Student-centric Model of Login Patterns: A Case Study with Learning Management Systems

Varun Mandalapu[1], Lujie Karen Chen[1], Zhiyuan Chen[1], Jiaqi Gong[2]
[1]University of Maryland Baltimore County, Baltimore, Maryland, USA, 21250
[2]The University of Alabama, Tuscaloosa, Alabama, USA, 35487
{varunm1, lujiec, zhchen}@umbc.edu, jiaqi.gong@ua.edu

ABSTRACT
With the increasing adoption of Learning Management Systems (LMS) in colleges and universities, research in exploring the interaction data captured by these systems is promising in developing a better learning environment and improving teaching practice. Most of these research efforts focused on course-level variables to predict student performance in specific courses. However, these research findings for individual courses are limited to develop beneficial pedagogical interventions at the student level because students often have multiple courses simultaneously. This paper argues that student-centric models will provide systematic insights into students' learning behavior to develop effective teaching practice. This study analyzed 1651 undergraduate student's data collected in Fall 2019 from computer science and information systems departments at a US university that actively uses Blackboard as an LMS. The experimental results demonstrated the prediction performance of student-centric models and explained the influence of various predictors related to login volumes, login regularity, login chronotypes, and demographics on predictive models. Our findings show that student prior performance and normalized student login volume across courses significantly impact student performance models. We also observe that regularity in student logins has a significant influence on low performing students and students from minority races. Based on these findings, the implications were discussed to develop potential teaching practices for these students.

Keywords
Student-centric Modeling, Learning Management Systems, Login Variables, Student Performance Prediction.

1. INTRODUCTION
Teaching and learning changed a lot in recent years with the increasing adoption of new computer-based teaching and learning technologies in educational institutions worldwide. As education and learning technology evolves with time, leveraging the technical advances to improve teaching practice and student learning will be a prominent research area. The most common technologies used by instructors to deliver course content include Learning Management System (LMS), Course Management Systems (CMS), and Learning Content Management Systems (LCMS) [1]. Even though these systems seem to be synonymous, they have their specific use in the education domain. LMS tools focus on communication, collaboration, content delivery, and assessment, whereas LCMS is similar to LMS with fewer administrative functions. CMS, on the other hand, will focus on the enrollment and performance of students. Of these three systems, LMS is the one that is best suitable for delivering learning strategy to students and is the primary focus of this study.

LMS systems provide a unique opportunity to administrators, and researchers to evaluate student data related to time spent on an activity, access times and day, grades, interactions, and many other useful student learning variables. The data logs collected by LMS systems are analyzed with scientific techniques published in the Educational Data Mining (EDM) domain. In their study, Romero and Ventura [2] described that current EDM methods rely on clustering and pattern recognition techniques to categorize students into various groups based on their interaction patterns. Categorization of students using clustering and pattern recognition supports instructors in making changes for a set of students. Teaching practices that impact the entire classroom can be evaluated using predictive analytics that tracks student learning and achievement from the vast amount of interaction data collected by LMS.

Existing research in Learning Analytics (LA) and EDM focused on developing highly accurate predictive models that can estimate student learning outcomes related to assignment scores, course grades, and drop-out probability [3,4]. These course-based predictive models provide early warning to student counselors or instructors associated with a specific course [5,6]. Even with considerable success in this area, many of the student performance prediction models have several shortcomings. One significant issue with course-based models is the bias introduced by teaching style and the type of course (descriptive, programming, mathematical etc.). This bias impacts these models' scalability across different courses and makes it difficult to understand the student level factors on their achievement. For example, if a student enrolls in five courses, developing models to study students' progress in these five courses independently is not realistic and gives different insights based on varying features and performances. Therefore, these modeling efforts are limited to reduce different biases introduced by instructor and the diverse amount of content made available in LMS.

Course level predictions are suitable for supporting instructor level decision making; however, if intervention is on student level behaviors such as study habits or self-regulation skills, it is beneficial to look at student-centered indicators so that interventions may be more targeted and cost-effective [7,8]. Developing student-centric models that analyze student LMS interactions across courses in a college/university setting will help

Varun Mandalapu, Lujie Chen, Zhiyuan Chen and Jiaqi Gong "Student-centric Model of Login Patterns: A Case Study with Learning Management Systems". 2021. In: Proceedings of The 14th International Conference on Educational Data Mining (EDM21). International Educational Data Mining Society, 263-274. https://educationaldatamining.org/edm2021/
EDM '21 June 29 - July 02 2021, Paris, France

address the issues with course specific models. This study is the first step in developing models that supports the identification of student level indicators.

For colleges that have a high penetration of LMS, LMS activity may give a holistic indicator of students' engagement level (behavior engagement specifically). We ask the question to what extent those holistic indicators predict student term Grade Point Average (GPA) performance in the future. To explore this, we specifically focus on student login related features as they can be generalized across courses and act as proxy variables for time management [51, 52]. It is also challenging to aggregate other features like discussions, readings, and assessments across courses compared to access related LMS variables. This study's data is drawn from Blackboard Learn, a commercial LMS software available for colleges and universities to deliver course content and assessments through internet-enabled computer systems. Most importantly, the data is drawn from all students in computer science and information systems at a large public university in the US during the Fall 2019 semester. In addition to student interactions from LMS, we also access demographic and prior student performance data from the university's student administration system to build and interpret downstream predictive models. The university's Institutional Review Board (IRB) approved this study, and all the student specific demographic and personal information are anonymized by following General Data Protection Regulation (GDPR) standards.

In this work, we focus on model predictions and explanations to understand student learning behaviors. First, we apply new methods to process student interaction data collected across different courses enrolled in a semester to build student-centric performance models based on machine learning principles. Secondly, we utilize a novel approach in local model explanations, correlation and regression to understand the impact of various features captured by LMS on student performance. One primary reason for using Locally Interpretable Model Explanations (LIME) is its ability to explain the relationship between predictor variables and predictions, especially the input variable's impact on the outcome. On the other hand, statistical correlation analysis will provide the relation between input predictors and the observed target variable. As correlation analysis does not consider the interaction effect between input variables, we also use a linear regression model to study the output variable's feature importance’s based on the model coefficients. To address the research gap discussed earlier, we explore the below three research questions.

RQ 1 How different student-centric machine learning models perform in predicting student end-of-term GPA?

RQ 2 How do student login and time interval pattern across courses influence student learning outcomes?

RQ 3 Is there a significant variability in feature importance for students coming from diverse demographics?

2. RELATED WORK

Universities and colleges around the world adopted LMS systems, such as Moodle and Blackboard, to provide onsite, hybrid, and online courses based on their capabilities to support communication, content creation, administration, and assessment [9, 10]. Besides the automation and centralization of various administrative tasks like creating and managing student accounts, creating syllabus, assignments, assessments, grading, etc., LMS systems assemble and deliver personalized learning materials and content quickly [11]. These systems also support the reusability of materials created by instructors. The systems also enable instructors to create content structures, deliver them in a sequence, maintain control access, organize group activities, track student activities, load and replace learning materials and provide feedback on assessments. With advanced database software developed by Oracle, IBM, and Microsoft that emphasize interconnectedness, data independence, and security, LMS systems employ various login roles based on user classification. These roles will permit instructors to create new content or privately address student issues and create discussion boards to capture student knowledge on specific topics.

