Predictive Analytics for Students' Performance Prediction

Palwinder Kaur Mangat, Kamaljit Singh Saini

Abstract: ***Personalized learning is being popular due to digitizations that enable a large number of technologies to support it. To predict students' learning abilities, it is necessary to estimate their behavior to know about their weaknesses and strengths. If it is possible for teachers to predict in advance at-risk and dropout students, they can plan more effectively to handle them. We are describing in this paper various intelligent tutoring systems with Educational Data Mining, Predictive Learning Analytics, prediction of at-risk students at an earlier basis, how this prediction task is done. We are describing various prediction models that can be used to predict students' behavior and how portable these predictive models are and the various risk prediction systems that are being used.***

Keywords: Predictive Learning Analytics, Intelligent Tutoring Systems, Student Risk Prediction, Risk Prediction Systems, EDM, Early Warning Systems (EWS).

I. INTRODUCTION

This research work is focused on intelligent tutoring system with educational data mining, predictive learning analytics, prediction models and risk prediction systems. With the modern technologies the way of teaching and studying have totally changed. Now the online and blended courses are being popular. Most of the institutes are teaching through the virtual learning environments. Using Learning Management Systems students' behavior is monitored. LMS can be used [1]. LMS helps in learning and teaching. In the previous paper, those students of K-12 students have been predicted that have poor academic performance. A framework to predict students with poor academic performance have been developed. Study on every level of curriculum have been done that is very useful for teachers to make effective interventions for helping students. Teachers can easily predict specific areas about students that need to be carefully observed. Historical data about students have been used about their academic performance to predict at-risk and no risk students [2]. Many educational institutions are using Learning Management System (LMS) and analyzing the massive amount of data store d in it to make decisions based on this. Administrators, instructors and students are taking benefit from this tool. This tool is very beneficial for the institutions that are fully online by identifying the at-risk students and providing personalized assistance to such students. The model that have been described for predicting students at risk is Gradual At-risk (GAR) and for automatically gathering data about students from LMS final early warning system (EWS) have been proposed [3].

In free, open online courses learners have different motivations for enrollment so average transfer of knowledge is acceptable, but in the field of corporate training where corporates spend millions for employees training their expectations are more which are not being satisfy by most of the training courses. So Weiyu Chen in his paper proposed a predictive analytics for improving learning outcomes of the learners [4] .With the advancement of digital technology, it had made possible to collect massive amount of data and analyze that data which is very beneficial for instructors, students and as well as administrators to. This data can be further used with predictive analytics to predict the performance of students.

II. ITS WITH EDM

After the appearance of Intelligent Tutoring Systems in1970s they became active in instructing and training instructors and students. ITS are educational software with artificial intelligence [5]. ITS use intelligent search agents for searching distributed knowledge bases and providing suitable knowledge to users. They use ontology environment which consist of two parts – first is ontology knowledge base systems and second is ontology layer. This multi-domain ontology consists of an ontology and multiple knowledge bases. There is a mapping relationship between multiple knowledge bases and between each ontology. Unnecessary technique details are hidden from the users and contents of the knowledge base are public to the users. [6].

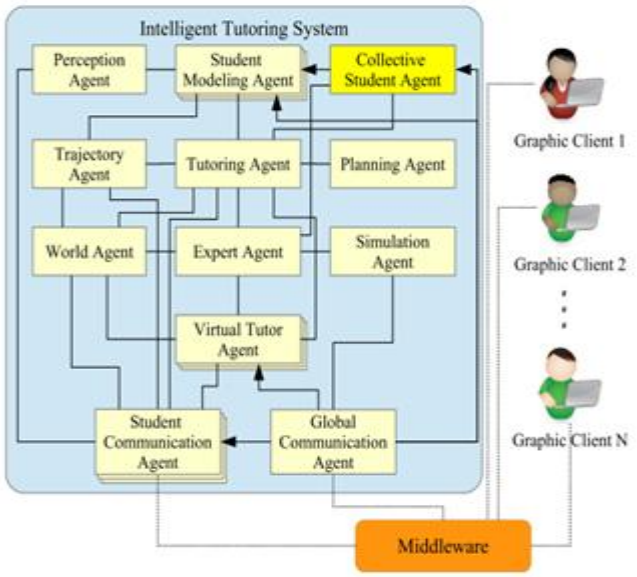

Figure 1 Intelligent Tutoring System ontology environment [6]

Revised Manuscript Received on August 29, 2020.

