INTERNATIONAL RESEARCH JOURNAL OF SCIENCE ENGINEERING AND TECHNOLOGY

ISSN 2454-3195

An Internationally Indexed Peer Reviewed & Refereed Journal

WWW.RJSET.COM
www.isarasolutions.com

Published by iSaRa Solutions

PREDICTION OF THE PERFORMANCE OF AT-RISK STUDENTS-AS SURVEY

1 Mrs.S.AMUTHA 2 .Dr.R.SUJATHA

1 Research Scholar Adaikalamatha Arts and Science College, Thanjavur

2 . Assistant Professor Department of Computer Science, Queens College of Arts & Science for Women Pudukottai.

ABSTRACT:

Numerous online learning platforms, including Massive Open Online Courses (MOOC), Virtual Learning Environments (VLEs), and Learning Management Systems (LMS), enable millions of students to pursue their interests in learning without regard to time or place restrictions. Online learning environments have numerous benefits, but they also have certain drawbacks, including low engagement, high dropout rates, low engagement, self-regulated behaviour, and forcing students to define their own goals. This study reviews various predictive model that assesses the issues encountered by at-risk students, enabling instructors to prompt intervention to persuade students to raise their study engagements and enhance their study performance. The prediction model can assist instructors in early identification of at-risk pupils for prompt intervention, preventing student dropouts. Our findings demonstrated that crucial aspects of online learning include the students' evaluation scores, the level of engagement (measured by clickstream data), and time-dependent variables.

Keywords: Predictive model, earliest possible prediction, at-risk students, machine learning, feedforward neural network, random forest, prompt intervention.

INTRODUCTION

The ability of educational systems to provide students with the necessary professional and career preparation for the 21st century has been a major source of worry during the past century.

Instructors can better comprehend and keep track of students' learning progress by evaluating and analyzing the student data produced by online learning platforms. The sooner student performance in virtual learning environments is identified, the easier it will be for the instructor to convince and caution students to keep them on the right path However, it would be more beneficial to forecast students' performance earlier in the course rather than after they have finished the course and taken the final test.

Early prediction of student performance can assist teachers in identifying students who require additional coursework, assignments.

Researchers have been able to create a number of predictive models to uncover hidden study patterns that explain the strengths and weaknesses of online students as a result of the development of artificial intelligence, machine learning, and deep learning techniques[1]. Researchers can utilize machine learning techniques to examine several factors that have a substantial impact on student dropout rates in an effort to lower dropout rates.

However, it is a difficult challenge to create a prediction model that can pinpoint the precise learning behavior of students earlier in the course by evaluating their behavior data. Machine learning (ML) techniques could assist in analyzing the variables that define students in an online learning environment where a lot of data is generated every day to produce results that better describe their learning behavior. As a result, machine learning may reveal information that is useful for both instructors and students [2][3].

By using learning approaches, predictive models can accurately depict pupils who are likely to stop studying, enabling teachers to take preventive action before dropout behavior happens. The main goal of our research study is to identify students who are at danger of dropping out as early as possible by using machine learning approaches to comprehend aspects related to students' learning behavior and how they interact with the Virtual Learning Environment.

The predictive model can help teachers intervene with students by sending them persuasional messages that urge them to stay on the correct path and prevent dropouts.

In earlier studies, researchers have employed a variety of feature selection strategies to increase the precision of performance prediction for students. Many studies manually choose the aspects from the originals that have a major impact on students' performance, drawing on the knowledge of education professionals [4][5]. These factors mostly consist of demographic data like gender, age, and nationality, as well as other aspects of prior academic success like test and curriculum scores. For automatic selection, many researchers also use feature selection techniques in machine learning. Gene Algorithm is another feature selection method that is frequently used by researchers to forecast student performance [1][6]. Some researchers employ dependency-based feature selection methods to build better subsets of original characteristics to increase prediction accuracy [7]. In their thorough application of five filtering techniques based on dependency, information gain, information gain rate, relief, and symmetric uncertainty, Xing et al. [8] determined the outcome of feature selection by taking the intersection of the outcomes of the five techniques. The findings of these studies demonstrate that feature selection techniques can significantly increase the precision of forecasting students' performance, but no comparative study of the many feature selection techniques employed in this sector exists.

