Students Performance Prediction in Online Courses Using Machine Learning Algorithms

Raghad Alshabandar
Department of Computer Science
Liverpool John Moores
University
Liverpool, United Kingdom
ragednazar@yahoo.com

Abir Hussain
Department of Computer Science
Liverpool John Moores
University
Liverpool, United Kingdom
A.Hussain@ljmu.ac.uk

Robert Keight
Department of Computer Science
Liverpool John Moores
University
Liverpool, United Kingdom
R.Keight@2015.ljmu.ac.uk

Wasiq Khan
Department of Computer Science
Liverpool John Moores
University
Liverpool, United Kingdom
w.khan@ljmu.ac.uk

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.

Keywords:
Massive Open Online Courses (MOOCs), Open University Learning Analytics Dataset (OULAD),Receiver Operating Characteristic(ROC).

I. 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 [1] [2].

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[4]. 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 [3] [5].

The cMOOCs are a new learning model based on connectivist learning theory [3][4]. 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 [3][4][5] posit the learning strategy of cMOOCs focused on a collaborative approach in which learning material combined remix, repurposable and provided, forwarded to other students.

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 [5][6].

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 [7], however few works have been done to examine the trajectories performance [8]. 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".

The reminder of this paper is organised as follows. Section 2 will give detailed information about prior works .In section 3, shows the methodology, including data give overview of features section , experiment setup and experiment results .The conclusion and future works are described in Section 4.

II. LITERATURE REVIEW

Student performance is a key indicator of measuring learning progress in a virtual learning environment. Researchers have adopted various methods to monitor performance [1]. In this section, the literature review in the relevant area is presented.

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 [9] [10]. 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 [10].

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[11] [12][13].

Students' marks in the first assessment and quiz scores in conjunction with social factors are used to predict students' final performance in online course [14]. 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[14].

The association between the Virtual Learning Environment (VLE) data and student performance has been investigated at the University of Maryland, Baltimore County (UMBC) [12]. 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 [12].

III. METHODOLOGY

A) 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[15].

The information related to students perfromnce 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

assessment results, such as the date of the submitted assessment and the assessment mark[15].

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.

B) Feature Extraction

The features extraction was undertaken in our previous work[16]. The VLE activities were used to generate behavioural features. For each student at a specific time t, two features are extracted: the number of sessions (o^t) and the number of clicks ((c^t). The behavioral features can be divided as either static or dynamic.

- **Static Behavioral features:** These are a set of behavioral features that correspond to student activities since the first time they engaged in the course till last day they quit the course. Let us consider the set $X \in \mathbb{R}^{T \times n \times m}$, in which $X_{t,i,j}$ represents the jth activity of the ith student at time t. S is a set of students denoted as an n-dimensional vector $[S_1 \ldots, Sn]$, where n is the number of students. Furthermore, M is defined as an m-dimensional vector that represents VLE learning activity types, M = [M1 …., Mm], where m is the number of learning activities that the ith student is assigned.

- **Dynamic Behavioral Features:** These are a set of behavioral features that vary over time. Let t be a sequence of disjointed time intervals, where $t \in [1,6]$. To represent all student activities at time t, we define the type of student's activity records as the vector $X_{t,i,j} = [X_{t,i,1}, X_{t,i,2}, \ldots, X_{t,i,m,}]$. Here j^{th} denotes learning activity that is undertaken at time t by student S_i, such that $j = 1$, the, m; where m is given as the number of learning activities.

C) 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[17]. 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.

In order to evaluate the student performance predictive model across two experiments, several metrics have been considered. In terms of the regression analysis, the Root Mean Square Error (RMSE) and R-squared(R^2) are used to predict students' assessment grades. With regards, to the classification analysis accuracy, specificity and sensitivity, F-Measure, Receiver Operating Characteristic(ROC), and Area Under Curve (AUC) are employed to predict final student performance. Furthermore, tenfold cross-validation is used for both a regression and classification analysis; 50% of the dataset is selected to train the algorithms. A further 50% of the data is disjointed from the original dataset, and was used to evaluate the generalization errors for unseen data.
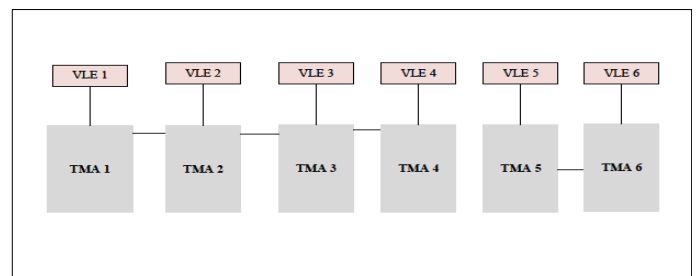

Figure 1.Student behavioral features based on TMA intervals

D) 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 [18]. It keeps recursively creating the

model until it gets a small 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[19][20].