Palwinder Kaur Mangat, Assistant Professor, Department of Computer Science, National College for Women, Machhiwara, Punjab, India.

Kamaljit Singh Saini, Assistant Professor, Department of Computer Science, National College for Women, Machhiwara, Punjab, India.

Intelligent Tutoring Systems being used with Educational Data Mining are Cognitive Tutor Algebra, Assistant that help students in solving math's problems. Educational Data Mining is applied on these Intelligent Tutoring Systems to create models to detect students' feelings and behavior. For procedural training CandarmTutor is used. An ITS architecture is MAEVIF which is a multi-agent architecture and is used in learning environments. This architecture can be used to know about the learning status about the students. It also provides behavior prediction of students. The main disadvantage of this architecture is the detailed knowledge of the background of students to predict students' actions [7]. ITS is a web-based tutoring system which focus on the students' individual information. It helps students in collaborative learning where students can acquire knowledge by choosing from different learning. Through discussions students can answer the questions and also can ask too. Student model of ITS stores information about students as well as the feedback from the students. Student model can be categorized in two parts one analyze information and the other part share the information. This model provides a way of collaborative study, decision making and solving problems together. This model provides a better way to learning lessons and helps students in learning [8].Various types of ITS have different methods for modeling and for taking feedback of the students and it is also having various methods for how it is triggered to the students. Some important types of intelligent tutoring systems are- First system is the Model Tracing Tutors which take and check student responses on the defined problem. Model Tracing Tutors is of two types- Example tracing tutor and Cognitive tutor. Second system is the Cognitive Tutor which is for instruction purpose. It basically focuses on the knowledge of the instructors. Third system is the Dialogue Based Tutor which is capable to communicate with the students. It is having pre-ordered script [9]. It helps parents and instructors to take suitable intervention by predicting student's performance. Amrita Learning System is an Intelligent Tutoring System which helps tutors to predict the students' performance on the basis on assessments which can be formative or summative [10].

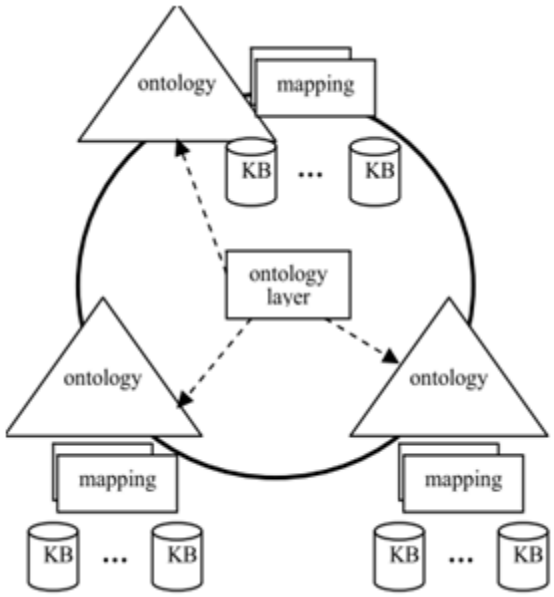

Figure 2 MAEVIF Architecture [7]

Agent technique is very beneficial in intelligent tutoring system. Agents can perform the role of leaders, students, teachers, collaborators, managers and valuators. Agent based ITS plays a very crucial role in real-time supervision, information filtering, analyzing studying and teaching, collaborative study and intelligent reasoning [11]. Multi-agent intelligent system consists of information agent, domain knowledge, pedagogical agent, user model and user interface. This system has best interaction between learner and the system because user models are built on users' knowledge levels and the learning styles [12].