In order to increase the precision of student performance prediction, we put forth the hybrid feature selection approach known as RnkHEU. This innovative feature selection technique first creates a set of candidate features using a Naive Bayes (NB) classifier before using a heuristic approach to produce the feature selection results.

LITERATURE SURVERY:

Online learning platforms start to produce a ton of data about student interaction, click-streams, courses, etc. Most literary studies focus on gathering data and forecasting students' success at the end of the semester. The findings from the studies were helpful in pinpointing the important factors that have the greatest impact on students' performance, but no recommendations on how to stop students from failing or dropping out were made. By examining characteristics and data from the very beginning of the course, a thorough predictive model could be created that would be useful in preventing failures and dropouts and enabling instructors to make effective interventions at the proper time.

Iqbal, Z., et al.[9 proposed an analysis of several studies that had been carried out to predict students' academic performance, either to simplify degree planning or to identify pupils at danger. In order to estimate how well students would succeed in algebra and bridge to algebra courses, Thai-nghe et al. suggested matrix factorization models. When there was little data and no task or background information for the pupils, factorization techniques were helpful. Data were divided into trainset and testsets. The information is derived from logs of conversations between pupils and computer-assisted tutoring programs. In order to predict student progress, Thai-nghe et al. expanded the research and employed tensor-based factorization. They developed a recommender system problem to model the issue of predicting student success and suggested tensor-based factorization methods to added the impact of student performance over time. As students interact with the system, logs of their successes and failures on exercises are saved. RBM performed better with rmsc=0. 3, mse=0. 09, and mae=0. 23.

Linear regression, logistic regression, random forest, K-Nearest Neighbor, and endpoint protection platform were utilized to monitor and forecast students' performance. Xu, Jie, and colleagues' ensemble-based progressive prediction method produced the best result with the lowest mean square error [10]. The restriction was stated as there was no discussion of an intervention approach and the courses predicted to the students were not implemented.

Martins, L. C. B., et al.'s gradient boosting machine, deep learning, and distributed random forest were the foundation for the study [11]. The deep learning model had a true positive rate of 71.1%, which was the highest. This study intends to demonstrate a classification using decision trees to foretell engineering course student avoidance in Brazil. Interaction with academics, the course syllabus, and problems with mental health were the biggest barriers to dropout.

A study on the level of engagement of students working on programming exercises in an intelligent tutoring system was proposed by Fwa, H. L., and Marshall, L.[12]. Students' head position, keystrokes, and activity logs were automatically recorded by the tutoring system and fed into a hidden markov model to determine their level of participation. When appropriate, intervention actions might be started automatically by the system thanks to the modeling of students' engagement on a moment-by-moment basis. This study was also one of the few studies that model involvement without using supervised machine learning methods or human data labeling.

Sckeroglu, B., et al. The study in [13] concentrated on student prediction without making any recommendations for improving the algorithm. Additionally, little to nothing had been done to demonstrate how machine learning had been applied to enhance taught outcomes while showing conceptual gaps. Furthermore, whereas empirical research has shown inconsistent results regarding which machine model was most effective in forecasting students' performance, the study merely offered the suggestion that the decision trees methodology was the best model to accomplish so. Hussain, et al. conducted research on an artificial neural network-based prediction model for student performance based on internal assessment. The artificial neural network used in this study produced a classification accuracy of 95. 34%. The artificial neural network used in this study produced a classification accuracy of 95. 34%. The performance of the artificial neural network was heavily influenced by the precision, recall, f-score, accuracy, and kappa statistics performance, which were determined as a statistical choice to find the best classification methods.

Gray, C.C. and Prekins, D. [14] suggested a study on current approaches and metrics for learning analytics that used the 97% accurate k-nearest neighbor method. Our research establishes a brand-new descriptive statistic for student attendance and develops a predictive model using cutting-edge machine learning tools and methodologies. The predictive model was only consistently applicable to students at Bangor University.