The results of the features selection across six intervals are listed in Table 1. It is based on RMSE criteria. The results reveal that students' assessments grades during the previous slice occupy the position of highest top features across all intervals.

The "session.homepage" and "sum_click.homepage" of the previous and current time slice are the most important features across all intervals. It also appears that "sum_click.forumng" and "session.forumng" have been selected as top features, only for intervals 1 and 2. In general, the activities ("homepage"," quiz"," subpage") robustly predicted the students' grades over all time intervals.

Table 1 Features selection (RFE)

Interval	Five Top Features
Interval 1	"session.homepage/T2"
	"sum_click.forumng/T2"
	"session.resource/T2"
	"session.forumng/T2"
	"sum_click.subpage.T2"
Interval 2	"Score.Assessment 1"
	"session.homepage/T3"
	"session.quiz/T3"
	"session.homepage /T2"
	"sum_click.forumng/T3"
Interval 3	"Score. Assessment 2"
	"sum_click.homepage/T3"
	"sum_click.subpage/T3"
	"session.homepage/T3"
	"session. subpage/T3"
Interval 4	"score. Assessment 3"
	"sum_click.quiz/T5",
	"session.quiz/T5",
	"sum_click.homepage/T5",
	"session.homepage/T5"
Interval 5	"score. Assessment 4",
	"session.quiz/T6",
	"sum_click.quiz/T5",
	"session.quiz/T5"
	" sum_click.homepage/T6"
Interval 6	"Score. Assessment 5",
	"session.homepage/T6",
	"sum_click.homepage/T6",
	"sum_click.homepage/T7"
	" session .homepage/T7"

E) Simulation results- Students Assessments Grades Model

The regression analysis has been applied to predict students' assessment grades over six-time intervals. Table 2 list the result of regression analysis .Two sets of features are considered. In the first set of features, dynamic behavioural features and student performance are used as predictors, while only the top five features are employed in the second set of features.

Various machine learning models are used in this experiment, including Random Forest (RF), Multi-Layer Perceptron (MLP) with two hidden layers, Neural Networks with single hidden layer (Nnet), Gradient Boosting Machine (GBM) and Generalized Linear Model (GLM).

The empirical results from the first set of features have acquired slightly better performance than the selected features. However, the Mlp and Rpart in the second set of features offer better results. The RMSE metric measures the difference between the predicted values and the actual observations. The lowest RMSE value demonstrates the better performance of the predictive model. In terms of the first set of features, RF obtains the best RMSE achieving values of 8.131 over interval 3. For the second set of features, the Gbm model gives the best RMSE, with a value of 11.230 over interval 1.

As can be seen, for both set of features the Gbm models over interval 1, 2, 4, 5, and 6 give the best RMSE, with averages values of 11-22. RF occupies the second-best model. The RMSE increased by 3% in Glm across all intervals and acquired an average value of 11-23. Mlp acquired the worst RMSE, with a value of 44.904 and 39.215 at interval 6 for both sets of features. In general, Mlp achieves the poorest RMSE across all intervals for both sets of features.

Table 2 Result for Predication Students' Assessments Grade*s* Model

Interval	ML	RMSE	R^2	Interval	ML	RMSE	R^2
1	Mlp	18.162	0.483	**1**	Mlp	24.324	0.17
	RF	11.395	0.790		RF	21.366	0.272
	Rpart	11.647	0.780		Rpart	22.457	0.198
	Glm	11.382	0.790		Glm	22.344	0.204
	Gbm	11.230	0.204		Gbm	*11.230*	0.274
2	Mlp	29.517	0.427	**2**	Mlp	26.374	0.558
	RF	18.221	0.640		RF	18.062	0.643
	Rpart	22.400	0.459		Rpart	22.628	0.444
	Glm	19.116	0.604		Glm	18.831	0.612
	Gbm	18.145	0.644		Gbm	18.321	0.634
3	Mlp	20.564	0.641	**3**	Mlp	27.574	0.405
	RF	**8.131**	**0.937**		RF	16.225	0.764
	Rpart	13.760	0.18		Rpart	24.304	0.434
	Glm	15.376	0.773		Glm	23.106	0.489
	Gbm	11.506	0.873		Gbm	22.600	0.512
4	Mlp	18.205	0.768	**4**	Mlp	16.733	0.790
	RF	13.904	0.853		RF	13.905	0.853
	Rpart	16.264	0.799		Rpart	16.264	0.799
	Glm	15.354	0.821		Glm	15.787	0.811
	Gbm	13.892	0.853		Gbm	14.074	0.849
5	Mlp	29.665	0.428	**5**	Mlp	26.271	0.54
	RF	19.900	0.716		RF	19.259	0.737
	Rpart	25.001	0.553		Rpart	25.001	0.553
	Glm	21.708	0.663		Glm	21.831	0.659
	Gbm	19.871	0.718		Gbm	19.478	0.728
6	Mlp	44.904	0.147	**6**	MLP	39.215	0.067
	RF	21.529	0.688		RF	22.794	0.650
	Rpart	25.259	0.569		Rpart	25.919	0.546
	Glm	22.058	0.673		Glm	23.332	0.633
	Gbm	21.425	0.691		Gbm	22.761	0.650

