Enhancing Student Engagement in VLEs for Academic Success: Machine Learning Approach

Jordan Sberlo
y.sberlo@student.utwente.nl
University of Twente

ABSTRACT

With the increasing prevalence of Virtual Learning Environments (VLEs) in education, understanding how student engagement impacts academic success has become essential. This study investigates the correlation between various engagement metrics and dropout rates in courses given through VLEs, utilizing the KDD Cup 2015 dataset from XuetangX University. Logistic regression (LR) analysis identified key engagement features correlating with academic outcomes, which were then used to train a Gradient Boosting Machine (GBM). The GBM model demonstrated high performance with an accuracy of 88%, surpassing that of LR. To further explore practical applications, a web app prototype was developed, integrating the predictive model and visualizations of student engagement data. This prototype was evaluated by a sample of students and teachers from the University of Twente. Survey results indicated that both students and teachers found the application easy to use and beneficial for enhancing engagement and academic performance. Notably, the prediction model and visual analytics features were highly valued by participants. This research underlines the potential of using machine learning and visualized analytics to improve educational outcomes in VLEs, providing valuable insights for educators and learners alike.

Virtual Learning Environments, Student Engagement, Machine Learning, Data Processing, Dashboard.

1 INTRODUCTION

Education has made a significant transition into the virtual world in the 21st century. As online courses and programs become more popular, new educational tools and methodologies must be developed to address the challenges they bring. Classical education has evolved over centuries but generally shares common characteristics in structure, pedagogy, and classroom experience. It takes place in a physical setting, typically limited to a couple dozen students per class. There is an emphasis on daily routines and a scheduled timetable, discipline, rules and regulations, and a highly social environment, where peer interactions, student-teacher interactions, and community events occur daily.

The emergence of the internet has introduced a new educational space enabled by what are called Virtual Learning Environments (VLEs). Nowadays, there is a wide variety of professional and academic courses offered entirely through VLEs, for example "Combinatorics and Algorithms Design" or "Exploring Psychology's Core Concepts" by Tsinghua university [35][36]. Even traditional institutions incorporate VLEs into their classical educational systems. VLEs enable anyone with an internet connection to access an increasing amount of knowledge and education at any moment, regardless of geographical constraints. It is clear that this revolutionizes the educational system, and as such, adaptations must be made.

1.1 Related Work

Despite the wide recognition of VLEs as an important opportunity within educational practices, research shows there are still many questions as to how students' needs might be satisfied, as evident by very high dropout rates[15]. Emphasis is placed on feelings of isolation, attributed to the geographical distance between teacher and student, and on the difficulties found for managing the learning itself, due to the flexibility of timetables and the access to many sources of information. The authors stress that learning in this educational context requires a greater level of motivation from the student than does education undertaken on-site [38][3]. In online education, authors such as Filtcher and Miller[13] indicate motivation as the most important determinant factor for the students academic performance. Moreover, their research also supports that VLEs require the students to take greater responsibility in their studies, that is, undertaking actions for the monitoring and regulation of their own learning. It is also emphasized that, although they are many and various, the virtual learning environments' interactive tools, in themselves, do not guarantee quality of the educational process. It is necessary to recognize that pedagogical work mediated through the use of VLEs is not a matter of transposing teaching strategies adopted in on-site conditions to the virtual environment. Educational situations which fail to consider the specific characteristics of online education create a space for undesirable results, such as procrastination, dropping out and de-motivation on the part of the students [6].

It is therefore imperative that these relatively new learning environments are thoroughly examined to identify potential shortcomings and advantages, how such advantages can be leveraged and shortcomings mitigated such that students are provided with a more holistic learning experience. Interestingly, the rise of VLEs also brings access to a new abundance of data concerning users interactions with the platforms, providing insights into the effectiveness of course design choices and students' learning behaviors outside of the classroom. A significant amount of research was done in this field in recent years, attempting to enhance the learning experience in VLEs. Some authors like Daukilas (2008)[9] and Stella (2004)[33] related their research to quality assurance and the improvement of teaching methods in VLEs. Authors such as Lin (2009)[23] and Dringus (2005)[11] focused on data mining to discover and assess general themes and plagiarism in discussion forums. Some research explored classification methods which allow to establish students' learning styles and grouping them based on their behavior, using data obtained during the course. Several models were suggested for classification of students by different authors (Graf and Kinshuk 2006)[16] (Özpolat and Akar 2009)[27]. Schiaffino (2008)[32] presented an intelligent agent called "eTeacher" that provides personalized assistance to students. eTeacher observes students' behaviors while they are taking the course and automatically builds a profile that comprises of the student's learning style and information about their performance, such as exercises done, topics studied and exam results. Dynamic analysis of students data is valuable in enabling us to gain a deeper understanding of learning styles, recognize patterns and trends, identify weak points in courses and deliver adaptive feedback and personalized attention to both students and teachers[28].

