Prediction of Learner’s Performance in Adaptive E-Learning System using Learning Analytics

Soni Sweta¹, Shalini Mahato², Laxmi Kumari Pathak³

¹Department of Computer Science & Engineering Amity University Jharkhand, Ranchi, soni.sweta16@gmail.com
² Department of Computer Science & Engineering Amity University Jharkhand, Ranchi, smahato@rnc.amity.edu
³Department of Computer Science & Engineering Amity University Jharkhand, Ranchi, lkpathak@rnc.amity.edu
E-mail: soni.sweta16@gmail.com

Abstract. The main objective of Learning Analytics is to collect, interpret and investigate the information for setting proper co-relation to improve the students’ learning experiences. As the popularity and demand of learning analytics are on rise, higher education is continuously moving from offline to online E-Learning Educational System. It has been further promoted on account of COVID-19. Learning pedagogies, evaluation measures and feedback measures have also been changed accordingly. The aim of this paper is to discuss the efficacy of learning analytics in online educational system using e-learning platform. R programming language with GGPLOT2 is used as Visualization tool to focus on gain insights of adaptation of learning analytics. The results show the better computational performance in terms of predicting students to improve their learning achievements and mitigating risk of failures.

1. Introduction

Learning Analytics have huge recognition and acceptance among teachers and analysts in higher education[1]. Using Educational Data Mining (EDM) and Learning Analytics (LA) in E-learning System, we can develop learner’s centric model that is useful in taking decision and making strategy for continuously ensuring improvement of the organization[2].The primary idea of learning analytics is very interesting. Its ideas and methods can be used by educators to predict their learner’s outcomes by using EDM in advance and take right decision. It is very useful for schools, colleges and universities for taking better decisions, accumulating huge data and its information and applying analytical tools to maximize decision making capability in terms of resource utilization, students’ achievements and outcomes[3]. Learning Analytics is an interdisciplinary research area because of its linkages with virtual learning and digital teaching that is useful in EDM, e-learning system and its statistics. The huge data generated during online teaching sessions are gathered, examined and then analysed the patterns and behavioural trends to prepare feedbacks on behavioural traits[4]. These mined "noteworthy knowledge" supports the teaching and learning and recommends personalized and customized information. Online virtual education systems support the future growth and development in virtual university and adapt to handle the uncertainty like COVID-19 pandemic[6].Higher education needs to enhance infrastructure and networking; so that more accountable financing models and guaranteed educational system can be built in view of operational effectiveness, accepting adoptive changes and competitive at all levels. E-learning education not only manage extravagant overheads, but also understand the actual needs of students who comprehend the requirements of the organization[5].

In this challenging digital era, more institutions use techniques for Learning Analytics so as to acquire discoveries on the research and development for analysing some interesting patterns and finding for betterment an advancement of learning process. Further, it is a suitable tools for analysing pattern in students learning behaviour to provide the support and feedback to educators or tutors[7]. The utilization of Learning Analytics in Universities is to improve the inter institutional relationship, co-operation and collaboration, which motivate for development of agenda for large group of students and teacher community[8].
This paper’s Second section is about review literature on LA in higher education, explained the proposed methodology in third section and discussion followed by conclusion discussed in fourth and fifth section, respectively.

2. Literature Review on Learning Analytics in Higher Education

We thoroughly examined various papers on LA in the related research work and gained insights into the pattern and trend. The description about LA and its increasing popularity has discussed in Horizon Report 2012 by Johnson et al[9], as “the estimation, gather, examination and visualizing of data and information about students and their usefulness for ability of comprehension and upgrading learning and the environments in which it happens”. In another way, we can express as “the utilization of intelligence, learner-centric information, and analysis models to find new patterns among hidden information to predict the learning outcomes and give the feedback” in a journal[10]. A model was discussed for the identification of students at risk based on their risk parameters, and mitigation of students risk to enhance learning awareness and benefit both students as well as teachers[11]. This paper discussed the current scenario of LA and revealed its effectiveness in higher education[12]. In this paper, dynamic testing for the development of cognitive skills and its prediction is very important in comparison to the existing set of learner’s characteristics[7]. In this paper, learners’ behaviours were identified and matched with their profile and their attributes by developing a model on the basis of algorithms similarity computation and Jaccard coefficient which helped in the improvement and optimization e-learning[13]. A review of the developments in the field of educational data mining, learning analytics are explained in big data scenarios briefly and with some models used are: multilayer perception (MLP) network, multiple linear regression, radial basis function (RBF) network and support[8]. The authors have developed a learning analytics model in which dynamic interaction among different stakeholders is proposed with the help of visual analytics. It mainly focused on personalized learning and supported to increase retention rate[6].

