Enhancing Predictive Analytics for Students' Performance in Moodle: Insight from an Empirical Study

Dynil Duch, Madeth May, Sébastien George

Received: 6 July 2024 | Revised: 3 September 2024 | Accepted: 20 September 2024 | Published online: 26 September 2024

RESEARCH ARTICLE

Journal of Data Science and Intelligent Systems
2024, Vol. 00(00) 1–13
DOI: 10.47852/bonviewJDSIS42023777

BON VIEW PUBLISHING

Enhancing Predictive Analytics for Students' Performance in Moodle: Insight from an Empirical Study

Dynil Duch[1,2,*], Madeth May[1] and Sébastien George[1]

[1]*LIUM Computer Science Laboratory, University of Le Mans, France*

[2]*Institute of Digital Research & Innovation, Cambodia Academy of Digital Technology, Cambodia*

Abstract: This paper explores the application of learning analytics in predicting students' performance within Moodle, a widely used learning management system. The study focuses on measurable academic progress and outcomes, aiming to assist educators in early identification and resolution of issues to boost student productivity and success. Our approach began with a literature review to identify predictive attributes for student performance. We then collected and analyzed data from a year-long study involving 160 students at the Cambodia Academy of Digital Technology. The dataset included attendance, interaction logs, quiz submissions, task completions, assignments, time spent on courses, and outcome scores. We utilized these data points to train and evaluate various classifiers, identifying the random forest classifier as the most effective. A predictive algorithm was developed using the coefficients from this classifier, tailored for practical application in educational settings. Our analysis confirmed significant correlations between the identified attributes and prediction accuracy, enhancing the algorithm's efficacy. A follow-up survey with the same participants one year later provided further validation, affirming the predictive indicators' effectiveness in improving academic performance. This comprehensive approach demonstrates the robustness of our findings and underscores the potential of predictive analytics in enhancing educational outcomes.

Keywords: deep analytics, knowledge extraction, student performance prediction, data mining, data normalization, data visualization, data indicators

1. Introduction

In recent years, the use of learning management systems (LMS) has grown significantly, primarily for their ability to manage and organize digital information related to teaching and learning. Educational institutions utilize LMS platforms such as Moodle, an open-source system widely adopted in the education sector, to facilitate the creation, distribution, and management of online learning materials [1–4]. In addition to supporting teaching practices, various studies have employed data mining techniques to predict students' performance (SP), such as Hussain et al. [5], Bisri et al. [6], Albahlil [7], and Pallathadka et al. [8]. These studies have demonstrated that utilizing data from a learning environment and efficient predictive techniques can assist teachers in evaluating SP. For instance, a teacher can identify areas where students may be struggling and implement targeted interventions to improve student outcomes. Thus, the ability to predict SP can significantly impact learning practices in general, as it allows for more personalized and adaptive learning experiences that can better meet the needs of individual students.

However, identifying the crucial data attributes (hereafter referred to as "key attributes" or "attributes") from online learning activities is challenging due to the large amount of information in the LMS [9], especially when the goal is to obtain a more accurate SP prediction [10, 11]. Moreover, predicting SP and its associated issues have always been a considerable concern in education. For example, both the nature and quality of data strongly impact the efficiency of a predictive approach. Additionally, the outcomes of using predictive techniques heavily rely on teachers' technical skills.

The research question in this study is: how can teachers employ predictive techniques to evaluate SP effectively? This question explores the potential benefits and limitations of using predictive analytics in the educational context and identifies best practices for implementing these techniques to support student learning. By addressing this research question, educators and researchers can better understand the concrete applications of predictive analytics in the classroom, contributing to more effective teaching practices and improved student outcomes.

To address this research question, we have set two main objectives:

1) Identify the key attributes in LMS data that are relevant for predicting students' performance. A number of research has identified the specific attributes for predicting SP as presented later in Section 2. However, there is a need for more identification of various attributes that could help educators or teachers with the selection process. Our research will conduct a literature review by analyzing various data points, such as grades, assessment results, engagement metrics, demographic

*Corresponding author: Dynil Duch, LIUM Computer Science Laboratory, University of Le Mans, France. Email: dynil.duch.etu@univ-lemans.fr

information, and teacher pedagogies, to determine which factors could be used for predicting SP.
2) Propose an appropriate prediction algorithm to forecast SP based on the critical attributes identified in the first objective. This includes using predictive algorithms and data mining techniques to analyze the data and identify patterns and trends that can be utilized to predict future SP.

To test our research question, we have developed the following hypotheses:

Hypothesis 1: There is a statistically significant relationship between student attendance (measured by the number of modules completed) and academic performance (measured by the final grades) in Moodle. This hypothesis investigates how consistent attendance influences SP prediction.
Hypothesis 2: The extent of student interaction with the LMS (measured by the number of interaction logs) has a statistically significant impact on academic performance (measured by the final grades). This hypothesis aims to explore whether higher levels of engagement with the LMS positively influence SP.
Hypothesis 3: The number of quizzes and tasks submitted by students statistically impacts their final academic performance. This hypothesis examines whether students' active participation in quizzes and tasks within the LMS significantly predicts their overall performance.

We conducted our initial tests using data from the Cambodia Academy of Digital Technology (CADT), as outlined in Section 4. This dataset provided a comprehensive basis for our empirical study, allowing us to analyze various student performance indicators. To further validate our findings and ensure the robustness of our model, we incorporated a follow-up survey. This survey was designed to capture additional insights and perspectives directly from the students, offering a practical validation of the predictive indicators identified in our research. The combination of these data sources helped us to confirm our hypotheses, the effectiveness of our model, and its applicability in real-world educational settings.

The rest of the paper presents our approach to identifying a wide range of key attributes and appropriate prediction algorithms. Our hypotheses will help us better understand the impact of attributes we have collected from CADT. The second section provides an overview of the most relevant related works. We examine critical attributes for predicting SP and explore the technology context and methodology approach in Sections 3 and 4. We discuss our experimental results in Section 5 and conclude this paper by highlighting significant areas we have been working on since completing the tests we conducted with our first three hypotheses.

2. Literature Review

This section is dedicated to a comprehensive review of current research on predicting student outcomes using machine learning and data mining techniques in educational settings that exploit LMS data, such as Moodle. The studies we mentioned here explore various aspects of SP, such as academic performance, retention, and learning behaviors. The studies also cover different techniques, including decision trees, neural networks, logistic regression, support vector machines, Naïve Bayes, and random forests.

A meta-study by Felix [12] reviewed 42 papers using data mining techniques to predict student outcomes. The study found that decision trees, neural networks, and logistic regression were the most commonly used techniques. However, it also highlighted challenges such as the need for extensive and diverse datasets, lack of standardized evaluation metrics, and potential ethical concerns related to using sensitive student data. Similarly, Namoun and Alshanqit [13] systematically reviewed 62 papers that used data mining and machine learning to predict student outcomes as a proxy for student SP. The review stated that existing studies mainly focused on the course level, using predictors such as previous academic performance, demographic data, and course-related variables. Machine learning algorithms, including decision trees, neural networks, support vector machines, Naïve Bayes, and random forests, were found to accurately predict student outcomes, with some studies reporting prediction accuracies of over 90%. However, the review also noted limitations, such as the lack of validation of the models on new datasets and limited explanatory power.

In another study, Felix et al. [14] used a dataset of 1,307 students' activity logs in a course, including variables related to student interactions in forums, chats, quizzes, activities, time spent on the platform, and grades. They built a predictive model of student outcomes using Naïve Bayes, decision trees, multilayer perceptron, and regression algorithms, with the Naïve Bayes model performing the best with an accuracy of 87%. The study found that the number of interactions with the system, attendance, and time spent on the platform were essential variables in predicting student outcomes. Nevertheless, the study was limited to a single course and did not consider other factors influencing student outcomes, such as prior knowledge or motivation.

