Students' Performance in Learning Management System: An Approach to Key Attributes Identification and Predictive Algorithm Design

Dynil Duch[1,2][a], Madeth May[1][b] and Sébastien George[1][c]
[1]LIUM, Le Mans Université, 72085 Le Mans, Cedex 9, France
[2]Institute of Digital Research & Innovation, Cambodia Academy of Digital Technology, Phnom Penh, Cambodia

Keywords: Deep Analytics, Knowledge Extraction, Student Performance Prediction, Moodle, Random Forest Classifier, Predictive Algorithm, Data Normalization, Student Engagement Activities, Student Achievement.

Abstract: The study we present in this paper explores the use of learning analytics to predict students' performance in Moodle, an online Learning Management System (LMS). The student performance, in our research context, refers to the measurable outcomes of a student's academic progress and achievement. Our research effort aims to help teachers spot and solve problems early on to increase student productivity and success rates. To achieve this main goal, our study first conducts a literature review to identify a broad range of attributes for predicting students' performance. Then, based on the identified attributes, we use an authentic learning situation, lasting a year, involving 160 students from CADT (Cambodia Academy of Digital Technology), to collect and analyze data from student engagement activities in Moodle. The collected data include attendance, interaction logs, submitted quizzes, undertaken tasks, assignments, time spent on courses, and the outcome score. The collected data is then used to train with different classifiers, thus allowing us to determine the Random Forest classifier as the most effective in predicting students' outcomes. We also propose a predictive algorithm that utilizes the coefficient values from the classifier to make predictions about students' performance. Finally, to assess the efficiency of our algorithm, we analyze the correlation between previously identified attributes and their impact on the prediction accuracy.

1 INTRODUCTION

In recent years, the use of Learning Management Systems (LMS) has grown in popularity primarily for their ability to manage and organize digital information related to teaching and learning. Educational institutions use 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 (Costaa et al. (2015)). In addition, to support teaching practices, various studies have employed data mining techniques to predict students' performance (SP) in Moodle and other LMS platforms, such as Félix et al. (2018). These studies have demonstrated that making use of 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. Therefore, 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, it is challenging to identify the crucial data attributes (hereafter referred to as "key attributes" or "attributes") from online learning activities because of the large amount of information in the LMS (Venkatachalam et al. (2011)), especially when the goal is to obtain a more accurate SP prediction (Albreiki et al. (2021)). Moreover, predicting SP and its associated issues have always been a considerable concern in education (Ramesh et al. (2013)). To state an example, both the nature and quality of data strongly impact the efficiency of a predictive approach. Then, the outcomes of using the 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 aims to explore the poten-

[a] https://orcid.org/0000-0002-7857-5811
[b] https://orcid.org/0000-0002-8527-7345
[c] https://orcid.org/0000-0003-0812-0712

Duch, D., May, M. and George, S.
Students' Performance in Learning Management System: An Approach to Key Attributes Identification and Predictive Algorithm Design.
DOI: 10.5220/0012754200003756

In *Proceedings of the 13th International Conference on Data Science, Technology and Applications (DATA 2024)*, pages 285-292
ISBN: 978-989-758-707-8; ISSN: 2184-285X

tial benefits and limitations of using predictive analytics in the educational context and identify 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, ultimately contributing to more effective teaching practices and improved student outcomes.

In the context of this research, SP refers to the measurable outcomes of a student's academic progress and achievement, particularly in relation to their use of an online LMS. It includes factors such as grades, test scores, completion of online activities and assignments, engagement with digital learning materials and resources within the LMS. It is essential to distinguish academic performance from performance in a broader sense. For example, while academic performance focuses specifically on a student's success in school or university, performance in a broader sense could refer to a range of activities or tasks in various settings. It is also crucial to note that SP is not solely focused on positive outcomes. In addition to measuring success, SP can also be used to identify specific learning activities or overall experience in using the LMS so that teachers can take appropriate actions in both assistance and evaluation tasks.