LMS platforms enable students to access learning material in various formats, such as pdf, PowerPoint presentations, video lectures, and audio files. The systems also track student activity related to content downloads, access timestamps to display student progress in learning to instructors [12]. LMS also provides both asynchronous and synchronous communication for students to interact with instructors and encourages group activities. Combining the tools provided by LMS with innovative learning strategies like self-directed learning, small group instructions, and collaborative learning with instructor interventions, a wide variety of activities can be developed for individual, small groups, or larger classes [12,13]. Given the simplicity and convenience of accessing online materials through LMS systems, it is not surprising to see high student satisfaction scores for courses delivered through LMS.

In recent years, the data sets related to student learning activities have drawn significant attention from researchers in academic communities to develop possible solutions to address student retention and academic success issues. This type of work has been called learning analytics and focuses on student activities such as navigating lecture materials, what information is accessed, how long it takes to complete an activity, and how students transform the information in learning materials into measurable learning [14, 15]. Multiple commercial resources like SPSS, google analytics, Stata and Nvivo can build predictive models on data captured by LMS to assess student drop-out probabilities to develop targeted learning courses or model collective learning behaviors. Since most instructors deliver course assessments and material through LMS, they can track student activity by processing a digital footprint during every online interaction captured by system log files.

2.1 Learning Analytics Research

LMS systems have the ability to capture large data streams related to user interactions through which administrators and instructors can develop methods to improve the learning experience. The collection, analysis, and reporting of data about learning activities on web-enabled learning platforms to assess student academic progress, predict performance, and identify potential issues that need attention is the central proposition of emerging fields like learning analytics and educational data mining [16, 17]. Outcomes derived from learning analytics aim to gain insights about student learning behaviors, real-time information about institutional practices and support the designing of personalized courses in CMS. Although there are huge data stores in universities and colleges that can be used to make data-driven decisions to support optimal use of both pedagogical and economic resources, to date there has been minimal application of this data in higher education [18].

2.1.1 Student Engagement and Frequency of LMS Use

An LMS system records student interaction details related to logins, number of posts written on discussion threads, time spent on lecture materials, total downloads, etc., in their log files. These logs can be analyzed to generate reports that help teachers to observe student progress at a granular level. Once there are enough student records collected in LMS, they can be used to develop computational models to predict future student performances. Multiple works in EDM and LA studied the relation between the usage of LMS and student academic achievements. Vengroff and Bourbeau's [19] study showed evidence that providing additional material in LMS benefited students at the undergraduate level. They also conclude that students who used LMS regularly did better in exams than their peers who have minimal interactions. In their research, Dutt and Ismail [20] observed that tracking resources students interact with on LMS supports developing new strategies that make learning easier and enhance learner progress. Their work also focused on analyzing thresholds related to student interaction features like self-assessment tests, time spent on exercises, discussion forums, and performance outcomes. Another study by Lust et al. [21] explored the usage variations in different tools used by students on LMS, such as time on web-link, time on web-lectures, time on a quiz, time on feedback, postings on discussion board, and messages read. The results from this study heavily contributed to the development of adaptive and innovative recommendation systems. In their work, Hung and Zhang [22] also found patterns based on six indices that represent student effort: Frequency of accessing the course material, number of LMS logins, total interactions in discussion threads, number of synchronous discussions, number of posts read, and final grades in a course.

While exploring a link between student online activity on LMS and their grades, Dawson et al. [23] observed a significant difference in the number of online sessions accessed, total time spent, and the number of posts in discussion forums between high and low performing students. Another study by Damainov et al. [24] developed a multinomial logistic regression model based on time spent in LMS. This study found a significant relationship between student time spent and grades, especially in students who attained lower grades between D and B. Instead of using time spent online, other works focused on the frequency of course material access within LMS. A study by Baugher et al. [25] found that regularity in student hits is a reliable predictor of student performance compared to the total number of hits. In their study, Chancery and Haque analyzed student interaction logs of 112 undergraduate students and found students with low LMS access rates obtained lower grades than their peers with higher access rates. This study was complemented by Biktimirovan and Klassen [26] that reported a strong relationship between student hit consistency and success. Their study counted access to various LMS activities and found that homework solution access is the only strong predictor of student performance. However, these studies are primarily descriptive rather than predictive.

2.1.2 Instructional Design and Student Participation

Online teaching strategies are primarily dependent on instruction design as each mode of interaction - student/instructor, student/student, and student/content have their own positive impacts on student progress. A study by Coldwell et al. [27] focused on the relationship between student participation in a fully online course and their final grades. They found a positive relationship between student participation and final grade. Dawson et al. [23] examined the impact of various LMS tools and found a highly positive correlation between discussion forum activity and student success. They observed more than 80% of interactions occurred in the discussion forum, which is the primary interaction tool in LMS. Another study by Greenland [28] found that asynchronous communication is the primary form of all online course interactions. Nandi et al. [29] found an increasing number of posts in discussion forums close to assignment and exam deadlines. They also found a high correlation between exam scores and online class participation throughout the semester, especially in high-achieving students.

All the studies discussed above adopted log files from LMS systems to extract unbiased details from activity and performance to identify a relationship between independent interaction variables and student grades. Most of the discussed studies are based on univariate analysis focusing on a single variable or a set of highly impactful variables of a single course or similar courses on student outcomes. However, student performance is a highly complex area in education to measure or understand, especially across various courses offered on-campus in a university setting. Most of the authors discussed above noted the need for more in-depth works to investigate student performance across courses and based on multiple variables. These studies also lack an explanation about variables used in their studies to track student performance, and it is evident that the authors selected LMS variables based on their belief that these variables are highly correlated with student scores.

2.1.3 Social Factors in Analytics

Factors that influence student academic performance have been the focus of researchers in LA and EDM domains for many years. It still remains an active area of education research, indicating the complex problem in measuring and modeling learner processes, especially in tertiary education. Positive learning characteristics have a significant positive impact on learner engagement improvement in multiple ways. The dispositional language specifies learning as a combination of self-regulation, learning inclinations, motivation, behavioral patterns, interactions, and cognitive ability. In their study, Buckingham et al. [30] proposed a combination of self-reported data gathered in surveys with student interaction data generated by LMS to study individual student performance, learning processes, and group interactions. These social analytics depend primarily on student self-reported data to develop toolkits that support a specific learning type, especially in courses with high diversity [31]. However, our study focuses on objective identification of student success based on data that LMS captures. We will also identify the crucial variables from predictive model output for various student groups based on their diverse backgrounds (race, gender, and student status).

