

Utilizing Interaction Metrics in a Virtual Learning Environment for Early Prediction of Students' Academic Performance

Saša Brdnik*, Vili Podgorelec and Tjaša Heričko

Faculty of Electrical Engineering and Computer Science, University of Maribor, Koroška cesta 46, Maribor, Slovenia

Abstract

A plethora of data on students' activities and interactions is acquired through the use of virtual learning environments. The analysis of gathered data and the prediction of withdrawal and academic success has been an ongoing point of discussion in the field of Learning Analytics. The main aim of this paper is an early prediction of students' academic performance in terms of the final grade, while also targeting the early identification of students at risk of failing the course. We introduce new handcrafted features encompassing interaction patterns of students' virtual learning environment usage, and highlight their predictive powers in monthly predictions. The experiment showed that a combination of the proposed features improved the Root Mean Square Error by up to 7.2 grade points compared to a model without the proposed features. Improvement of up to 0.52 was also observed in the coefficient of determination. Proposed model's accuracy of predicting students at risk of failure reached 74% at the end of April.

Keywords

virtual learning environments, academic performance, at-risk students, interaction behavior, interaction patterns, machine learning, regression

1. Introduction

Virtual learning environments (VLEs) have become all-important tools for educational institutions in the recent past [1]. By facilitating access to various teaching-learning materials and enabling learning without spatial or temporal limitations [1, 2], they provide support for both distance and classroom teaching [3]. Through the use of VLEs, the abundance of data capturing learners' activities and interactions is collected from the log records [3, 4]. The gathered data can be analyzed to mine learners' behaviors and extract patterns, especially those concerning distinct groups of learners [4, 5, 6]. Among others, applying such learning analytics helps identify students possibly at risk of failing to complete the course, which is a major challenge confronting the education sector [7]. This represents the first step toward addressing the problem of at-risk students by attaining a meaningful understanding of students' behavior and their learning outcomes, as well as assisting in the designing of data-driven corrective strategies and timely interventions aimed to help improve the students' academic performance [6, 8, 9, 10].

SQAMIA 2022: Workshop on Software Quality, Analysis, Monitoring, Improvement, and Applications, September 11–14, 2022, Novi Sad, Serbia

*Corresponding author.

✉ sasa.brdnik@um.si (S. Brdnik); vili.podgorelec@um.si (V. Podgorelec); tjasa.hericko@um.si (T. Heričko)

🆔 0000-0001-6955-7868 (V. Podgorelec); 0000-0002-0410-7724 (T. Heričko)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

The analysis of data collected by VLEs has received much research attention in the last decade, for its potential to impact and improve the teaching-learning process and address significant challenges in education, including identifying students at risk of failing or withdrawing from their respective courses. Many research endeavors have applied machine learning-based approaches, which confirmed that data from VLE activity logs present good predictors for forecasting students' academic performance [1, 3, 4, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16]. Identifying potentially at-risk students accurately while a course is still in progress is crucial to providing timely help aimed at minimizing the student failure rate [7, 16]. This paper aims to predict students' final grades and risk of failure as early as possible during a course, based on their interactions with the VLE. New features were also proposed to improve the prediction model. Motivated by finding significant predictors for the early prediction of student academic success, we used the Open University Learning Analytics (OULA) dataset to extract students' interaction patterns in VLE. We investigated how the interaction patterns captured by the proposed handcrafted features relate to academic performance. Then, we approached the prediction problem from a regression perspective – the final mark in points was predicted for each student, instead of a reasonably simple pass-fail prediction. The proposed predictive model was constructed during different points in the course's timeline from data available until that point, instead of using all data known after the course had already finished.

In summary, the key contributions of this work are threefold: (1) We present an insightful prediction of student academic performance in terms of final mark forecast at a fine-grained level, i.e., month-wise granularity, focusing on the early identification of students at risk of failing the course. (2) By introducing new handcrafted features encompassing interaction patterns of students' VLE usage, we highlight their predictive powers, and provide an additional understanding of the characteristics of at-risk students, which builds on findings from previous work. (3) The proposed features, on average, improve the Root Mean Square Error (RMSE) of predicted grades by 3.6 in all observed months, with improvement of up to 7.5 grade points in particular months. Compared to the models without the proposed features, our model improved the explained proportion of observed variation in grades, with a mean R^2 improvement of 0.245.

The rest of the paper is organized as follows. Section 2 presents a literature review of related work. The methodology is outlined in Section 3. Section 4 reports the results of the experiment. In Section 5, the results are discussed and compared to approaches from related work. Section 6 concludes the paper with summarized remarks and proposed future research directions.

