
ISSN: 2321-2152
IJMECE
International Journal of modern
electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

STUDENT PERFORMANCE PREDICTION IN ONLINE COURSES USING MACHINE LEARNING

[1] Mr.K.LAKSHMAN REDDY, [2] KARISINAPU VIJAYA RAJU
[1](Associate Professor), MCA, S.V.K.P & Dr K.S. Raju Arts & Science College(A) Penugonda, W.G.District,Andhra Pradesh, klreddyin@gmail.com
[2]PG, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College(A) Penugonda, W.G.District,Andhra Pradesh, vijayarajukarisinapu@gmail.com

ABSTRACT

Advances in Information and Communications Technology (ICT) have increased the growth of Massive open online courses (MOOCs) applied in distance learning environments.Various tools have been utilized to deliver interactive content including pictures, figures, and videos that can motivate the learners to build new cognitive skills. High ranking universities have adopted MOOCs as an efficient dashboard platform where learners from around the world can participate in such courses. The students learning progress is evaluated by using set computer marked assessments. In particular, the computer gives immediate feedback to the student once he or she completes the online assessments. The researchers claim that student success rate in an online course can be related to their performance at the previous session in addition to the level of engagement. Insufficient attention has been paid by literature to evaluate whether student performance and engagement in the prior assessments could affect student achievement in the next assessments. In this paper, two predictive models have been designed namely students' assessments grades and final students' performance. The models can be used to detect the factors that influence students' learning achievement in MOOCs. The result shows that both models gain feasible and accurate results. The lowest RSME gain by RF acquire a value of 8.131 for students assessments grades model while GBM yields the highest accuracy in final

students' performance, an average value of 0.086 was achieved.

1.INTRODUCTION

Massive Open Online Courses (MOOCs) is one of the most widespread e-learning platforms. The MOOCs present the course using digital tool materials in various forms such as visual, audio, video and plain text. Most students prefer using video lectures to understand the contents of lessons over thoroughly reading plain text documents. The interactive video in the MOOCs could reduce students' stress, help them to feel relaxed and learn quickly.

MOOCs can be classified into two distinct types mainly, connectivist Massive Open Online Courses (cMOOCs) and eXtended Massive Open Online Courses (xMOOCs).The xMOOCs are learning paradigm based on the principles of cognitivist behaviorist theory. The structure of the courses is similar to the traditional course where the syllabus consists of a set of video lectures and a set of multiple choice quizzes in addition to the final exam. The video lectures featuring the course instructor reviewing the content of the previous online lesson are released weekly. The participants can watch and pause the video at their own pace. Moreover, the students can socially interact with other participants and the instructor through posting in discussion forums. The instructors usually post questions, provide task solutions and reply to student questions via these discussion forums; as a consequence the discussion forums play a vital role in enhancing the course quality and make online sessions collaborative and engaging.

The cMOOCs are a new learning model based on connectivist learning theory . With the connectivism approach, the instructor would not provide the actual learning material; the students get the course syllabus by asking the questions and sharing this information with other participants. References posit the learning strategy of cMOOCs focused on a collaborative approach in which learning material combined remix, repurposable and provided, forwarded to other students.

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

With cMOOCs, it is impossible to involve expertise to assess the students' knowledge whereas in xMOOCs, university lecturers can evaluate the students' knowledge through the use of computer-marked assessment feedback. In particular, the computer gives
immediate feedback to the student when he completes the online assessment.

The learner, upon successful completion, will be awarded their certification in xMOOCs. The cMOOCs do not include a formal assessment. Hence, universities are not considered cMOOCs as an official course.