Machine learning (ML) is today's most rapidly growing technical field, intersecting computer science and statistics, and is at the very core of artificial intelligence and data science. Recent progress in ML has been driven by the development of new algorithms and theory, decreasing costs of computation and the ongoing explosion in availability of data[19]. In the context of online education, ML has been applied to help recognize behavioral trends in students, and their expected outcome based on that. For example, Vivek and Neha (2016)[10], used decision trees and regression to generate predictions of students success in passing the course, based on their click-through rate in the VLE. Research done by Fisnik et al. (2018)[8] focuses on the prediction rates of different ML techniques, gives a comprehensive analysis of the challenges in the field, and discusses the various reasons behind the starkly high dropout rates observed in online courses. Some of the difficulties attributed to ML techniques are; Lack of structured sample data, data variance and imbalance, student schedule related challenges and the lack of a standard for creating and representing engagement data.

1.2 Area of Focus

This research focuses on the analysis of student engagement data in VLEs, utilizing ML techniques to provide insight of the significance of different engagement metrics, for example, click-through rate of different elements on the platform (watching videos, solving problems etc.) for academic performance. Additionally, it's been recognized that such experiments have been done in the past, however, this research aims to not only inquire about the implications of the data, and the accuracy of predictions, but also how can these findings be effectively integrated to improve the outcome of online courses. Whereas major technical improvements have been achieved in the field, there exists a gap in how to make practical use of these tools and ultimately provide better learning and teaching experiences in VLEs. Therefore this research will attempt to bridge that gap and assess the usefulness and desirability of such tools among teachers and students alike. There are many ways in which these tools can be implemented, in this research, a Hi-Fi prototype of a web-app is developed. The app features a dashboard layout with ranging analytics of students activity and engagement, and a ML based assessment of each individual student, given their engagement data.

As mentioned before, one of the challenges in this field is the availability of publicly accessible datasets. MOOC (Massive Open Online Course) platforms are often reluctant to publish the data due to confidentiality and privacy concerns. Since we are interested in training ML models, a large dataset is required. The two largest datasets currently available are OULAD (Open University Learning Analytics Dataset) published by The Open University (2017) [34], and KDD-Cup 2015 (Knowledge Discovery and Data-Mining) made available by Tsinghua University, originally for the purpose of the KDD competition organized by ACM (2015) [20][14]. The KDD-Cup 2015 was chosen to facilitate this research due to it's larger size, and better variety, that is, while OULAD only provides logs of clicks, KDD specifies what element has been clicked, out of 7 possible element types; problem, video, access, wiki, discussion, navigate, page-close. The set contains 8,157,277 logs of students on the platform, with 120,542 enrolled to 39 of the different courses offered by the XuetangX platform (founded by Tsinghua University). There is no demographic information or data that indicates which languages the courses were given in. In this research we will explore the use of the Gradient Boosting Machine (GBM) model in effectively achieving similar, or better results than logistic regression (LR). GBM is chosen due to it's high training speed and high performance in binary classification [4][1].

In order to gain a deeper understanding of students' behavior on VLEs, its relationship to their subsequent results, and how such data can help us improve the educational experience, this paper explores the following questions:

(1) **RQ1 Regression:** What are the relationships between different types student engagement metrics in VLEs, for example, amount of problems solved, and students dropouts?
(2) **RQ2 Accuracy:** How accurately can machine learning algorithms like Logistic Regression and Gradient Boosting Machine, predict student dropouts?
(3) **RQ3 Application:** To what extent can a web application that securely integrates visualized data analysis and machine learning tools, be useful to students and teachers in VLEs?

2 METHODOLOGY

2.1 Participants

For answering the last research questions, a usability test is conducted for the prototype. Since the prototype will be developed with different authorizations for students and teachers, the usability tests will be conducted on the two groups separately. On deciding the sample size for each group, authors like Nielsen (1993)[26] argued that 5 participants are enough for finding the vast majority of problems in usability testing, with diminishing returns when going beyond that. However, a more recent study by Faulkner (2003)[12] shows that in some cases, the amount of problems revealed by user testing with 5 participants was at

55%, and argued against its sufficiency. In the same study Faulkner showed that with 10 participants they were able to find a mean of nearly 95% of problems with a minimum of 82%. Therefore it was decided to aim for 10 students and 10 teachers. Out of the desired 20, only 8 students and 5 teachers from the University of Twente (UT) enrolled as participants for this study, all from the BMS and EEMCS faculties. The acceptance criteria are: Students: (1) They are currently enrolled at the UT. (2) They have experience with at least one course that was given in VLEs. Teachers: (1) they are currently staff members at the UT. (2) they have experience with giving at least one course mediated completely through VLEs.

In the student group all participants were in the 18-24 age group, 25% were Dutch and the rest from varying nationalities with 87.5% of them being male. In the teacher group 80% of the participants were in the 35-50 age group, 40% were Dutch and the rest from varying nationalities with 80% of them being male.