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

1. Identify the key attributes in LMS data that are relevant for predicting students' performance. Many researchers have 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 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:

1. **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.
2. **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.
3. **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 use data from the Cambodia Academy of Digital Technology (CADT) to conduct our first tests, as detailed in section 3. It is worth mentioning that the access and the use of students' data by all research partners from France are regulated. Our research complies with both Cambodian local regulations and European GDPR. All personal and sensitive information has been stripped of all data samples we use in our research.

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. We discuss our experimental results in section 4 and conclude this paper by highlighting significant areas that we have been working on since completing the tests we have conducted with our first three hypotheses.

2 RELATED WORK

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 Félix et al. (2018); Namoun and Alshanqiti (2020) systematically reviewed 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, limited explanatory power, lack of standardized evaluation metrics, and potential ethical concerns related to using sensitive student data.

In another study, Felix et al. (2019) 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 (2018) 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. (2022) 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 (Quinn and Gray (2019)), the likelihood of student's success in a course (Arizmendi et al. (2022)), or predicting student grades using both academic and non-academic factors (Yağcı (2022)). Some studies have also focused on predicting student outcomes in specific contexts, such as interaction logs (Brahim (2022)), assessment grades, and online activity data (Alhassan et al. (2020)), or based on teacher pedagogies (Trindade and Ferreira (2021)).

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. 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 3.2.1.

3 METHODOLOGICAL APPROACH

3.1 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 1. 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 1 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 at-

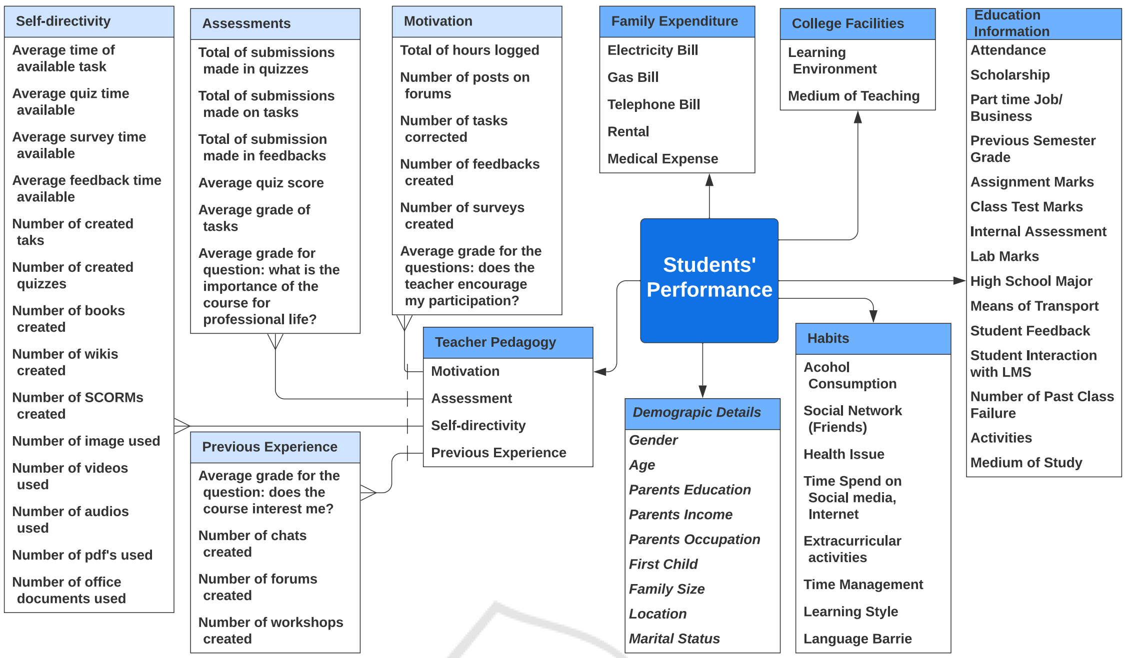

Figure 1: The attributes for predicting SP.

tributes 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 3.2.1 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.

3.2 Predictive Approach

This research uses data mining techniques and predictive algorithms within the Moodle environment. The methodology adapted for this research is divided into several steps, as shown in Figure 2.