With rapid advancements in technology, artificial intelligence has recently become an effective approach in the evaluation and testing of student performance in online courses. Many researchers applied machine learning to predict student performance in however few works have been done to examine the trajectories performance. As a result, educators could not monitor the real-time students learning curve. Two sets of experiments are conducted in this work. In the first set of eperiements, regression analysis is implemented for estimation of students' assessment scores. The student past and current activities in addition to past performance are employed to predict student outcome. In the second set of experiments, supervised machine learning method has been utilized to predict long-term student performance. Three types of candidate predictors have been considered firstly behavioral features, followed by temporal and demographic features. The proposed models offer new insight into determining the most critical learning activity and assist the educators in keeping tracking of timely student performance. To the best of our knowledge, student performance has been evaluated in online course using only two targets: "success" and "fail". Our model predicts the performance with three-class labels "success", "fail" and "withdrew".

2.LITERATURE SURVEY

During my literature review study, to define the research problem as well as objectives, I would like to mention that

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

I went through a lot of research papers and web application statistics. Few of paper's summarized details are as follows. In this first research study, author used feature selection technique to reduce the number of feature form the large attribute set. In this paper author use ASSISTments platform dataset which is a web based teaching system developed at Worcester Polytechnic institute and used with 4th to 10th grade math students. In this paper author used technique to remove irrelevant, redundant or noisy data. In this paper author used various classification algorithm and ranker algorithm to find top most contributed attribute and removed the less appropriate attribute. This helps to speeds up the process of data mining and improves its performance parameters such as predictive accuracy. In this research paper author used three different approaches. Cross tabulation analysis, Feature selection and balancing imbalance data. Features selection method is used to select those attribute which are highly affected dependent variables. Classification tree is built considering all available attributes. This method finds out all possible splits that can occur for each indicator variable at each node. The search stops when the split with the largest imprudent in goodness of fit is found. A few element choice calculations are connected and includes positioning higher in numerous calculations are chosen. In this way 15 vital parameter are chosen from unique 77 attributes. Misbalancing issue is resolved by using data balancing and rebalancing algorithm specifically SMOTE(Synthetic Minority Over sampling technique). Ten fold cross validation is used for establishing training and testing data from original data. This data set is prepared in three categories. First category contains data with all 77 attributes. Next category contains data with 15 important attributes. Last category contains balanced data after applying rebalancing technique in weak. This paper provides the hybrid approach for outlier detection. They used two algorithms: K-mean and Neural Network. The proposed method use Integrating Semantic Knowledge (SOF- Semantic outlier factor) for outlier detection.This method detects the semantic outlier.

This technique identifies the semantic anomaly. The main motive of this research was to reduce the number of outliers in clusters as well as data by improving the cluster formulation methods so that outlier rate reduces. It also decreases the error and improves the accuracy. The result showed that the hybrid algorithm performs better than that of genetic k-means. This proposed strategy manages content and date dataset that has not been executed before using genetic kmeans. [This research study describes the various approaches such as Neural network, K-Nearest Neighbour, Bayesian Classifier, Fuzzy Logic and decision tree classification Algorithms for implementation of intrusion Detection system. With the help of this paper, it is clear that the data mining methods are used to perform the intrusion detection system But this paper don't describe which technique is best for all of these.

3. EXISTING SYSTEM

The Factor Analysis Model (FAM) was proposed to predict the student's performance in Intelligent Tutoring System (ITS) taking into consideration the difficulty level of assessments based on Item Response Theory concept. The difficulty level of tasks can infer measurement of the correlation between the student's performances and assessment questions. To compute the probability of a student solving a task correctly, a set of predictor variables are defined in the FAM including the number of opportunities presented to the student at each task, the duration spent on each step and the difficulty level of each question or latent variable. The results reveal that incorporating the latent variables into the estimates of student performance can significantly enhance the model.

To measure how the activities of learners could impact their learning achievement in MOOCs, the researchers found that Learning Analytics (LAs) in conjunction with machine learning, are effective tools that offer the potential to trace student knowledge. The researchers demonstrated that machine learning could help the educator in providing cohort information about the learning process, furnishing researchers with the ability to both visualise and analyse the information obtained from each tier of the learner. Thus, an accurate predictive model can be acquired in such courses. Students' marks in the first assessment and quiz scores in conjunction with social factors are used to predict students' final performance in online course.