F) Simulation results- Final Students Performance Model

The classification analysis results for the second experiment presented as follows. The same set of machine learning classifiers in the previous experiment are used in this case study.

As can be seen in a Table 3 all classifiers obtain similar ideal results, the highest performance achieved by Gbm with the value of 0.868 while RF, Nnet producing the value of 0.854, achieved the lowest accuracy. Table 4 shows the class "Withdrawn" acquired the best accuracy of all Classifiers reaching an average value of 0.99 whereas the class "Fail" gives the lowest performance, with an approximate range of accuracy between 0.76-0.80.

The sensitivities are high overall classifiers for class "Withdraw" and "Pass". The best sensitivity achieved by Rpart reported the values of 0.99 and 0.92. The class "Fail" gained very low sensitivities across all classifiers. This is expected since the number of records with target class "Fail" are limited hence, the algorithm cannot learn well. With regards, to true negative instance, the Gbm and Nnet produce the best result, specificity =0.998 for class "Withdrawn". The poorest specificity gained by Rpart for class "Pass" obtained the values of 0.81. As can be seen, the best F1-Measure gained by Gbm yielded a value of 0.993, 0.864, 0.772, for the class "Withdrawn"," Pass" and "Fail" respectively. The lowest F1-Measure is shown for Rpart with the value of 0.67 over class "Fail".

ROC is used in this study to choose a decision threshold value for the true and false positive rate across each class. Figure 2 lists ROC curves. Overall, a range of AUC values between 0.99-0.82 for all classes was obtained. As previously mentioned, the demographic behavioral and temporal features in classification analysis were combined. In this model the total numbers of variables are 35. As a result, the predictive model may suffer from the overfitting issue. In this case, we compare classifiers results in terms of train and test error which could give an indication of the overfitting problem.

Figure 3 displays the result of overfitting evaluation. It can be observed that training and test error are low for all classifiers. The lowest test and train error was obtained by Gbm. The RF ,Nnet obtained a similar test error with an approximate percentage of 14%. The training errors are slightly higher in these classifiers. The largest error was acquired by the Mlp model. Although all models fit well for most classifiers, Mlp suffers from overfitting.

Table 3 Accuracy Result for Final Students Performance model

Classifier	Accuracy
Mlp	0.858
RF	0.854
Rpart	0.862
Gbm	0.868
Nnet	0.854

Table 4 Results for Final Students Performance Prediction Model

Classifier	Performance Metrics				
MLP	ACC.	F1	Sens.	Spec.	AUC
Pass	0.858	0.850	0.892	0.824	0.916
Fail	0.782	0.690	0.631	0.932	0.886
Withdrawn	0.993	0.992	0.989	0.996	0.996
RF	ACC.	F1.	Sens.	Spec.	AUC
Pass	0.855	0.843	0.844	0.866	0.924
Fail	0.808	0.713	0.712	0.904	0.892
Withdrawn	**0.995**	0.993	0.991	0.990	0.995
Rpart	ACC.	F1	Sens.	Spec.	AUC
Pass	0.866	0.865	0.923	0.810	0.867
Fail	0.767	0.671	0.582	0.952	0.821
Withdrawn	0.997	0.991	0.996	0.992	0.997
Gbm	ACC.	F1	Sens.	Spec.	AUC
Pass	0.872	0.864	0.903	0.841	0.925
Fail	0.802	0.722	0.665	0.939	0.900
Withdrawn	0.994	0.993	0.991	0.998	**0.997**
Nnet	ACC.	F1	Sens.	Spec.	AUC
Pass	0.856	0.8471	0.870	0.843	0.925
Fail	0 .795	0.7045	0.670	0.920	0.900
Withdrawn	0.994	0.9934	0.991	0.998	0.998