2.2 Apparatus and materials

To answer the first two questions of this research, we make use of the KDD-Cup 2015 dataset. In order to understand the procedure, it is importat that we first clarify the structure of the dataset.

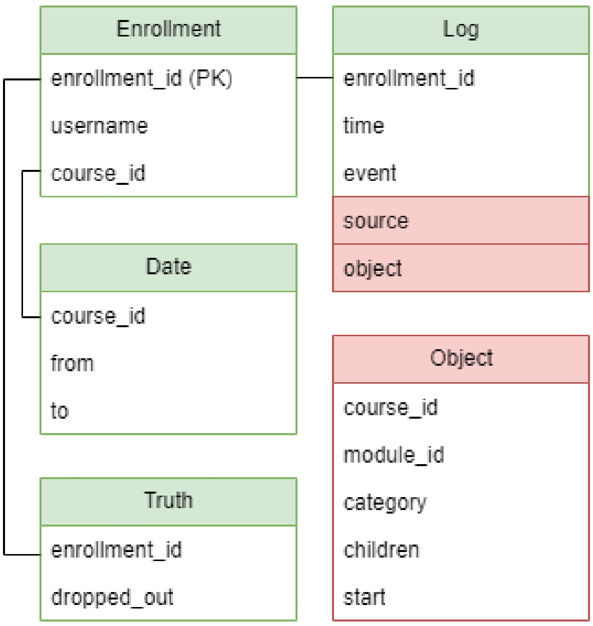


Figure 1: Schema

Let us detail the content of each table (see figure 1);
Object - Interestingly, this table does not reference to the object column in table Log and due to lack of documentation it's been excluded from this research.
Enrollment - Each entry is represented by a unique enrollment-id, indicating which usernames are enrolled in which courses.
Date - Each entry indicates the start (from) and end (to) date of every course.
Log - Each entry is a behavior record "event" of a specific enrollment-id, categorized by one of 7 options: problem, video, access, wiki, discussion, navigate, page-close. Columns source and object were dropped for irrelevancy.
Truth - Each entry is a pair of enrollment-id and a binary value, where 1 indicates the student dropped out from the course and 0 means they successfully passed the course.

The IDE that was used in this research is Visual Studio Code. Jupyter notebook was chosen as a platform due to its modularity and organizational advantages. The main programming language used was python, with the help of the following libraries: numpy[17] and pandas[24] for data processing and analysis, matplotlib [18] for plotting the results, and sklearn [7] for the LR and GBM algorithms, as well as producing their metrics.

In order to answer the third research question of this paper, a small scale HiFi prototype was implemented in figma for the participants to interact with. Subsequently, a short survey was compiled in Google forms, containing demographic questions, multiple choice questions, slider questions and open ended questions to assess user experience and perceived usefulness of the app [37][31].

2.3 Procedure

Feature Engineering - Using the Pandas library, the dataset was first pre-processed to filter incomplete entries and extreme outliers. A new data-frame was created to concentrate the features for fitting the models later in this research. Features were sorted for each enrollment-id, that included the following columns: (number of) access, page-close, problem, video, discussion, navigation, wiki, which were all counted from the Log table. While the dataset only records timestamped events, these timestamps can provide more insight into students behaviour. Two new features were calculated and added to the data-frame:

average_timestamp (normalized) - This feature represents the relative average time a student was active within the course duration. It is calculated as follows:
Step 1 - Calculate average timestamp:

$$avg\_timestamp = \frac{\sum_{i=1}^{n} timestamp(i)}{n}$$

Where **n** is the number of recorded events for that user. To normalize this variable we calculate the following values.
Step 2 - Calculate course duration in seconds:

$$course\_duration = (to - from).total\_seconds()$$

Step 3 - Calculate time since start:

$$time\_since\_start = (avg\_timestamp - from).total\_seconds()$$

Step 4 - Normalize:

$$normalized\_avg\_timestamp = \frac{time\_since\_start}{course\_duration}$$

active_days (normalized) - This feature represents the proportion of days a student was active relative to the total course duration in days and is calculated as such:
Step 1 - Count active days:

Jordan Sberlo
y.sberlo@student.utwente.nl
University of Twente

Achieved in a multi-step process of merging of data-frames, for the complete snippet please refer to the git repository [29].

Step 2 - Calculate course duration in days:

$$course\_duration\_days = (to - from).days$$

Step 3 - Normalize:

$$normalized\_active\_days = \frac{active\_days}{course\_duration\_days}$$

Since we are looking for regression coefficients between continuous variables and a binary outcome (pass or fail), the relationship between our independent variables (engagement metrics) and our dependent variable (dropped_out) will be calculated using logistic regression [25].