3. Methodology and Execution

3.1 Learning Process

Authors Campbell and Oblinger in 2007[14] said that, there are five steps of learning analysis process, shown in Fig. 1. Capturing real time data and gathering information from various sources like virtual learning environment, LMS. It contains Blog, Forum and Chat etc like features. There is a reporting mechanism in which the collected data are used to make a model to enhance the students’ progress. Before making any prediction, first we need to visualize the data set which gives the relationship among variables then we make a model based on high impact variables. Predicting, the information is utilized to recognize the predictor for learner’s success and outcomes. Further, it is used for making decision of courses which utilized by the decision maker of the organization[15]. Outcomes from data analysis process are useful for interventions in acting steps e.g providing extra support to those learners who may have more chances of failure or at risk in the course. In refining, collected information are used for continuous support and improvement in teaching learning process. All the stakeholders need to discuss the idea and opportunities on how to incorporate all these possibilities in their research and development. Even though the main aim of LA is to focus on learning process, but its result shows the benefits of all types of stakeholders associated with the system.

Figure 1. The five steps of the Learning Analysis process.
3.2 Different Stakeholders

According to Romero and Ventura in 2013[9] there are four classes of stakeholders which were divided on the basis of their own perspective, aim and objectives. Learners Group supports the learner with adaptive feedback, personalization, customization of the responses according to their preference, the learning performance improvement. Teacher Group, which understands students’ learning process, reflects on teaching pedagogy and performance, understands cognitive and behavioral aspects. Researchers use the right educational data mining technique to find the hidden information which is beneficial for decision making strategy and evaluation of learning effectiveness in different challenging environment. Administrators, measure educational growth, do proper resource utilization, and generate educational offers.

In our research, we have considered a new type of stakeholder named as Mentor. Mentor bridges the gap between learners and teachers, parents and teachers and learners and institutions and counsels them to mitigate the learning gap. Mentor does mentor all the time during the course of their mentee. It will improve the overall personality of mentee and their performances.

3.3 Task to be performed - Algorithm and Implementation

We consider some factors in e-learning system which highly impact on achievement and outcomes of learners like using R programming language with DYPLYR package along with corr function. Then we get the significant value which show which variable have high impact over performance of students. Once we find then we compare it with the threshold value. Those students have less score then threshold value they will be at high risk. The threshold values will be decided by the faculty. It may vary course to course because of different courses have different difficulty levels as provided by the institution.

We have taken 140 students as a sample for e-learning courses. For academic achievements we have analysed their marks obtain in quiz, project-based assignments, and participation in forum. For checking the engagement level, we analysed their behavioural patterns like time stamp, no. of attempts, time spend on completing the session. The academic marks obtained in multiple modes of exams and their behavioural values as mentioned in table 1 and LMS tool given below in figure2.

![Moodle Virtual Learning Management Tool](image-url)

**Figure 2.** Moodle Virtual Learning Management Tool.

The SVM is a supervises learning method used to separate two group classification problems by classification algorithms. The support vector machines technique is developed by Cortes & Vapnik [15]. It is used to separate two groups of data using hyper plane as defined by the support vectors. The separating hyper plane is decision boundaries which classify one type of data in one side and another type of data in another side. Many hyper planes can be drowning for the separation of data, but optimal line defines data in a better way. It helps in finding maximum and minimum of the margin between two classes of training data set to find out position of data in X-Y coordinates which is shown in figure 3.

Real world problems are generally non-linearly separated in which SVM separates data and minimizes misclassification. In this research work, there are huge amount of nonlinear data generated by LMS; so
SVM is the best way to classify these training datasets into two groups one is called risky and another one is non risky classes.

Figure 3. Support Vector Machine Technique.

4. Result and Discussion
In this research work, we have focussed mainly on time variant data features collected during e-learning course where 140 students were participated in C++ course. In which data related to learning objects like multiple choice test were conducted to measure theoretical knowledge, project work-Assignments were conducted to measure practical knowledge and semester end exams were conducted to see the differences in performance over a period of times. Participation in forum and chat also gave the interactive and engagement level/ sensitivity of students. Weightage of test is shown in table -1

20% project-work assignment,10% forum/group activity for engagement level,20% for MCQ test, 30 % for end semester exams, 20% for demographic data in which we have 10% weight and 5 % weights for both communication medium and parental income. If students score more than threshold limit of 50% his or her chances will be higher to fall into non risky otherwise it will be risky. We can configure same data set with decision tree algorithm for comparing studies and find that SVM predicts with high percentage of accuracy as compared with decision tree.

\[ \text{Accuracy} = \frac{\text{true\_predicted\_at\_non\_risk} + \text{true\_predicted\_at\_risk}}{\text{total\_number\_of\_learners}} \]  

Table 1. Learner’s features from data set for training and testing purpose

| Based on | Features                      | Value range | Threshold value (%) |
|----------|-------------------------------|-------------|---------------------|
| Time Dependent | Multiple choice question | 0-30        | 15                  |
|           | Project work assignment      | 0-20        | 10                  |
|           | End semester exam            | 0-20        | 10                  |
|           | Participation in forum/chat  | 0-10        | 5                   |
| Demographic | Area (Rural, Urban)         | 0-10        | 5                   |
|           | Communication Medium (Basic, high, proficient) | 0-5 | 2.5 |
|           | Parental Income (Low, Medium, High) | 0-5 | >0 |
Table 2. Learner’s risk and non-risk cases examined

| Class    | Marks obtained in percentage |
|----------|------------------------------|
|          | MCQ Test | End Term Test | Project-work assignment | Participation in Forum/Chat |
| Risky    | 12       | 7             | 6                        | 2                           |
| Non-Risky| 21       | 16            | 18                       | 6                           |

Figure 4. Accuracy results on the 140 students of the online course with both the method.

