahang (UMP) in Malaysia[22]	Personal beliefs, and family support		interest, study behaviour, and engaging time	Support Vector Machine (SVM) Rapid-Miner SVM produced high accuracy (83%)

IV CURRENT TRENDS

While this survey covers educational prediction studies performed from past few years there are newer studies that are revealed in 2020. Many of the approaches rely on ancient machine learning ways, like SVM, call trees, supply regression, and Naïve mathematician. However, there are some new data processing ways that are explored, like Structural Equation Modeling and probabilistic neural networks. though deep neural nets have seen a growing quality within the machine learning community, notably with applications to tongue process, they're still not adopted within the EDM literature. this is often in all probability thanks to their would like of terribly giant coaching knowledge, whose sourcing is problematic in instructional contexts[5]. With regard to the sort of options accustomed perform the predictions, demographical features are still usually used. Interestingly, exploitation such options shows promising results. This involves make sure that prediction of students' performance remains a really actively researched downside, whose current solutions will still be improved, and that the factors that largely influence educational outcomes and therefore may be accustomed predict future performances are still not wide understood[6].

V CONCLUSION

Educational data processing is a locality that holds exciting opportunities for researchers and practitioners all round the world. This field assists in rising institutional effectiveness by supporting deciding and enhancing student learning to succeed in visible and measurable targets. This paper provides a chic literature review on the prediction of educational achievement in teaching for the past few years with the ultimate aim of providing researchers and academic planners with data to help them once making an attempt to carry out associate EDM resolution. This paper unconcealed that a substantial quantity of researchers has been performed in analyzing and predicting tutorial performance. It showed that classification and regression algorithms can be used with success to predict students' performance in each course and degree level. It will be seen that the majority of the reviewed EDM analysis within the past decade has been completed victimization the open supply machine learning computer code. we have a tendency to found that the most used ways for predicting action area unit call tree algorithms, On the opposite hand, neural networks, support vector machines and K-nearest neighbour weren't of times used as compared to the remainder of the DM ways. We so conclude that additional analysis is required 1st to deepen our understanding of the contribution of every historically used feature, and second, to increase the set of options and methodologies for additional rising this prediction accuracies.