Training the Models - For training the models, the dropped_out field from the Truth table was appended for each entry in the data-frame. The set was then split into training(0.8) and testing(0.2) sets for fitting and evaluating the performance of the LR and GBM models. The regression coefficients were then generated from the LR model, using the sklearn library. a regression coefficient quantifies the change in the log odds of the dependent variable (e.g., the probability of an event occurring) for a one-unit change in the independent variable, holding all other variables constant. Finally both the LR and GBM models were evaluated for overall accuracy as well as precision, recall and F1-score for each class, that is, 1 for dropping out and 0 for passing the course. The metric we will be focusing on is F1-score which is calculated from the confusion matrix as follows:

$$F1Score = 2 \times \frac{Percision \times Recall}{Percision + Recall}$$

Where

$$Percision = \frac{TruePositives}{TruePositives + FalsePositives}$$

And

$$Recall = \frac{TruePositives}{TruePositives + FalseNegatives}$$

The GBM was later used in implementing the prototype that serves as a means to answer the third research question of this paper.

Application and Testing - Using the analysis and the GBM model, a web application prototype was designed and developed following HCI practices from the book Human-Computer Interaction: Fundamentals and Practices (GJ Kim 2015)[21] with emphasis on user experience, accessibility and security[30]. The app has authorization for two types of users, teacher and student.

While students have only access to their own profiles (see figure 3), the teacher portal includes a complete user management system that allows the user to filter by course and have access to all profiles of students that are enrolled in their courses (see figure 2a). Filtering by course also displays analytics relevant to the course as a whole, namely Total Interactions Per Day, illustrated by a line graph with amount of interactions (y axis) against day of the course (x axis), and Average Interactions, represented by a pie chart with proportional values, and exact average values displayed in the legend to the right (see figure 2a). When selecting a specific student, the teacher is given further analytics, relevant to that student(see figure 2b). Interactions Per Day (bottom left) similar to the previous graph but only representing that specific student, Cumulative Active Days (top left), which is represented by a line graph with cumulative active days (y axis) against day of the course (x axis), Interactions (top right) which shows a comparison between the course averages seen in figure 2a, and the students' interactions. Finally, on the bottom right, the teacher can see the students "Status", represented by either "OK" or "AT RISK", in accordance with the prediction generated by the GBM model, when given that students' engagement metrics, where "OK" refers to passing the course and "AT RISK" refers to dropping out.

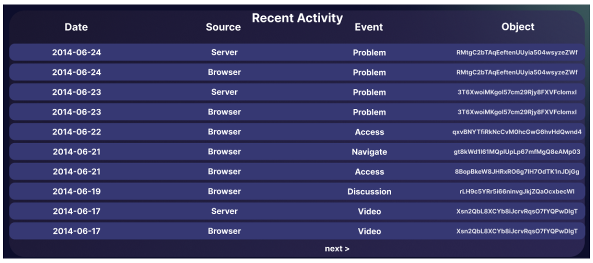
(a) Figure 2: Teacher Portal

(b) Figure 2: Teacher Portal

The student portal only gives access to analytics for the logged user, similar to what the teacher sees. However, it also features a course buddy system where students can request to follow each-other and see their friends' status (see figure 3). This feature was added in an attempt to reinforce accountability and increase the probability of passing the course, as indicated in the research by Carnoy et al. (2002 [5]) . The students can view their dashboard for any course they are enrolled in by selecting the course on the top right. Another difference is the location of their own status which is on the top right, next to the course.

When viewing a student profile, users can scroll down and view the activity history of that student (see figure 4).

Figure 3: Student Portal



Recent Activity

Date	Source	Event	Object
2014-06-24	Server	Problem	RMtgC2bTAqEeftenUUyia504wsyzeZWf
2014-06-24	Browser	Problem	RMtgC2bTAqEeftenUUyia504wsyzeZWf
2014-06-23	Server	Problem	3T6XwoiMKgol57cm29Rjy8FXVFclomxl
2014-06-23	Browser	Problem	3T6XwoiMKgol57cm29Rjy8FXVFclomxl
2014-06-22	Browser	Access	qxvBNYTfiRkNcCvM0hcGwG6hvHdQwnd4
2014-06-21	Browser	Navigate	gt8kWd1l61MQplUpLp67mfMgQ8eAMp03
2014-06-21	Browser	Access	8BopBkeW8JHRxRO6g7IH7OdTK1nJDjGg
2014-06-19	Browser	Discussion	rLH9c5YRr5i66ninvgJkjZQaOcxbecWI
2014-06-17	Server	Video	Xsn2QbL8XCYb8iJcrvRqsO7fYQPwDlgT
2014-06-17	Browser	Video	Xsn2QbL8XCYb8iJcrvRqsO7fYQPwDlgT

next >

Figure 4: Activity Window

The participants were then given the opportunity to interact with the application for 5 minutes respective to their roles (teachers with the teacher portal and students with the student portal) in part of a moderated usability testing. Finally, the participants were asked to fill out a survey to assess their experience and perceived usefulness of the application (appendix A) [31]. The survey was given in person on an individual basis. We will refer to the questions as follows.

