https://doi.org/10.36315/2023v2end016
Education and New Developments 2023

PREDICTING STUDENT PERFORMANCE FROM MOODLE LOGS IN HIGHER EDUCATION: A COURSE-AGNOSTIC APPROACH

Ricardo Santos, & Roberto Henriques
Nova IMS Information Management School, Universidade Nova de Lisboa (Portugal)

Abstract

The institutional adoption of learning management systems (LMS) aims to improve educational outcomes and reduce churn through student engagement with educational content. Modern LMS record all student interactions and store them as activity logs that encode patterns of learning behaviour. Previous research has shown that insights derived from log data can detect students at risk of failing in a single or a few courses, but comprehensive institution-wide surveys are few and far between.
The work presented herein uses machine learning to create predictive models to identify students at risk or excellent students using the Moodle logs generated by a sample of 9296 course enrollments at a Portuguese information management school. 31 candidate features were extracted to create and train different predictive models. Model performance was evaluated through 30 repetitions of Stratified K-Fold Cross-Validation, using the area under the receiver operating characteristic (ROC) curve (AUC) and the F1-score. All experiments were repeated with the addition of the average of the intermediate grades obtained by the student in the course as a 32nd candidate feature.
The results suggest that features extracted from Moodle logs are good predictors of students at risk, as indicated by the 0.752 AUC score achieved by Random Forest. The addition of intermediate grades significantly improves the predictive performance, leading to an AUC score of 0.922 and F1-Score of 0.693 for the best classifier, Gradient Boosting. However, the performance for identifying excelling students was comparatively lower, with an AUC score of 0.781 and F1-Score of 0.567 for Gradient Boosting. Future work should focus on exploring the implementation of an early warning system that can assist educators in identifying students in need while there is still time to provide feedback and develop corrective measures.

Keywords: *Student performance, learning management systems, higher education, classification, machine learning.*

1. Introduction

The potential role of higher education institutions (HEI) in promoting prosperity and sustainability in communities and society at large is widely recognised by scholars (Žalėnienė & Pereira, 2021) and policymakers worldwide have, throughout time, made efforts to democratise and increase flexibility in access to tertiary-level education in their countries (OECD, 2022). However, increases in student enrollment have also brought a plethora of new challenges to HEI, with the decreased ability of educators to track and monitor the progress of each individual being among the main ones (Clancy & Goastellec, 2007; Macfadyen & Dawson, 2010).

Learning management systems (LMS) are digital tools with close to ubiquitous adoption by HEI whose primary purpose is facilitating the engagement of students with the educational content, whether it is accessing course materials or communicating remotely with educators (Coates, James, & Baldwin, 2005; Walker, Lindner, Murphrey, & Dooley, 2016). Modern LMS keep timestamped records of every student interaction with the system, referred to as clickstream data, which educators and researchers use to track student progress and provide personalised support (Bernacki, Chavez, & Uesbeck, 2020; Macfadyen & Dawson, 2010).

Clickstream data has gathered the interest of researchers and educators that attempt to predict student performance since the mid-2000s. In 2006, Calvo-Flores, Galindo, Jiménez, & Pérez (2006) extracted features from the LMS logs of 240 students attending a course at a Spanish university to train an artificial neural network (ANN) that achieved 80.2% accuracy when predicting whether a student would pass or fail. While arguing that clickstream data could play a role in the early identification of students at risk, Macfadyen and Dawson (2010) used a Logistic Regression (LR) to correctly identify 80.9% of the

students at risk while maintaining an accuracy of 73.7%. Throughout time, research in student performance prediction has branched into different niches. In the first niche, works like Zacharis (2015, 2018) were mainly focused on making predictions using models trained from data from a single course. In recent years, the work presented in Bernacki et al. (2020) went a step beyond identifying struggling students in a course, but it also allowed some of them to receive timely feedback and outperform the struggling students that did not receive that feedback. In a second relevant niche, works like Gašević, Dawson, Rogers, and Gasevic (2016) or Conijn, Snijders, Kleingeld, and Matzat (2017) skeptically explored the possibility of creating models trained on data from multiple courses and eventually argued against the use of general models due to their poor performance against course-specific models. The research works in the third niche mainly focused on general course-agnostic models that could be applicable in multiple contexts. For example, Romero, Espejo, Zafra, Romero, and Ventura (2013) and Tsiakmaki, Kostopoulos, Kotsiantis, and Ragos (2020) obtained average accuracies of 66% and 86.1% respectively, using models trained on data from 7 different courses. In addition, there is also a set of works that achieves outstanding performances with models trained with data from more than 600 courses (Baneres, Rodriguez, & Serra, 2019; Riestra-González, Paule-Ruíz, & Ortin, 2021). Despite these efforts, literature on institution-wide surveys that aim to predict student performance is scarce. Moreover, most works focus solely on identifying students at risk, rather than identifying students with high potential.