In a true online learning environment, Ortigosa, a., et al.[15] researched the work done to promote student dropout risk prevention. To achieve this, we created predictive models based on the c5.0 algorithm using information gathered over five years from more than 11,000 pupils. Then we created spa, an early warning system that generated both static and dynamically updated early dropout-risk estimates using these models. Additionally, it encourages the documentation of the ensuing retention-focused interventions for later research. Spa was currently in its fourth semester of continuous use after going into production in 2017. To forecast the dropout risk of more than 5,700 students, it has calculated more than 117,000 risk scores. In order to continue assessing the effectiveness of the retention measures used and the real-world performance of the generated prediction models. When designing techniques to increase retention elsewhere, we thought that our experience—including the insights discovered while putting spa into production—would be helpful to the community.

N. Wu, L. Zhang,et al. proposed predicting dropout in Massive Open Online Courses[16]; we proposed a deep neural network model that combined support vector machine, long short-term memory network, and Convolutional Neural Network. In addition to taking into account the effects of the sequential relationship between student conduct and class imbalance on dropout, our model's effective feature extraction method also reinforced the performance of dropout prediction..

Numerous studies using a range of statistical techniques looked at the effect of learner variables on Massive Open Online Course (MOOC) retention rates. Imran, A. S., et al.'s study [17] on predicting and explaining student dropout used a feed-forward deep neural network with an accuracy rate of >90%. Prediction was made following the conclusion of the course.

Liao, S.N., et al.'s [18] study of the predictions of student outcomes at universities and colleges looked at all levels: retention at the institution level, completion/graduation at the program level, success at the course level, and understanding at the module level within a course using support vector machine with 70% accuracy. varying predictors had been used at each level due to the varying timelines and outcomes of interest. Modeling approaches and the methods used to assess them also differ greatly. For many years, there had been extensive research in the field of forecasting student outcomes. Prior research had a strong emphasis on modeling student performance to increase retention rates and forecast exam results.

Hussain, M., et al.[19] investigated the difficulty of online learning in predicting students. The goal of this study was to foresee the challenges that students will face in a following session of a digital design course. We examined the information recorded by the machine-learned algorithm-based Digital Electronics Education And Design Suite (DEEDS) system, which is a Technology-Enhanced Learning (TEL) system. Artificial Neural Networks (ANNS), Support Vector Machines (SVMs), Logistic Regression, Nave Bayes Classifiers, and Decision Trees were among the machine-learned techniques that achieved the greatest accuracy of 75%. We would anticipate instructors reporting improved student performance during the following session because ANNs and SVMs could be readily integrated into the existing system. The study's limitations were stated as the low model accuracy and lack of student dropout prediction.

. Lee, S. and chung, J. Y. et al. presented Random Forest, Boosted Decision Trees (BDT), with BDT having the best accuracy of 99% [20]. The goal of the current study was to enhance the functionality of a dropout early warning system by (a) addressing the class imbalance problem using synthetic minority oversampling techniques (smote) and ensemble machine learning methods, and (b) assessing the trained classifiers using both Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves. A small network data source was employed, and not all features were taken into account for building the predictive model.

Chung, J., Y. and Lee, S.'s study [21]used random forest for the prediction of student dropouts. Numerous strategies make use of ensemble-based techniques. An ensemble-based strategy was created by J.Xu et al. [22] by merging a two-layered structure with numerous base predictors and a cascade made up of ensemble predictors. Based on their performance on performance exams, this methodology was developed to forecast pupils' future performance. A clustering technique based on a latent factor was utilized to build the base predictors. The precision of the prediction model's feature weight calculations has the potential to be poor.

Ai-Shabandar, R, et al. [23] offered a study based on student withdrawal and learnt achievements that employed a gradient boosting model with 95% accuracy. The study's main aim was identifying at-risk individuals for early, intense intervention in online courses. As one of the study's drawbacks, temporal characteristic was not taken into account, and no further datasets were used to validate the predictive model.

Figueroa-Canas, J. and Sancho-Vinusea [24] proposed using Decision Trees (DT) with more than 90% accuracy to identify dropout-prone students sooner in online statistical courses. Both the training set and the validation set were made up of students enrolled in the same academic year for the approach.