Likert scale questions:
Q1 - I think the system was easy to use (1 - very difficult, 5 - very easy).
Q2 - How would you rate the overall experience on the app? (1 - very poor, 5 - excellent).
Q3 - What is your opinion about the organization of information on the screen? (1 - very poor, 5- excellent).
Q4 - How useful would this product be for you as a student/teacher? (1 - not at all, 5 - very useful)
Q5 - Would you feel more engaged in your learning/teaching using this application? (1 - not at all, 5 - very much)

Multiple choice question:
Q6 - What features do you find useful? (leave empty if none)
Teacher: analytics, prediction model, user management.
Student: analytics, prediction model, buddy system.

Open questions:
Q7 - What is the one thing you wish the app could do that it doesn't already?
Q8 - Briefly describe what improvements you would suggest for this web app and why.

These questions were selected based on the the study made by Law et al. (2014)[22]. The authors highlight the issues with UX research and examine the various ways in which UX is being measured. Since experts in the field of HCI have not reached a consensus on the definition of UX, there is no widely accepted standard for measuring it. The authors emphasize the two disparate stances on how UX should be studied (i.e. qualitative versus quantitative) and that they are not necessarily compatible or can even be antagonistic. However, a rather comprehensive review on the recent UX publications (2011)[2] shows that UX research studies have relied primarily on qualitative methods. In the paper, UX qualities and constructs were classified based on their measurablity. Among the measurable, a select few were chosen for this research, namely, engagement, flow, interest and satisfaction.

3 RESULTS

We are first interested in finding the **regression coefficients** between different types of user engagement metrics, and students academic result. In other words, what's the statistical relationship between a students activity during a course, and their end result (pass / fail the course). the LR model generated the following coefficients:

Feature Regression		
Temporal	active days	-10.016202
	average timestamp	-2.346308
Numerical	page_close	-0.023201
	wiki	-0.016154
	navigate	-0.011618
	access	-0.003678
	problem	-0.001803
	discussion	0.0022
	video	0.010275

Figure 5: Regression Coefficients

It is noteworthy that we are examining the regression between engagement metrics and **dropout** (see figure 5). A negative correlation suggests a positive relationship with passing. The highest correlation was observed in active_days at -10.016202. This suggests that as the active_days increase, the log odds of a student dropping out decrease significantly. In simpler terms, students who are active for a greater proportion of the course duration, are much less likely to drop out. Although not to the same degree, average_timestamp also shows a strong relationship which indicates that students who engage more as the course progresses (closer to the end) are less likely to drop out.

Interestingly, all of the numerical features show marginal relationships, with page_close, wiki, navigate, access, and problem having negative coefficients, indicating engagement with such activities are associated with lower likelihood of dropping out. Whereas engaging in increasing

Jordan Sberlo
y.sberlo@student.utwente.nl
University of Twente

amount of discussion or video activities seem to correlate with higher likelihood of dropping out.

Next, many features were removed for weak statistical significance and potential redundancy, Only features with coefficients of -0.02 or less were selected: [active_days , page_close, avg-timestamp]. Both LR and GBM were refitted with the adjusted dataset and yielded the results:

Logistic Regression

	PERCISION	RECALL	F1-SCORE
0	0.79	0.49	0.61
1	0.88	0.96	0.92
Accuracy			0.87
Macro Avg.	0.83	0.73	0.76
Weighted Avg.	0.86	0.87	0.85

Gradient Boosting Machine

	PERCISION	RECALL	F1-SCORE
0	0.77	0.58	0.66
1	0.9	0.95	0.92
Accuracy			0.88
Macro Avg.	0.83	0.77	0.79
Weighted Avg.	0.87	0.88	0.87

Figure 6: Classification Report

By **accuracy**, we refer to the combined F1-score. The LR results showed a total accuracy of 0.87 and a weighted average of 0.85. Dropout(1) F1-score is at 0.91 whereas pass(0) F1-score is at 0.6. The GBM model results showed slightly higher performance with a total accuracy of 0.88 and weighted average of 0.87. Dropout(1) F1-score is at 0.92 whereas pass(0) F1-score us at 0.66. While both models have very similar true positive rate for for dropout(1), GBM manages to provide a higher true positive rate for passing(1), 0.58 against 0.49 attained by LR.

