International Journal of Computer Trends and Technology
ISSN: 2231 – 2803 / https://doi.org/10.14445/22312803/IJCTT-V68I12P105
Volume 68 Issue 12, 21-25, December 2020

Review Article

Analysis of Academic Performance in massive Open Online Courses (Moocs) Using Process Mining

Mahesh T R[1], B Mohan Kumar Naik[2]

[1]*Research Scholar, Department of Computer Science & Engineering, JJTU, Rajasthan, India*
[2]*Research Supervisor, Department of Computer Science & Engineering, JJTU, Rajasthan, India*

Received Date: 14 October 2020
Revised Date: 22 November 2020
Accepted Date: 24 November 2020

Abstract - *The purpose of this paper is to provide a survey of academic performance in massive open online courses (MOOCs) to improve students' learning experience. Due to the large volume of data in educational databases of Student's data, e.g., weekly evaluation grades or points, demographic variables such as age, ethnicity, and sex and weekly interaction data based on event logs, e.g., video lecture interaction, submission time of assignment solution, amount of time spent weekly driven this design, Automated Student performance prediction is a very important task. This study compares the four distinct logistic regression techniques for machine learning classification, Naïve Bayes (NB), (LR), random forest (RF), and K-nearest neighbor, to track the performance of students every week and to predict their overall performance.*

While MOOCs provide a versatile learning platform, they are prone to early dropouts and low completion rates. This research focuses on a data-driven plan to enhance students' learning experience and dramatically reduce the dropout rate. Early forecasts based on individuals' involvement will help educators provide students who are currently struggling in the course with proper support.

Keywords - *Prediction, MOOCs, Machine learning, Learning analytics, Process mining, Education data mining*

I. INTRODUCTION

Huge Open Online Courses (MOOCs) are free online courses accessible to anyone who registers for free. MOOCs offer an inexpensive and versatile way of learning new skills, progressing your career, introducing yourself to the latest technology, and delivering quality education. For a number of purposes, millions and billions of people worldwide use this tool, including skills advancement, job growth, changing jobs, college preparations, supplementary learning, lifelong learning, eLearning & training for companies, further training, and more. In making the whole world learn, MOOCs have drastically changed a lot.

Particularly among student communities, massive open online courses (MOOCs) have become very popular because they can register for prestigious universities' courses around the world. MOOCs offer a really strong forum for learning that attracts many learners with various objectives and motives. The three founders of the MOOCs network, Coursera, edX, and Udacity, are then closely followed by others around the world, such as Swayam in India, Miriada and Spanish MOOC in Spain, Khan Academy in North America, FutureLearn in England, Open2Study in Australia, Fun in France, Veduca in Brazil.

By centralizing global capital and restructuring the learning environment to bring it closer to learners' fingertips, the MOOC environment has revolutionized the education system. MOOCs offer open access to courses for anybody interested in learning how to access the Internet, unlike conventional higher education learning environments. It offers free and open access to advanced, high-quality courses consisting of video tutorials, reading materials, quizzes, assignments, problem sets, and platforms for constructive discussions to facilitate the learning process and learning communities' growth.

While there are significant enrollments in MOOCs courses, the completion and retention rate of persistent students is very low, sometimes less than 20% (Kizilcec et al., 2013). In one of the highly critical MOOCs, low retention rates of students, it is important to predict the drop out so that measures can be taken to retain students by assisting them in their learning activities.

To make pedagogical strategies during the course to enhance student learning, how do we learn from the log history data? Campbell et al. (2007) identified five phases. Namely, capture, record, anticipate, act, and refine, as a central theme in LA. When data documenting student experiences with the course is captured, analysts work to make predictions for pedagogical intervention, which is increasingly refined.

II. RELATED WORK

Several works (Ye and Biswas, 2014: Several papers (Ye and Biswas, 2014: Marquez-Vera et al., 2013; Martinho et al., 2013; Bayer et al., 2012; Manhães et al., 2014; Simon et al., 2006; Watson et al., 2013) have proposed techniques for Education Data Mining (EDM) as the way forward to assist students in predicting academic failure. In this research, the emphasis is on presenting the forecast of students' success through the traces they leave when pursuing a course. The objective is to apply data mining/machine learning algorithms to students' data as students pass through a course to predict which students are at risk or are more likely to fail to fulfill the course requirements.

Such students' awareness will encourage teachers to conduct different early intervention types or provide additional and more personalized assistance as preventive measures. Many studies have reported and achieved promising results in predicting learners who are likely to fail in a given course. The data used for predicting non-academic knowledge in some of these tests, many of which require extra effort to obtain. The research has used process mining to improve existing features found in the literature.