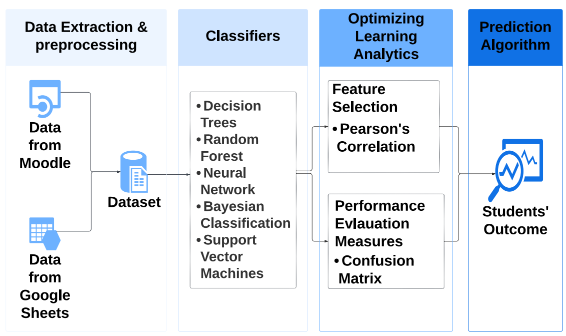

Figure 2: The predictive approach.

3.2.1 Data Extraction and Preprocessing

The dataset used in our research comprises two primary sources: the Moodle LMS and Google Sheets. The Moodle data has around 1000 students enrolled in both short course training and a bachelor program, totalling 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. 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. Afterwards, 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. 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 are attendance, number of interaction log, total quiz submitted, total task submitted, total assignment submitted, time spent on course, and outcome score.

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

3.2.3 Classifier

To predict student outcomes and select the best classifier, several classification methods are used with our dataset for comparison. As found in our studies in related work, the most commonly used data mining classifiers include Decision Trees, Random Forest, Neural Networks, Naive Bayes, and Support Vector Machine. In this research, we have made comparisons of those five classifiers with our dataset.

3.2.4 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 Aguiar et al. (2014), 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 Félix et al. (2018).

3.2.5 Student's 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.

$$sp = \sum_{i=1}^{N} C_i \frac{x_i - min_i}{max_i - min_i} \tag{1}$$

where

$$\begin{cases} sp, x, C \in \mathbb{R} \\ 0 \leq sp, C \leq 1 \\ \sum_{i=1}^{N} C_i = 1 \end{cases}$$

sp the student's performance
x_i the value of attribute i
C_i the coefficient of attribute i

max_i the maximum value of attribute i
min_i the minimum value of attribute i
N the number of attributes

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.

4 EXPERIMENTAL RESULTS

The experimental results of our research on assisting teachers in using predictive techniques to evaluate the student's performance at CADT are as follows:

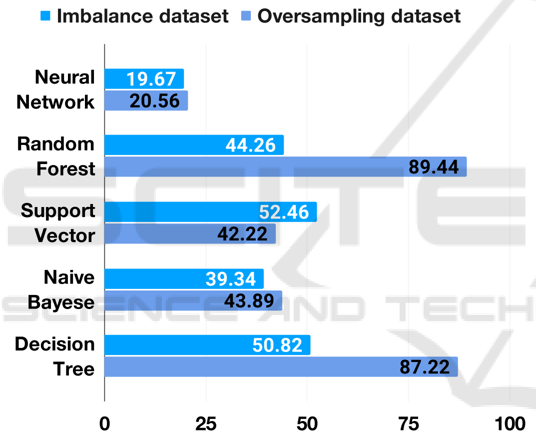

Figure 3: The accuracy of each classifier.

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 3, 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%.

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 4, 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.

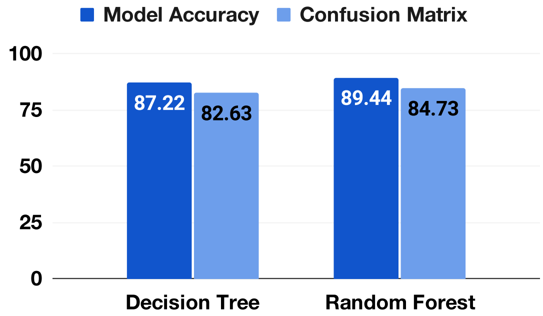

Figure 4: The comparison of classifiers accuracy of Random Forest and Decision Tree.

4.1 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 5. 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.