2. Literature Review

Predicting students at risk of failing a course with a classification approach has been explored by [3, 4, 7, 9, 12, 13, 17], namely using C4.5 Decision Tree, K-Nearest Neighbors, Naïve Bayes, Support Vector Machines, and Neural Networks. A regression approach is less common in similar studies, and can be observed in [11, 18], most using Logistic Regression. Other focuses in the field are predicting students at risk of dropout [19, 20, 21], and students who fail to finish their academic obligations in time [10]. Some studies gathered and presented their datasets [11, 13, 16, 22, 23], while most have conducted experiments on publicly available datasets [1, 3, 4, 6, 7, 9, 10, 12, 15] Related work differs in the scope of the predictions, with studies

focusing on generalized predictions [11, 12, 22] and on specific course predictions [13, 24]. Most studies analyze data from the entire course timeline [12, 1], while others focus on predictions in selected periods [4, 14, 15, 16]. Prediction models generally use students' demographic data, VLE logs, assessment scores, and previous academic records. The raw VLE data are in the form of simple logs. Researchers, therefore, depend on different features extracted from them. Features vary from studies; the most common are daily or weekly summarizations of activities in VLE, e.g., the sum of all the daily clicks in a VLE [14], the daily number of sessions [18], and the number of interactions with each activity type [15]. We observed an opportunity for exploring new metrics derived from basic log data. Similarly, Mubarak et al. [21] presented an interesting feature capturing the number of active days per week, while Waheed et al. [6] included the sum of clicks on different content types before and after the course.

Classification studies focused on student success (or failure) recognition at different times in the semester report high accuracy of their predictions. Accuracy is defined as the number of correct predictions divided by the total number of predictions, while precision refers to the number of true positives divided by the total number of positive predictions. On the OULA dataset, Kuzilek et al. [8] conducted weekly predictions of at-risk students, and reported 47% precision of prediction at the start of the semester and 90% precision at the end of the semester. Aljohani et al. [4] reported accuracy of around 80% by week 5, 90% by week 10, and 93% by the end of the semester. Al Azawei et al. [15] conducted daily predictions and reported 70% accuracy on day 53 (week 9), and 91% accuracy on the day before the end of the semester. Herrmannova et al. [14] achieved an F-score of 0.80 on week 4 and 0.93 on week 8. Our study differs from existing work with the presentation of new features obtained from the OULA dataset, and using regression for student success prediction, offering some more insight into the predicted range of students' final grades at different time frames through the semester. A review of selected related work using VLE logs for predicting student success is presented in Table 1.

3. Methodology

This study uses an experimental design. The experiment was conducted in Python 3.8.5 in a Jupyter notebook 3.8.5 environment, using Pandas 1.2.4, Numpy 1.19.2, Sklearn 0.23.2, and Matplotlib 3.3.2. Statistical analyses were performed using Jamovi 1.2.27.

3.1. Dataset

The OULA dataset [17] was used for experimental analysis. The publicly available dataset provides anonymized data about courses, students' demographics, their interaction with a VLE, and their academic performance. The dataset was gathered by the Open University in the UK in the academic year 2013/2014. It comprises seven courses, four semesters, and 32,593 students altogether. For our analysis, the course *DDD* was selected, as it is the only one with available final exam scores. Students could obtain grades between 0-100, with 40 set as a threshold for a positive grade. The course consists of Tutor Marked Assessment (TMA) and a Final Exam, each worth 50% of the final grade. The TMA grade is calculated from six assessments, weighing 5%, 10%, 10%, 25%, 25%, and 25% of the grade, respectively. The course was organized in the fall semester, starting in October 2013, with the final exam in May 2014.

Table 1
Overview of related work

Study	Prediction	Features	Time Interval	Results
[1]	Course pass rate	Demographics, VLE log	Entire course	Acc=78.5%
[3]	Student success/failure	Demographics, VLE log	Daily	Acc=88%
[4]	Students at risk of failing	VLE log	Weekly	Acc _{Week 5} =80%, Acc _{Week 10} =90% Acc _{Week 25} =93%
[6]	Students at risk of failing and withdrawing	Demographics, VLE log	Semester quarters	Acc=84%–93% Acc _{At-risk} =88%
[7]	Students at risk of failure and dropout	VLE log	Milestone	Acc _{Week 10} =75%
[8]	At-risk students (failing, submitting assessments)	Demographics, VLE log, course data, grades in other courses	Weekly, milestones	Prec _{Semester Start} =47% Prec _{Semester End} =90%
[9]	Students at risk of failing	Demographics, VLE log, assessment data, date of registration	Daily	F ₁ =0.25–0.71
[10]	Students at risk of not completing on time	VLE log, assessment data	Milestones	F ₁ =0.98
[12]	Students at risk of failure and marginal students	Demographics, VLE log	Entire course	Acc _{At-risk} =92.2%–93.8% Acc _{Marginal} =91.3%–93.5%
[14]	Students at risk of failing	Demographics, VLE log	Weekly	F _{1, Week 4} =0.80 F _{1, Week 8} =0.93
[15]	Student final score	Demographics, VLE log, student scores	Milestones	Acc _{Day 53} =70% Acc _{Day 207} =91%