The study of Hirokawa [15] used machine learning methods like random forests, support vector machines, and decision trees to forecast academic achievement. The study unveiled that previous academic records were essential for predicting academic performance, followed by the student's gender and age. At the same time, other attributes, such as extracurricular activities and family background, had a lesser impact. Yet, the study showed some limitations, as it did not focus on data from LMS activities and excluded influence factors such as teacher pedagogies.

In the same context, Gaftandzhieva et al. [16] used a machine learning algorithm to predict students' final grades in an Object-Oriented Programming course using data from Moodle LMS activities. The study found that the random forest algorithm had the highest prediction accuracy of 78%, and attendance was strongly correlated with final grades. The study's weaknesses, however, included its limited sample size and singular course focus. Other studies have used data mining and machine learning to predict the likelihood of students dropping out of a course [17], the likelihood of student's success in a course [18], or predicting student grades using both academic and non-academic factors [19, 20]. Some studies have also focused on predicting student outcomes in specific contexts, such as interaction logs [21], assessment grades [22], demographic data [23], previous grade [8, 24], and online activity data [25], or based on teacher pedagogies [26].

Thus far, the studies we cover demonstrate the potential of data mining and machine learning techniques to predict various student outcomes in educational settings. However, they also highlight challenges such as the need for extensive and diverse datasets, the lack of validation of models on new datasets, the lack of study on predicting SP at the program level, and potential ethical concerns related to using sensitive student data. Furthermore, the focus on a single course at the course level did not lead the educators to make a final decision on overall students' performance at the end of the program or academic year. Meanwhile, predicting students' performance at the program level is demanded [27, 28]. Based on this observation, our research examines the critical attributes in LMS data relevant to predicting SP at the program level. Focusing on the variables collected from Moodle, our research investigates

the relationship between student engagement activities and SP, presented later in Section 4.2.1.

While studies we outlined here have identified specific attributes used to predict SP, there is still a need to identify various related attributes. In an effort to close this gap, our study seeks to comprehensively understand the various variables by thoroughly examining existing research. This approach allows us to identify the key attributes from the CADT's Moodle to enhance data mining techniques for effective and accurate SP prediction. Through this research effort, we aim to provide valuable insight into how institutions can better understand student performance across their programs and make necessary improvements to enhance student learning outcomes. Additionally, we hope to contribute to advancing the field of educational data mining, allowing educators and policymakers to effectively support at-risk students and promote educational justice by improving overall educational quality.

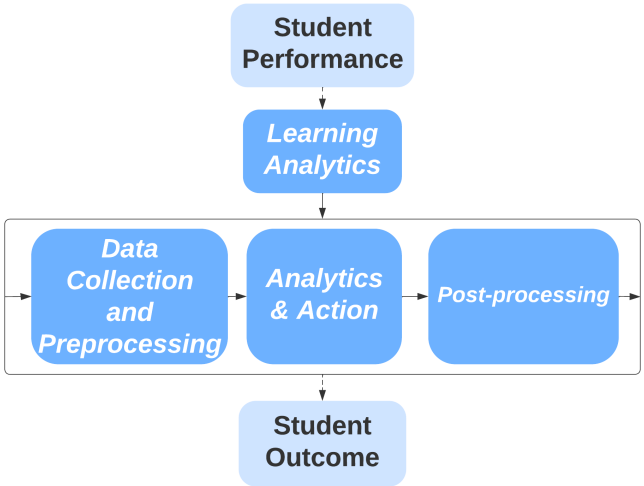
3. Technological Context

The diagram in Figure 1 shown the relationship between students' performance and the three parts of learning analytics (LA): data collection and data preprocessing, analytics and actional, and postprocessing. SP is at the beginning of the diagram, highlighting its importance as a critical input to the LA process. Indeed, LA uses data from SP and other data sources to gain insights into the learning process and develop actionable interventions for output student outcomes.

3.1. Learning analytics

LA is a vital area of education technology that has become increasingly important in recent years, especially in online learning environments [29]. As defined by Chatti et al. [30], LA refers to collecting, analyzing, and presenting data from learners and their learning context to gain insights into the learning process and improve the learning experience. By collecting and examining data from diverse sources, such as SP, behavior, and interactions, LA can help educators make informed decisions to enhance teaching and learning strategies.

Figure 1
The student-centric learning analytics diagram



The implementation of LA in our research was carried out through the following three steps: 1) data collection and preprocessing, 2) analytics and action, and 3) post-processing.

1) **Data Collection and Data Preprocessing:** The first step in our LA is collecting data from Moodle LMS and Google Sheets. The preprocessing of this data is crucial as it transforms raw data into a usable format that can be analyzed. This step includes data cleaning, integration, transformation, reduction, user and session identification, and path completion.
2) **Analytics and Actional:** The second part of our study involves applying analytical techniques to data to identify patterns, trends, and relationships. The insights gained from our analysis are used to develop interventions that support individual learners.
3) **Post-processing:** The third part involves evaluating the effectiveness of the LA process. We measured the accuracy and effectiveness of the learning models, evaluating the impact of interventions and assessing the overall quality of the LA process. For example, we participated in determining unique attributes required for further interaction, identifying new indicators/metrics, modifying the analysis variables, and choosing a new analytics method.

In summary, LA plays a crucial role in our research work, guiding us to develop technological solution to improve students' learning experiences and outcomes.

3.2. Students' performance

SP is a vital aspect of the LA process, which employs data-driven methods to analyze and interpret information related to academic performance [13]. By leveraging various data mining techniques, we discovered patterns of student behavior that were used to enhance SP in LMS. Moreover, we analyzed diverse student data (as detailed in Section 4.2.1), such as grades, attendance, time spent on LMS, number of interaction logs to LMS, total of assignments submitted, total of quizzes submitted, and total of tasks submitted. We used this information to identify the critical attributes influencing SP and its impactful factors. Our intention is not only to improve learning outcomes but also to develop effective educational practices.

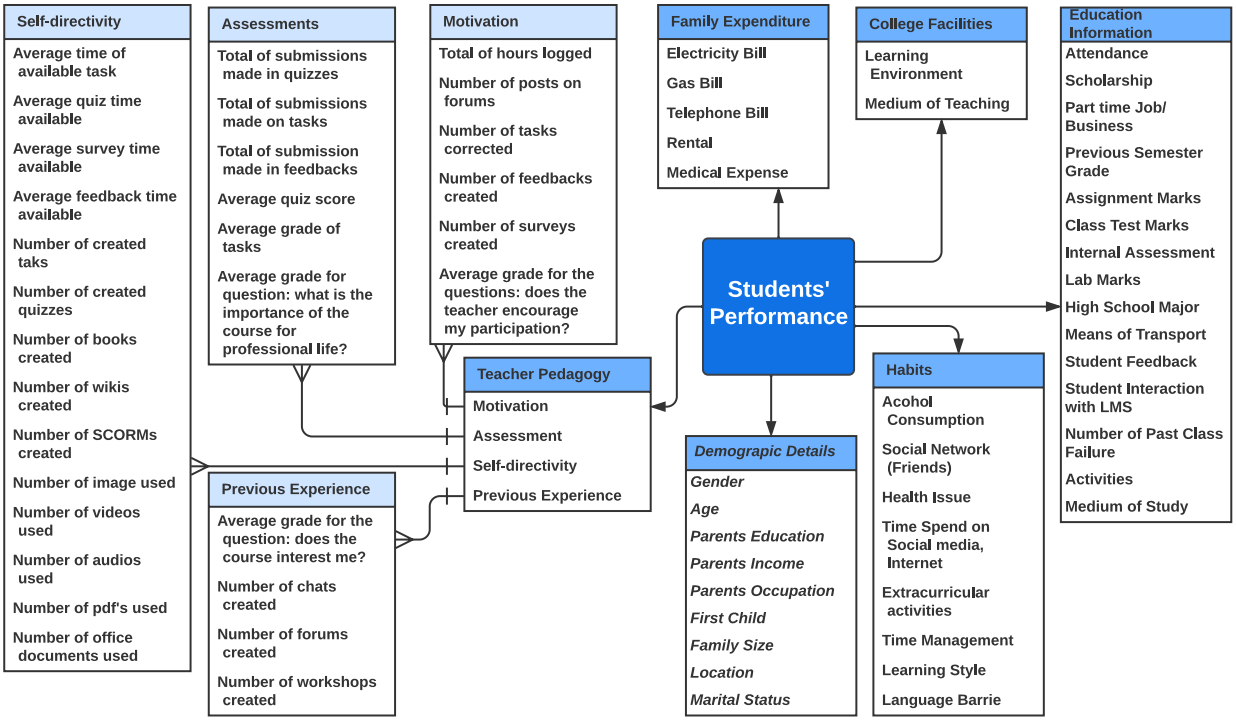
The primary challenge we face is determining which factors are most strongly correlated with SP at all levels of analysis. Indeed, by gaining a better understanding of the factors that influence SP, researchers and educators can collaborate to develop targeted interventions and support LMS that suit both individual teaching and learning experiences. To address this technical challenge, we explored various data points and we selected appropriate prediction algorithms to forecast SP. Our methodological approach is detailed in the following section.