This study investigates the potential of data from the Moodle LMS to predict and identify students who are at risk and, in a parallel problem, identify students who are excelling. The research question being addressed is: *Can institution-wide clickstream data from a Learning Management System accurately predict and identify at-risk and high-potential students in higher education institutions?* To answer this question, we compare the performance of different course-agnostic predictive models trained on features extracted from Moodle logs obtained from courses taught at a Portuguese information management school. The results suggest that features extracted from LMS logs exhibit predictive potential and can contribute to future development of more generalisable early warning systems, pedagogical strategies and support systems in higher education institutions.

The remainder of this paper is structured as follows: the following section presents the data and methods used in this study. The third section presents and discusses our main findings, followed by the conclusion and reccomendations for future work.

2. Methodology

2.1. Data

This work analysed data from 138 courses taught at a Portuguese information management school collected during the 2020/2021 academic year. The data comprised 9296 course enrollments by 1590 unique students and included the Moodle logs, intermediate grades, and final course grades associated with each enrollment. The final grades of each student were used to create two binary variables. The first variable classified students as being *at risk* (1) if they scored less than 11 out of 20 in a course, with the remaining students being *not at risk* (0). The second variable classified students as *excelling* (1) if their final course score exceeded the 85th percentile of the course and *not excelling* (0) otherwise. Table 1 showcases the distribution of courses, students, and enrollments according to the program level. Notably, undergraduate level courses had close to 50% (911 out of 1872) of the enrollments labeled *at risk* across all program types despite only representing a third of the total number of enrollments. Moreover, the largest proportion of excelling students was found in courses taught in master's level programs.

Table 1. Overview of the characteristics of courses and students per program level.

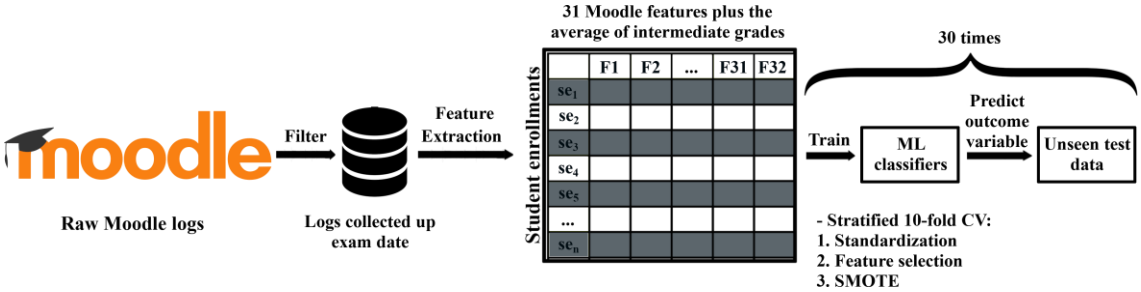
Program level	Courses	Unique Students	Student enrollments	Enrollments per course	Students at risk	Excelling students
Undergraduate	55	409	3387	61.58	918	769
Master's	62	872	5013	80.85	833	1543
Postgraduate	21	325	896	42.67	121	262
Total	**138**	**1606**	**9296**	**67.36**	**1872**	**2574**

2.2. Data analysis

Figure 1 illustrates the experimental design followed for each classification problem. The work was divided into two stages and, unless explicitly noted otherwise, all data manipulation and analysis were performed in Python (McKinney, 2018) using Scikit-learn (Pedregosa et al., 2011). In the first stage, Moodle logs were preprocessed and converted into a dataset suitable for training various machine learning classifiers, with 31 candidate features extracted per student enrollment. The second stage

involved creating and training different machine learning classifiers using the dataset created in the first stage to address the two classification problems. Ten traditional machine learning classification algorithms were trained for each classification problem: K-Nearest Neighbors (KNN), LR, Naïve Bayes (NB), Classification and Regression Tree (CART), ANN, Support Vector Machines (SVM), RF, Extremely Randomised Trees (ExtraTrees), Adaptive Boosting (AdaBoost) and Gradient Boosting (GBoost). Model performance was initially evaluated using the average area under the receiver operating characteristic (ROC) curve (AUC), with the best models being evaluated by the F1-score, a metric computed from precision and recall. All performances reported herein refer to the average model performance across 30 repetitions of training with Stratified 10-Fold Cross-Validation.

Figure 1. Overview of the experimental approach adopted for each classification problem.