Finally, we are interested in evaluating the usefulness of an **application** that integrates the classification enabled by GBM, and visual analytics. The survey results:

			min	max	mean	std error
Teachers	User Experience	Q1	4	5	4.2	0.2
		Q2	3	5	3.8	0.2
		Q3	3	5	3.6	0.2
	Usefulness	Q4	3	5	4	0.4
		Q5	2	5	3.6	0.5
Students	User Experience	Q1	4	5	4.3	0.2
		Q2	4	5	4.1	0.1
		Q3	3	5	4.1	0.3
	Usefulness	Q4	2	5	3.4	0.3
		Q5	2	5	3.9	0.4

Figure 7: Questions 1 - 5

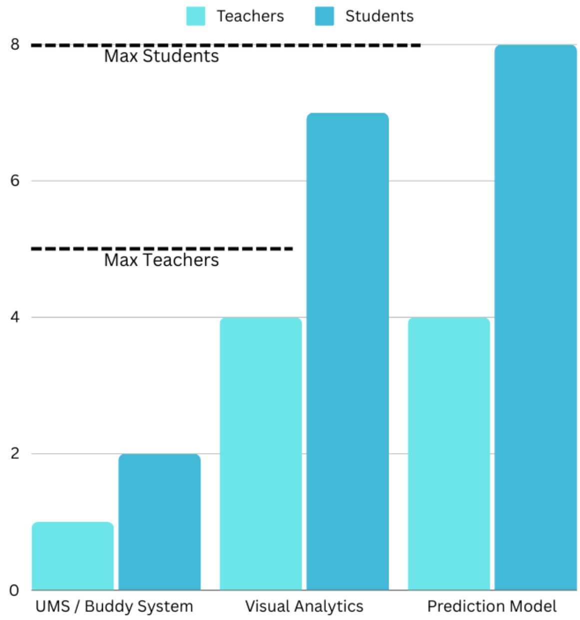


Figure 8: Question 6

In figure 7 we can observe an overall satisfaction with the user experience, with ease of use rated at 4.2 by teachers and 4.3 by students, organization and design rated at 3.6 by teachers and 4.1 by students, and overall experience of using the platform at 3.8 by teachers and 4.1 by students. The question "How useful would this product be for you as a student/teacher?" was rated 4 by students and 3.4 by teachers, whereas the question "Would you feel more engaged in your learning/teaching using this application?" was rated 3.6 by students and 3.9 by teachers. This indicates that students perceive a higher level of usefulness from the system than teachers, but teachers would feel more engaged in their teaching, using this platform, than students feeling more engaged with their learning. In all cases we attain a positive score for usefulness and engagement in both groups (3 being the threshold).

In figure 8 we observe 100% of the students and 80% of the teachers, found the prediction model useful. 87.5% of the students and 80% of the teachers, found the visual analytics useful. Only 20% of teachers found the user management system useful and 25% of students found the buddy system useful. This suggest both the analytical tools and prediction model were highly desirable by participants of both groups. However, the usefulness of other features such as the buddy system and user management system was significantly lacking.

Lastly, for the full results of the open ended questions (Q7 and Q8) please refer to Appenidx A. Some notable answers are: "System to share/view note made by student according to its course/section within the course" for Q7 and "For the main page: Get the information for students at risk on the first page". Some suggestions were made for a better representation of the data, such as using bar charts for interactions and using histograms for interactions per day.

4 DISCUSSION

4.1 Findings

This research reveals the significance of regular engagement with course materials for achieving academic success. Prompting students to action throughout the duration of the course can highly increase their likelihood of passing. Moreover, it seems that engaging students, especially on the later parts of the course, will have a greater effect on their end result. frequency of engagement possess a much stronger relationship to success, than the quantity of engagement.

Machine learning algorithms prove to be a reliable tool for detecting potential drop outs in VLEs. However, there are still major issues that hold them back from achieving higher performance, such as a lack of a standard for representing user engagement data, and its unstructured tendency. These issues make it difficult to consistently apply and improve the use of machine learning in this field. Still, both students and teachers showed interest in utilizing it for enhancing their educational experience, and deemed it useful to them.

The web app prototype demonstrates a practical application of the research findings, providing a user-friendly interface for both students and teachers. The positive feedback from usability testing suggests that such tools can effectively enhance the learning experience by offering personalized insights and fostering increased engagement. Teachers were especially positive of the usefulness of such an application which could indicate this area has potential in improving academic results.

4.2 Limitations

Dataset Issues
As previously mentioned, there is a severe deficit in availability of datasets of this type, and a lack of consensus on how it should be represented. The KDD-Cup dataset was released 9 years ago which arrises the question of its relevancy in the wake of a rapidly changing virtual environment. It also lacks demographic information with could contribute to bias in the results. Lastly, the application of logistic regression relies on the assumption that there is little or no multicollinearity among the independent variables. This means that the independent variables should not be too highly correlated with each other, something that cannot be definitively concluded from this dataset.

Prototype and Scope
Having a relatively broad scope for this research meant compromise in detailed analysis of each of the research questions. There is more to be explored, especially in the area of application. Some of this is indicative in the improvements suggested by participants for Q8 (see Appendix A). Many improvements could be made in areas such as data visualization, interface design and available features.

4.3 Future Work

While this research explored engagement data in VLEs laterally, and provided a wide understanding of the subject matter, it also raised several questions. How should data of this type be collected and represented? An attempt to answer this question can provide future research to build upon and significantly improve our understanding of what affects students academic results. It can also help foster a methodology for engineering more significant features, while reducing their redundancy, thereby enabling ML models to achieve better performance.