4. Methodological Approach

To identify key attributes significantly effective for predicting SP and to choose the best performance predictive approach, our study takes several steps leveraging LA:

1) **Literature Review:** We comprehensively reviewed existing literature to identify key factors influencing student outcomes.
2) **Data Collection:** Based on the attributes identified in the literature review, we collected data from CADT's Moodle and Google Sheets, focusing on student engagement and performance. We then selected a subset of the most relevant attributes for predicting SP.

Figure 2
The attributes for predicting SP



3) **Classification Method Development:** We utilized these attributes to develop a classification method using a random forest classifier, which was identified as the most effective among various classifiers. We also applied oversampling techniques to address data imbalance and fine-tune the classifier.
4) **Performance Evaluation:** We assessed the performance of the predictive classifier, making necessary adjustments to enhance its accuracy.
5) **Attribute Coefficients:** After training the random forest classifier, we derived the coefficients for each attribute and applied them to our SP algorithm for outcome prediction.
6) **Regression Analysis and Survey:** We performed a regression analysis to validate our hypotheses and surveyed to assess the effectiveness of our predictive indicators qualitatively.

4.1. Attributes identification

The attributes identified in our study are derived from an extensive literature review as previously presented in Section 2. We dedicated time to scrutinize various papers that use different attributes to predict academic outcomes, success, or dropout rates. We uncovered six composite attributes commonly used in predictive models: Demographic Details, Education Information, Family Expenditure, Habits, College Facilities, and Teacher Pedagogy, as illustrated in Figure 2. These attributes guide teachers, providing a foundational understanding to navigate the numerous attributes available in an LMS and their relationships.

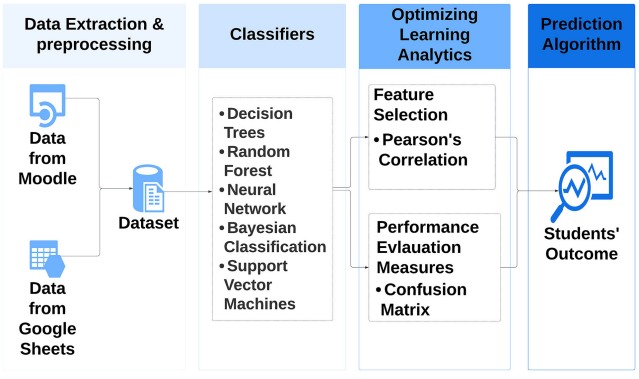
It is crucial to note that Figure 2 does not represent an exhaustive list of all possible predictors of SP. Instead, it serves as a starting point to facilitate decision making for teachers in selecting pertinent attributes from the multitude available in LMS when aiming to predict SP. The intention is to offer researcher and institutions the ability to decide when and which attributes to use based on relevance, practicality, and ease of interpretation within their specific context.

After a thorough and rigorous process of attribute identification, we have gained valuable knowledge to focus our effort on developing our predictive model. Specifically, attributes presented later in Section 4.2.1 have been prioritized from the broader set defined in our model. This strategic selection aims to provide a more practical and targeted approach for both teachers and the research team at CADT and in France.

4.2. Predictive approach

This research uses data mining techniques and predictive algorithms within the Moodle environment. The methodology adopted for this research is divided into several steps, as shown in Figure 3.

Figure 3
The predictive approach



4.2.1. Data extraction and preprocessing

The dataset used in our research comprises two primary sources: the Moodle LMS and Google Sheets. The Moodle data have around 1000 students enrolled in both short course training and a bachelor program, totaling around 5 million records. By conducting a preliminary analysis, we kept only the bachelor's degree data because the data from the short course training does not contain any assessment score or final score, which is the critical target variable. Moreover, the students from short course training had less interaction with Moodle. Our study aimed to predict SP at the program level. Among all registered students, many were enrolled in short courses or workshops, which did not provide a comprehensive view of their performance within a program. Short courses typically focus on specific topics and may only cover some aspects required to evaluate a student's overall performance in a program. To ensure that our analysis and predictions were based on a thorough understanding of SP in the learning environment, we focused on 160 students representing a diverse range of courses, including Linear Algebra, Discrete Mathematics, Probability and Statistics, C Programming Language, Visual Art, Soft Skills, and Information Technology Essentials. Because the short course training data could not be used, we selected only 2 million records from the bachelor program.

At CADT, instructors develop content in various formats, such as videos and files uploaded to Moodle. It allows students to engage with the material online before attending physical classes. In these classes, instructors can review the lessons, conduct activities to ensure comprehension, and address questions. Assessments are conducted both in-person and online via Moodle. All student interactions with Moodle, including viewing content, participating in discussions, and completing assessments, are recorded in the Moodle log. These logs are then queried to extract relevant data for analysis.

Queries were used to count the number of records in each attribute to obtain data at the program level for both the first and second terms of the first-year program. Finally, we obtained a dataset with 310 records from Moodle. Afterward, another set of 310 records was collected from Google Sheets, representing this time the final scores of students enrolled in both terms of their first year of the bachelor's degree, aligning with the data collected from Moodle during the academic year 2022. The dataset of two terms represents two program levels. It also incorporates Hypothesis Video Player scores, which measure student engagement in interactive video activities. The attributes in the dataset provide information on various aspects of a student's engagement and performance in all courses. The attributes are collected and counted with queries as shown in Table 1, which details the attributes and their descriptions with queries for acquisition.