Understanding learned material and enhancing instruction had huge potential. But just creating novel models by ai and machine learning experts that predict labeled data more precisely won't bring about the promised. We proposed that learned engineering teams used their interdisciplinary knowledge to create explanatory learner models that offered interpretable and practical insights in addition to precise prediction. In order to construct better explanatory learner models that promote learnt science, produce better pedagogical practices, and clearly increase student learning, we advise learned engineering teams, shared infrastructure, and funder incentives. by Carolyn P. Rose et al. [25] Traditional machine learning models weren't very good at generalizing.

Long Short Term Memory (LSTM) was a flexible model that could be utilized for a variety of tasks, including language translation, and Bayesian Knowledge Tracing (BKT), a method that is frequently used for student modeling. In this study, we explicitly compared three models on two different types of student modeling tasks: post-test score prediction and learning gains prediction: BKT, its variation Intervention-BKT (IBKT), and LSTM. In addition, although earlier research on student learning frequently used skill/knowledge components specified by subject-matter experts, we included an automatic skill discovery approach (SK) to all three models. This method comprises a nonparametric prior over the exercise-skill assignments.

Since desire, aptitude, and learned habits are just a few of the many influences on student cognitive processes, modeling them was extremely difficult. Computer-based learning environments offer a vast array of features and resources, significantly complicating the work of monitoring student learning. Any computer-based educational program that allows for adaptability and personalization was built from the ground up using an accurate student model. In earlier study, student modeling has been thoroughly and widely investigated. The model hypermeters were manually adjusted, and when the course was over, intervention strategies were applied.

Based on conditional inference trees and a large German data set containing a wide range of features of student life and studied courses, random forests by behr, a, et al. [26] were used to predict university dropout. By modeling students' progression from school (pre-study), through the study-choice phase (decision phase), through the first semesters of university (early studied phase), dropout decision is treated as a binary classification (graduate or dropout), with a focus on extremely early prediction of student dropout. We compare the predictive performance of the three models, and we find that adding factors from the first analyzed events significantly improves performance, with an auc (area under the curve) of 0.86. The study's limitations were stated as not taking into account aspects of student happiness (wishes and needs).

We give a summary of the relevant reviews that have been suggested in the literature while highlighting the important distinctions between this effort and the current reviews. The scope and perspective of the subject varied significantly between the existing reviews. a deep neural network was used to predict learnt metrics in videos for MOOCs courses by mubarak a. a. et al [27]. The author employed the Long Short-Term Memory networks, Artificial Neural Network, SVM, and Logistic Regression algorithms, with the LSTM algorithm having the highest accuracy (93%). The limitation was stated because the current model only uses learners' interactions with films; it does not include learners' full range of learned activity patterns.

J. Berens and K. Schneider recommended a study on student attrition on [28] , saying that it was crucial to know the underlying causes of attrition and which students were in danger of quitting. In order to forecast student achievement and serve as the foundation for a focused intervention, we developed An Early Detection System (EDS) using administrative student data from a public and a private institution. Regression analysis, neural networks, decision trees, and the adaboost algorithm were utilized by the educators to identify student traits that set probable dropouts apart from graduates. At the end of the first semester, the state institution's prediction accuracy was 79%, while the private university of applied sciences' prediction accuracy was 85%. The accuracy increased to 90% for the state university and 95% for the private university of applied sciences after the fourth semester.

Summary of researched studied used ml techniques for student dropout and performance prediction along with the particular algorithms used, their performance, problem addressed and limitations.

Research Studies related to the use of ML techniques in creating Predicting Models	**Algorithm used and Performance achieved**	**Problem Addressed**	**Limitations**
Iqbal ,z et al 2017 [9]	Collaborative filtering (cf) , matrix factorization (mf), and restricted boltzamamn machines (rbm)techniques RBM showed better performance having RMSC=0.3,MSE=0.09,MAE=0.23	Grade prediction of students using ML techniques	Features related to students motivation were not included A limited data sets was used the early prediction was not rendered
Xu, jie et al. Aug	Linear regression Logistic regression	Tracking and predicting students	Courses prediction to the students was