Popular mode of education is full time studies where the students can directly interact with the students. The tutor provides teaching material to students and make their assessment through summative and formative exams. The most important part of education is tutor due to that computerized tutoring systems are providing the students with the services of the tutor. The aim of these intelligent tutoring systems is to simulate tutors with the use of artificial intelligence. The main parts of intelligent tutoring systems are student, teacher and the expert. The basic architecture of ITS consist of students, communication module, additional components and the basic components like student module, expert module and the tutoring module [13].

III. STUDENTS' GROWTH, PERFORMANCEVAND PREDICTIVE LEARNING ANALYTICS

Students' growth is described as change in students' achievements in studies in a certain period of time. For example, students taking grade 1 to grade 2 in examinations. It is also sometimes referred as 'value-added'. Based on this several models have been developed. Student performance is students' educational performance that is measured by their scores. Student growth is measured in total scores SGTS and subject score SGSS. SGTS measures students overall performance and SGSS measures performance in subjects [14].

Learning analytics are best way to analyze students' behavior and predict at-risk students and help them. Clickers are a source of data for LA which are used to promote interactions in classrooms. But clickers are costly because these require response devices for students and receiver. The cost-effective way to use clickers is with use of students own mobile devices. By using the web-based services data can be obtained at low cost [15]. LA are data driven and offline and online learning can be used for theoretical arguments. study shows that same set of variables cannot be used in multiple courses to predict students' behavior [1].

Predictive learning analytics is a research area which help instructors to improve the course quality. It provides help to instructors in predicting student drop-off rates, exam performance and grades in exams which have great impact on learning experience. Most Predictive learning analytics have been developed for MOOC. But now the online courses are collecting behavioral characteristics of students and their Social Learning Networks. With the help of click stream and SLN, PLA methods and models for short – courses can be designed [4].

Predictive analytics not helps only in students' performance prediction they also provide a way to find dropout students and the preventive measures to stop them from leaving on-line courses. Students dropout patterns are found out with the recording of progress of students. Then this data is analyzed and patterns are found and interventions are suggested to prevent the students to leave the study. This is done with the intelligent tutoring systems [16].

Student dropout rate in Massive Open On Line course is also a great problem, so it is necessary to predict those students who have low performance and can dropout the course in the future. So, the association analysis has been used in the paper to predict students at risk doing online courses. Apriori principal is the best to use with this algorithm which reduces computational complexity. Depending on the type of the learner whether the learner is at ease to learn by listening only audio, by watching video or by practicing the assignments, suggestion is given to them for the way of studying. But it is completely depends on the learner to follow the instructions of predictive analytics or not [17]. If the students who are going to dropout are predicted in advance measures can be taken by the instructor to improve his /her study related problems.

This scenario is not visible in MOOC only it is also present in higher educational institutions. Students doing degree programs also drop out the course sometimes. So the institutes seek for such predictive analytics that can predict at risk students at early stage so that desired interventions can be taken to improve students' performance and to decrease their dropout rate. So, the data and assignment scores about students were taken from virtual learning environment. This was further analyzed to check the student at risk who can left the course. comparative analysis of four predictive models was done where the models were Random Forest, K- Nearest Neighbor, Naïve Bayes and Linear Discriminant Analysis [18]. As learning analytics are very crucial in predicting students' performance they can be further used to measure the institutional strategic goals and help in the achievement of these goals [19].

IV. STUDENT RISK PREDICTION

For predicting the academic risk of students, the steps have been followed- firstly the data have been described, secondly the problem has been defined and thirdly the specific approach to solve the problem have been described. Data have been collected from GCP School and data about students and teachers have been recorded in tables in database and data have been collected about tests about students on state and national level. Based on these tests subtest level performance risk and strand level performance risks are predicted [2].