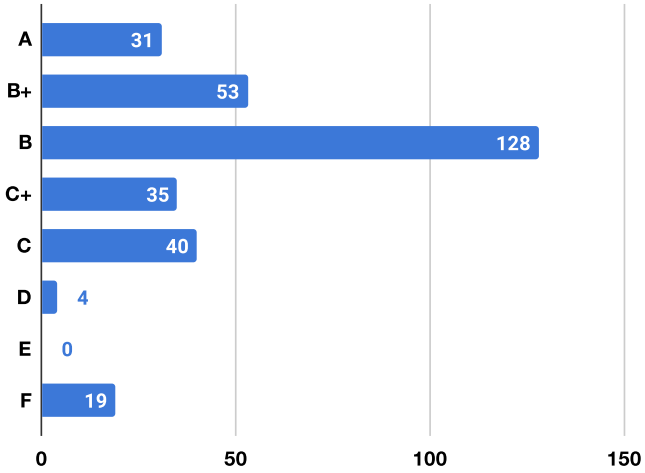
Table 1
Attributes and their descriptions with queries for acquisition

Attribute	Descriptive	Query (how to acquire)
Attendance	It represents the number of modules in all courses that a student has completed.	In Moodle, it is calculated by joining the mdl_course_modules table and mdl_course_modules_completion, and counting all the rows with completion-state!=0, grouped by user_id and course_id.
Number of interaction log	It counts the number of interactions a student has had with all courses.	The attribute is calculated by counting all the rows in the mdl_logstore_standard_log table, grouped by user_id and course_id.
Total quiz submitted	It means the number of quizzes a student has submitted in all courses.	It is acquired by joining the mdl_course_modules table and mdl_course_modules_completion, counting all the rows with completion-state!=0 and module=16, grouped by user_id and course_id. Note that module=16 means that the module is a quiz.
Total task submitted	It represents the number of tasks a student has submitted in all courses.	It is calculated by joining the mdl_course_modules table and mdl_course_modules_completion, counting all the rows with completion-state!=0 and module=1 or 16 or 37, grouped by user_id and course_id. Note that module=1 means that the module is an assignment, module=16 means that the module is a quiz, and module=37 means that the module is an hvp.
Total assignment submitted	It is the number of assignments a student has submitted in all courses.	The attribute is obtained by joining the mdl_course_modules table and mdl_course_modules_completion, counting all the rows with completion-state!=0 and module=1, grouped by user_id and course_id. Note that \module=1 means that the module is an assignment.
Time spent on course	It is the total count of hours a student spent interacting with a course.	It is calculated by summing all seconds between each record in mdl_logstore_standard_log that is less than 3600 s, order by user_id and course_id.
Outcome score	It is a numeric measure of a student's academic performance after completing the first term and the second term of the first-year program.	The outcome score is typically calculated by taking the weighted average of the final scores of each course in the term. The weight assigned to each course is based on various factors such as credit hours, difficulty level, or course importance. The outcome score is a significant metric used in academic and employment contexts to evaluate a student's academic performance and potential.

4.2.2. Data handling and GDPR

A grading system was used to translate the outcomes score, ranging from 0 to 100, into grades A to F. In any case, no student received an E as shown in Figure 4.

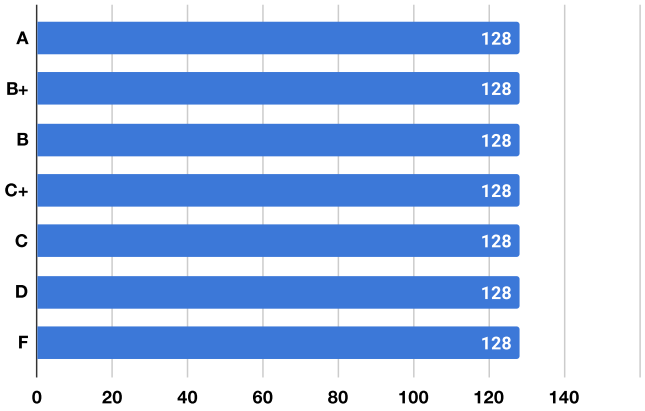
Figure 4
Distribution of grades



The majority class in grade B is significantly larger than the other grades, and this imbalance in the data caused our classifier to become biased in favor of the prediction. To address this issue, we employed the random oversampling method to enhance the model categorization. Using this method, we add instances from the minority class to the dataset randomly, with replacement, to calculate the model's accuracy. We obtained the final data for 896 records after applying the oversampling procedure as shown in Figure 5.

We then normalized the data by scaling all attributes to the same range between 0 and 1 as shown in Table 2 because the coefficients of each attribute in the SP prediction algorithm depend on the attribute values' scale. Normalizing the data ensures that all attributes are given equal weight in calculating the performance score and that the score is not biased toward any particular attribute. Additionally, normalizing the data makes it easier to compare the performance of different students, as the scores are all on the same scale.

Figure 5
Distribution of grades by using the random oversampling technique



As we move further with data analysis containing students' learning activities and outcomes, it is important for us to point out that handling data ethically and protecting personal data privacy are crucial aspects of our research. Compliance with the General Data Protection Regulation is a cornerstone of our research methodology. Ethical considerations are at the forefront, with participants fully informed about the nature of their involvement, their rights, and the procedures in place for data management. Transparency is maintained through clear communication, and participants can request the deletion of their data. An ethics committee, including research team members, oversees and approves all aspects of our research design and execution. We adhere to the guidelines and regulations set forth by relevant authorities to ensure ethical and responsible data handling. For the protection of the privacy of the participants, all personal identifiers such as name, contact details, and identification numbers are removed from the data before analysis. In addition, the data are stored securely with limited access to authorized personnel only. We also obtained informed consent from the participants before collecting any data, and we ensured that the data collection process did not cause any harm to them.

4.2.3. Selection feature

Feature selection is an essential step in predicting SP as it allows for the identification of the most relevant attributes that contribute to the outcome. Several feature selection methods are available, such as the Pearson correlation coefficient, Spearman's rank correlation coefficient, mutual information, and recursive feature elimination. Each technique offers advantages and disadvantages depending on the dataset and the problem being addressed.

Our study employs the Pearson correlation coefficient as a feature selection method for two reasons:

1) The Pearson correlation coefficient is commonly used to quantify the linear connection between two continuous variables, and it is appropriate for our dataset containing continuous features.
2) The Pearson correlation coefficient gives a clear and comprehensible estimate of the degree and direction of the link between attributes and the target variable.

Although alternatives such as Spearman's rank correlation coefficient and mutual information can handle nonlinear relationships and are more robust against outliers, they are more complex to compute and may not provide easily interpretable results. By employing the Pearson correlation coefficient, the information related to the education information category, specifically engagement activities, as found in collected data from CADT's Moodle. Therefore, our study will investigate the relationship between student engagement activities and performance using the attributes already mentioned in Section 4.2.1. It is worth reminding that these attributes are significant predictors of student outcomes in the existing literature. However, there may be differences between the attributes discussed in the literature and those in our dataset. Nonetheless, our primary goal was to identify and analyze attributes that could be feasibly obtained from Moodle while still providing valuable insights into SP. In addition, by focusing on Moodle-based data, we aimed to develop a model easily applied and adapted in similar LMS environments. Plus, the selected Moodle-based attributes still capture essential aspects of SP. Our findings can contribute to the

Table 2
The CADT's sample dataset

Attendance	Number of interaction logs	Total quiz submitted	Total task submitted	Total assignment submitted	Time spent on course	Grade
0.279412	0.265866	0.000000	0.166667	0.304348	0.239303	F
0.382353	0.798456	0.333333	0.333333	0.521739	0.772971	A
0.397059	0.421098	0.333333	0.309524	0.478261	0.260056	B+
0.397059	0.482847	0.333333	0.309524	0.478261	0.317422	A
0.161765	0.325043	0.000000	0.214286	0.391304	0.253804	B

broader understanding of factors that impact academic success in online learning environments.

4.2.4. Classifier

To predict student outcomes, we evaluated several classification methods to determine the most suitable for our dataset. The classifiers considered include decision trees, random forests, neural networks, Naive Bayes, and support vector machines. These methods were chosen based on their widespread use and effectiveness in similar contexts, as identified in related work.

Selection Criteria:

1) **Classifier Suitability:** We assessed each classifier's ability to handle the specific data types in our study, including continuous and categorical attributes.
2) **Flexibility and Adaptation:** We prioritized classifiers that could adapt to the diverse attributes and relationships in the data, ensuring they could model the complex nature of student performance.
3) **Scalability:** We considered how well each classifier performs with large datasets, ensuring they can handle the volume of data efficiently and effectively.
4) **Interpretability:** We evaluated the ease with which each classifier's results and decisions could be interpreted by educators and researchers, aiming for a balance between predictive accuracy and comprehensibility.