2.1.4 Multivariate Analysis to Predict Student Success

Even though there is a common agreement about the purpose of learning analytics, there are still several varying opinions on what data needs to be collected and analyzed to improve teaching and learning processes. A study by Agudo-Peregrina et al. [32] argued that it is highly complex to identify the net contribution of various interactions to the learning processes. Their findings show that peer interaction between students has a lower influence than student-teacher interaction, which contradicts earlier studies that showed high importance for student peer interactions. A study by Dominquez et al. [33] utilized multiple variables like LMS logins, time stamps, and content access flags captured in a biology course to predict student grade at the end of course completion. The results show that the algorithm predictive accuracy is at 50% in subsequent semesters. Lerche and Keil's [34] recent study utilized Moodle log data from 369 students enrolled in three online courses across three semesters to predict their scores at the end of the term for each course. Their regression results related to predicting student scores in a course at the end of the semester varied from 0.17 to 0.6 for all three courses. This broad range of performance across courses is due to varying variables utilized in each course based on the course structures. Studying the difference in instructional design, variables in extracted data, statistical inferences, predictive modeling used, interpreting model outcomes and pattern observations, etc., might explain the inconsistencies in results shown in earlier studies.

Data captured by LMS systems became prominent in LA and EDM circles as they capture student interactions in non-intrusive and ready-to-use settings. Several studies were discussed earlier in this research that utilized the LMS data to develop models that track student progress. However, it is still challenging to build highly accurate models that predict student learning outcomes across courses and understand the impact of different variables captured by LMS. Another significant gap in earlier research is their inability to predict student performance across courses in a given semester. One primary issue in predicting student performance across a semester is to find methods that aggregate student LMS variables across courses. This research shows methods to address the research gap found in earlier studies.

In this study, we approach the problem of tracking student achievement by developing student-centric models that build on aggregated LMS interaction variables collected across a semester irrespective of student year and course. One unique aspect of our work is related to the study of model performance on longitudinal student data. We develop models that predict student end-of-term GPA based on four cumulative periods in a semester. This work also focuses on explaining the impact of different aggregated LMS variables on various student groups categorized based on performance, race, gender, and student type. The importance of features is explained by adopting correlation statistics for univariate importance, a regression model for interaction effect, and LIME for model-based yet model agnostic explanations.

3. DATA & FEATURE SET

3.1 Dataset

For this study, we chose undergraduate student data captured by LMS in Fall 2019 from a large public university in the United States. These students were part of either Information Systems (IS) or Computer Science (CS) departments. The students from these departments were chosen as the instruction format and courses are closely aligned in both of them. Blackboard system is predominantly used as an LMS to deliver course material, assessment, and grading. The student demographic data captured by a standalone Student Information System (SIS) is used to categorize students based on different demographic variables. A total of 1651 students were enrolled in these two departments in the Fall 2019 semester. Based on student distribution, we categorized students into three ethnicities: White, Asian, and Minority. This study also researches student performance based on their admit types, such as four-year regular student or transfer student. The demographics of student data are provided in the below table 1. This study was approved by IRB and sensitive student data was de-identified based on GDPR standards.

Table 1. Student demographics

Demographic	Student Count
Total Students (N)	1651
No of unique courses	440
No of unique course instructor combinations	638
Male : Female	1302 (79%) : 369 (21%)
White : Asian : Minority	630 (38%) : 495 (30%) : 526 (32%)
4 – Year : Transfer	976 (59%) : 675 (41%)
Full Time : Part Time	1446 (88%) : 205 (12%)
IS : CS	934 (57%) : 717 (43%)
1st Yr : 2nd Yr : 3rd Yr : 4th Yr	115 (7%) : 329 (20%) : 515 (31%) : 692 (42%)
<= 3 : 4-5 : >5 (Courses enrolled)	298 (18%) : 1035 (63%) : 318 (19%)

3.2 Feature Extraction

We explored various LMS features related to student logins, content accesses, time spent, discussion posts, assignment submissions, and time intervals based on earlier literature. While exploring these features, we identified that only three features could be commonly extracted from different courses: Student Login Counts, Time intervals & prior knowledge.

One of the significant challenges while building a student-centric model on LMS data is to extract aggregated features that are least biased. As Blackboard's content is dependent on instructor and course, it is crucial to mitigate the variations caused by these factors on aggregate student variables. This work employs multiple statistical measures to mitigate these issues. The details are explained in the below sub-sections.

3.2.1 Normalized Login Volume

Earlier studies identified that student performance prediction is strongly dependent on the volume of student logins. One challenge with counting the student logins in Blackboard is its inability to find which course they accessed during each login. Also, calculating the total login count introduces a hidden bias as courses with more content on Blackboard prompt students to login more often than other courses with less content and flexible deadlines. To mitigate this issue, our work followed the below steps to extract student login features.

1. Extract all courses enrolled by all students in IS and CS.
2. Count the total number of logins for all students irrespective of their department in these extracted courses.
3. Calculate the Z-scores of student logins in each course. The reason for doing this is to mitigate the bias introduced by variations in the absolute count of logins as course logins vary a lot between students. Z-scores provide a value that helps understand if student logins are higher or less than average logins in a specific course.

4. Once the z-scores are calculated for all courses, we extract a vector of login z-scores for each student based on their enrolled courses.

5. As predictive models do not take vectors of variable length as input, this work extracts seven significant statistics from the login vector: mean, median, minimum, maximum, standard deviation, skewness, and kurtosis.

3.2.2 Login Regularity

Apart from student login volumes, the regularity between logins also provides valuable insights into student achievement as regularity is related to self-regulation capabilities. In this work, we utilize an entropy-based method to extract features that define student login regularity in each course. In information theory, entropy is used to define uncertainty or randomness [48]. Entropy measure will explain if student's logins are regular (less random) or irregular (more random). Based on this concept, if the entropy value is high, then a student has an irregular login pattern, and if the entropy value is low, the student has a regular login pattern. The steps to calculate student regularity features are given below.

1. Extract all course accesses with timestamps for every student in IS and CS.

2. Calculate the difference between timestamps. This difference will give a vector of time intervals for each course enrolled by a student.

3. Calculate entropy using the KL estimator with the k-nearest neighbor method proposed by Kozachenko and Leonenko [45]. KL estimator uses k-nearest neighbor distances to compute the entropy of distributions. The reason for adopting this method instead of Shannon entropy is based on the time interval vector's continuous characteristic [46].

4. Once the entropies are calculated, we get a vector of entropies for each student based on the number of enrolled courses. We then calculate the seven statistics similar to student logins: mean, median, minimum, maximum, standard deviation, skewness, and kurtosis.

3.2.3 Login Chronotypes

Studies in chronobiology and chronopsychology showed variation in different individual active periods at different times of the day [41, 42]. These studies classify an individual into either morning type or evening type based on their high activity time. For example, if an individual is highly active in the morning compared to the evening, they are considered morning type and vice versa. Inspired by this work in human psychology, this work divides a day into four-time bands T1 (12 AM to 6 AM), T2 (6 AM to 12 PM), T3 (12 PM to 6 PM), and T4 (6 PM to 12 AM) and extract student logins based on these four time bands. In addition to this, this work also extracts the logins on weekdays and weekends to study their influence on student performance.

1. Count the number of logins during each time band and on weekdays and weekends for each course.

2. Calculate the mean of login count vector for each of these time bands and weekday/weekend.

3. Normalize the login count with the number of courses enrolled by an individual student. This normalization will mitigate the bias introduced by the number of courses enrolled across the student cohort.