Two predictive models were introduced. In the first model, logistic regression was used to predict whether students gained a normal or distinction certificate. In the second predictive model, logistic regression was also used to predict if students achieved certification or not. The results indicated that the number of peer assessment is the most effective feature for acquiring a distinction. The average quiz scores were considered the most reliable predictor for earning a certificate. The accuracy of distinction and normal models were reported with the percentage of 92.6% for the first model and 79.6 % for the second model, respectively.

The association between the Virtual Learning Environment (VLE) data and student performance has been investigated at the University of Maryland, Baltimore County (UMBC). LA used through the implementation of the Check My Activity (CMA) tool. CMA can be defined as an LA tool, which compares students VLE activities with other activities and provides lecturers frequent feedback of students' emotional states. The results showed the students who engage with the course frequently are more likely to earn mark C or higher than those who did not regularly engage.

DISADVANTAGES

In the existing work, the system is poor performance in which the assessments are not Tutor Marked Assessment (TMA), Computer Marked Assessment (CMA).

This system is less performance due to Lack of Massive open online courses (MOOCs).

4. PROPOSED SYSTEM

✓ **DATA DESCRIPTION**

The OULAD dataset was captured from the Open University Learning Analytics Dataset (OULAD)repository. The open university in the UK delivers the online course in various topic for undergraduate and postgraduate students in the period between 2013-2014. The main composite table called "studentInfo" is linked to all tables.The"studentInfo "table includes information relevant to students' demographic characteristics. The information related to students performance are collected in "Assessments" and Student Assessment tables. The table "Assessments" contains information about the number, weight and the type of assessments required for each module. In general, each module involves a set of assessments, followed by the final exam. The assessments are Tutor Marked Assessment (TMA), Computer Marked Assessment (CMA). The final average grade is computed with the sum of all assessments (50%) and final exams (50%). The "Student Assessment" table involves information relating to student and the assessment mark. The "Student Registration" table contains information about the date the students registered and unregistered in a particular module. The overall date is measured by counting numbers of unique days that students interact with courses until the course ends. In Open University online courses, students are able to access a module even before being a student of the course; however, it is not possible to access the course post-course closure date. The students' information related to their interaction with digital is store in learners Virtual Learning Environment dataset.

- ✓ **STUDENTS PERFORMANCE MODEL**

Two sets of experiments are conducted in this case study. In the first experiment, the dynamic behavioral features are considered to predict student performance, while the static behavioral attributes are employed in the second experiment. The problems are formulated as classification and regression. The regression setting is considered when we aim to predict students' assessments grades, whereas classification setting is utilised when we seek to predict final student performance in the entire course. It is considered a multi-class problem where the target class is whether students pass, fail or withdraw from courses. Early grade prediction could help educators deliver timely intervention support and additional learning materials to help students who have low scores. As discussed previously, the student should participate in five CMA assessments and six TMA assessments, in addition to the final exam. The assessments should be handed in within a specific period. Due to the TMA assessment weighing 45% of the final result, while the CMA assessment weighs only 5%, our temporal analysis is based on the submission date of the TMA. In first set experiments, student performance is predicated in a timely manner, as can be seen, in Figure 1 the course is subsequently the into six-time intervals, corresponding with assessment submission dates. The student behavioral records are distributed according to the assessment date. The student performance in prior assessments with their interaction behavior is considered in this analysis.

In the second set of experiments, we evaluated the trajectories student performance by aggregate the student's behavioral activities across the six-time slices into a single time slice. The behavioral features, demographic features and temporal features are used as input variables. We did not account for past assessments grade, and final exam mark as target class is computed based on these features. The dataset contains 4004 records where the proportion of "fail", "withdrawn" and "pass" classes are 28%t,40% and 32% respectively.