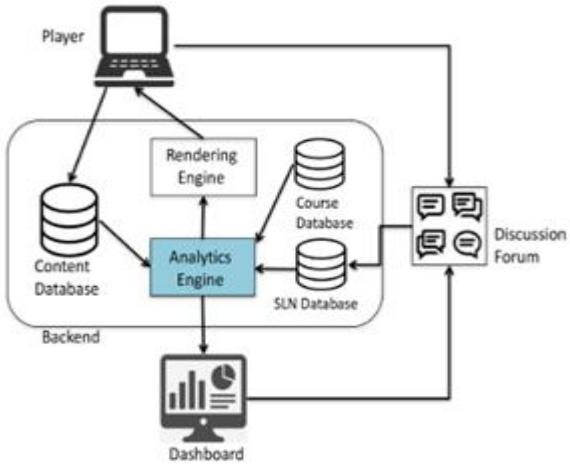

Figure 3 System for prediction [4]

Prediction of students' educational performance is one of the objectives of Educational Data Mining (EDM) to predict at risk students who are going to fail in a course. Clickers are a way to promote interaction in classroom. Students engagement increases their subject knowledge, level of comprehension and also the teacher insight about students' difficulties. Due to availability of mobiles at low costs and free cloud-based tools clickers are becoming more popular. Learning analytics cycle's final stage is intervention which is the most challenging both for teachers and students because it requires metacognitive skills. That is a real problem for students who are already weak and face problem with intervention and teachers also have to spend more time on such students to take remedial actions [15].

Predictive analytics to improve students' attrition used intervention strategy. This consist of peer-to-peer communication and its implementation is known as campaign. Four campaigns were held in each semester to increase students satisfaction towards the university [20].

Prediction of students' performance is done in three categories - End of course prediction, Mid-term prediction and Day by day prediction.

V. PREDICTION MODELS FOR EARLY WARNING SYSTEMS AND THEIR PORTABILITY

Development of predictive models is increasing at a high speed from the last decade to predict students' performance. The main part in this type of research is to explore those variables that are best suited for the model. Different classification algorithms have been analyzed through proposed predictive models like Decision Tree, Hierarchical Mixed models, Naïve Bayes, K- Nearest Neighbors, Support Vector Machine, Neural Network Models, Logistic Regression and Additive Regressive Trees [3].

In on line short courses for prediction of outcomes the system developed have four parts- 1. Course player, 2. Dashboard, 3. Backend, 4. The discussion forum.

Each course in this system is divided into different modules which contain one or more units. Then this course is delivered to the learners. Player is being used to deliver the course via units which contain one or more content files.

These files can contain various sources of learning lecture notes which can be in pdf format, ppts and video lectures. The Dashboard is divided into overview, engagement and content tabs. Discussion forum is divided in to threads. This system is integrated with an open source discussion forum platform NodeBB. All the learning resources are stored in course database and the learner's activities are stored in content database. The job of the rendering engine is to fetch information from the content database after the work of analytics engine that performs the computations on this database. The work of SLN database is to send the results to the dashboard for visualization [4].

Early Warning Systems (EWS) are predictive models for finding at risk students and taking measures to improve their performance before the final exams. Following figure [3] describes the various services that contribute to run the predictive models. Dashboards provides analytical information to instructors, students. Two more services are also their messaging service and feedback service. Two data sources are being used in this model data mart and CAR. Data mart is to train the predictive models and CAR is to collect grades of current semester of students. Based on the scheduled time and availably of data a scheduler triggers each predictive model [3]. Following are some of the models-

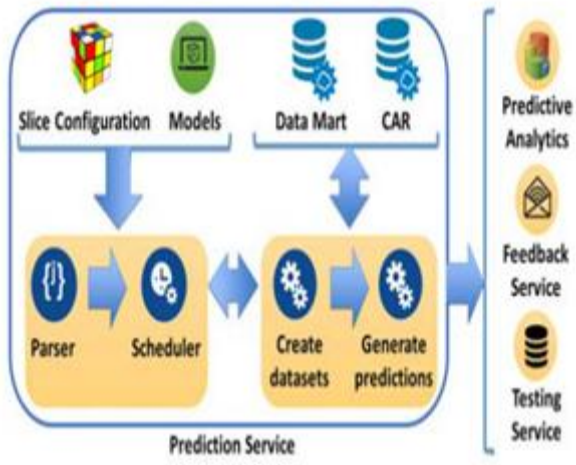

Figure 4 Components of EWS [3]

A. Collective Student Model

This model is based on the historical data of the students. Students are categorized into different categories depending on their performance in practical assignments. Clustering technique is applied on students logs to complete the assignments in groups [7].