Another important question worth exploring is, how can these tools be implemented and optimized for students and teachers? This is an avenue future research can greatly benefit from. It appears there's demand among teachers and students for tools of this type and further inquires in HCI and UX can help match this demand with appropriate solutions that can contribute to real world improvement in academic results.

4.4 Conclusion

This study delved into the analysis of student engagement data in Virtual Learning Environments (VLEs) using Machine Learning (ML) techniques, focusing on understanding the relationship between various engagement metrics and academic performance. Through the application of Logistic Regression (LR) and Gradient Boosting Machine (GBM) models on the KDD-Cup 2015 dataset, we identified significant correlations between the frequency of student activities on VLEs and their dropout rates. The findings underscored the importance of active participation throughout the course duration in reducing the likelihood of dropout. Moreover, the development and usability testing of a Hi-Fi prototype web application provided valuable insights into how predictive analytics and visualized data can enhance both teaching and learning experiences in VLEs. While the prototype received positive feedback on usability and usefulness from students and teachers alike, highlighting the potential for such tools to improve engagement and educational outcomes, future research could explore constructing broader and more suitable datasets, and refine predictive models to further optimize personalized learning experiences in online education.

4.5 Acknowledgments

Lastly, I extend my appreciation to Dipti K. Sarmah and Mahboobeh Zangiabady for their professional support and guidance that lead me in the right direction and significantly contributed to the depth and scope of this study.

4.6 Use of Generative AI

The use of generative AI during this research was limited to ChatGPT 3.5 for two purposes: (1) Debugging - during feature engineering and data analysis it was used for finding information about the libraries and their documentation and help solve errors through TraceBack. (2) Vocabulary - during writing of this paper it was used to find a synonyms and transition words in order to avoid repetition and maintain reader interest.

Jordan Sberlo
y.sberlo@student.utwente.nl
University of Twente

REFERENCES

[1] Adeniyi J Adewale, Irina Dinu, and Yutaka Yasui. Boosting for correlated binary classification. *Journal of computational and graphical statistics*, 19(1):140–153, 2010.

[2] Javier A Bargas-Avila and Kasper Hornbæk. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2689–2698, 2011.

[3] Andrea Carvalho Beluce and Katya Luciane de Oliveira. Students' motivation for learning in virtual learning environments. *Paidéia (Ribeirão Preto)*, 25:105–113, 2015.

[4] Candice Bentéjac, Anna Csörgő, and Gonzalo Martínez-Muñoz. A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54:1937–1967, 2021.

[5] Martin Carnoy and Susanna Loeb. Does external accountability affect student outcomes? a cross-state analysis. *Educational evaluation and policy analysis*, 24(4):305–331, 2002.

[6] Kuan-Chung Chen and Syh-Jong Jang. Motivation in online learning: Testing a model of self-determination theory. *Computers in human behavior*, 26(4):741–752, 2010.

[7] David Cournapeau. Sikit-Learn. https://scikit-learn.org/stable//, 2024.

[8] Fisnik Dalipi, Ali Shariq Imran, and Zenun Kastrati. Mooc dropout prediction using machine learning techniques: Review and research challenges. In *2018 IEEE global engineering education conference (EDUCON)*, pages 1007–1014. IEEE, 2018.

[9] Sigitas Daukilas, Irma Kaciniene, Daiva Vaisnoriene, and Vytautas Vascila. Factors that impact quality of e-teaching/learning technologies in higher education. *Quality of Higher Education*, 5:132–151, 2008.

[10] Vivek Doijode and Neha Singh. Predicting student success based on interaction with virtual learning environment. In *Proceedings of the southeast SAS users group conference, Bethesda, MD, USA*, pages 16–18, 2016.

[11] Laurie P Dringus and Timothy Ellis. Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education*, 45(1):141–160, 2005.

[12] Laura Faulkner. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. *Behavior Research Methods, Instruments, & Computers*, 35:379–383, 2003.

[13] Carol Filcher and Greg Miller. Learning strategies for distance education students. *Journal of Agricultural education*, 41(1):60–68, 2000.

[14] Philippe Fournier-Viger. The KDD Cup 2015 Dataset. https://data-mining.philippe-fournier-viger.com/the-kddcup-2015-dataset-download-link/, 2015.

[15] Marcela Gerogina Gomez-Zermeno and Lorena Aleman De La Garza. Research analysis on mooc course dropout and retention rates. *Turkish online journal of distance education*, 17(2), 2016.

[16] Sabine Graf and Prof Kinshuk. An approach for detecting learning styles in learning management systems. In *Sixth IEEE International Conference on Advanced Learning Technologies (ICALT'06)*, pages 161–163. IEEE, 2006.

[17] Jim Hugunin. Numpy. https://numpy.org/, 2024.