✓ **FEATURES SELECTION**

As our aim is to investigate the impact of student performance in the previous assessment into the following assessment, features selection is only consider for the first set of experiment. Recursive Feature Elimination (RFE) is utilized in this case study. Rrecursive feature elimination is one of the most popular wrapper feature selection approaches. RFE can be defined as an optimisation algorithm based on backwards selection and resampling techniques . It keeps recursively creating the model until it gets a small

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

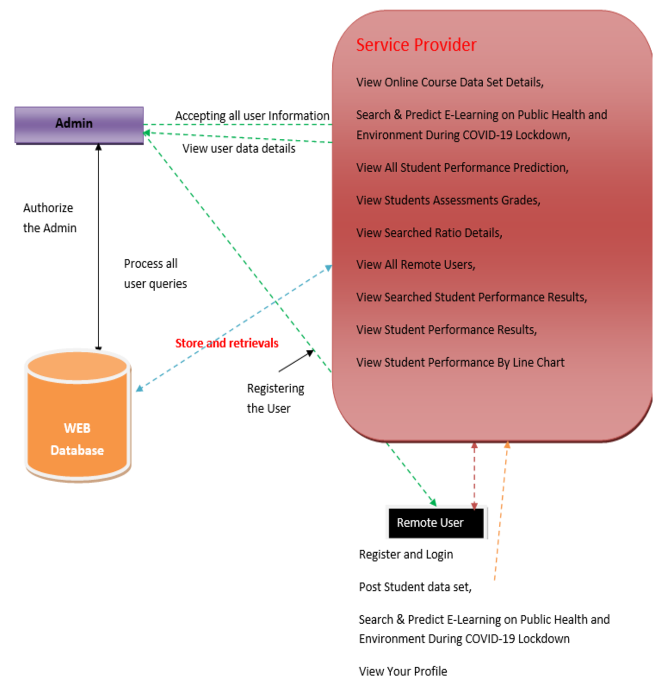
number of features. The data set is partitioned into train and bootstrap samples with different elements. At each iteration, the algorithms are chosen as the most important features. To assess the probability of ranking features, the new model that includes the most essential predictors is retained until all are exhausted.

ADVANTAGES

- Students' marks in the first assessment and quiz scores in conjunction with social factors are used to predict students' final performance in online course.
- Learning Analytics (LAs) in conjunction with machine learning, are effective tools that offer the potential to trace student knowledge.

5. SYSTEM ARCHITECTURE

Architecture Diagram



6.MODULES INVOLVED

MODULES SERVICES

- **SERVICE PROVIDER**

In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as View Online Course Data Set Details,Search & Predict Student Performance Details ,View All Student Performance Prediction, View Students Assessments Grades,View Searched Ratio Details,View All Remote Users,View Searched Student

Performance Results,View Student Performance Results,View Student Performance By Line Chart.

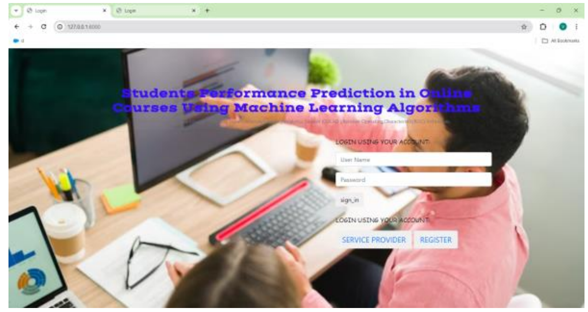
- **VIEW AND AUTHERIZE USERS**

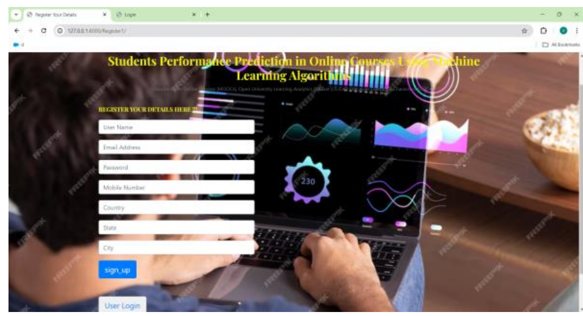
In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorizes the users.