We systematically compared the classifiers by applying these criteria to select the one best suited for predicting student performance based on our dataset. This comparison ensures that the chosen classifier aligns with the study's objectives and provides reliable predictions.

4.2.5. Performance evaluation measures

To evaluate the performance of our classifier, we employed the confusion matrix, which is a widely used as an evaluation measure in machine learning for summarizing classifier performance. It provides information about the number of true positive, true negative, false positive, and false pessimistic predictions made by the classifier. According to Aguar et al. [31], this information is presented in a matrix format, where each row represents the actual class, and each column represents the predicted class. The entries in the matrix provide insight into the classifier's ability to predict each class and its tendency to misclassify instances. This evaluation measure is crucial in unbalanced datasets, such as in this case, where the number of failing students is much smaller than that of successful students [12].

4.2.6. Students' performance predictive algorithm

We have defined the algorithm formula for predicting SP as a mathematical equation that uses a combination of various student attributes and their corresponding coefficient values to calculate a predicted score. The formula starts by taking the sum of all attribute values and then normalizing each value by dividing it by the range of possible values for that attribute. Then, each normalized attribute value is multiplied by its corresponding coefficient value, representing the weight or importance of that attribute in the overall prediction. Finally, the sum of all these weighted attribute values is divided by the sum of all the coefficients to arrive at a final performance score. This score can then be used to classify students as likely to succeed or likely to struggle in their studies.

$$sp = \sum_{i=1}^{N} \frac{C_i \frac{x_i - \min_i}{max_i - min_i}}{\sum_{i=1}^{N} C_i}, \quad where \begin{cases} sp, & x, \quad C \in R \\ 0 \leq sp, & C \leq 1 \end{cases} \tag{1}$$

$\frac{x_i - min_i}{max_i - min_i}$ is a normalization technique to scale attribute values to a standardized range.

sp the student's performance
x_i the value of attribute i
C_i the coefficient of attribute i which derived through the model training process
max_i the maximum value of attribute i
min_i the minimum value of attribute i
N the number of attributes

In our algorithm formula, the sum of all coefficient values provided by each classifier's result equals 1. In this case, we can update the formula accordingly.

$$sp = \sum_{i=1}^{N} C_i \frac{x_i - \min_i}{max_i - min_i}, \quad where \begin{cases} sp, & x, \quad C \in R \\ 0 \leq sp, & C \leq 1 \\ \sum_{i=1}^{N} C_i = 1 \end{cases} \tag{2}$$

This approach for predicting SP is based on mathematical modeling and statistical analysis principles; by using a weighted average calculation incorporating multiple variables, the algorithm can provide a more accurate and reliable prediction of a student's likely performance. Moreover, the normalization of attribute values ensures that all variables are treated equally, regardless of their scales or units of measurement.

To make it easily understandable, let's start with an example of the SP of approximately 0.63, which falls between 0 and 1. We could interpret it as an above-average performance according to our chosen attributes and coefficients. Values closer to 1 may indicate better predicting algorithm performance, while those closer to 0 suggest poorer performance.

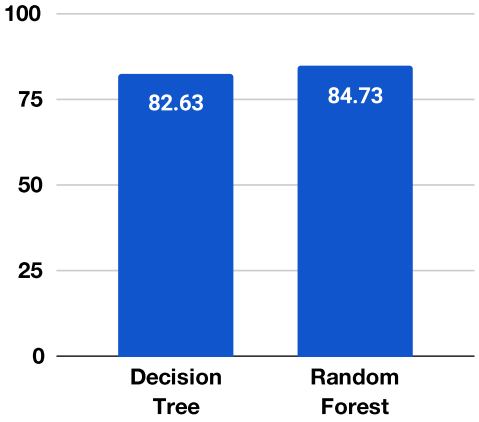
5. Experimental Results

This section details our study's experimental approach and outcomes in predicting SP using LA. The experimental process began with identifying key attributes relevant to predicting SP through a thorough literature review. This was followed by data

collection from Moodle at the CADT, focusing on various indicators such as attendance, interaction logs, quiz submissions, and final scores. We evaluated several classifiers after data preprocessing, including handling missing values and normalization. The Random Forest classifier was identified as the most effective through accuracy assessment. To ensure the robustness of our findings, the classifier was refined using oversampling techniques to address class imbalances. The model's predictive capability was further validated through a follow-up survey with the same student cohort, affirming our predictive indicators' practical relevance and effectiveness in educational settings. The experimental results of our research on assisting teachers in using predictive techniques to evaluate the student's performance at CADT are as follows:

> The data from CADT are analyzed with five distinct classifiers (as previously stated in Section 4.2.4) to examine the dataset and assess each model's accuracy. Based on the outcome shown in Figure 6, the best two classifiers are the Decision Tree classifier, with an accuracy of 87.22%, and the Random Forest classifier achieved a higher accuracy of 89.44%.

Figure 7
The comparison of classifiers accuracy of random forest and decision tree



While accuracy is a common and intuitive metric for evaluating classifier performance, it has limitations, especially in class imbalances or unequal misclassification costs. Performance evaluation metrics, such as confusion matrices, provide a more detailed and nuanced understanding of how well a classifier is performing. Next, we used confusion matrix to evaluate the classifiers. Based on the results in Figure 7, we demonstrated that both Random Forest and Decision Tree offer good accuracy. Therefore, for the purpose of comparison, we selected the prediction classifiers, which include Decision Tree and Random Forest. However, the latter was chosen for predicting SP due to its superior performance.

5.1. Attribute coefficients

At the end of the classifier performance evaluation, we obtained the coefficients of feature values for each attribute to illustrate its importance in predicting SP when using decision trees and random forest classifiers, as depicted in Figure 8. The coefficients indicate the weight each attribute has in the classifiers. For example, the coefficient values for the number of interaction log, and time spent on course are higher for both classifiers, indicating that these attributes are essential and impactful in determining SP. On the other hand, the coefficients for the attendance, total quizzes submitted, total tasks submitted, and total assignments submitted are relatively low because they are already factored into the final grade. This indicates that these attributes have little importance in influencing SP prediction, even if some teachers might utilize them outside of Moodle.

Figure 6
The accuracy of each classifier

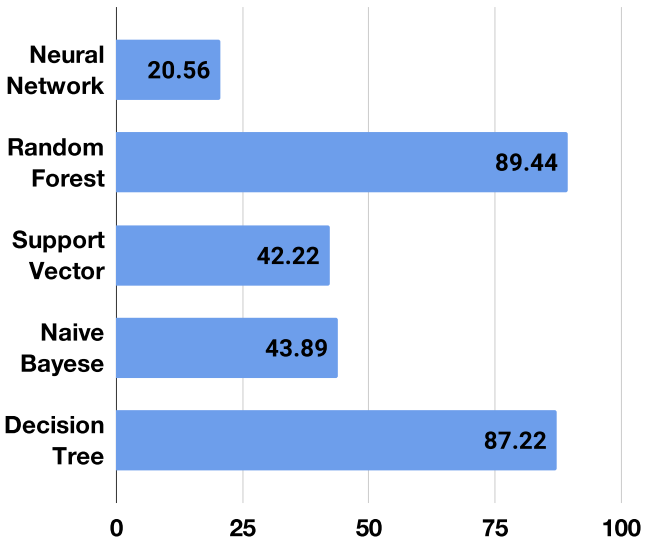
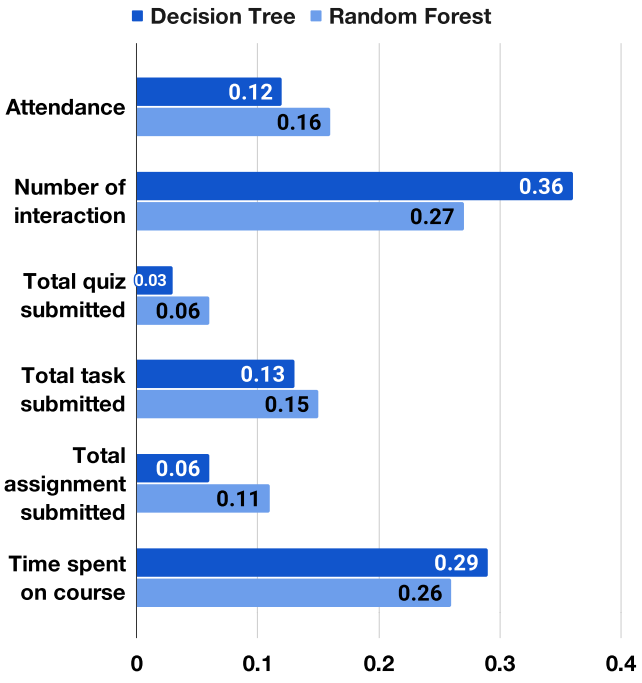