Khobragade (2015) proposed a system based on induction and decision tree rule to predict students' academic failure using the decision tree, Naive Bayes, and Classifiers. The data used for the classifications included social, academic, and background details from the students. Such knowledge has been accumulated through surveys. A total of 11 featured features were used for prediction after implementing the feature selection algorithm. Based on the accuracy, the classifiers were then checked. The best way was offered by Naive Bayes Classifier with over 87 percent. Nevertheless, knowledge was gathered through time-consuming surveys and included methods that render general predictions and do not take early prediction into account.

The research suggested by Ahmed et al. (2015) offers an approach where EDM techniques predict first-year students' academic progress in a computer science course. The Decision tree, Naive Bayes, and the classification based on rules are used. The data used during the study course, demographic data, past academic records, and other family-related information are included. Other techniques were outperformed by Rule-based classifiers and produced 71 percent prediction accuracy.

Another research conducted by Boongoen (2017) at Thai University used a link-based cluster ensemble technique as a data transformation prediction framework. Many state-of-the-art techniques to minimize dimensionality have been compared in the study. Ye and Biswas (2014) used and extended standard characteristics for MOOC analysis with greater granularity to make more accurate predictions for drop pout and performance. The study was carried out using some comprehensive temporal features, and standard features were extended, standard features were improved, such as when some assessment started during the week or when the first lecture was watched, using data collected from video lectures, weekly quizzes, and peer evaluations from ten-week courses.

The findings showed that these characteristics improved the accuracy of the cast relative to current research. It was found that a strong indicator is a time when a student begins the peer evaluation assignment. Once the peer evaluation score was available, the performance of the prediction grew. Research indicates that the students who watched the video and did not take quizzes were the ones who mostly dropped out. Overall findings indicate that more precise temporal characteristics and more quantitative knowledge improved early prediction accuracies and false alarms compared to using only assessment score features.

Bydzovska (2016) suggested an approach that utilizes the course's features and previous grades to predict student results. Two different approaches were used. In the first approach, classification and regression were used to predict performance using academic-related data and social behavior data from students. With the small number of students, the findings were important. In the second approach, collective filtering techniques were used to predict student performance based on similarity. Classification algorithms support the vector machine, decision tree, component, IBI, RF, Naive BAyes, and rule-based classifier. Collective filtering methods were used in the second process, where the support vector machine produced the best predictions that were further improved by adding social behavior data.

Cambruzzi et al. (2015) used learning analysis to estimate the dropout rate in distance education. They developed a framework that predicts dropouts and encourages integrating pedagogical methods and textual analysis to reverse proven dropout trends. The method could predict to drop out with 87 percent accuracy, and then they drop out by 11 percent by introducing special pedagogical.

III. CLASSIFICATION METHODS

In education, suitable classification approaches have been used and are suitable for imbalanced data sets. The following are the Machine Learning algorithms used in the experiments. Naive Bayes, RF, KNN, and LR. In the following sub-sections, the classification methods employed are briefly explained.

Logistic regression: LR is a parametric approach used in classifications where, based on training data, a sigmoid function is calculated. This approach is based on the premise that a logistic distribution is accompanied by the likelihood of an occurrence occurring. The distribution is characterized as follows:

P(result=Pass/X)=11+e-Xtβ

Where a set of measurements is XTβ= β0+ β1x1 + β2x2+ + βnxn and X, X= [x1,x2,...,xn].

The input space is split into two regions using this function. New cases are categorized according to the region they belong to. The distribution is in the form of an "S," implying that the difference at the extreme ends will not affect much when opposed to the difference around the center. To represent the probabilities, LR is bounded by 0 and 1. High probabilities and low probabilities also reflect the upper and lower portions of "S," respectively. b. In the literature, this technique is commonly used to estimate retention with high precision (Lin and Reid, 2009; Mertes and Hoover, 2014; Veenstra et al., 2009; Dunn and Mulvenon, 2009).

A simple supervised procedure, which is a special type of discriminant analysis, is Naive Bayes. Based on the Bayes theorem, using the evidence obtained from observed data returns the likelihood of prediction. This methodology is based on two assumptions: all attributes are conditionally independent and contribute equally to the class's final result. There are no hidden attributes that can influence the prediction process. The Naive Bayes classifier assigns the class value with the highest conditional probability to each case. This technique has been used extensively since the 1950s and is a very common tool, especially in text mining. In cases where even the attributes are not independent, Naive Bayes performs remarkably well. In many studies in the field of education data mining, this approach is used (Pittman, 2008; Zhang et al., 2010; Khobragade, 2015; Nandeshwar et al., 2011; Mashiloane and Mchunu, 2013; Sharma and Mavani, 2011; Costa et al., 2017; Ahmad et al., 2015).