Acc = Accuracy, Prec = Precision, F₁ = F-score

3.2. Data Preparation and Feature Computation

Data preprocessing included several strategies and techniques, with the main goal of selecting data for analysis and computing features from the data. First, tables with the course, assessments, and student data were filtered to the values connected to the *DDD* course. For each assessment, we joined the data (grades and submission dates) from the table with student assessments. Missing values were converted to zero. From the scores and assessment weights, first, the TMA was calculated, following the formula $\sum_{i=1}^n weight_i \times score_i$, with n being the number of assessments. The final exam and TMA scores were used to calculate the final score. The resulting dataset was then joined with the students' demographics. At this point, 1,803 students were observed from *DDD* course. Due to the high unregistration (i.e. student withdrawal) rate, only students who remained registered during the semester and submitted at least one assignment or took the final exam were kept. As we were not interested in student performance on a course-granular level, students were removed who had previously already attended the course. The final dataset included data for 777 unique students. Categorical data was encoded as numerical by converting it into dummy variables. The earliest date was extracted of a student's interaction with a VLE. VLE interaction dates were transformed into DateTime format, and information about the date, weekday, and month of each VLE activity were added to the dataframe. The VLE activity logs were then grouped separately in months and weekdays for each student. The

sum of clicks on each weekday and each month were calculated, as well as weekday (Mon–Fri) activity percentage, compared to weekend (Sat, Sun) activity. The overall sum of clicks and clicks on a particular activity type were calculated for the whole semester and each month, including the activities in September before the course started. We were interested in observing patterns of inactivity in the VLE. Thus, we included features for the number of days when a student was active in a VLE from the start of the semester, the number of inactive days since the start of the semester, as well as the highest number of consecutive active and inactive days during the observed time period, which were calculated as the longest intervals of consecutive days students interacted (or had not interacted) with the VLE. Eight copies of the dataset were created, where the metrics were recalculated for observations on the last day of each month between September and April. The final datasets included 24 to 68 features, with the number varying due to the available assessment grades and the number of previous months. Features that have, to the best of our knowledge, not been observed in related work on the OULA dataset, are presented in Table 2. Additionally, normalization of all inputs with MinMaxScaler was conducted for prediction, which scaled and translated each feature so that it was in the given range on the training set (i.e., between 0 and 1), enduring features with a higher value (e.g., number of clicks) do not, inherently, have higher importance than features with lower values (e.g., the first day of interaction).

Table 2
Proposed novel interaction metrics

Metric	Description
<i>WeekdaysClicksRatio</i>	Represents the ratio of VLE activity (clicks) on weekdays (Mon–Fri), compared to the weekends (Sat, Sun) in the observed time period.
<i>DailyClicksPreSemester</i>	Mean number of daily clicks in the VLE before the start of the semester.
<i>NrOfDaysActive</i>	Represents the number of days a student interacted with the VLE since the start of the semester. The opposite metric was also computed for inactive days.
<i>LongestNrOfDaysActive- Consequently</i>	Represents the number of consecutive days a student was active in the observed time period. The opposite metric was also computed for inactive days.

3.3. Exploratory Data Analysis

Our dataset includes 467 (60%) male and 310 (40%) female students. Most students (570 or 73%) were in the age bracket of 0–35 years, 202 (26%) were aged between 35 and 55, and only five were older than 55. Most students (395 or 51%) achieved A level or equivalent education before taking the course; another 259 (33%) had education lower than A level, 114 (15%) had higher education qualifications, 3 had no formal qualification, and 6 had post-graduate qualifications. Based on the university’s performance standards thresholds, students were divided into seven brackets (bare fail: 0–14, fail: 15–29, bare fail: 30–39, pass 4: 40–54, pass 3: 55–69, pass 2: 70–84, pass 1: 85–100). Out of 777 students, 124 failed the class. By exploring students’ activity, measured with the sum of clicks in VLE, and their success in passing the course, we found out the following. The hypothesis for normal distribution of the number of clicks in the VLE was rejected with the Kolmogorov-Smirnov (KS) test ($p < .001$, $D = .199$). The Man-Whitney U (MW-U) test showed a significant difference ($U = 12763$, $p < .001$) between the number of VLE clicks of