2017 [22]	Random forest kNN EPP. Ensemble based Progressive prediction showed the best result having the lowest mean square error	performance using ML techniques	not carried out No intervention technique was discussed
Martins , l ,c,b,et al . Dec.2017 [11]	Gradient boosting machine, deep learning , distributed random forest The deep learning model achieved the highest true positive rate of 71.1%	Earliest possible prediction of students attrition	There were no study-related data for the first semester.
Fwa , h,l.and Marshall ,l 2018 [12]	Hidden Markov model (hmm) leave-one-out cross-validation (LOOCV)	Modeling programming students using ML techniques	Techniques related to students engagement were not used and implemented The early prediction was not rendered
Sckeroglu , b et al Mar'2019 [13]	Support vector regression (svr) LSTM ,SVM, Gradient boosting classifiers (gbc) , ANN.	Students perform classification and prediction using ML techniques	The study did not address the earliest possible identification of at-risk students the data set size was very small
Gray ,c,c. and prekins ,d.[14] Apr'2019	K-nearest – neighbor with 97% accuracy	Identifying possible failing students at week of the fall semester	The predictive model was stable only applicable to bangor university students

Ortigosa , a, et al Apr' 2019[15]	C5.0 algorithm with more than 85% accuracy	Identification of real life challenges of the early dropout prevention system	Comprehensive real world performance of the prediction model and effectiveness of retention action is needed
N. Wu, L. Zhang,et al may'2019 [16]	CLMS-net. Combination of convolutional neutral network,long short-term memory network , and support vector machine accuracy = 91.55%	Predicting dropout in MOOCs	The validity of the predictive model on more datasets are needed
Liao , s,n, et al Jun'2019 [18]	Support vector machine with AUC = .70	Identifying students at-risk of performing poorly in courses	The study did not predict at risk students earlier in course
Hussain , m, et al Jun' 2019 [2]	Artificial neural network (ann) and support vector machine (svm) achived the highest accuracy of 75%	Predicting students difficulties in online learning	Students droupout prediction was not performed model accuracy was low
Lee , s .and chung ,j.y.et al [20] July 2019	Random forest , boosted decision trees (bdt) with bdt having the highest accuracy 99%	Improving the performance of dropout prediction model using the ml-based early warning system	A limited sets data base was used all features were not included in creating a predictive model
Chung , j , y.andlee,s,[21] 2019	Random forest (RF) with 95% accuracy	Students binary classification	Predictive model suffers from potential in accuracy in calculating the weights of the features
AI-shabandar , r, et al .[23] 2019	Gradient boosting model with 95% accuracy	Identification of at risk students with intensive earlier intervention in online courses	Temporal feature were not considered the predictive model was not validated with additional datasets

Figueroa-canas,j and sancho-vinusea[24] 2019	Decision trees (dt) with more than 90% accuracy	Identifying dropout-prone students earlier in online statistical course	The methodology used both training set and validation set of the students enrolled in the same academic year
Imran , a, s, et al 2019 [17]	A feed forward deep neural network with accuracy >90%	Predicting and explaining students dropout	Prediction is done after the course completion
Rose , Carolyn P et al 2019 [25]	Traditional machine learning models do not generalize well	Understanding why ML models alone are not the solution	ML models are not interpretable and actionable
Behr , a, et al .[26] Feb'2020	Random forest with auc (area under the curve) of 0.86	Binary classification , modeling students dropout	Students satisfaction (wishes and needs) features were not considered
Mubarak a. a . et al Jul' 2020 [27]	LSTM, ANN, SVM, logistic regression. LSTM with highest accuracy 93%	Using a deep neural network to predict learning analytics in MOOCs courses videos	The current model only employs learners interactions patterns with videos A complete learning activity pattern of learners is missing
] J. Berens, K. Schneider, Dec. 30, 2018 [28]	Machine Learning Methods: Regression analysis, neural networks, decision trees, and the AdaBoost algorithm	Early detection of students at risk Predicting student dropouts using administrative student data and machine learning methods	Administrative student data for early detection of students at risk

DISCUSSION

Few works actually attempted to anticipate student dropout, despite the fact that several works had demonstrated the viability of explaining student dropout. In this study, we employed machine learning approaches that could automatically identify significant features. With the appropriate model, it is possible to forecast student dropout rates and explain the variables that were probably helpful in the forecast.