This work also utilizes the demographic and prior performance measured by GPA features captured by the SIS system. These features were listed in below table 2.

Table 2. Student demographic features

Demographic	Values
Start GPA (Prior Performance)	Cumulative GPA available till the start of semester
Gender	Male & Female
Ethnicity	White, Asian & Minority
Student Year	Freshman, Sophomore, Junior & Senior
Admit Type	Regular & Transfer
Enrollment Type	Full time & Part time
Student Age	Continuous variable

4. METHODOLOGY

The methodology section details the predictive modeling approach to predict student end-of-term GPA in fall 2019. In addition to this, we also describe the correlation-based LIME method to explain the features that contribute to model predictions. The workflow of developing student-centric models is depicted in figure 1.

4.1 Predictive Modeling

This work studied five of the most common regression models for comparison purposes. The selected models include Generalized Linear Model (GLM), Decision Tree (DT), Support Vector Regressor (SVR), Random Forest (RF), and Gradient Boosted Regressor (GBR). As model hyperparameter influences their predictive performance, we utilized a grid search mechanism to select multiple parameters to predict with high accuracy. We also adopted a feature selection method based on a multi-objective evolutionary algorithm in addition to hyperparameter search. This feature selection algorithm evaluates each feature set based on pareto-optimal that balances model complexity and accuracy. The details of models and hyperparameter search criteria are discussed below.

Generalized Linear Model: GLM is an extension of traditional linear models that fits input data by maximizing the log-likelihood. The regularization parameter is set so that the hyperparameter search space looks for an alpha value that fits between ridge and lasso regression. An alpha value of 1 represents lasso regression, and an alpha value of 0 represents ridge regression. This study searched for the best alpha value using a grid search between 0 and 1 in increments of 0.1.

Decision Tree: The decision tree algorithm is a collection of linked nodes intended to estimate the numerical target variable. Each node in the tree represents a rule used to split on an attribute value. The node uses a least-squares criterion to minimize the squared distance between the average value in a node when compared to the actual value. The hyperparameter search space for this algorithm evaluates both maximal depth and pruning. The maximal depth value varies between 1 and 100 in increments of 10. Pruning will make the DT algorithm use multiple criteria like

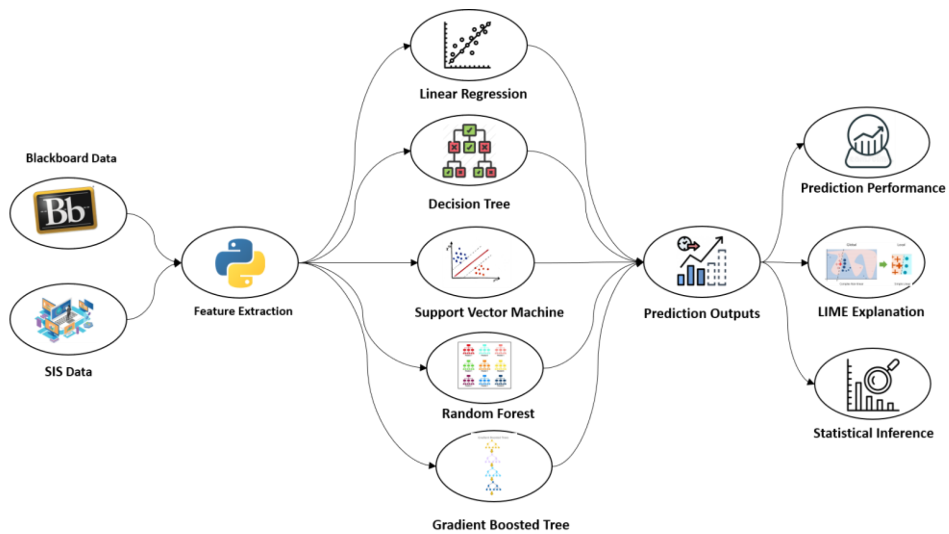

Figure 1: Student-centric Model Workflow

minimal gain, minimal leaf size, and pruning alternatives to decide the stopping criterion.

Support Vector Machines: The SVM used in this study is built based on Stefan Reupping's mySVM [47]. This algorithm will construct a set of hyperplanes in a high dimensional space for regression tasks. A good hyperplane is decided based on the functional margin. The hyperparameter search space focused on both dot and radial kernel functions with a C (SVM complexity) value range between 10 and 200. The kernel gamma function is set for a radial kernel with a range of 0.005 and 5 with three logarithmic increments.

Random Forest: A RF model builds an ensemble of decision trees on bootstrapped datasets. The splitting criteria are similar to a decision tree. The regression outcome is the average of the observed train data GPA present at that end node. We only tuned the number of trees hyperparameter to reduce the time complexity of the execution. The number of tree searches varied between 10 and 1000 trees in 10 linear steps.

Gradient Boosted Tree: The GBT model builds multiple regression trees in a sequence by employing boosting method. By sequentially applying weak learners on incrementally changed data, the algorithm builds a series of decision trees that produce and an ensemble of weak regression models. As GBT is a non-linear model, we search hyperparameters related to the number of trees, learning rate, and maximal depth. The number of tree values varies between 1 and 1000 in five quadratic increments, the learning rate varies between 0.001 and 0.01 in five logarithmic increments, and the maximal depth parameter varies between 3 and 15 in three logarithmic increments.

4.2 LIME Explanation

The concept of Locally Interpretable Model Explanations (LIME) was introduced to explain the predictions made by black-box models that deal with classification problems. LIME explains each prediction made by a complex model by training a surrogate model locally [35]. However, this earlier methodology is not scalable to deal with categorical variables, tabular data, and regression problems. In this work, we adopt the correlation-based LIME method available in RapidMiner to explain machine learning models' predictions [36, 37, 38].

1. Perturb data in the neighborhood of each sample in the dataset. The number of simulated samples can be user-defined. A higher number of simulated samples will provide higher accuracy of explanations but at the cost of more run times.
2. Make predictions using the ML model for all the simulated samples around each original sample in the dataset.
3. Calculate the correlation between each feature in the dataset and the target variable.
4. The features that have a positive correlation are considered supporting features, and features with negative correlation with predicted outputs are referred to as contradicting features.

As LIME provides feature importance value for each feature at each sample, we aggregate the importance value for all samples to build global importance for each variable. The significant advantage of this method compared to traditional global importance methods is its flexibility. As model global importance's are calculated across all samples in the data, the LIME based feature importance's can be calculated for subsets of data. This flexibility provides users with a deeper understanding of each feature's role for different sets of populations present in a dataset.

In addition to applying the LIME methodology, this work also studies univariate and multivariate feature importance on student performances by applying correlation and linear regression methods. The student dataset used in this study is divided into multiple subsets containing different student groups based on various demographics. A correlation value is calculated between input features and student end-of-term GPA. This value provides us with an intuition about the impact of various features on student performances related to different demographics. As correlation only provides independent variable importance on student performance, we also adopt a linear regression model to explore the variation of feature importance based on coefficient values. Applying a linear regression model will also consider the interaction effect between input features to fit the outcome variable.