An optimization is performed at each stage to pick the function and the numeric threshold or group of values in each node if the attribute is categorical, which will generate the lowest G value if used to divide the node. This method continues until the Gini index in any node can not be reduced. A classification tree with entirely homogeneous lower nodes in the final production.

KNN is a method of classification that, by computing a distance metric, estimates the class for each new instance using the k-closest instances from the training set. Class probabilities for the new instance are calculated in each class as the proportion of neighbors set for training. Links are broken randomly or by including in the equation the k + 1 nearest neighbor. K is the number of neighbors, an important parameter to consider when this approach is used. A small value leads to an increase in the likelihood of over-fitting, while too big a value induces a high-bias classification. In a large number of classification problems, this simple algorithm has been successful (Gray et al., 2014; Minaei-Bidgoli et al., 2003; Mayilvaganan and Kalpanadevi, 2014; Yukselturk et al., 2014).

IV. RESULTS AND DISCUSSIONS

Four machine learning methods on the data set were used for prediction. The 1.1. table. Shows the characteristics gained from the data collection of MOOCs: regular features.

Table 1.Features obtained from MOOCs data set

S.No.	Feature	Explanation
1	Age	Age over the years
2	Education	Maximum qualification
3	Gender	Male/female/Null
4	The average score in week-specific quizzes	The average score in week-specific quizzes
5	Amount of attempted quizzes	Average quiz attempt for a specific week of the week
6	Quiz lag	The period from first to last quiz occurrence
7	Lecture lag	The period between the first and last lecture activities
8	Attended total Lectures	In a particular week, total lectures attended
9	Count of Video activity	Activity counts during video conferencing (pause, play, stop, etc.)
10	Efforts in terms of seconds	Total time spent in a specific week
11	Average weekly quiz score	The average score in quizzes of a specific week
12	Number of quizzes attempted	Average attempt for the quiz in a particular week
13	Quiz lag	The period from first to last quiz occurrence
14	Lecture lag	The period from first to last lecture occurrence
15	Full lecture attended	In a particular week, total lectures attended
16	Count of Video Operation	Activity counts during video conferencing (pause, play, stop, etc.)

The comparative results of machine learning algorithms' efficacy on the data set using process mining features and mean ranks of classifiers from highest (1) to lowest(N) are shown in Table 1.2 below.

Table 2. comparative results

Data set	LR	RF	NB	K.NN
Week-1	0.831	0.817	0.829	0.816
Week-2	0.831	0.833	0.861	0.796
Week-3	0.842	0.852	0.872	0.808
Week-4	0.87	0.845	0.871	0.825
Week-5	0.878	0.892	0.878	0.844
Week-6	0.854	0.889	0.88	0.835
Week-7	0.865	0.868	0.879	0.828
Week-8	0.879	0.886	0.89	0.848
Rank(mean)	2.56	2.0	*1.43*	4

Educational data mining gives an interpretation of educational data. Most of the EDM reports, however, have used conventional techniques for data mining. To achieve high prediction accuracy, this work identifies an opportunity to combine process mining approaches. To predict student results, the use of features obtained from the process mining approach is novel.

The Coursera MOOC was taken as a case study concentrating on predicting the Student's success through the traces left while pursuing a course. Data mining/machine learning algorithms have been applied to student data produced every week as students progress through a course to predict which students are at risk or are more likely to fail to meet course requirements.

V. CONCLUSION AND FUTURE SCOPE

A comparative analysis of four techniques was done in this report (LR, RF, Naive Bayes, and KNN). The influence of process mining features on the effectiveness of the above techniques has been analyzed. The findings indicate that the study methods are capable of predicting student success at an early level. The effectiveness of some techniques has improved through the incorporation of process mining characteristics with conventional features. LR and Bayes Naive.

For the data set, classifiers outperform other techniques in a statistically significant manner. Also, using an RF classifier, the significance of features was calculated. The results show that process mining features were among the top ten significant features for all data sets; however, the most critical features were features related to weekly time spent and video watching activities.