All parties profit when students may be predicted and helped at various points throughout the course. It gives teachers the chance to help dropout-prone pupils and to intervene when it's most effective to change how they behave in class. In the current study, we developed a number of machine-learned techniques and generated predictive models for forecasting student performance.

It was necessary for the retention action to be effective and for the prediction model to perform thoroughly in the real world. More datasets were required to test the predictive model's validity. A small network data source was employed, and not all features were taken into account for building the predictive model. The utilization of student engagement strategies was not put into practice. The early forecast did not come true. The data set size was quite little, hence the study did not address the early possible identification of pupils who were at risk. The needs and wants of the students were not taken into consideration. The courses that were predicted to the pupils never happened. No method of intervention was mentioned.

Only a very little amount of the information from the aforementioned survey [9,13,16,20] may be utilised, according to all of the researchers. A sizable data set has been captured and used after analysis by me. The significance of each activity in relation to how well the pupils perform will be looked at in the future.

CONCLUSION, LIMITATION AND FUTURE WORK:

Both students and instructors can benefit by anticipating student needs and stepping in to help them at certain points throughout the course. It gives teachers the chance to help students who are at risk of dropping out and to intervene when it is most effective to change the way they study. In the current work, we suggested multiple predictive models for forecasting students' performance based on demographics factors, demographics + click stream variables, and demographics + click stream + assessment variables. These models were trained on several ML and DL algorithms. For forecasting students' performance at the various lengths of course, the RF predictive model with the highest performance scores was ultimately chosen. An effective prediction model can help teachers identify at-risk kids early and convince them to modify their study habits. The assessment and click stream variables had the most effects on the students' final results out of all the other factors.

Overall, the results of the RF predictive model demonstrated effectiveness in the earliest possible prediction of the performance of at-risk students. Such data-driven studies can assist VLE administrators and instructors in the formulation of the online learning framework which can contribute to the decision-making process. We also deemed that more in-depth studies are required to evaluate various online activities in the OULAD. Particularly, how various early intervention techniques can be implemented in the online learning environment to encourage students to keep on the right track. In the future, we plan to examine the activity-wise significance having a prominent influence on the students' performance by modeling textual variables relating to students' feedbacks by utilizing deep learning models and natural language processing techniques.

REFERENCES:

1. K. S. Rawat and I. Malhan, "A hybrid classification method based on machine learning classifiers to predict performance in educational data mining", *Proc. 2nd Int. Conf. Commun. Comput. Netw.*, pp. 677-684, 2019.

2. M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi and S. Ali, "Using machine learning to predict student difficulties from learning session data", *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 381-407, Jun. 2019.

3..O. E. Aissaoui, Y. E. A. El Madani, L. Oughdir and Y. E. Allioui, "Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles", *Procedia Comput. Sci.*, vol. 148, pp. 87-96, Jan. 2019.

4..A. J. Lekan, "Student academic performance prediction using support vector machine," 2019.

5.A. Qazdar, D. Mammas, C. Chihab, and R. E. Barhim, "A machine learning algorithm framework for predicting students performance: a case study of baccalaureate students in Morocco," *Education and Information Technologies*, vol. 24, no. 1, 2019.

6.H. Turabieh, "Hybrid machine learning classifiers to predict student performance," in *Proceedings of the 2nd International Conference on new Trends in Computing Sciences (ICTCS)*, Amman, Jordan, October 2019.

7.D. Sobnath, T. Kuduk, I. U. Rehman, and O. Isiaq, "Feature selection for uk disabled students' engagement post higher education: a machine learning approach for a predictive employment model," *IEEE Access*, vol. 8, 2020.

8.W. Xing, R. Guo, E. Petakovic, and S. Goggins, "Participation-based student final performance prediction model through interpretable Genetic Programming: integrating learning analytics, educational data mining and theory," *Computers in Human Behavior*, vol. 47, no. jun, pp. 168–181, 2015.

9. Z. Iqbal, J. Qadir, A. Noor Mian and F. Kamiran, "Machine learning based student grade prediction: A case study", *arXiv:1708.08744*, 2017

10. J. Xu, K. H. Moon and M. van der Schaar, "A machine learning approach for tracking and predicting student performance in degree programs", *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 5, pp. 742-753, Aug. 2017.