Standardisation, feature selection, and Synthetic Minority Oversampling Technique (SMOTE) (Chawla, Bowyer, Hall, & Kegelmeyer, 2002) were performed independently for each fold. The choice of which features to keep in each classification problem was made by a multi-layered feature selection process that required a feature to be found relevant by a minimum of four of the following eight algorithms: Recursive Feature Elimination (Guyon, Weston, Barnhill, & Vapnik, 2002) in its simple and with cross-validation forms, Ridge Regression, Lasso Regression, ElasticNet Regression, and the application of SelectFromModel to LR, Random Forest (RF), and Light Gradient Boosting Machine (Ke et al., 2017), with the latter not having a Scikit-learn implementation. Moreover, the experimental procedure was repeated on a modified dataset that also featured the average of the student's intermediate grades. A brief description of all features used in this study can be found in Table 2.

Table 2. List and description of the 32 candidate features extracted from the source data.

Features	**Description**
Total clicks (n)	Number of clicks made in the course
Clicks (% of course total)	Number of clicks made relative to total clicks of all students in the course
Online sessions (n)	Number of online sessions
Clicks/session (n)	Total clicks / Online sessions
Clicks/day (n)	Total clicks/ number of days
Forum clicks (n)	Number of clicks on the course forum
Discussions viewed (n)	Number of discussions and course forum posts viewed
Forum posts (n)	Number of posts and replies in discussions and course forum
Folder clicks (n)	Number of clicks on folders
Resources viewed (n)	Number of course educational resources viewed
URLs viewed (n)	Number of clicks on external links
Course clicks (n)	Number of clicks on course pages
Assessments started (n)	Number of assessments and quizzes started
Assignments viewed (n)	Number of assignment page views
Assignments submitted (n)	Number of assignments submitted
Submissions (% of course total)	Number of submissions relative to total submissions made in the course
Total time online (min)	Sum of the duration of all online sessions undertaken by the student
Aver. duration of online sessions (min)	Total time online / Online sessions
Largest period of inactivity (h)	Largest temporal interval between consecutive online sessions
Days with 0 clicks (n)	Difference between the number of days and days with at least one click
Days with 0 clicks (% of period)	Percentage of Days with 0 clicks
PercCourse_1Login	Percentage of course duration at the 1st login by the student in the course
PercCourse_2Login	Percentage of course duration at the 2nd login by the student in the course
….	Percentage of course duration at the nth login by the student in the course
PercCourse_10Login	Percentage of course duration at the 10th login by the student in the course
Average of intermediate grades	Average of the intermediate grades obtained by the student in the course

3. Results and discussion

Table 3 presents the average performance of the classifiers used in this study. Model selection was primarily performed using AUC score, with F1-score being a secondary criterion. The models selected from the initial screening stage for each experiment and classification problem are highlighted in bold.

Table 3. Average AUC, Accuracy and Recall scores obtained by each classification algorithm.

	Students at risk				**Excelling Students**			
	Moodle		**Moodle + Intermediate grades**		**Moodle**		**Moodle + Intermediate grades**	
	AUC	**F1-score**	**AUC**	**F1-score**	**AUC**	**F1-score**	**AUC**	**F1-score**
KNN	0.714	0.435	0.839	0.570	0.587	0.421	0.623	0.447
LR	0.707	0.422	0.783	0.518	**0.616**	**0.432**	0.634	0.450
NB	0.677	0.392	0.716	0.443	0.600	0.397	0.608	0.414
ANN	0.712	0.432	**0.898**	**0.666**	0.584	0.404	0.673	0.476
CART	0.678	0.413	0.843	0.654	0.569	0.378	0.708	0.545
SVM	**0.727**	**0.445**	**0.894**	**0.642**	**0.618**	**0.429**	0.690	0.498
RF	**0.752**	**0.460**	**0.921**	**0.693**	**0.621**	**0.389**	**0.756**	**0.563**
AdaBoost	0.704	0.418	**0.906**	**0.658**	0.607	0.391	**0.755**	**0.553**
GBoost	**0.742**	**0.421**	**0.922**	**0.693**	**0.616**	**0.332**	**0.781**	**0.567**
ExtraTrees	**0.724**	**0.432**	**0.897**	**0.640**	**0.626**	**0.436**	0.720	0.526

Before adding the average intermediate grades as a feature, RF achieved the highest AUC (0.752) and F1-score (0.460) for identifying students at risk. For the identification of excelling students, the best classifier was ExtraTrees with an AUC of 0.626 and an F1-score of 0.436. Adding average intermediate grades improved performance, making GBoost the best classifier for both problems. GBoost had an AUC score of 0.922 and a F1-score of 0.693 for students at risk and an AUC score of 0.781 and a F1-score of 0.567 for excelling students. Interestingly, the classifiers with the highest AUC score always had the highest F1-scores, even among the non-selected classifiers. However, this trend was not consistent when comparing between all models.