5. RESULTS

This results section is divided into three subsections based on the three research questions we are focusing on in this study. The first subsection will detail various predictive models' performance on longitudinal student interaction data collected during the fall 2019 semester. The second subsection will detail the importance of student logins and regularity on performance predictions based on LIME methodology. The final subsection will discuss the importance of input features based on correlation and regression methods.

5.1 How different student-centric machine learning models perform in predicting student end-of-term GPA?

The five machine learning models adopted in this study were evaluated using a five-fold cross-validation method. In this method, the student data is divided into five equal folds at a student level. In every iteration, four of the five folds are used for model training, and one fold is used for model testing. The machine learning models are evaluated based on two performance metrics: R squared (R^2) and Root Mean Squared Error (RMSE). The output performance metrics are the average of five test fold performances.

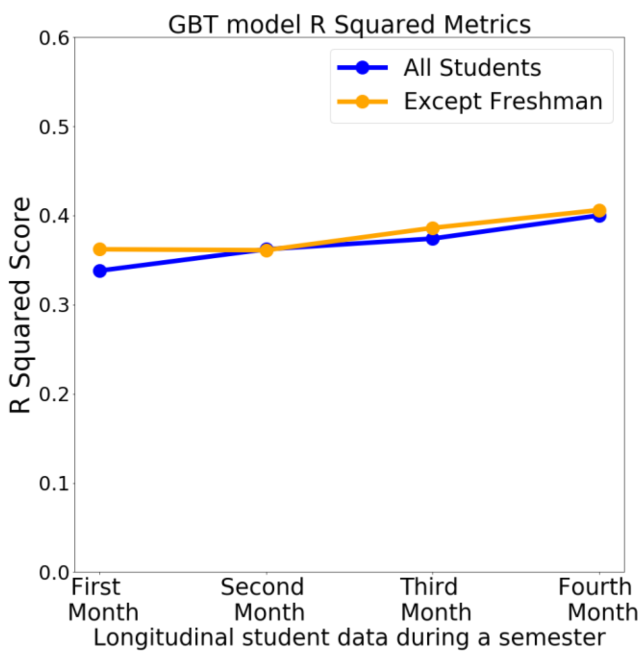

Figure 2. Compare performances of GBT model on different longitudinal datasets

In this study, we divided a semester into four parts to understand the impact of longitudinal interaction data across the semester on predictive model performances. This analysis will support the amount of data needed to balance predictive performance and early detection for interventions. The performance metrics evaluated on these four cumulative datasets will help understand the amount of student data needed to make accurate predictions. Tables 3, 4, 5, and 6 present the machine learning models' results evaluated on four cumulative datasets. While differentiating student performance based on multiple longitudinal datasets, we also study algorithms' performance without Freshman student data. This differentiation is to study the impact of missing start GPA feature values for first-year students as most of the full-time regular students in US universities start in the Fall semester.

Table 3. Student features from start to end of first month

Model	R^2		RMSE	
	All Students	Except Freshman	All Students	Except Freshman
GLM	0.213	0.249	0.657	0.633
DT	0.266	0.270	0.638	0.633
SVM	0.216	0.324	0.666	0.607
RF	0.332	0.353	0.607	0.588
GBT	**0.338**	**0.362**	**0.602**	**0.581**

Table 4. Student features from start to middle of semester

Model	R^2		RMSE	
	All Students	Except Freshman	All Students	Except Freshman
GLM	0.257	0.266	0.67	0.628
DT	0.263	0.295	0.67	0.618
SVM	0.195	0.315	0.705	0.609
RF	0.360	0.352	0.621	0.591
GBT	**0.362**	**0.361**	**0.622**	**0.586**

Table 5. Student features from start to end of third month

Model	R^2		RMSE	
	All Students	Except Freshman	All Students	Except Freshman
GLM	0.25	0.266	0.644	0.626
DT	0.255	0.255	0.658	0.650
SVM	0.335	0.344	0.612	0.597
RF	0.371	0.386	0.589	0.575
GBT	**0.374**	**0.386**	**0.588**	**0.572**

Table 6. Student features from start to end of semester

Model	R^2		RMSE	
	All Students	Except Freshman	All Students	Except Freshman
GLM	0.251	0.269	0.644	0.625
DT	0.246	0.274	0.657	0.641
SVM	0.320	0.289	0.616	0.627
RF	0.387	0.410	0.585	0.564
GBT	**0.400**	**0.406**	**0.575**	**0.562**

From the above tables, we observe that the GBT model performed better than the other four models based on the tradeoff between R squared and RMSE values. We also observe that there is no significant difference in student end-of-term GPA prediction with and without freshman details. This might be due to less sample

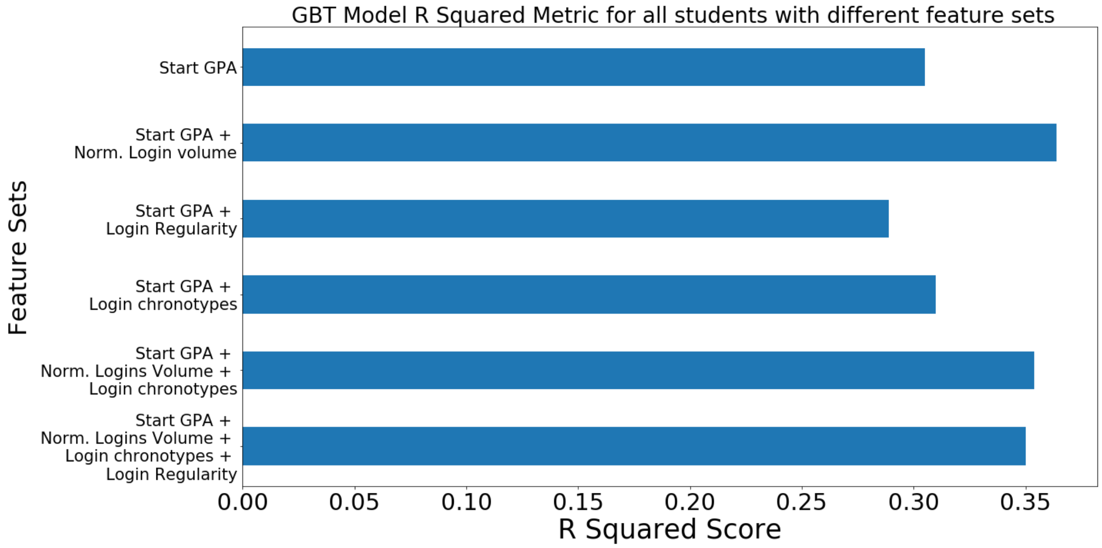

Figure 3. Compare performances of GBT model on different input feature sets.

size (7%) related to freshman cohort. From figure 2, it is also evident that there is a gradual increase in the performance of GBT model as we add data to predictive models as the semester progresses. Even though there is an increase in performance if we add all data captured during the semester, it doesn't help much for real-world interventions as activities that effects student performances will be completed by the end of the semester. Based on this understanding, we focus on data captured until the middle of the semester for feature importance study.