students who passed and those who failed. Successful students interacted more with the VLE (N=653, M=1338, SD=1357), while students who failed the course interacted less (N=124, M=432, SD=345). The hypothesis for normal distribution of the date of the first interaction with VLE in the groups of successful and unsuccessful students was rejected with the KS test ($p < .001$, $D = .24$). The MW-U test also showed a significant difference ($U = 28865$, $p < .001$) in successful (N=653, M=-20.3, SD=6.5) and unsuccessful students (N=124, M=-17.5, SD=7.1) based on the date of their first interaction with the VLE, with 0 being the first day of the semester. Further insight into VLE interaction is presented in Figure 1, displaying the sum of students' daily clicks in correspondence with course events. Correlation between students' interaction with the VLE and the beginning of the semester is visible, as well as their increased interaction before the final exam in May. Students' average activity in the VLE during the week was highest on Thursdays and Fridays, while the least activity is observed on Tuesdays. The trend of increased clicks on Thursdays and Fridays and the lowest activity on Tuesdays was similar in groups of successful and unsuccessful students. On average, students interacted with the VLE more over the weekend than on Monday, Tuesday, or Wednesday. The mean share of weekday clicks through the semester was 68.5%, meaning almost one-third of interactions in the VLE were conducted at the weekends. Observing all VLE activities in the semester by activity type, students most commonly clicked on forums (M=427), homepage (M=303), subpages (M=170), and outcontent (M=145).

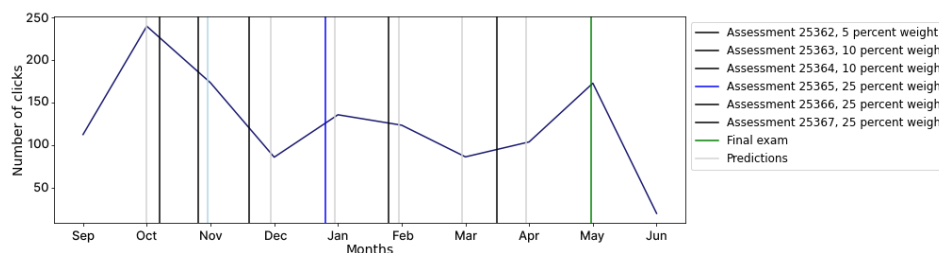


Figure 1: Daily VLE activity and course calendar

We observed differences in academically successful and unsuccessful students through the proposed metrics. The mean values of the percentage of VLE clicks over weekdays, average daily VLE clicks before the semester, the number of active days, and the number of consecutively active days, were re higher for successful students. On the contrary, the mean values of the overall number of inactive days and the overall number of consecutively inactive days were higher in the group of unsuccessful students. Comparing the ratio of weekday clicks, the mean number was higher for successful students ($M = .688$) than unsuccessful ($M = .665$), suggesting they were more active during the weekdays than at the weekend. The hypothesis for the normal distribution of the weekday clicks ratio in groups of successful and unsuccessful students was confirmed with the KS test ($p = .187$, $D = .0401$). The assumption of variances' homogeneity was violated using Levene's test ($F = 19.7$, $p < .001$). Thus, Welch's T-test was used, and no significant difference ($t(111) = -1.86$, $p = .065$) was observed between the weekday clicks ratio in the groups of successful and unsuccessful students. Further, the average daily pre-semester clicks were observed. The hypothesis for the normal distribution of average pre-semester clicks in groups of successful and unsuccessful students was rejected with the KS test ($p < .001$, $D = .229$).