3.1. Discussion

Per the nomenclature adopted by Gašević et al. (2016), a classifier exhibits acceptable discriminative capabilities if it achieves an AUC score greater than 0.7. While exclusively using Moodle logs, 9 out of 10 classifiers met this threshold when identifying students at risk. However, when identifying excelling students the best performances did not go beyond poor discrimative capabilities (with ExtraTrees achieving 0.626 AUC). Nonetheless, the results demonstrate that features extracted from LMS have discriminative power on their own even if they do not encapsulate all of the information that would reasonably be accessible to an educator when making the prediction, as is the case of intermediate grades obtained throughout the course.

The addition of the intermediate grades led to substantial bumps in discriminative performance in both classification problems. For students at risk, RF, AdaBoost and GBoost achieved AUC scores greater than 0.9, with ANN, Extratrees and SVM nearly reaching this benchmark as well. For identifying excelling students with intermediate grades, 5 classifiers met the 0.7 benchmark with RF, AdaBoost and GBoost having AUC scores greater than 0.75. These results are consistent with other works that have found intermediate grades to be among the influential predictors of performance (Conijn et al., 2017; Riestra-González et al., 2021).

Overall, features extracted from LMS clickstream exhibit the potential to help educators identify either students at risk or excelling students. That potential can be enhanced by combining the LMS features with other data from other reasonably accessible sources of data, as is the case with the partial grades obtained throughout the course.

4. Conclusion

The work presented herein uses LMS log data collected from a Portuguese information management school to create models that predict student performance. The findings show that LMS data exhibits good discriminative power in, at least, the identification of students at risk. Future research could explore whether the discriminative power seen in the analysis can be extended to early identification of students of interest, and potentially implement these models into a customized version of Moodle for real-time identification.

References

Baneres, D., Rodriguez, M. E., & Serra, M. (2019). An Early Feedback Prediction System for Learners At-Risk Within a First-Year Higher Education Course. *IEEE Transactions on Learning Technologies*, *12*(2), 249–263. https://doi.org/10.1109/TLT.2019.2912167

Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting achievement and providing support before STEM majors begin to fail. *Computers & Education*, *158*, 103999. https://doi.org/10.1016/j.compedu.2020.103999

Calvo-Flores, M. D., Galindo, E. G., Jiménez, M. C. P., & Pérez, O. (2006). Predicting students' marks from Moodle logs using neural network models. *Current Developments in Technology-Assisted Education*, *1*(2), 586–590.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Clancy, P., & Goastellec, G. (2007). Exploring Access and Equity in Higher Education: Policy and Performance in a Comparative Perspective. *Higher Education Quarterly*, *61*(2), 136–154. https://doi.org/10.1111/j.1468-2273.2007.00343.x

Coates, H., James, R., & Baldwin, G. (2005). A critical examination of the effects of learning management systems on university teaching and learning. *Tertiary Education and Management*, *11*, 19–36.

Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting Student Performance from LMS Data: A Comparison of 17 Blended Courses Using Moodle LMS. *IEEE Transactions on Learning Technologies*, *10*(1), 17–29. https://doi.org/10.1109/TLT.2016.2616312

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002

Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, *46*(1–3), 389–422.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., … Liu, T.-Y. (2017). LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, *30*, 3149–3157.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

OECD. (2022). *Education at a Glance 2022: OECD Indicators*. OECD. https://doi.org/10.1787/3197152b-en

Riestra-González, M., Paule-Ruíz, M. del P., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. *Computers & Education*, *163*, 104108. https://doi.org/10.1016/j.compedu.2020.104108

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), 135–146. https://doi.org/10.1002/cae.20456

Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Transfer Learning from Deep Neural Networks for Predicting Student Performance. *Applied Sciences*, *10*(6), 2145. https://doi.org/10.3390/app10062145

Walker, D. S., Lindner, J. R., Murphrey, T. P., & Dooley, K. (2016). Learning Management System Usage: Perspectives From University Instructors. *Quarterly Review of Distance Education*, *17*(2), 41–50.

Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education*, *27*, 44–53. https://doi.org/10.1016/j.iheduc.2015.05.002

Zacharis, N. Z. (2018). Classification and Regression Trees (CART) for Predictive Modeling in Blended Learning. *International Journal of Intelligent Systems and Applications*, *10*(3), 1–9. https://doi.org/10.5815/ijisa.2018.03.01

Žalėnienė, I., & Pereira, P. (2021). Higher Education For Sustainability: A Global Perspective. *Geography and Sustainability*, *2*(2), 99–106. https://doi.org/10.1016/j.geosus.2021.05.001