Figure 8
The coefficient of each attribute of decision tree and random forest



5.2. Correlation analysis of attributes for predicting SP

Once we identified the key attributes and selected the classifier, we analyzed the correlation between different attributes. According to the findings in Figure 9, a strong positive correlation between attendance and total quiz submissions ($r = 0.82$) suggests that students who attend more regularly are likely to engage in quiz activities. Similarly, the strong positive correlation between the number of interaction log and time spent on the course ($r = 0.88$) indicates that students who spend more time on the system tend to have higher interaction logs. Moreover, the strong positive correlation between total task submitted and total assignment submitted ($r = 0.87$) implies that most tasks students submit are assignments. On the other side, there is a moderate positive correlation between attendance and number of interaction log ($r = 0.66$), total quiz submitted and number of interaction log($r = 0.61$), total quiz submitted and time spent on course ($r = 0.57$), and attendance and time spent on course ($r = 0.53$). This indicates that these attributes are positively related.

These results provide valuable insights into the connections between different attributes, which can help understand patterns and potential influences on student performance or engagement in the course.

5.3. Verification of hypotheses through regression analysis results

The results we have presented thus far can serve as a valuable guide for teachers, helping them to select key attributes and understand the relevance, degree of importance, and impact of each attribute in predicting SP. The last step in our study is to examine the independent variables' P-value and confidence interval (CI) to verify our hypotheses using the regression analysis results as shown in Table 3.

Figure 9
The correlation of each attribute

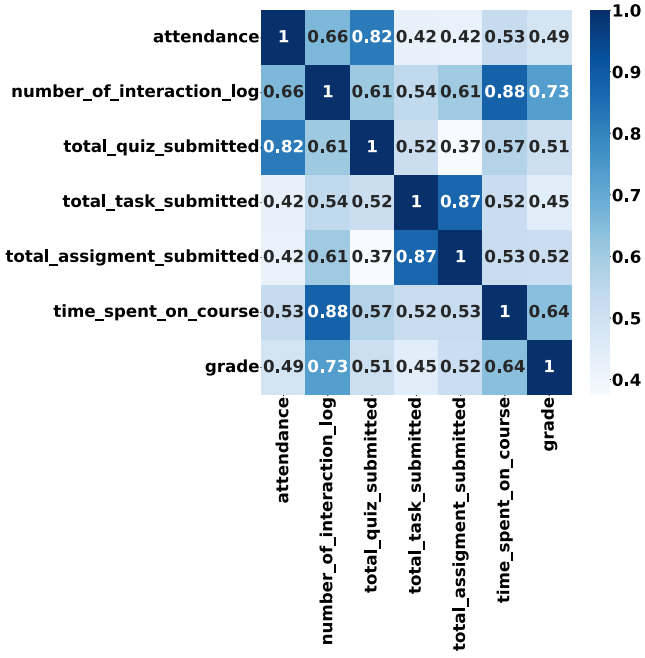


Table 3
Regression results

Hypothesis	*t*-statistic	*P*-value	CI (95%)
H1	16.782	0.000	[2.269, 2.870]
H2	32.168	0.000	[5.133, 5.800]
H3	24.816	0.000	[3.899, 4.569]

Hypothesis 1: According to the regression results in Table 3, the *P*-value for the independent variable attendance with the dependent variable grade is 0.000. It indicates a statistically significant relationship between attendance and SP. On top of that, the confidence interval of [3.899, 4.569] suggests that the effect of attendance on the total quiz submitted is positive and significant.

Hypothesis 2: Based on the regression results in Table 3, the *P*-value for the independent variable number of interaction log with the dependent variable grade is 0.000, indicating a statistically significant relationship between the two variables. Furthermore, the confidence interval of [5.133, 5.800] shows that the effect of number of interaction log on grade is positive and significant.

Hypothesis 3: The *P*-value for the independent variable time spent on course with the dependent variable grade is 0.000, showing a statistically significant relationship between the two variables. However, the confidence interval of [3.899, 4.569] indicates that the effect of time spent on the course on SP is positive and significant.

In conclusion, all three hypotheses are supported by the regression results, with all independent variables showing a statistically significant impact on their respective dependent variables. Our findings suggest that the utilization of these key attributes and the prediction algorithm can assist teachers at CADT in assessing SP and identifying those at risk of not performing well.

5.4. Survey analysis

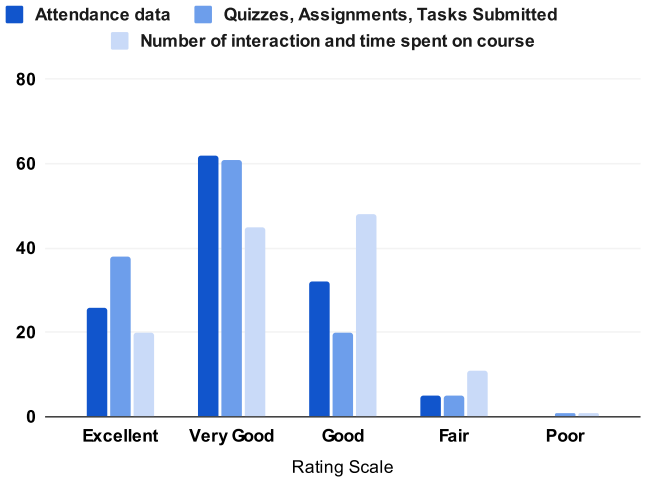
In December 2023, one year after the initial study, we conducted a follow-up survey with the same cohort of 160 students from CADT. The survey, conducted over seven days, successfully reached 125 participants. Although our predictive model, based on a trained classifier, had already identified key indicators for predicting student performance, this survey was essential as it allowed us to validate our model's findings with real-world student feedback, providing a practical perspective on the theoretical predictions. By focusing on quantitative data analysis, we could compare the students' perceptions with the statistical significance of the identified indicators.

Descriptive statistics were used to summarize and interpret the survey responses, offering a clear overview of student engagement and performance. The survey results provided substantial evidence to confirm our hypotheses, reinforcing the effectiveness of the identified indicators. This dual-validation approach not only affirmed the robustness of our predictive model but also demonstrated its practical implications in a real educational setting.

5.4.1. Usefulness indicators

The survey allowed us to gain valuable insights into the effectiveness of the predictive indicators we previously identified. Focusing exclusively on quantitative data analysis, we employed descriptive statistics to summarize and interpret the survey responses, providing a comprehensive overview of the students' engagement and performance.

Figure 10
Usefulness indicators



The survey asked students to rate the importance of various indicators in predicting their academic success. These indicators were grouped into three categories as shown in Figure 10, mirroring the attributes for which we had previously defined coefficients based on our trained classifier:

1) **Attendance Data:** This was highly rated by the majority of students, reinforcing its significant coefficient value in our model and highlighting its critical role in predicting academic performance.
2) **Number of Quizzes/Assignments/Tasks Submitted:** This group was viewed as a vital indicator of student performance, aligning perfectly with its substantial predictive weight in our model. The high rating confirms that frequent submissions are a clear sign of student engagement and progress.
3) **Number of Interaction Logs to LMS/Time Spent on Course:** Students rated this category as highly important, making it a significant predictor in our model. The correlation between time spent on the LMS and academic success is evident and strongly supported by the survey results.