The MW-U test showed a significant difference ($U=30219$, $p<.001$) in daily VLE clicks pre-semester between the observed groups. Before the semester, successful students had interacted daily more with the VLE ($N=653$, $M=4$, $SD=5.7$) than unsuccessful students ($N=124$, $M=2.3$, $SD=3.1$). The mean daily number of pre-semester clicks for successful students ($M=4$) was almost double compared to unsuccessful students ($M=2.3$). Additionally, the KS test rejected the hypothesis for normal distribution of the overall number of active days in groups of successful and unsuccessful students ($p<.001$, $D=.0995$). The MW-U test showed a significant difference ($U=10376$, $p<.001$) between the overall number of active days of students who passed and those who failed. Successful students interacted with the VLE on a more frequent daily basis ($N=653$, $M=101$, $SD=50.9$), while students who failed the course ($N=124$, $M=42.3$, $SD=24.8$) interacted with the VLE less often in terms of daily accesses. The KS test rejected the hypothesis for normal distribution of the overall number of consecutive active days in groups of successful and unsuccessful students ($p<.001$, $D=.0261$). The MW-U test showed a significant difference ($U=-9.34$, $p<.001$) between the overall number of consecutive active days of successful and unsuccessful students. Successful students interacted consecutively more with the VLE ($N=653$, $M=11.1$, $SD=14.7$), while students who failed the course ($N=124$, $M=5$, $SD=3.5$) interacted with the VLE in larger gaps. Oppositely, successful students had shorter consecutive inactive periods of VLE usage ($N=653$, $M=22.2$, $SD=11.9$), while unsuccessful students had longer ($N=124$, $M=98.6$, $SD=66.1$). The KS test rejected the hypothesis for normal distribution of the overall number of consecutive inactive days in the two groups of students ($p<.001$, $D=.0209$). The MW-U test showed that the difference was significant ($U=12.8$, $p<.001$)

3.4. Prediction Task

Linear regression was used for prediction. Due to the small sample size, 5-fold cross-validation with scikit KFold was used to improve the estimate of the model performance. RMSE and R^2 were obtained for each period, calculated as $RMSE = \sqrt{\frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (y_i - \hat{y}_i)^2}$, where y_i is the predicted value of the i -th sample, and \hat{y}_i is the corresponding true value, and unadjusted $R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$, where \hat{y}_i is the predicted value of the i -th sample and y_i is the corresponding true value for total samples.

To compare the impact of the proposed metrics on the performance, three other models were created – M_B containing grades and VLE metrics, used most commonly in related work; M_C containing demographics, VLE metrics, and grades; M_D containing demographics and grades. A detailed overview of the features included in each model is presented in Table 3. Prediction models were compared monthly, as times indicated in Figure 1, with mean RMSE and R^2 calculated from all folds for each monthly prediction. The evaluations were conducted after the end of each month on days 0, 31, 61, 92, 123, 151, 182, and 212. Accuracy in predicting failure was an important point of interest. It was calculated manually with the same formula as recall as proposed in [24], as $Accuracy_{(Fail)} = \frac{TP}{TP+FN}$, where TP (True positive) is the number of students who failed and were identified and FN (False Negative) the number of students who failed but were not identified.

Table 3

Features used in each model, predictions for each model were made monthly, from Sep-Apr

Model	Features
$M_{A \text{ Sep-Apr}}$	Average Pre-semester Daily Clicks, First Date of VLE interaction (num. of days before/after the first day of the semester), Longest Number of Days Active Consequently for Each Passed Month (Sep-Apr) Longest Number of Days Inactive Consequently for Each Passed Month (Sep-Apr), Number of Active Days for Each Passed Month (Sep-Apr) Number of Inactive Days for Each Passed Month (Sep-Apr), Overall Longest Number of Days Active Consequently, Overall Longest Number of Days Inactive Consequently, Overall Number of Active Days, Overall Number of inactive Days, Sum of All VLE clicks, Sum of Clicks for Each Passed Month (Sep-Apr), Sum of Clicks for Each VLE type, TMA Grades, Weekday Percentage Clicks
$M_{B \text{ Sep-Apr}}$	First Date of VLE interaction Sum of All VLE clicks, Sum of Clicks for Each Passed Month (Sep-May), Sum of Clicks for Each VLE type, TMA Grades
$M_{C \text{ Sep-Apr}}$	Age Band, Disability, Gender, Highest Education, Studied Credits, Sum of Clicks for Each Passed Month (Sep-May), Sum of Clicks for Each VLE type, TMA Grades
$M_{D \text{ Sep-Apr}}$	Age Band, Disability, Gender, Highest Education, Studied Credits, TMA Grades

4. Experimental Results

The predictions of M_A are visualized in Figure 2. We can observe a tendency in the improvement of predictions through time as more data were introduced to the model. A detailed overview of RMSE and R^2 for data for each time period compared to the scores of other models are represented in Table 4. In April, after gathering all TMA, the model was capable of explaining 86% of the variability in final grades' predictions, with an RMSE of 8.0, which, in most cases, was within one grade bracket. M_A was relatively successful in predicting the variability in data, even before the start of the semester (RMSE=14.2). In the first months, we observed the highest prediction errors in students whose actual grade fell below the passing threshold of 40 points. Compared to other models, M_A covered more variance in predicted final grades in all the observed months. The mean RMSE was also lower in all observed months in M_A compared to other models. The accuracy of grade predictions for students failing the test was relatively low, which is visible from Figure 2, where the grades around the 40-point threshold were the least accurately predicted, as well as from Table 4. In April, a month before the final exam, 74% of the failing students were detected.