5.2 How do student login and time interval pattern across courses influence student learning outcomes?

To answer research question 2, we adopted a stepwise feature addition study that inputs features by adding one by one into the model and evaluates the performance based on R square and RMSE values. This study is performed on student data collected until the middle of the semester as models developed during this stage will help identify student level indicators and give enough time to deploy interventions that improve student performance. We first start with inputting student Start GPA (Cumulative GPA till the start of Fall 2019 semester) as start GPA showed a high correlation with end-of-term GPA based on our preliminary analysis. We then add normalized login volumes, login regularity, and login chronotypes in a step by step method. Figure 3 shows the R squared performance metric of student-centric models with different input variables.

From figure 3, we observe that students start GPA with normalized student login volumes across courses adds more predictive power to machine learning models. This observation is also supported by earlier studies [39, 40] that showed the importance of student login counts on student course grades and score predictions. Another observation is related to the importance of adding student self-regulation capability based on login regularity measured using entropy statistic. Based on figure 3, we observe that adding login regularity features with student login features and start GPA adds slightly more predictive power compared to model with only login regularity and start GPA features. In addition to these observations, we also observed that login counts based on login chronotypes with start GPA did not add much predictive power to machine learning models. From these results, we also imply that student aggregated login volumes might be adding the same information as login chronotypes.

5.3 Is there a significant variability in feature importance's for students coming from diverse demographics?

One limitation of using the earlier mentioned model-based feature importance study is its inability to explain each feature's importance on different student cohorts. To address this issue and understand the importance of login volumes and regularity features on different student groups, we adopt three approaches: one based on LIME, the second based on correlation analysis, and the third based on linear regression.

5.3.1 LIME based importance's

LIME based approach extract feature importance at the local level, also called local fidelity. By applying the LIME method explained in the methodology section, we extract feature importance's for different student groups categorized based on their demographics.

From figure 4, we can observe that cumulative student GPA at the start of the semester is an important feature to predict student end-of-term GPA. Student login volumes are the second important feature set for model predictions on different student demographics. This study's focus is also on student self-regulation capability measured by the regularity of logins (entropy). We observe that for students with GPA values less than 2, the regularity of logins feature played a key role compared to a student with a higher GPA. This observation also holds for students from minority ethnicity. One implication from these observations This observation suggests that introducing teaching practices that guide LMS use and time management will significantly impact students with low GPA and from a minority race. Start GPA played a slightly less significant role in transfer students than regular students as transfer students join in different years and their cumulative GPA might not be available at the start of the semester, similar to freshman.

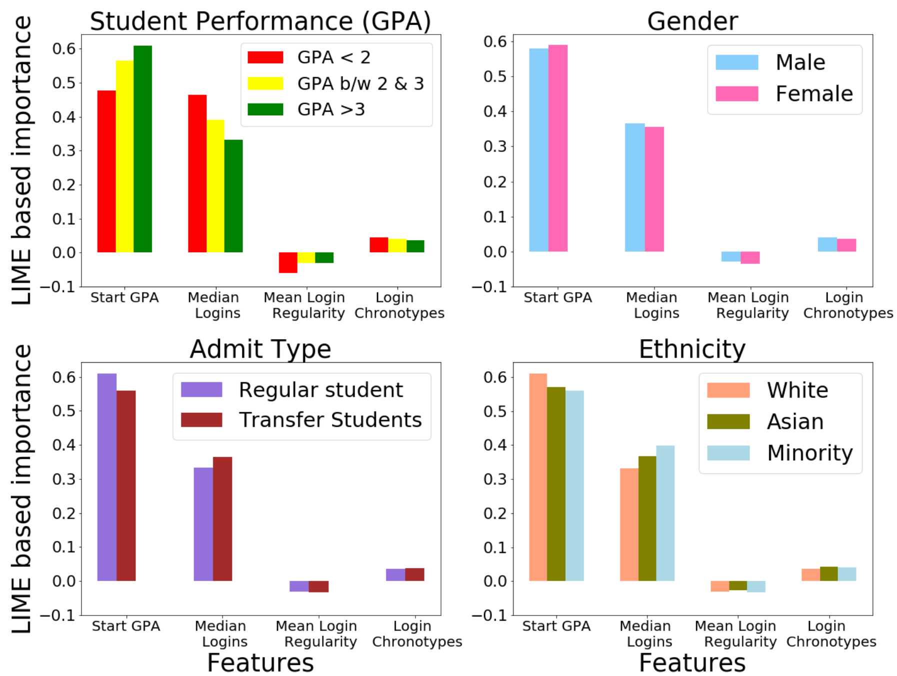

Figure 4. LIME importance's for different student groups divided based on GPA, ethnicity, admit type and gender.

Even though there is a huge imbalance in the number of male and female students present in the dataset, we do not observe any significant difference in feature importance's between these two genders. One limitation of the LIME method is related to global importance's. The importance's showed by LIME at the local level do not necessarily correspond to global importance's. Based on this limitation, we can infer which feature is essential for different students' groups but not quantify them as the importance's calculated in this study are the aggregate of importance's provided by LIME for each individual student.

5.3.2 Correlation based Feature Importance's

As earlier feature importance methods showed a significant impact of login volumes and login regularity measured by entropy statistic to predict student performance, we adopt Pearson correlation statistic to infer this relationship for different student groups. To do this, we create subsets of student data based on different groups: student GPA, gender, ethnicity, and admit type.

From figure 5, we observe that the student logins count and regularity in logins is highly significant for a student with a GPA lower than 2. We can also observe that as the entropy increases, the GPA reduces. This observation holds true as regularity in student logins represents their self-regulation capabilities. Earlier research showed that students with good self-regulation capabilities perform better in class [49, 50]. For other student groups divided based on gender and admit type, there is no significant variation in the importance of logins and entropy on student performances.

Even though the absolute values of correlation observed in figure 5 are not very strong, the comparison between different groups helps understand which features are significant for students from different demographics. In addition to this, we also observe a similar pattern in LIME based importance's discussed in earlier

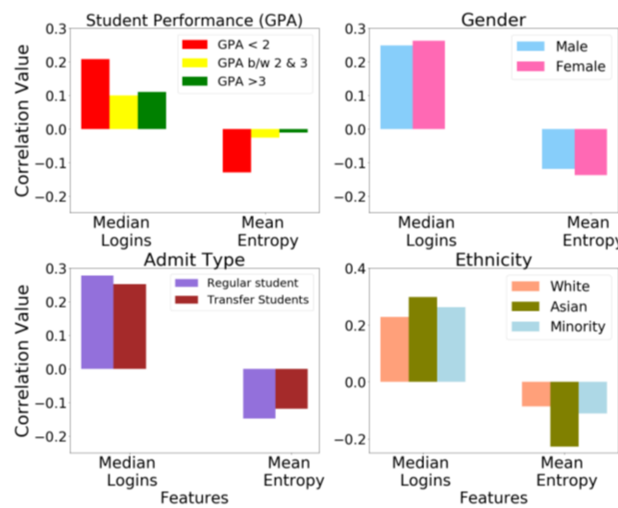

Figure 5. Correlation values for different student groups divided based on GPA, ethnicity, admit type and gender.

sections. We can infer that LIME based method also scales well for global feature importance in this study.