REFERENCES

1. A Systematic Review of Deep Learning Approaches to Educational Data Mining Antonio Hernández-Blanco ,¹ Boris Herrera-Flores ,² David Tomás ,³ and Borja Navarro-Colorado
2. Amin Zollanvari, RefikCaglarKizilirmak, Yau Hee Kho, Daniel Hernández-Torrano. (2017). "Predicting students' GPA and Developing Intervention Strategies Based on Self-Regulatory Learning Behaviours." In proceedings IEEE Access 5: 23792-23802.
3. Ashraf, M., Zaman, M., Ahmed, M., & Sidiq, S. J. (2017). Knowledge discovery in academia: a survey on related literature. *Int. J. Adv. Res. Comput. Sci.*, 8(1).
4. An, A., Cercone, N., & Chan, C. (1997). Integrating rule induction and case-based reasoning to enhance problem solving. In J. Carnegie, M. U. Carbonell, & University of S. Siekmann (Eds.), *Second international conference on case-based reasoning* (pp. 499–458). Springer,
5. Berlin. Berens, J., Oster, S., Schneider, K., & Burghoff, J. (2019). Early detection of students at risk: Predicting student dropouts using administrative student data and machine learning methods. *Schumpeter School of Business and Economics*, 11(3), 1–32.
6. Bhutto, S., Siddiqui, I. F., Arain, Q. A., & Anwar, M. (2020). Predicting students' academic performance through supervised machine learning. In *ICISCT 2020: 2nd international conference on information science and communication technology*. Institute of Electrical and Electronics Engineers Inc.
7. Clark, M. (2013). An introduction to machine learning with applications in R. Retrieved 22 August 2018 from http://web.ipac.caltech.edu/staff/fmasc/home/astro_refs/ML_inR.pdf.
8. Geng, M. (2006). A comparison of logistic regression to random forests for exploring differences in risk factors associated with stage at diagnosis between black and white colon cancer patients. Master Degree Thesis. Department of Biostatistics, University of Pittsburg.
9. Ha, S., Bae, S., & Park, S. (2000). Web mining for distance education. In *IEEE international conference on management of innovation and technology*.
10. Harrington, P. (2011). Machine learning in space: Extending our reach. *Machine Learning*, 84(3), 335–340.
11. Jie Xu, KyeongHo Moon, Mihaela van der Schaar. (2017) "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs". In *Proc IEEE journal of selected topics in signal processing* 11: 742 – 753.
12. Kaushal, A., & Shukla, M. (2014). Comparative analysis to highlight pros and cons of data mining techniques- clustering, neural network and decision tree. *International Journal of Computer Science and Information Technologies*, 5(1), 651–656.
13. Ngehe, N. T., Janecek, P., & Haddawy, P. (2007). A comparative analysis of techniques for predicting academic performance. In *2007 37th Annual frontiers in education conference—Global engineering: knowledge without borders, opportunities without passports* (pp. T2G-7–T2G-12), IEEE.
14. Neha, K., & Sidiq, S. J. (2020). Analysis of Student Academic Performance through Expert systems. *International Research Journal on Advanced Science Hub*, 2, 48-54.
15. Pena-Ayala, A. Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications* (2013), <http://dx.doi.org/10.1016/j.eswa.2013.08.042>
16. Sidiq, S. J., Zaman, M., & Butt, M. (2018). An empirical comparison of classifiers for multi-class imbalance learning. *International Journal of Data Mining And Emerging Technologies*, 8(1), 115-122.
17. Shaojie Qu, Kan Li, Shuhui Zhang, and Yongchao Wang. (2018) "Predicting Achievement of Students in Smart Campus". In *Proc IEEE Access*.6: 2169-3536.
18. Sidiq, S. J., Zaman, M., Ashraf, M., & Ahmed, M. (2017). An empirical comparison of supervised classifiers for diabetic diagnosis. *Int J Adv Res Comput Sci*, 8(1), 311-315..
19. Simon Fong, Yain-Whar Si, Robert P. Biuk-Aghai. (2009) "Applying a Hybrid Model of Neural Network and Decision Tree Classifier for Predicting University Admission", in *Proceedings 7th International Conference on Information, Communications and Signal Processing (ICICS)*.
20. Sidiq, S. J., & Zaman, M. (2020). A Binarization Approach for Predicting Box-Office Success. *Solid State Technology*, 63(5), 8652-8660.
21. [21]Suhass Athani, Sharath A Kodli, Mayur N Banavasi P. G. Sunitha Hiremath. (2017). "Student Academic Performance and Social Behaviour Predictor using Data Mining Techniques." In *Proc International Conference on Computing, Communication, and Automation (ICCCA2017)*.
22. [22]Sembiring, S., Zarlis, M., Hartama, D., Wani, E., & Magister, P. (2011). Prediction of student academic performance by an application of data mining techniques. In *International conference on management and artificial intelligence* (Vol. 6, pp. 110–114).

23. [23] Tomar, D., & Agarwal, S. (2013). A survey on data mining approaches for healthcare. *International Journal of Bio-Science and Bio-Technology*, 5(5), 241–266.
24. [24]T. Barnes, M. Desmarais, C. Romero, and S. Ventura, presented at the 2nd Int. Conf. Educ. Data Mining, Cordoba, Spain, 2009.
25. [25]V.O. Oladokun, A.T. Adebajo, and O.E. Charles- Owaba. (2008) “Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course.” In *Proc Pacific journal of science and technology* 9.
26. [26]Yu-Wei, C., & David, C. (2015). *Machine learning with R cookbook: Explore over 110 recipes to analyze data and build predictive models with the simple and easy-to-use R code*. Birmingham: Packt Publishing.
27. [27]Yadav, S. K., Bharadwaj, B., & Pal, S. (2011). Data mining applications: A comparative study for predicting student’s performance. *International Journal of Innovative Technology and Creative Engineering*, 1(12), 13–