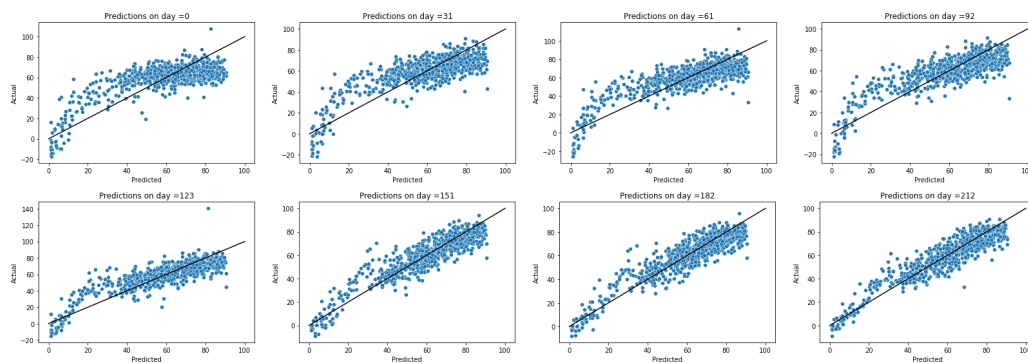
**Figure 2:** Monthly predictions

Table 4

Performance evaluation of the prediction models

Model	Metric	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
M_A	RMSE _{Mean}	14.2	13.1	12.7	12.2	11.0	8.8	8.7	8.0
	R ² _{Mean}	0.54	0.61	0.64	0.67	0.73	0.83	0.83	0.86
	Acc _(Fail)	51%	51%	53%	56%	63%	69%	69%	74%
M_B	RMSE _{Mean}	21.4	19.7	17.6	15.4	14.8	10.5	10.5	8.8
	R ² _{Mean}	0.02	0.12	0.31	0.47	0.43	0.74	0.75	0.83
	Acc _(Fail)	0%	12%	34%	41%	55%	72%	73%	76%
M_C	RMSE _{Mean}	21.6	20.1	18.1	15.9	15.1	10.6	10.4	8.8
	R ² _{Mean}	0.04	0.08	0.27	0.43	0.41	0.74	0.75	0.83
	Acc _(Fail)	0%	15%	32%	44%	55%	70%	71%	75%
M_D	RMSE _{Mean}	21.7	19.5	17.6	15.3	12.3	9.8	9.7	8.4
	R ² _{Mean}	0.01	0.16	0.31	0.47	0.67	0.78	0.78	0.84
	Acc _(Fail)	0%	10%	27%	44%	54%	69%	69%	75%

The correlation between features was observed with Spearman's correlation coefficient for M_A . As expected, there was a moderate to a strong positive correlation between the final result and the assessment grades (correlation coefficient of TMAs was 0.44, 0.46, 0.65, 0.76, 0.79, and 0.77 for assessments, respectively). The ratio of weekday clicks had a very weak correlation with the final score (0.05), and the mean number of daily clicks before the start of the semester had a weak correlation (0.29). A moderate negative correlation was observed between the longest number of consequently inactive days and the final score. The negative correlation increased through the months, with the highest correlations observed in March (-0.46), April (-0.49), and May (-0.53). A weak positive correlation was observed between the longest number of consequently active days and the final score, again increasing through the months, with the highest correlations observed in March (0.32), April (0.32), and May (0.35). The number of inactive days in each month also correlated negatively with the final score, though the correlation was weak to moderate, increasing slowly from October (-0.33) to April (-0.49) and May (-0.51). Oppositely, a positive correlation was observed with the number of active days observed each month, again increasing from October (0.33) to April (0.49) and May (0.51). The exact contribution of proposed metrics is visible in the differences of the RMSE and R² scores in all the observed months between M_A and M_B , where the only difference between them was in the introduction of proposed metrics in M_A .