5.3.3 Regression Modeling for Feature Importance

One significant limitation of earlier methods is their inability to capture interaction effects as feature importance might change in the presence of other features. To study the interaction effects, we apply a linear regression model on different categories of student login data collected till the middle of semester. These student categories were divided based on GPA, gender, admit type and ethnicity of students. Even though linear regression models are applied on all features discussed in earlier sections, we only report the coefficients of median login volume and mean login regularity in table 7, as these variables are the focus of this study. From table 7, we observe that login volumes, and login regularity features are following similar direction for students with lower GPA and students from minority ethnic backgrounds as observed in the LIME and correlation based analysis. There are some discrepancies in other observations as there is no statistical significance (high p values) for coefficients in these cases. Another reason for focusing on student from these two groups is their higher attrition rates found in earlier studies [43, 44]. Studying these groups closely will help develop targeted interventions in the future.

Table 7. Regression coefficients (Significance marked with *)

Student Demographic	Student Groups	Median Logins Coefficient	Mean Login Regularity Coefficient
GPA	**GPA <= 2**	**0.171***	**-0.398***
	GPA >2 & <= 3	-0.013	0.200
	GPA >3	0.065	-0.002
Gender	Male	0.135	0.130
	Female	-0.021	-0.157
Admit Type	Regular	0.399	-0.004
	Transfer	0.611	0.191
Ethnicity	White	0.204	-0.029
	Asian	-0.085	0.201
	Minority Race	**0.201***	**-0.153***

6. DISCUSSION & CONCLUSION

There is a growing interest in building models that capture student behavioral patterns while using LMS systems to predict their performance. Earlier research showed that building efficient models based on LMS data to predict student performances is not a simple task as multiple learning and demographic factors impact student learning processes. Although earlier research in EDM and LA tried to address different issues related to student performance tracking, there is still a gap in developing models that accurately predict overall student performance and explain underlying factors that improve their academic performance. As a step in this direction, this study presents a student-centric modeling approach based on aggregated LMS features to predict and explain the reasons behind varying student performances. This context is both relevant and timely given the increase of LMS adoption and a need for efficient and interpretable model development.

6.1 Key Contributions

One primary contribution in this study is the development of student-centric models on aggregated student LMS login data that are least biased towards the diverse course contents and instructor teaching styles. Using the feature extraction methods developed in this study, we were able to build efficient GBT model that is able to predict student end-of-term GPA with an average R squared of 0.37 across the semester. Furthermore, models built at different durations of a semester showed only slight improvement in predictive performance after crossing a specific duration (middle of the semester). This observation helps develop models in the middle of the semester to estimate student performance accurately.

In addition to developing student-centric models, this study also focused on understanding the impact of various LMS features on student performances. Earlier studies in this domain primarily focused on volume of logins. In this work, we also studied the impact of login regularity measured by entropy statistics on student performance by implementing LIME explanation, correlation, and linear regression methods. From our interpretation studies, we observed that students who login regularly into the LMS system have a positive relationship with performance improvement. This observation is highly significant for underperforming students (GPA < 2) and students from minority races.

We also found no significant difference in the impact of LMS features on Male and Female students. This observation is valid as LMS features used in this study are captured objectively rather than subjectively. This observation also holds for regular and transfer students.

Our study also extracted student interaction features based on concepts in chronobiology and chronopsychology to understand if there is a student performance variation based on different chronotypes. From the results, we observed no significant difference in performance. The impact of these features is negligible in the presence of aggregated student login volume.

6.2 Applications & Limitations

Student performance tracking is a complex process as it depends on multiple dimensions and facets. Developing student-centric models to predict student performance models helps student counselors and educational administrators design student level interventions that attract students' attention. Also, developing predictive models that estimate students' overall performance in the middle of the semester will make them aware of their predicted end-of-term performance. These predictions might act as an external intervention to improve their performance in the remaining part of the semester. By understanding the difference in the impact of LMS features on students from different demographics, researchers and administrators can build more personalized instructional methods that are suitable for diverse student cohorts.

There were also some limitations in this study. The predictive performance achieved by using aggregate features across different courses enrolled by students is moderate at best. It would be more helpful to explore ways to improve the performance of these models. One possibility is to add other features that target independent content access durations, mid-semester assessments, and other external factors. One major challenge that needs to be addressed in our future studies is to find an effective method to aggregate content level features across different courses enrolled by a student. The dataset used in this study is extracted in a single semester and students from two departments that are closely related to each other. To understand if the findings in this study are scalable to other undergraduate students, we will extend these models to students from various departments in the university.

To conclude, we built student-centric models to predict student performances that supports the development of student level interventions. We then use the LIME explanations to study LMS features' importance on student performance prediction. Finally, we study the univariate and multivariate feature importance's using correlation and regression methods and assess them with the feature importance's extracted in LIME method.

7. REFERENCES

[1] Ninoriya, S., Chawan, P. M., & Meshram, B. B. (2011). CMS, LMS and LCMS for elearning. International Journal of Computer Science Issues (IJCSI), 8(2), 644.

[2] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert systems with applications, 33(1), 135-146.

[3] Plak, S., Cornelisz, I., Meeter, M., & Van Klaveren, C. (2019, March). Early Warning Systems for More Effective Student Counseling in Higher Education–Evidence from a Dutch Field Experiment. In Proceedings of the SREE Spring 2019 Conference, Washington, DC, USA (pp. 6-9).

[4] Beck, H. P., & Davidson, W. D. (2001). Establishing an early warning system: Predicting low grades in college students from survey of academic orientations scores. Research in Higher education, 42(6), 709-723.

[5] Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2016). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. IEEE Transactions on Learning Technologies, 10(1), 17-29.

[6] Mozahem, N. A. (2020). Using learning management system activity data to predict student performance in face-to-face courses. International Journal of Mobile and Blended Learning (IJMBL), 12(3), 20-31.

[7] Riestra-González, M., del Puerto Paule-Ruíz, M., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. Computers & Education, 163, 104108.

[8] Chiu, Y. C., Hsu, H. J., Wu, J., & Yang, D. L. (2018). Predicting Student Performance in MOOCs Using Learning Activity Data. J. Inf. Sci. Eng., 34(5), 1223-1235.

[9] Alokluk, J. A. (2018). The Effectiveness of Blackboard System, Uses and Limitations in Information Management. Intelligent Information Management, 10(06), 133.

[10] Berechet, L. D., & Georgescu, M. The Road from Blackboard Learning Management System to Moodle Learning Management System in Modern Universities.

[11] Shchedrina, E., Valiev, I., Sabirova, F., & Babaskin, D. (2021). Providing Adaptivity in Moodle LMS Courses. International Journal of Emerging Technologies in Learning (iJET), 16(2), 95-107.

[12] Ghilay, Y. (2019). Effectiveness of learning management systems in higher education: Views of Lecturers with different levels of activity in LMSs. Ghilay, Y.(2019). Effectiveness of Learning Management Systems in Higher Education: Views of Lecturers with Different Levels of Activity in LMSs. Journal of Online Higher Education, 3(2), 29-50.