a) Mlp Classifier

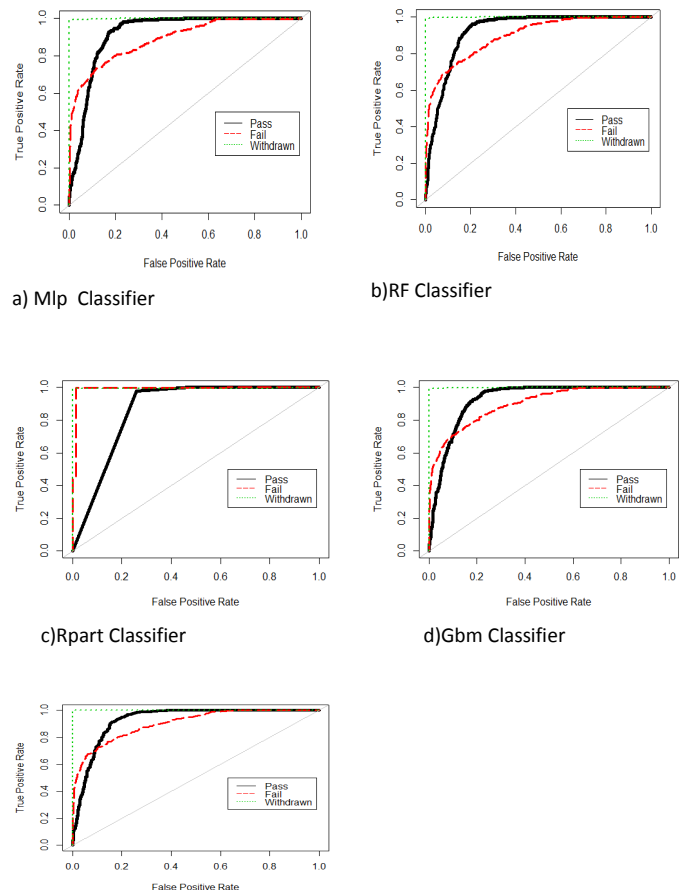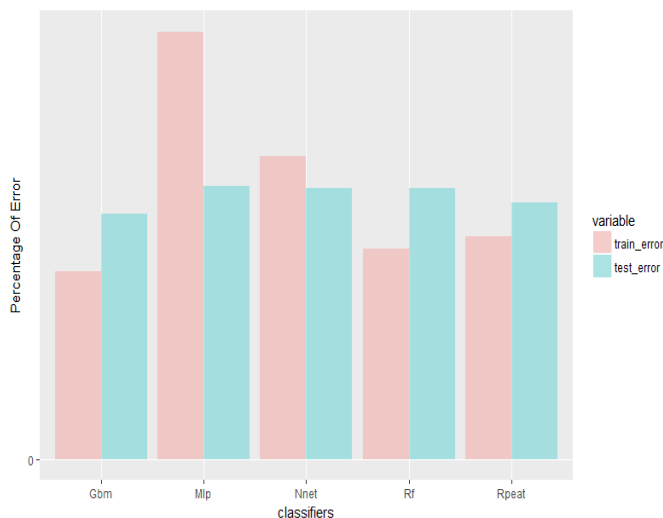b)RF Classifier

c)Rpart Classifier

d)Gbm Classifier

Figure 2. *Roc Curve*

Figure 3. Comparing Computational Training and Test Time

G .Discussion

The regression and classification analysis are considered in This case study. The result of predicting students' assessments grades model shows that best RMSE and R^2 acquire over interval 3 while worst RMSE and R^2 are shown for intervals 5 and 6.This could be attributed to a number of students record over interval 3 than other interval as a result; The algorithm can learn more and model will fit well. In contrast to, the lowest number of training example shown over interval 6.

Final student performance predictive model revels ideal sensitivities and specificities for all classifiers Although, the sensitivities and specificities are balanced for all classifiers over class "Withdrawn" and "Pass", the specificities are higher than corresponding sensitivities for class "Fail". This is due to; the database is skewed in favor of chosen the majority classes "Withdrawn" and "Pass". In this case, predict withdrawal students are more priority than predict success and fail students as it is worthwhile to predict students who withdraw from the entire course than students who stay engaged with the course.

The primary reason the machine learning models obtain higher performance in classification than regression is the relevant to the type of features sets. As such, in classification analysis, the static behavioral features in conjunction with the temporal features and demographic features are used as input variables in the prediction of final student performance model while only dynamic behavioral features employed to estimate students' assessments grades. The temporal features that include the date of student registration and un-registration in an online course are robust predictors features that adequately impact on student achievement. It could be impossible to combine the temporal features with behavioral features with respect to regression analysis as the database include student temporal information for the entire period.

IV. 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.

References

[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 Analyse : 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.