5. Discussion

Different models were deployed to predict the final grades of the students based on their behavior, demographic characteristics, and VLE engagement patterns. The model with the proposed metrics, created only with data from VLE interactions, performed best in the case of the observed course in terms of RMSE and R², while the accuracy of predicted failure was higher than comparing models until January. While the accuracy of failure prediction in M_B ,

M_C , and M_D was very low in the first months, it improved visibly after February, reaching up to 3% higher accuracy than M_A . Comparing R^2 between the models, we can observe the predictions of M_A approximated the real data points better than the other three models up to January. From February to April, M_D was comparable to M_A , with the latter having 0.02 to 0.05 lower R^2 . Comparing RMSE, M_D reached the most comparable values from January to April, with 0.4 to 1.3 higher RMSE values compared to M_A .

Though not directly comparable, the classification approach to predicting at-risk students in related work reached a high accuracy; 84% accuracy at 20% of the course (day 50, mid November) and up to 94% accuracy at the end of the course [2], or 70% accuracy at the beginning of the semester and 94% at the end [4], with [12] also reported 93% accuracy at the end of the semester for the observed *DDD* course. Various classification approaches yielded better results compared to linear regression in observed the course data in similar timeframes, suggesting linear regression is not the optimal model in this case. The accuracy of predicting failure in students at the end of the semester presented by Rivas et al. [1] (78,5%) and Wang et al. [7] (75%) was only slightly higher than the accuracy obtained with M_A at the end of April (74%), one month before the end of the semester. Using the regression approach offered more insight into the exact predicted grade, offering students and educators more opportunities to react during the semester. Admittedly, the limitations of linear regression can be observed in this case. As the final score was calculated based on some features, a degree of multicollinearity could not be avoided (i.e., correlation of several independent variables in a model). The highest correlation between the computed metrics was observed in a score of assessment 5 (ID=25366) and the final score. Furthermore, due to the use of linear regression, the predictions for outliers were less accurate. Comparison of different prediction approaches in search of the optimal algorithm is out of the scope of this work; however, in the observed case, the use of ensemble models or multilayer perceptron might improve the model's performance.

We observed increasing negative correlations between monthly inactive days and the final score, which can be explained by the lower weight of the first TMAs, as, until the end of November, students can achieve up to 25% of their final TMA grade with three out of six finished assessments. In later months, the weights of TMAs increase, and with them, the importance of interaction with material available in the VLE. Similarly, an increasing negative correlation between the longest number of consequently inactive days and the final score can be interpreted that the inactivity in the first months of the semester can be improved during the semester. However, due to the rising weight of TMA's, inactivity in the last months of the semester contributes greatly to a student's lower final score. We can interpret regular students' activity (measured with consequent active days) in VLE as the prerequisite for a positive grade, as it does not actively contribute directly to their final grade. However, the lack of consequent activity has noticeable negative consequences. Correlations between the six new proposed metrics and the final score in M_A suggest that metrics measuring the daily clicks pre-semester, the number of active and inactive days, and the longest number of consequently active and inactive days reach weak to moderate correlation with the student's final score, while the metric measuring the percentage of weekday clicks achieved very weak correlation, and was overall less suitable for prediction. The findings suggest that it is meaningful for learning analytic researchers to take into account the possibilities of calculating new, custom features from the available raw data.

6. Conclusions and future work

This study aimed to predict students' final grades by converting the problem into monthly formats and observing the performance evaluation based on different features presented to the model. The primary motivation of this paper was to draw attention to the exploration of previously unused metrics on an existing dataset, and observation of their impact on prediction. The metrics which contributed to this discussion on the OULA dataset are the mean daily number of pre-semester VLE clicks, the number of active and inactive days, and the number of consecutive active and inactive days. Main limitations of this work are: (1) Narrow focus and validation on one course, with a chance that the proposed features prove less useful on large-scale or generalized data; (2) Used linear model is simple, and can be optimized further to improve performance, especially around the passing threshold; (3) Limited practical implications, as withdrawn students and students retaking the course were not included. In future work, the introduction of various methods for identifying the use of the proposed features in different courses will be explored, as well as testing the proposed features with classification models.

Acknowledgments

The authors acknowledge the financial support from the Slovenian Research Agency (Research Core Funding No. P2-0057).