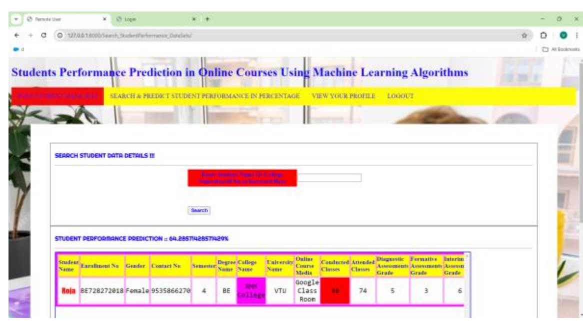
- **REMOTE USER**

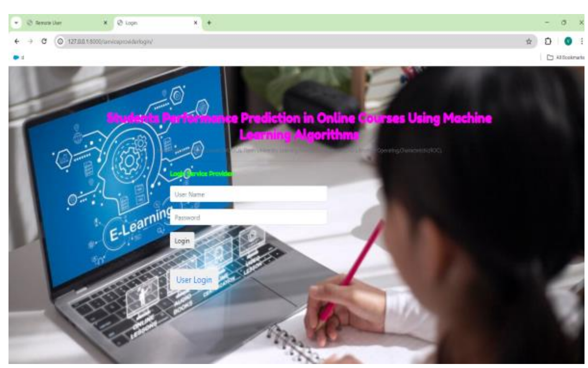
In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user will do some operations like Post Student data set, Search & Predict Student Performance Details,View Your Profile

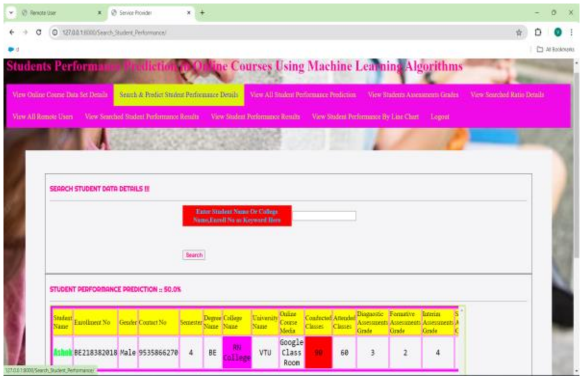
7. OUTPUTSCREENS











8. CONCLUSION

Two sets of exterminates have been carried out in this study using regression and classification analysis. The results of predicting students' assessments grades model show that the students' performance in a particular assignment relies on students' mark in the previous assignment within single Courses. The researchers conclude that students' prior grade point average (GPA) with a low mark is considered as a significant factor of withdrawal from the next course in the traditional classroom setting, Both conventional classroom setting and virtual class share similar characteristic in term of the effective of pervious performance into student learning achievement in the future.

The final student performance predictive model revealed that student engagement with digital material has a significant impact on their success in the entire course. The findings' results also demonstrate that long-term students' performance achieves better accuracy than students' assessments grades prediction model, due to the exclusion of temporal features in regression analysis. The date of student deregistration from the course is a valuable predictor that is significantly correlated with student performance. With the regression analysis, the data does not provide the last date of students' activity prior to undertaken assessments. The findings' results have been recommended to take into account the temporal features on predicting of subsequent assessments grades.

Future research direction involves the use of temporal features for predicting students' assessments grades model. With temporal feature time series analysis will be untaken, might be more advanced machine leering will be utilized.

9. REFERENCE

[1] K. F. Hew and W. S. Cheung, "Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges," *Educ. Res. Rev.*, vol. 12, pp. 45–58, 2014.

[2] H. B. Shapiro, C. H. Lee, N. E. Wyman Roth, K. Li, M. Çetinkaya-Rundel, and D. A. Canelas,"Understanding the massive open online course(MOOC) student experience: An examination of attitudes, motivations, and barriers," *Comput. Educ.*, vol. 110, pp. 35–50, 2017.

[3] J. Renz, F. Schwerer, and C. Meinel, "openSAP: Evaluating xMOOC Usage and Challenges for Scalable and Open Enterprise Education.," *Int. J. Adv. Corp. Learn.*, vol. 9, no. 2, pp. 34–39, 2016.