[18] John D. Hunter. Matplotlib. https://matplotlib.org/, 2024.

[19] Michael I Jordan and Tom M Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260, 2015.

[20] KDD. KDD Cup Archives. https://kdd.org/kdd-cup, 2016.

[21] Gerard Jounghyun Kim. *Human-computer interaction: fundamentals and practice.* CRC press, 2015.

[22] Effie Lai-Chong Law, Paul Van Schaik, and Virpi Roto. Attitudes towards user experience (ux) measurement. *International Journal of Human-Computer Studies*, 72(6):526–541, 2014.

[23] Fu-Ren Lin, Lu-Shih Hsieh, and Fu-Tai Chuang. Discovering genres of online discussion threads via text mining. *Computers & Education*, 52(2):481–495, 2009.

[24] Wes McKinney. Pandas. https://pandas.pydata.org/, 2024.

[25] Scott Menard. *Applied logistic regression analysis*, volume 106. Sage, 2002.

[26] Jakob Nielsen and Thomas K Landauer. A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 206–213, 1993.

[27] Ebru Özpolat and Gözde B Akar. Automatic detection of learning styles for an e-learning system. *Computers & Education*, 53(2):355–367, 2009.

[28] Saulius Preidys and Leonidas Sakalauskas. Analysis of students' study activities in virtual learning environments using data mining methods. *Technological and economic development of economy*, 16(1):94–108, 2010.

[29] Jordan Sberlo. KDD Cup 2015 Feature Engineering and Data Analysis - Repository. https://github.com/Zum0shi/KDDCup-Dashboard, 2024.

[30] Jordan Sberlo. Virtual Learning Environment Web Application For Students and Teachers. https://www.figma.com/proto/x0ZeD2gKeDCMnBo9c5xGD1/Dashboard?node-id=237-76&t=PxkMhaKRMig9Z33N-0&scaling=scale-down&content-scaling=fixed&page-id=0%3A1, 2024.

[31] Jordan Sberlo. Web Application Survey For Participants. https://forms.gle/buhFVAb39VNkJvGSA, 2024.

[32] Silvia Schiaffino, Patricio Garcia, and Analia Amandi. eteacher: Providing personalized assistance to e-learning students. *Computers & Education*, 51(4):1744–1754, 2008.

[33] Antony Stella and Arumugham Gnanam. Quality assurance in distance education: The challenges to be addressed. *Higher education*, 47(2):143–160, 2004.

[34] The Open University. Open University Learning Analytics Dataset. https://analyse.kmi.open.ac.uk/open_dataset, 2017.

[35] Tsinghua University. Combinatorics and Algorithms Design. https://www.xuetangx.com/course/1219129KC0028920intl/4048105?channel=i.area.home_course_ad, 2020.

[36] Tsinghua University. Exploring Psychology's Core Concepts. https://www.xuetangx.com/course/thu07111000658intl/3995164?channel=i.area.home_course_ad, 2020.

[37] Wake Forest University. How Do I Analyze My Survey Results? https://ir.wfu.edu/large-survey/how-do-i-analyze-my-survey-results/, 2024.

[38] Saijing Zheng, Mary Beth Rosson, Patrick C Shih, and John M Carroll. Understanding student motivation, behaviors and perceptions in moocs. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 1882–1895, 2015.

Appendix A: Questions 7 - 8

What is the one thing you wish the app could do that it doesn't already? (optional)

Obviously it only works for online engagement. I have no idea on how this could be extended to all activities.

More in depth information about parts of the course, so not only overall over the whole course

Based on my statistics predict failing rates in other courses that im not participating in atm.

System to share/view note made by student according to its course/section within the course

Briefly describe what improvements you would suggest for this web application and why. (optional)

A % based prediction instead of a binary one, grade estimation and possible suggestions on which specific tasks/applications can help me the most

Each individual section could be bigger and more readable, there is no reason to fit everything in 1 screen.

I would say the pie chart is not the most correct way of showing the interaction information.
A bar chart with a background bars to show the averages would be better as it would be easier to compare the numbers, not just the distribution of actions.

When you enter the tab with the info abt the student I would put his name somewhere top left maybe, visible.
Idk if it s necessary to have your own name at the top all the time.

For the main page: Get the information for students at risk on the first page,
also in case one of the course is OK (and then possibly presented instead of the one at risk).

sorting of students by at-risk level

- Graphs not necessary, status OK/RISK/... with a mention of e.g. not many hours in system would be sufficient and more clear.
- Doubts about reliability per course, especially in bsc when there are many conflicting study units/deadlines.

explanation of the prediction, confidince intervals

I think that visualization for Interaction per day can be more clear by using Histogram

It might not be the best idea as a student to allow all my friends to see if I am at risk or doing okay,
even if I accept their request. It would be nice if this feature was optional

Arrangement of data is overwhelming, especially for students who are not experienced with reading graphs.

A more simple and defined interface would improve the usability