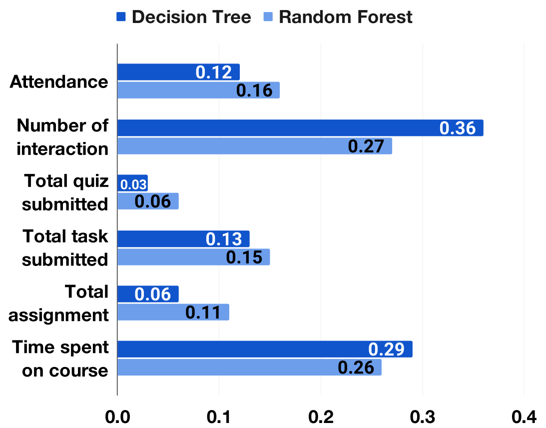

Figure 5: The coefficient of each attribute of Decision Tree and Random Forest.

4.2 Verification of Hypotheses Through Regression Analysis Results

The results we have analyzed and presented in the previous sections can serve as a valuable guide for teachers, assisting them not only in selecting the key attributes but also in understanding 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 1.

Table 1: Regression results.

Hypo.	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 1, 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 1, 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. As illustrated in Table 1, 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. 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.

5 CONCLUSION AND FUTURE WORK

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 has been used in our first attempt to gain a better understanding of predictive analytics in education, which has become increasingly important. On top of that, 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.

The major contribution of our research effort 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. 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.

REFERENCES

Aguiar, E., Chawla, N., Brockman, J. B., Ambrose, G. A., and Goodrich, V. (2014). Engagement vs performance: using electronic portfolios to predict first semester engineering student retention. *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*.

Albreiki, B., Zaki, N., and Alashwal, H. (2021). A systematic literature review of student' performance prediction using machine learning techniques. *Education Sciences*.

Alhassan, A., Zafar, B. A., and 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.

Arizmendi, C. J., Bernacki, M. L., Raković, M., Plumley, R., Urban, C. J., Panter, A., Greene, J. A., and Gates, K. M. (2022). Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. *Behavior Research Methods*, 55:3026 – 3054.

Brahim, G. B. (2022). Predicting student performance from online engagement activities using novel statistical features. *Arabian Journal for Science and Engineering*, 47:10225 – 10243.

Costaa, C., Alvelosa, H., and Teixeiraa, L. (2015). Centeris 2012-conference on enterprise information systems the use of moodle e-learning platform : a study in a portuguese university.

Felix, I., Ambrosio, A., Duilio, J., and Simões, E. (2019). Predicting student outcome in moodle. In *Proceedings of the Conference: Academic Success in Higher Education, Porto, Portugal*, pages 14–15.

Félix, I. M., Ambrosio, A. P., da Silva Neves Lima, P., and Brancher, J. D. (2018). Data mining for student outcome prediction on moodle: a systematic mapping. *Anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE 2018)*.

Gaftandzhieva, S., Talukder, A., Gohain, N., Hussain, S., Theodorou, P., Salal, Y. K., and Doneva, R. (2022). Exploring online activities to predict the final grade of student. *Mathematics*.

Hirokawa, S. (2018). Key attribute for predicting student academic performance. In *International Conference on Education Technology and Computer*.

Namoun, A. and Alshanqiti, A. (2020). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences*.

Quinn, R. J. and Gray, G. (2019). Prediction of student academic performance using moodle data from a further education setting. *Irish Journal of Technology Enhanced Learning*.

Ramesh, V., Parkavi, P., Ramar, K., Thai-Nghe, N., Busche, A., Schmidt-Thieme, L., Carol, Bastin, P., Thiyagaraj, Yosuva, S., and Arulkumar, V. (2013). Predicting student performance: A statistical and data mining approach. *International Journal of Computer Applications*, 63:35–39.

Trindade, F. R. and Ferreira, D. J. (2021). Student performance prediction based on a framework of teacher's features. *International Journal for Innovation Education and Research*.

Venkatachalam, S. T., Doraisamy, S. C., Golzari, S., Norowi, N. B. M., Sulaiman, M. N. B., Liu, H., Motoda, H., Setiono, R., Zharo, Z., and Maben, A. F. (2011). Identifying key performance indicators and predicting the result from student data. *International Journal of Computer Applications*, 25:45–48.

Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9.