The study's significance is the use of process mining to enrich the features, and the results show that overall efficiency is statistically significantly improved using process mining characteristics. The missing values and the small size of the data are the drawbacks of this analysis. Such limitations in the data sets are recognized in this study; however, this subset indicates student engagement and completion from a real-world MOOC environment. The data sets were primarily used to demonstrate the efficacy of data mining techniques using process mining features.

REFERENCES

[1] Ahmad, F., Ismail, N.H. and Aziz, A.A. The prediction of students' academic performance using classification data mining techniques, Applied Mathematical Sciences, 9(129) (2015) 6415-6426.

[2] Bayer, J., Bydzovská, H., Géryk, J., Obsivac, T. and Popelínský, L ., Predicting dropout from social behavior of students, Proceedings of the 5th International Conference on Educational Data Mining, (2012) 103-109.

[3] Benbassat, G. , An essay towards solving a problem in the doctrine of chances. La Reconnaissance Automatique De La Parole, Rev. Laryngol. Otol. Rhinol., 111 (1990) 4 389-392.

[4] Bergner, Y., Kerr, D. and Pritchard, D.E. Methodological challenges in the analysis of MOOC data for exploring the relationship between discussion forum views and learning outcomes, Proceedings of the 8th International Conference on Education Data Mining, (2015) 234-241.

[5] Boongoen, N.I.-O.T. Improved student dropout prediction in Thai University using an ensemble of mixed-type data clusterings, International Journal of Machine Learning and Cybernetics, 8 (2) (2017) 497-510.

[6] Breiman, L. (2001), Random forests, Machine Learning, Vol. 45 No. 1, pp. 5-32.Bydžovská, H. , A comparative analysis of techniques for predicting academic performance, Proceedings of the 9th International Conference on Educational Data Mining, (2016)306-311.

[7] Cambruzzi, W., Rigo, S.J. and Barbosa, J.L.V. Dropout prediction and reduction in distance education courses with the learning analytics multi trail approach, Journal of Universal Computer Science, 21(1), (2015) 23-47.

[8] Campbell, J.P., DeBlois, PB and Oblinger, D.G. Academic Analytics, Education Review, 42(10) (2007) 40-57.

[9] Clow, D. , An overview of learning analytics, Teaching in Higher Education, 18 (6) (2013a) 683-695.

[10] Clow, D. , MOOCs and the funnel of participation, Proceedings of the Third International Conference on Learning Analytics and Knowledge, (2013b) 185.

[11] Cortes, C. and Vapnik, V. , Support-vector networks, Machine Learning, 20(3) (1995) 273-297.

[12] Costa, E.B., Fonseca, B., Santana, M.A., de Araújo, F.F. and Rego, J. , Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses, Computers in Human Behavior, 73, (2017) 247-256.

[13] Demšar, J. , Statistical comparisons of classifiers over multiple data sets, Journal of Machine Learning Research, 7(1) (2006) 1-30.

[14] Dunn, K.E., and Mulvenon, S.W. , A critical review of research on formative assessments: the limited scientific evidence of the impact of formative assessments in education, Practical Assessment, Research and Evaluation, 14(7) (2009) 1-11.

[15] Ebner, M., Lienhardt, C., Rohs, M. and Meyer, I. , Microblogs in higher education – a chance to facilitate informal and process-oriented learning?, Computers & Education, 55(1) (2010) 92-100.

[16] Gorunescu, F. , Data Mining: Concepts and Techniques, 12 (2011), Springer Science & Business Media.

[17] Gray, G., McGuinness, C. and Owende, P. , An application of classification models to predict learner progression in tertiary education, IEEE International Advance Computing Conference, (2014) 549-554.

[18] Hechenbichler, K. and Schliep, K. , Weighted k-nearest-neighbor techniques and ordinal classification, Discussion Paper 399, Collaborative Research Center 386.

[19] Khobragade, L.P. (2015), Students' academic failure prediction using data mining, 3 5, pp. 2321-7782.

[20] Kizilcec, R.F., Piech, C. and Schneider, E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses, Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (LAK '13), (2013)170-179.

[21] Lin, J.J.J. and Reid, K.J. , Student retention modeling: an evaluation of different methods and their impact on prediction results, Research in Engineering Education Symposium, (2009)1-6.

[22] Manhães, L.M.B., da Cruz, S.M.S. and Zimbrão, G. , WAVE: an architecture for predicting dropout in undergraduate courses using

EDM, Proceedings of the 29th Annual ACM Symposium on Applied Computing, (2014) 243-247.