5. Conclusion
This paper presents a method of prediction of learner’s performance in adaptive e-learning system using one of the machines learning techniques i.e. support vector machine. The proposed method used learner’s demographic data along with data extracted from different activees performed by learners and learning process during the continuation of the course. It takes log data from the learning management system and finds patterns from the adaptive course to estimate where learners are lagging and how the proposed method helps them to enhance their performance dynamically. The SVM technique used to determine the prior demographic information and dynamics inputs during adaptive e-learning in terms of feedbacks, special assistances as personalised recommendations helped learners to prevent failures and enhance performance.

6. Future Aspects
In future other machine learning techniques may be applied to fine tune the model in terms of accuracy, precision, and sensitivity.

7. References
1. Jo IH, Yu T, Lee H, Kim Y. "Relations between student online learning behavior and academic achievement in higher education: A learning analytics approach", Lect Notes Educ Technol. Emerging Issues in Smart Learning, Springer-Verlag Berlin Heidelberg, 2015,pp 275-287 DOI 10.1007/978-3-662-44188-6
2. S. Sweta and K.Lal, “Adaptive and Personalized Intelligent Learning Interface (APIE-LMS) in e-learning System,” International journal on Applied Engineering and Research: ISSN 0973-4562 Volume 10, Number 21 (2015) pp: 42488-92.
3. Peña-Ayala A. "Learning analytics: A glance of evolution, status, and trends according to a
proposed taxonomy", Vol. 8, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. Wiley-Blackwell; 2018.

4. Lang C, Siemens G, Wise A, Gašević D,"Research A. Handbook of Learning Analytics SOCIETY for LEARNING", ISBN: 978-0-9952408-0-3,2017, DOI: 10.18608/hla17, 2017; www.solarresearch.com

5. Baker, R., & Siemens, G. (2014)."Educational Data Mining and Learning Analytics", In R. Sawyer (Ed.), The Cambridge Handbook of the Learning Sciences (Cambridge Handbooks in Psychology, pp. 253-272).Cambridge: doi:10.1017/CBO9781139519526.016

6. S. Sweta and K.Lal. “Personalised Adaptive Learner Model in E-Learning System Using FCM and Fuzzy Inference System” International Journal of Fuzzy Systems (IJFS), Springer, ISSN No 2199-3211, 2016 doi: 10.1007/s40815-017-0309-y.

7. S. Sweta and K. Lal, “Learner Model for Automatic detection of Learning Style Using FCM in Adaptive e-Learning System” International Organization of Scientific Research (IOSR) JCE eISSN -2278-0661, vol. 18, Issue 2, Ver.1 pp: 256–270, 2016.

8. Tsai Y-S, Gasevic D."Learning Analytics in Higher Education-Challenges and Policies”, Proceedings of the Seventh International Learning Analytics & Knowledge ConferenceMarch 2017 Pages 233–242, doihttps://doi.org/10.1145/3027385.3027400.

9. Romero-Zaldivar V, Pardo A."Monitoring student progress using virtual appliances: A case study",Computers & Education,Volume 58, Issue 4,2012,Pages 1058-1067,ISSN 0360-1315, https://doi.org/10.1016/j.compedu.2011.12.003.

10. Leitner P., Khalil M., Ebner M."Learning Analytics in Higher Education—A Literature Review", In: Peña-Ayala A. (eds) Learning Analytics: Fundaments, Applications, and Trends. Studies in Systems, Decision and Control, vol 94. Springer,Cham..2017 https://doi.org/10.1007/978-3-319-52977-6_1

11. Larrabee Sønderlund A, Hughes E, Smith J." Student engagement and wellbeing over time at a higher education institution", A systematic review. Br J Educ Technol. 2019;vol.50(5),pp 2594–618. DOI:https://doi.org/10.1371/journal.pone.0225770

12. Sønderlund AL, Hughes E, Smith JR."The efficacy of learning analytics interventions in higher education-A systematic review", British Journal of Educational Technology, Vol 50 No 5 pp 2594–2618 2019, DOI: 10.1111/bjet.12720

13. Liang, Kun & Zhang, Yiyong & He, Yeshen & Zhou, Yilin & Tan, Wei & Li, Xiaoxia.." Online Behavior Analysis-Based Student Profile for Intelligent E-Learning", Journal of Electrical and Computer Engineering. 2017. 1. T. DOI: 10.1155/2017/9720396

14. Campbell J, DeBlois P,"Academic analytics: A new tool for a new era", EDUCAUSE Review, vol. 42, no. 4,pp 40-57, 2007; https://er.educause.edu/articles/2007/7/academic-analytics-a-new-tool-for-a-new-era

15. Cortes C, Vapnik V. Support-vector networks. Mach Learn 20(3), pp: 273–97, 1995 Sep;20(3), DOI: https://doi.org/10.1007/BF00994018