B. Gradual At-Risk Model

The Gradual At-Risk model is a collection of predictive models. Theses predictive models are sub models which are based on the grades of students in a specific course. The objective of the model is to predict at risk students. This model is called gradual because the assessment activity is graded by applying the respective sub model. Weka software was used for data processing and operations and for classification J48 classification algorithm was used. Analysis of GAR model was done by the Naïve Bayes classifier [3].

Overlay Model, Buggy Model and Cognitive Model are student model for predicting student performance.

To reduce the time to complete a course, universities need predictive models that can predict in advance the successful and unsuccessful students. For this purpose, three predictive models were developed that works on the previous data of the students taken at the time of enrollment, at the end of the first semester and at the end of the academic first year [21]. Students success predictive models can be created on the basis of the selected input parameters at the time of their admission in the institution. This will decrease the dropout rate of students [22] .

Portability of the prediction models means whether the same model can be used by the different courses or different institutions. this issue was recognized with the initiative of the Open Academic Analytics Initiative. The objective of this was to develop an open source model for predicting students' performance. Study shows that even a number of predictor variables are being used and an institution is using a single LMS the portability of model will be low [1].

VI. RISK PREDICTION SYSTEM

Open university or distance learning institutions prefer analytics technique to improve students experience. The knowledge about students learning activities is held in various data sources. The aim is to analysis this data which contains final exams and tutor marked assessments for predicting student failure from the virtual learning environment. VLE is used to deliver contents to the students and data related to their clicks is recorded. On the basis of this data students' performance drop and final outcome is found out [23]. In merged students record data is data about students that is taken from school data warehouse. Prediction task controller contains task lists and configuration file which provides input to it. Prediction task is in tasks list table. It is read by the task controller and model is built. Model building block is invoked for each task that is in tasks list table by the prediction task controller. Then risk prediction block performs tasks like – prepare task data extracts data, process task data performs processing on the data and handles missing data, score model and load prediction model for prediction, prediction output provides the output. Finally, this output is sent to prediction output table [2].