11. L. C. B. Martins, R. N. Carvalho, R. S. Carvalho, M. C. Victorino and M. Holanda, "Early prediction of college attrition using data mining", *Proc. 16th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, pp. 1075-1078, Dec. 2017.

12.H. L. Fwa and L. Marshall, "Modeling engagement of programming students using unsupervised machine learning technique", *GSTF J. Comput.*, vol. 6, no. 1, pp. 1-6, 2018.

13..B. Sekeroglu, K. Dimililer and K. Tuncal, "Student performance prediction and classification using machine learning algorithms", *Proc. 8th Int. Conf. Educ. Inf. Technol.*, pp. 7-11, Mar. 2019.

14. C. C. Gray and D. Perkins, "Utilizing early engagement and machine learning to predict student outcomes", *Comput. Edu.*, vol. 131, pp. 22-32, Apr. 2019.
15. A. Ortigosa, R. M. Carro, J. Bravo-Agapito, D. Lizcano, J. J. Alcolea and O. Blanco, "From lab to production: Lessons learnt and real-life challenges of an early student-dropout prevention system", *IEEE Trans. Learn. Technol.*, vol. 12, no. 2, pp. 264-277, Apr. 2019.
16. N. Wu, L. Zhang, Y. Gao, M. Zhang, X. Sun and J. Feng, "CLMS-net: Dropout prediction in MOOCs with deep learning", *Proc. ACM Turing Celebration Conf.*, pp. 1-6, May 2019.
17.A. S. Imran, F. Dalipi and Z. Kastrati, "Predicting student dropout in a MOOC: An evaluation of a deep neural network model", *Proc. 5th Int. Conf. Comput. Artif. Intell. (ICCAI)*, pp. 190-195, 2019.
18.S. N. Liao, D. Zingaro, K. Thai, C. Alvarado, W. G. Griswold and L. Porter, "A robust machine learning technique to predict low-performing students", *ACM Trans. Comput. Edu.*, vol. 19, no. 3, pp. 1-19, Jun. 2019.
19.M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi and S. Ali, "Using machine learning to predict student difficulties from learning session data", *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 381-407, Jun. 2019.
20. S. Lee and J. Y. Chung, "The machine learning-based dropout early warning system for improving the performance of dropout prediction", *Appl. Sci.*, vol. 9, no. 15, pp. 3093, Jul. 2019.
21. J. Y. Chung and S. Lee, "Dropout early warning systems for high school students using machine learning", *Children Youth Services Rev.*, vol. 96, pp. 346-353, Jan. 2019.
22. J. Xu, K. H. Moon and M. van der Schaar, "A machine learning approach for tracking and predicting student performance in degree programs", *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 5, pp. 742-753, Aug. 2017.
23. R. Al-Shabandar, A. J. Hussain, P. Liatsis and R. Keight, "Detecting at-risk students with early interventions using machine learning techniques", *IEEE Access*, vol. 7, pp. 149464-149478, 2019.
24. J. Figueroa-Ca nas and T. Sancho-Vinuesa, "Predicting early dropout student is a matter of checking completed quizzes: The case of an online statistics module", *Proc. LASI-SPAIN*, pp. 100-111, 2019.
25. C. P. Rosé, E. A. McLaughlin, R. Liu and K. R. Koedinger, "Explanatory learner models: Why machine learning (alone) is not the answer", *Brit. J. Educ. Technol.*, vol. 50, no. 6, pp. 2943-2958, 2019.
26. A. Behr, M. Giese and K. Theune, "Early prediction of university dropouts—A random forest approach", *J. Nat. Stat.*, vol. 1, pp. 743-789, Feb. 2020.
27. A. A. Mubarak, H. Cao and S. A. M. Ahmed, "Predictive learning analytics using deep learning model in MOOCs' courses videos", *Edu. Inf. Technol.*, vol. 6, pp. 1-22, Jul. 2020.
28. J. Berens, K. Schneider, S. Görtz, S. Oster and J. Burghoff, "Early detection of students at risk—Predicting student dropouts using administrative student data and machine learning methods", *CESifo Group Munich CESifo Work. Paper 7259*, pp. 1-41, Dec. 2018,