The graphical representation of these results showcased a clear alignment between the students' perceptions and our model's coefficients, providing compelling evidence of the robustness and accuracy of our predictive framework. The consistency between the students' views and our analytical model underscores the practical relevance and reliability of the identified indicators in forecasting student performance.

5.4.2. Verification of hypotheses through survey results

The survey questions were carefully designed to confirm the following hypotheses, as shown in Table 4:

In the previous section, we used regression analysis to test and confirm our hypotheses regarding the relationship between various student engagement indicators and their academic performance (SP). The findings shown in Table 4 enable us to further confirm these hypotheses.

The first question illustrates the effectiveness of attendance data on academic performance. The survey results strongly support Hypothesis 1, showing a significant correlation between consistent attendance and higher academic performance. They indicate that students who regularly attended and completed modules tended to achieve better final grades. This reinforces our regression analysis findings, confirming that attendance is critical to predicting academic success.

Table 4
Hypotheses, survey questions, and response rates

Hypothesis	Survey question	Number of responses	
H1	Q1: How important do you think attendance is in predicting your academic performance?	Excellent	26
		Very good	62
		Good	32
		Fair	5
		Poor	0
H2	Q2: How significant do you find your interactions with the LMS influence your grades?	Excellent	20
		Very good	45
		Good	48
		Fair	11
		Poor	1
H3	Q3: In your opinion, how important are the quizzes and tasks you submit in predicting your academic performance?	Excellent	38
		Very good	61
		Good	20
		Fair	5
		Poor	1

The second question focuses on the effectiveness of the number of interactions with the LMS on grade. This survey results confirms Hypothesis 2. The survey data indicate that students who frequently interacted with the LMS exhibited improved academic performance.

The third question examines the effectiveness of quizzes and tasks on academic outcome. The positive impact of these factors aligns with our regression analysis results, validating that higher levels of engagement and active participation in quizzes and tasks significantly enhance students' overall academic outcomes.

To sum up, these survey results provide extra evidence supporting our hypotheses and demonstrate the practical implications of our findings. The alignment between the regression analysis and survey data underscores the importance of the attributes previously identified as key predictors of student performance in Moodle. Indeed, the empirical validation we provided here can contribute to aspects beyond SP prediction. For instance, by understanding student performance and obtaining pertinent data for the analysis process, educators can adapt and personalize interventions and support strategies for a more targeted approach to student success. We hope that the integration of these findings into educational settings, where data-driven decision-making is a powerful tool, will transform our teaching practices, leading to better academic performance and well-being for students.

6. Discussion, Conclusion, and Future Work

6.1. Discussion and conclusion

Our research aimed to assist teachers in identifying key attributes and choosing prediction algorithms to evaluate students' performance. Following a rigorous literature review, our research effort also included an empirical study. Indeed, data from an authentic learning situation at CADT have been used in our research effort to gain a better understanding of predictive analytics in education, which has become increasingly important.

Through our analysis, we demonstrated that two key factors, student attendance and interaction with the LMS, statistically impact student performance. Specifically:

1) **Student Attendance:** Consistent attendance was strongly correlated with higher academic performance, measured by the number of modules completed. This finding underscores the importance of encouraging regular participation in online learning activities.

2) **Student Interaction with the LMS:** The extent of engagement, as indicated by the number of interaction logs, also significantly impacted academic outcomes. This suggests that fostering an interactive and engaging online learning environment is crucial for enhancing student success.

In addition, through our analysis, we demonstrated the identified key attributes, as listed in Section 4.2.1, have a statistically significant impact on student performance. We also determined that random forest models are effective in predicting student performance. To further validate these findings, we conducted a follow-up survey with the same cohort of students. The survey results not only confirmed the importance of the attributes identified and trained in our model but also reinforced our hypotheses, providing a double confirmation of our predictive framework's robustness and practical applicability.

The major contribution of our research can be summarized in two aspects. First, whereas previous studies have primarily emphasized the identification of factors influencing student outcomes, our research goal is to explore and explain the impact of the identified key attributes and their complex relationships. Second, the initial iteration of our predictive algorithm is designed for both course and level programs, while existing approaches from our literature review mainly focused on the course level for predicting students' outcomes. The first data analysis batches helped us not only determine what and how important attributes are in predicting student performance but also understand in which areas we can improve teaching practices and support students from predicting their academic performance.

It is also essential to acknowledge the advantages, disadvantages, and potential applications of our research findings:

The advantages of our research are notable and multifaceted:

1) **Improved Student Support:** By accurately predicting student performance, educators can intervene early to provide targeted support and resources to students at risk of underperforming.
2) **Data-Driven Decision Making:** The use of predictive analytics enables data-driven decisions, enhancing the effectiveness of educational strategies and policies.
3) **Scalability:** The model can be applied to large datasets, making it suitable for institutions with significant student populations.
4) **Comprehensive Analysis:** The integration of various data sources, such as Moodle logs and Google Sheets, allows for a holistic view of student engagement and performance.

On the other hand, our research does have some inherent disadvantages:

1) **Data Quality:** The accuracy of predictions is highly dependent on the quality and completeness of the data. Missing or inaccurate data can impact the model's performance.
2) **Privacy Concerns:** Handling sensitive student data requires strict adherence to privacy regulations and ethical standards to protect individuals' information.
3) **Model Interpretability:** Complex models like random forest can be challenging to interpret, making it difficult for educators to understand the rationale behind predictions without additional tools like LIME or SHAP.

6.2. Future work

The potential applications of our research are extensive and promising:

1) **Early Warning Systems:** Implementing the predictive model as part of an early warning system to identify and support at-risk students before they fail.
2) **Personalized Learning:** Tailoring educational content and teaching methods to individual students based on their predicted performance and learning needs.
3) **Resource Allocation:** Optimizing the allocation of educational resources, such as tutoring services and academic counseling, to areas where they are most needed.
4) **Curriculum Development:** Informing curriculum design and development by identifying which aspects of the program are most strongly associated with student success.

However, it is essential to point out that our research has some limitations and further research and validation are needed. Current and future work in this area include:

1) Expanding the Sample Size and Validation: We are currently conducting a study that covers a bigger dataset from our partners in France. We also expect to validate the results of our study by comparing them to other studies and testing the model on new data to ensure its ability to generalize well to unseen data.
2) Implementing the Model: Our next challenge will be a real-time application. For that, we are studying the possibilities of integrating explainable AI methods like LIME or SHAP. As a matter of fact, we are interested in helping our lecturer colleagues interpret the model predictions.
3) Evaluation Metric: We acknowledge the value of using Cohen's Kappa for evaluating classifiers on imbalanced datasets. Thus, we decided to incorporate Cohen's Kappa as an additional evaluation metric, alongside the confusion matrix. Indeed, by using both evaluation approaches, we aim to provide a more robust and comprehensive assessment of the classifier's performance in predicting SP.

Overall, our research lays a foundation for advancing students' performance prediction, benefiting CADT and French partner universities and potentially impacting the wider educational community. The positive results suggest broader implications, influencing global educational practices and fostering a more data-informed and supportive learning environment.

Ethical Statement

This study does not contain any studies with human or animal subjects performed by any the authors.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data available on request from the corresponding author upon reasonable request.

Author Contribution Statement

Dynil Duch: Conceptualization, methodology, software, investigation, data curation, writing – original draft, writing – review & editing, and project administration. **Madeth May:** Conceptualization, methodology, validation, formal analysis, writing – review & editing, supervision, funding acquisition, and project administration. **Sébastien George:** Validation, writing – review & editing, supervision, project administration.