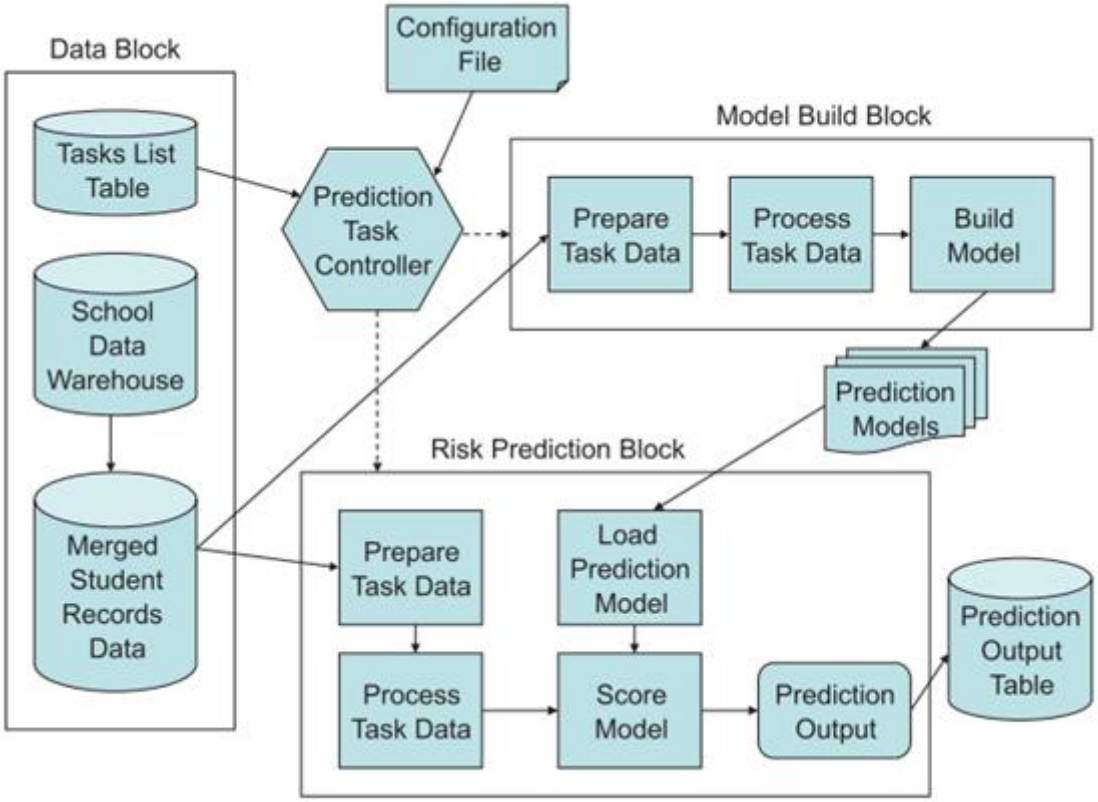

Figure 5 Risk Prediction System [2]

VII. CONCLUSION

In this paper we have described the intelligent tutoring system with their architecture and how it is beneficial in educational data mining. It has been described how intelligent agents are used in these systems and which methods are being used. These systems with predictive models are used in different types of courses like full time courses, on-line courses, distance education and part time courses. Predictive analytics predict students at risk and helps tutors to take interventions to improve the performance status of them. Predictive analytics help in performance prediction as well as to decrease dropout rates of students from different institutions or courses. Various prediction models can be used to predict students' behavior like collective student model, gradual at-risk model, Overlay Model, Buggy Model and Cognitive Model. Early warning systems are a good measure to predict at-risk students and to take actions to improve the grades or scores of students. Risk prediction is done with the risk prediction systems that are made of with different components. In the future, we would like to do comparative analysis of various prediction models using different classification algorithms.