[13] Aldiab, A., Chowdhury, H., Kootsookos, A., Alam, F., & Allhibi, H. (2019). Utilization of Learning Management Systems (LMSs) in higher education system: A case review for Saudi Arabia. Energy Procedia, 160, 731-737.

[14] Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. Computers in Human Behavior, 89, 98-110.

[15] Lang, C., Siemens, G., Wise, A., & Gasevic, D. (Eds.). (2017). Handbook of learning analytics. New York, NY, USA: SOLAR, Society for Learning Analytics and Research.

[16] Shum, S. B., & Crick, R. D. (2012, April). Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 92-101).

[17] Shum, S. B., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. Journal of Learning Analytics, 6(2), 1-9.

[18] Dawson, S., Heathcote, L., & Poole, G. (2010). Harnessing ICT potential. International Journal of Educational Management.

[19] Vengroff, R., & Bourbeau, J. (2006, February). In-class vs. On-line and Hybrid Class Participation and Outcomes: Teaching the Introduction to Comparative Politics Class. In annual meeting of the APSA Teaching and Learning Conference.

[20] Dutt, A., & Ismail, M. A. (2019, June). Can we predict student learning performance from LMS data? A classification approach. In 3rd International Conference on Current Issues in Education (ICCIE 2018) (pp. 24-29). Atlantis Press.

[21] Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In Search of user profiles. Computers & Education, 57(3), 2135-2144.

[22] Hung, J. L., & Zhang, K. (2012). Examining mobile learning trends 2003–2008: A categorical meta-trend analysis using text mining techniques. Journal of Computing in Higher education, 24(1), 1-17.

[23] Dawson, S., Heathcote, L., & Poole, G. (2010). Harnessing ICT potential. International Journal of Educational Management.

[24] Damianov, D. S., Kupczynski, L., Calafiore, P., Damianova, E. P., Soydemir, G., & Gonzalez, E. (2009). Time spent online and student performance in online business courses: A multinomial logit analysis. Journal of Economics and Finance Education, 8(2), 11-22.

[25] Baugher Varanelli Weisbord, D. A. E. (2003). Student hits in an internet-supported course: How can instructors use them and what do they mean?. Decision Sciences Journal of Innovative Education, 1(2), 159-179.

[26] Biktimirov, E. N., & Klassen, K. J. (2008). Relationship between use of online support materials and student performance in an introductory finance course. journal of education for business, 83(3), 153-158.

[27] Coldwell, J., Craig, A., Paterson, T., & Mustard, J. (2008). Online students: Relationships between participation, demographics and academic performance. Electronic journal of e-learning, 6(1), 19-30.

[28] Greenland, S. J. (2011). Using log data to investigate the impact of (a) synchronous learning tools on LMS interaction. In Proceedings of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE) Conference.

[29] Nandi, D., Hamilton, M., Harland, J., & Warburton, G. (2011, January). How active are students in online discussion forums?. In Proceedings of the Thirteenth Australasian Computing Education Conference-Volume 114 (pp. 125-134).

[30] Shum, S. B., & Ferguson, R. (2012). Social learning analytics. Journal of educational technology & society, 15(3), 3-26.

[31] Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. Computers in Human Behavior, 47, 157-167.

[32] Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. Computers in human behavior, 31, 542-550.

[33] Dominguez, M., Bernacki, M. L., & Uesbeck, P. M. (2016). Predicting STEM Achievement with Learning Management System Data: Prediction Modeling and a Test of an Early Warning System. In EDM (pp. 589-590).

[34] Lerche, T., & Kiel, E. (2018). Predicting student achievement in learning management systems by log data analysis. Computers in Human Behavior, 89, 367-372.

[35] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

[36] Hofmann, M., & Klinkenberg, R. (Eds.). (2016). RapidMiner: Data mining use cases and business analytics applications. CRC Press.

[37] Mandalapu, V., & Gong, J. (2019, January). Studying Factors Influencing the Prediction of Student STEM and Non-STEM Career Choice. In Proceedings of the 12th International Conference on Educational Data Mining.

[38] Mandalapu V., Gong, J., & Chen, L. (2021). Do we need to go Deep? Knowledge Tracing with Big Data. arXiv preprint arXiv:2101.08349.

[39] Kennedy, G., Coffrin, C., De Barba, P., & Corrin, L. (2015, March). Predicting success: how learners' prior knowledge, skills and activities predict MOOC performance. In Proceedings of the fifth international conference on learning analytics and knowledge (pp. 136-140).

[40] Stapel, M., Zheng, Z., & Pinkwart, N. (2016). An Ensemble Method to Predict Student Performance in an Online Math Learning Environment. International Educational Data Mining Society.

[41] Putilov, A. A., Sveshnikov, D. S., Puchkova, A. N., Dorokhov, V. B., Bakaeva, Z. B., Yakunina, E. B., ... & Mairesse, O. (2021). Single-Item Chronotyping (SIC), a method to self-assess diurnal types by using 6 simple charts. Personality and Individual Differences, 168, 110353.

[42] Romo-Nava, F., Blom, T. J., Guerdjikova, A., Winham, S. J., Cuellar-Barboza, A. B., Nunez, N. A., ... & McElroy, S. L. (2020). Evening chronotype, disordered eating behavior, and poor dietary habits in bipolar disorder. Acta Psychiatrica Scandinavica, 142(1), 58-65.

[43] Kupczynski, L., Gibson, A. M., Ice, P., Richardson, J., & Challoo, L. (2011). The impact of frequency on achievement in online courses: A study from a south Texas university. Journal of Interactive Online Learning, 10(3).

[44] Liu, F., & Cavanaugh, C. (2011). High enrollment course success factors in virtual school: Factors influencing student academic achievement. International Journal on E-learning, 10(4), 393-418.

[45] Kozachenko, L. F., & Leonenko, N. N. (1987). Sample estimate of the entropy of a random vector. Problemy Peredachi Informatsii, 23(2), 9-16.

[46] Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. Physical review E, 69(6), 066138.

[47] Ruping, S. (2000). mySVM-manual. http://www-ai. cs. uni-dortmund. de/SOFTWARE/MYSVM/.

[48] Gray, R. M. (2011). Entropy and information theory. Springer Science & Business Media.

[49] Kim, J. H., Seodaemun-gu, S., Park, Y., Song, J., & Jo, I. H. (2014). Predicting students' learning performance by using online behavior patterns in blended learning environments: comparison of two cases on linear and non-linear model. Korea, 120, 750.

[50] Park, Y., & Jo, I. H. (2015). Development of the learning analytics dashboard to support students' learning performance. Journal of Universal Computer Science, 21(1), 110.

[51] Jo, I. H., Kim, D., & Yoon, M. (2015). Constructing proxy variables to measure adult learners' time management strategies in LMS. Journal of Educational Technology & Society, 18(3), 214-225.

[52] Jo, I. H., Park, Y., Yoon, M., & Sung, H. (2016). Evaluation of online log variables that estimate learners' time management in a Korean online learning context. International Review of Research in Open and Distributed Learning, 17(1), 195-213.