References

- [1] A. Rivas, A. González-Briones, G. Hernández, J. Prieto, P. Chamoso, Artificial neural network analysis of the academic performance of students in virtual learning environments, *Neurocomputing* 423 (2021) 713–720.
- [2] M. Adnan, A. Habib, J. Ashraf, S. Mussadiq, A. A. Raza, M. Abid, M. Bashir, S. U. Khan, Predicting at-risk students at different percentages of course length for early intervention using machine learning models, *IEEE Access* 9 (2021) 7519–7539.
- [3] H. Heuer, A. Breiter, Student success prediction and the trade-off between big data and data minimization, in: D. Krömker, U. Schroeder (Eds.), *DeLFI 2018 - Die 16. E-Learning Fachtagung Informatik, Gesellschaft für Informatik e.V., Bonn, 2018*, pp. 219–230.
- [4] N. R. Aljohani, A. Fayoumi, S.-U. Hassan, Predicting at-risk students using clickstream data in the virtual learning environment, *Sustainability* 11 (2019).
- [5] L. Haiyang, Z. Wang, P. Benachour, P. Tubman, A time series classification method for behaviour-based dropout prediction, in: *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, 2018, pp. 191–195.
- [6] H. Waheed, S.-U. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, R. Nawaz, Predicting academic performance of students from vle big data using deep learning models, *Computers in Human Behavior* 104 (2020) 106189.
- [7] X. Wang, B. Guo, Y. Shen, Predicting the at-risk online students based on the click data distribution characteristics, *Scientific Programming* 2022 (2022).

- [8] J. Kuzilek, M. Hlosta, D. Herrmannova, Z. Zdrahal, J. Vaclavek, A. Wolff, Ou analyse: analysing at-risk students at the open university, *Learning Analytics Review* (2015) 1–16.
- [9] M. Hlosta, Z. Zdrahal, J. Zendulka, Ouroboros: Early identification of at-risk students without models based on legacy data, in: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference, LAK '17*, ACM, New York, NY, USA, 2017, p. 6–15.
- [10] S. H., K. Ravikumar, Student risk identification learning model using machine learning approach, *International Journal of Electrical and Computer Engineering* 9 (2019) 3872.
- [11] D. Gašević, S. Dawson, T. Rogers, D. Gasevic, Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success, *The Internet and Higher Education* 28 (2016) 68–84.
- [12] K. T. Chui, D. C. L. Fung, M. D. Lytras, T. M. Lam, Predicting at-risk university students in a virtual learning environment via a machine learning algorithm, *Computers in Human Behavior* 107 (2020) 105584.
- [13] Y.-H. Hu, C.-L. Lo, S.-P. Shih, Developing early warning systems to predict students' online learning performance, *Computers in Human Behavior* 36 (2014) 469–478.
- [14] D. Herrmannova, M. Hlosta, J. Kuzilek, Z. Zdrahal, Evaluating weekly predictions of at-risk students at the open university: Results and issues, in: *Proceedings of the European Distance and E-Learning Network 2015 Annual Conference*, 2015.
- [15] A. Al-Azawei, M. Al-Masoudy, Predicting learners' performance in virtual learning environment (vle) based on demographic, behavioral and engagement antecedents, *International Journal of Emerging Technologies in Learning (IJET)* 15 (2020) 60–75.
- [16] H.-C. Chen, E. Prasetyo, S.-S. Tseng, K. T. Putra, Prayitno, S. S. Kusumawardani, C.-E. Weng, Week-wise student performance early prediction in virtual learning environment using a deep explainable artificial intelligence, *Applied Sciences* 12 (2022).
- [17] J. Kuzilek, M. Hlosta, Z. Zdrahal, Open university learning analytics dataset, *Scientific data* 4 (2017) 1–8.
- [18] S. Palmer, Modelling engineering student academic performance using academic analytics, *International Journal of Engineering Education* 29 (2013) 132–138.
- [19] S.-U. Hassan, H. Waheed, N. R. Aljohani, M. Ali, S. Ventura, F. Herrera, Virtual learning environment to predict withdrawal by leveraging deep learning, *International Journal of Intelligent Systems* 34 (2019) 1935–1952.
- [20] S. Pongpaichet, S. Jankapor, S. Janchai, T. Tongsanit, Early detection at-risk students using machine learning, in: *2020 International Conference on Information and Communication Technology Convergence (ICTC)*, 2020, pp. 283–287.
- [21] A. A. Mubarak, H. Cao, W. Zhang, Prediction of students' early dropout based on their interaction logs in online learning environment, *Interactive Learning Environments* (2020) 1–20.
- [22] P. Jia, T. Maloney, Using predictive modelling to identify students at risk of poor university outcomes, *Higher Education* 70 (2015) 127–149.
- [23] F. Marbouti, J. Diefes-Dux, Heidi A. and Strobel, Building course-specific regression-based models to identify at-risk students, in: *2015 ASEE Annual Conference & Exposition, ASEE Conferences*, Seattle, Washington, 2015.
- [24] F. Marbouti, H. A. Diefes-Dux, K. Madhavan, Models for early prediction of at-risk students in a course using standards-based grading, *Computers & Education* 103 (2016) 1–15.