References

[1] Arifin, M., Eryani, I., & Farahtika, G. (2023). Students' perception of using Moodle as a learning management system in tertiary education. *AL-ISHLAH: Jurnal Pendidikan*, *15*, 5140–5152. https://doi.org/10.35445/alishlah.v15i4.3855

[2] Wu, J. (2024). E-learning management systems in higher education: Features of the application at a Chinese vs. European university. *Journal of the Knowledge Economy*, 1–31. https://doi.org/10.1007/s13132-024-02159-6

[3] Mustapha, A. M., Zakaria, M., Yahaya, N., Abuhassna, H., Mamman, B., Isa, A. M., & Kolo, M. A. (2023). Students' motivation and effective use of self-regulated learning on learning management system Moodle environment in higher learning institution in Nigeria. *International Journal of Information and Education Technology*, *13*, 195–202. https://doi.org/10.18178/ijiet.2023.13.1.1796

[4] Alomari, A. M. (2024). Perceptions of faculty members on using Moodle as a learning management system in distance education. *International Journal of Technology in Education and Science*, *8*, 75–110. https://doi.org/10.46328/ijtes.507

[5] Hussain, M. M., Akbar, S., Hassan, S. A., Aziz, M. W., & Urooj, F. (2024). Prediction of student's academic performance through data mining approach. *Journal of Informatics and Web Engineering*, *3*, 241–251. https://doi.org/10.33093/jiwe.2024.3.1.16

[6] Bisri, A., Supardi, S., Heryatun, Y., Hunainah, H., & Navira, A. (2025). Educational data mining model using support vector machine for student academic performance evaluation. *Journal of Education and Learning*, *19*, 478–486. https://doi.org/10.11591/edulearn.v19i1.21609

[7] Albahli, S. (2024). Efficient hyperparameter tuning for predicting student performance with Bayesian optimization. *Multimedia Tools and Applications*, *83*, 52711–52735. https://doi.org/10.1007/s11042-023-17525-w

[8] Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials Today: Proceedings*, *80*, 3782–3785. https://doi.org/10.1016/j.matpr.2021.07.382

[9] Zhang, X., Lee, V., Xu, D., Chen, J., & Obaidat, M. S. (2024). An effective learning management system for revealing student performance attributes. *arXiv Preprint: 2403.13822*.

[10] Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences*, *11*(9), 552. https://doi.org/10.3390/educsci11090552

[11] Nawang, H., Makhtar, M., & Hamzah, W. (2021). A systematic literature review on student performance predictions. *International Journal of Advanced Technology and Engineering Exploration*, *8*, 1441–1453. https://doi.org/10.19101/IJATEE.2021.874521

[12] Felix, I., Ambrósio, A. P., Lima, P. D. S., & Brancher, J. D. (2018). Data mining for student outcome prediction on Moodle: A systematic mapping. In *Brazilian Symposium on Computers in Education, 29*(1), 1393. https://doi.org/10.5753/cbie.sbie.2018.1393

[13] Namoun, A., & Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*, *11*, 237. https://doi.org/10.3390/app11010237

[14] Felix, I., Ambrosio, A., Duilio, J., & Simões, E. (2019). Predicting student outcome in Moodle. In *Proceedings of the Conference: Academic Success in Higher Education*, 14–15.

[15] Hirokawa, S. (2018). Key attribute for predicting student academic performance. In *Proceedings of the 10th International Conference on Education Technology and Computers,* 308–313. https://doi.org/10.1145/3290511.3290576

[16] Gaftandzhieva, S., Talukder, A., Gohain, N., Hussain, S., Theodorou, P., Salal, Y. K., & Doneva, R. (2022). Exploring online activities to predict the final grade of student. *Mathematics*, *10*(20), 3758. https://doi.org/10.3390/math10203758

[17] Quinn, R. J., & Gray, G. (2020). Prediction of student academic performance using Moodle data from a Further Education setting. *Irish Journal of Technology Enhanced Learning*, *5*(1), 1–19. https://doi.org/10.22554/ijtel.v5i1.57

[18] Arizmendi, C. J., Bernacki, M. L., Raković, M., Plumley, R. D., Urban, C. J., Panter, A. T., & Gates, K. M. (2022). Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. *Behavior Research Methods*, *55*, 1–29. https://doi.org/10.3758/s13428-022-01939-9

[19] Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*, 11. https://doi.org/10.1186/s40561-022-00192-z

[20] Arifin, M., Widowati, W., Farikhin, F., & Gudnanto, G. (2023). A regression model and a combination of academic and non-academic features to predict student academic performance. *TEM Journal*, *12*, 855. https://doi.org/10.18421/TEM122-31

[21] Brahim, G. B. (2022). Predicting student performance from online engagement activities using novel statistical features. *Arabian Journal for Science and Engineering*, *47*, 10225–10243. https://doi.org/10.1007/s13369-021-06548-w

[22] Luo, Y., Han, X., & Zhang, C. (2024). Prediction of learning outcomes with a machine learning algorithm based on online learning behavior data in blended courses. *Asia Pacific Education Review*, *25*, 267–285.https://doi.org/10.1007/s12564-022-09749-6

[23] Baker, R. S., Esbenshade, L., Vitale, J., & Karumbaiah, S. (2023). Using demographic data as predictor variables: A questionable choice. *Journal of Educational Data Mining*, *15*, 22–52. https://doi.org/10.5281/zenodo.7702628

[24] Kukkar, A., Mohana, R., Sharma, A., & Nayyar, A. (2023). Prediction of student academic performance based on their emotional wellbeing and interaction on various e-learning platforms. *Education and Information Technologies*, *28*, 9655–9684. https://doi.org/10.1007/s10639-022-11573-9

[25] Alhassan, A., Zafar, B., & Mueen, A. (2020). Predict students' academic performance based on their assessment grades and online activity data. *International Journal of Advanced Computer Science and Applications*, *11*(4), 185–194. https://doi.org/10.14569/IJACSA.2020.0110425

[26] Trindade, F. R., & Ferreira, D. J. (2021). Student performance prediction based on a framework of teacher's features. *International Journal for Innovation Education and Research*, *9*, 178–196. https://doi.org/10.31686/ijier.vol9.iss2.2935

[27] Ognjanovic, I., Gasevic, D., & Dawson, S. (2016). Using institutional data to predict student course selections in higher education. *The Internet and Higher Education*, *29*, 49–62. https://doi.org/10.1016/j.iheduc.2015.12.002

[28] Issah, I., Appiah, O., Appiahene, P., & Inusah, F. (2023). A systematic review of the literature on machine learning application of determining the attributes influencing academic performance. *Decision Analytics Journal*, *7*, 100204. https://doi.org/10.1016/j.dajour.2023.100204

[29] Zhang, Y., Ghandour, A., & Shestak, V. (2020). Using learning analytics to predict students performance in Moodle LMS. *The Electronic Journal of Information Systems in Developing Countries*, *79*(1), 1–13. https://doi.org/10.1002/j.1681-4835.2017.tb00577.x

[30] Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*, 318–331. https://doi.org/10.1504/IJTEL.2012.051815

[31] Aguiar, E., Ambrose, G. A. A., Chawla, N. V., Goodrich, V., & Brockman, J. (2014). Engagement vs performance: Using electronic portfolios to predict first semester engineering student persistence. *Journal of Learning Analytics*, *1*(3), 7–33. https://doi.org/10.18608/jla.2014.13.3

How to Cite: Duch, D., May, M., & George, S. (2024). Enhancing Predictive Analytics for Students' Performance in Moodle: Insight from an Empirical Study. *Journal of Data Science and Intelligent Systems*. https://doi.org/10.47852/bonviewJDSIS42023777