[4] S. Li, Q. Tang, and Y. Zhang, "A Case Study on Learning Difficulties and Corresponding Supports for Learning in cMOOCs| Une étude de cas sur les difficultés d'apprentissage et le soutien correspondant pour l'apprentissage dans les cMOOC," *Can. J. Learn. Technol. Rev. Can. l'apprentissage la Technol.*, vol. 42, no. 2, 2016.

[5] S. Zutshi, S. O'Hare, and A. Rodafinos, "Experiences in MOOCs: The Perspective of Students," *Am. J. Distance Educ.*, vol. 27, no. 4, pp. 218–227, 2013.

[6] Z. Wang, T. Anderson, L. Chen, and E. Barbera, "Interaction pattern analysis in cMOOCs based on the connectivist interaction and engagement framework,"*Br. J. Educ. Technol.*, vol. 48, no. 2, pp. 683–699,2017.

[7] W. Xing and D. Du, "Dropout Prediction in MOOCs: Using Deep Learning for Personalized Intervention," *J. Educ. Comput. Res. p.0735633118757015.*, 2018.

[8] M. J. Gallego Arrufat, V. Gamiz Sanchez, and E.Gutierrez Santiuste, "Trends in Assessment in Massive Open Online Courses," *Educ. Xx1*, vol. 18, no. 2, pp. 77–96, 2015.

[9] B. J. Hao Cen, Kenneth Koedinger and Carnegie,

"Learning Factors Analysis – A General Method for Cognitive Model Evaluation and Improvement," in *In International Conference on Intelligent Tutoring Systems*, 2006, vol. 8, pp. 164–175.

[10] K. R. Koedinger, E. A. Mclaughlin, and J. Stamper, "Automated Student Model Improvement," in *Educational Data Mining, proceedings of the 5th International Conference on*, 2012, pp. 17–24.

[11] J. Sinclair and S. Kalvala, "Student engagement in massive open online courses," *Int. J. Learn. Technol.*, vol. 11, no. 3, pp. 218–237, 2016.

[12] J. Mullan, "Learning Analytics in Higher Education," London, 2016.

[13] P. and K. Al-Shabandar, R., Hussain, A.J., Liatsis, "Detecting At-Risk Students With Early Interventions Using Machine Learning Techniques," *IEEE Access*, vol. 7, pp. 149464–149478, 2019.

[14] S. Jiang, A. E. Williams, K. Schenke, M. Warschauer, and D. O. Dowd, "Predicting MOOC Performance with Week 1 Behavior," in *Proceedings of the 7th International Conference on Educational Data Mining (EDM)*, 2014, pp. 273–275.

[15] L. Analytics and C. Exchange, "OU Ansssalyse : Analysing at - risk students at The Open University,"

in *in Conference, 5th International Learning Analytics and Knowledge (LAK) (ed.)*, 2015, no. October 2014.

[16] R. Alshabandar, A. Hussain, R. Keight, A. Laws, and T. Baker, "The Application of Gaussian Mixture Models for the Identification of At-Risk Learners in Massive Open Online Courses," in *2018 IEEE Congress on Evolutionary Computation, CEC 2018 - Proceedings*, 2018.

[17] J.-L. Hung, M. C. Wang, S. Wang, M. Abdelrasoul, Y. Li, and W. He, “Identifying At-Risk Students for Early Interventions—A Time-Series Clustering Approach,” *IEEE Trans. Emerg. Top. Comput.*, vol. 5, no. 1, pp. 45–55, 2017.

[18] C. Yun, D. Shin, H. Jo, J. Yang, and S. Kim, “An Experimental Study on Feature Subset Selection Methods,” *7th IEEE Int. Conf. Comput. Inf. Technol. (CIT 2007)*, pp. 77–82, 2007.

[19] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Comput. Electr. Eng.*, vol. 40, no.1, pp. 16–28, 2014.

[20] J. S. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, “Investigation of the random forest framework for classification of hyperspectral data,” *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, 2005.