[23] MarquezVera, C., Morales, C.R. and Soto, S.V. , Predicting school failure and dropout by using data mining techniques, IEEE Revista Iberoamericana de Tecnologias del Aprendizaje, 8(1) (2013) 7-14.

[24] Martinho, V.R.C., Nunes, C. and Minussi, C.R. , Prediction of school dropout risk group using neural network, Federated Conference on Computer Science and Information Systems, (2013) 111-114.

[25] Mashiloane, L. and Mchunu, M. , Mining for marks: a comparison of classification algorithms when predicting academic performance to identify 'students at risk,'in Rajendra, P. and Kathirvalavakumar, T. (Eds), Lecture methods with an educational web-based system, 33rd Annual Frontiers in Education Conference, 1(T2A_13-T2A_18).

[29] Nandeshwar, A., Menzies, T. and Nelson, A. , Learning patterns of university student retention, Expert Systems with Applications, 38(12) (2011) 14984-14996.

[30] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Louppe, G., Prettenhofer, P., Weiss, R. , Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher , M., Perrot, M. and Duchesnay, É. (2011), Scikit-learn: machine learning in python, Journal of Machine Learning Research, 12(10) 2825-2830.

[31] Pittman, K. (2008), Comparison of data mining techniques used to predict student retention, Ph.D. thesis, Nova South eastern University.

[32] Surendiran, R., Rajan, K.P. and Sathish Kumar, M., 2010. Study on the Customer targeting using Association Rule Mining. International Journal on Computer Science and Engineering, 2(7), pp.2483-2484.

[33] Sharma, M. and Mavani, M. , Accuracy comparison of predictive algorithms of data mining: application in education sector'," Advances in Computing, Communication and Control, Springer, Berlin and Heidelberg, (2011) 189-194.

[34] Simon, F.S., Robins, A., Baker, B., Box, I., Cutts, Q., Raadt, M.D., Haden, P., Hamer, J., Hamilton, M., Lister, R., Petre, M., Sutton, K., Tolhurst, D. and Tutty, J. "Predictors of success in a first programming course," Proceedings of the 8th Australian Conference on Computer Education, 52 (2006) 189-196.

[35] Van der Aalst, W., Adriansyah, A. and Van Dongen, B. , Replaying history on process models for conformance checking and performance analysis, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(2) (2012) 182-192.

Notes in Computer Science (including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8284, Springer International Publishing, (2013). 541-552.

[26] Mayilvaganan, M. and Kalpanadevi, D. , Comparison of classification techniques for predicting students' cognitive skill in the education environment, IEEE International Conference on Computational Intelligence and Computing Research, (2014) 1-4.

[27] Mertes, S.J. and Hoover, R.E. , Predictors of the first-year retention in a community college, Community College Journal of Research and Practice, (2014) 38(7), 651-660.

[28] MinaeiBidgoli, B., Kashy, D., Kortemeyer, G. and Punch, W. (200 3), Predicting student performance: an application of data mining

[36] Van Der Aalst, W.M.P., Van Dongen, B.F., Ğunther, C., Rozinat, A., Verbeek, H.M. W. and Weijters, A.J.M.M. , Prom: the process mining toolkit, CEUR Workshop Proceedings, 489, Ulm, September 8 (2009).

[37] Veenstra, C.P., Dey, E.L. and Herrin, G.D. , A model for freshman engineering retention, Advances in Engineering Education, 1(3) (2009) 1-23.

[38] Watson, C., Li, FWB and Godwin, J.L. Predicting performance in an introductory programming course by logging and analyzing student programming behavior, Proceedings – 2013 IEEE 13th International Conference on Advanced Learning Technologies, (2013)319-323.

[39] Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G.J., Ng , A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, DJ and Steinberg, D. Top 10 algorithms in data mining, Knowledge and Information Systems, 14(1) (2008) 1-37.

[40] Ye, C. and Biswas, G. , Early prediction of student dropout and performance in MOOCs using higher granularity temporal information, Journal of Learning Analytics, 1(3) (2014) 169-172.

[41] Yukselturk, E., Ozekes, S., Türel, Y.K., Education, C., Ozekes, S., Türel, Y.K. and Education, C. , Predicting dropout student: an application of data mining methods in an online education program, European Journal of Open, Distance, and E-Learning, 17(1), (2014)118-133.

[42] Zhang, Y., Oussena, S., Clark, T. and Hyensook, K. , Using data mining to improve student retention in HE: a case study, Proceedings of the 12th International Conference on Enterprise Information Systems, 1, (2010) 190-197.