REFERENCES

1. R. Conijn, C. Snijders, A. Kleingeld, and U. Matzat, "Predicting Student Performance from LMS Data : A Comparison of 17 Blended Courses Using Moodle LMS," vol. 10, no. 1, pp. 17–29, 2017.
2. U. States, "On early prediction of risks in academic performance for students," vol. 59, no. 6, pp. 1–14, 2015.
3. D. Baneres, M. E. Rodríguez, and M. Serra, "An Early Feedback Prediction System for Learners At-risk within a First-year Higher Education Course," vol. 1382, no. c, 2019.
4. W. Chen, C. G. Brinton, D. Cao, A. Mason-singh, C. Lu, and M. Chiang, "Early Detection Prediction of Learning Outcomes in Online Short-Courses via Learning Behaviors," vol. 1382, no. c, pp. 1–14, 2018.
5. C. Ramos, C. Frasson, S. Ramachandran, and C. Ramos, "Introduction to the Special Issue on Real World Applications of Intelligent Tutoring Systems," IEEE Trans. Learn. Technol., vol. 2, no. 2, pp. 62–63, 2009.
6. S. U. N. Yu and X. U. Tianwei, "Intelligent Search Agents for Web-based Intelligent Tutoring Systems," pp. 1148–1151, 2008.
7. D. Riofr, J. Ram, and M. Berrocal-lobo, "Predicting Student Actions in a Procedural Training Environment," vol. X, no. c, pp. 1–13, 2017.
8. [8] L. I. U. L. Wu, M. W. Hua, I. E. College, and I. Technology, "Design an Applied Student Model for Intelligent Tutoring System," pp. 2–6.
9. B. I. Mohammad, S. I. Shaheen, and S. A. Mokhtar, "Novel Online Tutor Modeling For Intelligent Tutoring Systems," 2001.
10. G. Gutjahr, K. Menon, and P. Nedungadi, "Using an intelligent tutoring system to predict mathematics and English assessments," Proc. - 5th IEEE Int. Conf. MOOCs, Innov. Technol. Educ. MITE 2017, pp. 135–140, 2018.
11. S. Ke and X. Lu, "Study on intelligent tutoring system based on multi-agents," Proc. - 2010 6th Int. Conf. Nat. Comput. ICNC 2010, vol. 6, no. Icnc, pp. 2948–2952, 2010.
12. S. U. N. Yu, "A Multi-Agent Intelligent Tutoring System," pp. 1724–1728, 2009.
13. R. Rollande and J. Grundspenkis, "Representation of study program as a part of graph based framework for tutoring module of intelligent tutoring system," 2012 2nd Int. Conf. Digit. Inf. Process. Commun. ICDIPC 2012, pp. 108–113, 2012.
14. T. A. O. Ren and J. Xiao, "The Influence of Student Abilities and High School on Student Growth : A Case Study of Chinese National College Entrance Exam," vol. 7, 2019.
15. S. P. M. Choi et al., "International Forum of Educational Technology & Society Learning Analytics at Low Cost : At-risk Student Prediction with Clicker Data and Systematic Proactive Interventions Published by : International Forum of Educational Technology & Society Linked refe," vol. 21, no. 2, 2018.

16. R. M. M. F. Luis, M. Llamas-Nistal, and M. J. F. Iglesias, "Enhancing learners' experience in e-learning based scenarios using Intelligent tutoring systems and learning analytics: First results from a perception survey," Iber. Conf. Inf. Syst. Technol. Cist., pp. 1–4, 2017.
17. M. Srilekshmi, "Learning Analytics to Identify Students at-risk in MOOCs," pp. 194–199, 2016.
18. R. Umer, T. Susnjak, A. Mathrani, and S. Suriadi, "A learning analytics approach : Using online weekly student engagement data to make predictions on student performance .," 2018 Int. Conf. Comput. Electron. Electr. Eng. (ICE Cube), pp. 1–5, 2018.
19. J. Jonathan, S. Sohail, F. Kotob, and G. Salter, "The Role of Learning Analytics in Performance Measurement in a Higher Education Institution," 2018 IEEE Int. Conf. Teaching, Assessment, Learn. Eng., no. December, pp. 1201–1203, 2018.
20. E. Seidel, "Using predictive analytics to target and improve first year student attrition," no. 2011, 2017.
21. L. W. Santoso, "Early Warning System for Academic using Data Mining," 2018 Fourth Int. Conf. Adv. Comput. Commun. Autom., pp. 1–4, 2018.
22. S. Milinkovi and V. Vujovi, "Students ' Success Predictive Models Based on Selected Input Parameters Set," 2019 18th Int. Symp. INFOTEH-JAHORINA, no. March, pp. 1–6, 2019.
23. A. Wolff, Z. Zdrahal, and M. Pantucek, "Improving retention : predicting at-risk students by analysing clicking behaviour in a virtual learning environment," pp. 145–149, 2013.

AUTHORS PROFILE

Palwinder Kaur Mangat received Bachelor of Computer Applications degree from Panjab University, Chandigarh in 2004, and Master of Computer Applications degree from Punjab technical University, Jalandhar in 2007. Worked as Assistant Professor in Computer Science in National College for Women, Machhiwara and G.G.S. Khalsa College for Women, Jhar Sahib for 10 years. She is currently doing Ph.D. Computer Applications from Chandigarh University, Gharuan. She has attended many national conferences and presented her papers. Her research interests include data mining, network architecture, the Internet of Things, artificial intelligence, big data.