10th International Conference on Higher Education Advances (HEAd'24)
Universitat Politècnica de València, València, 2024
DOI: https://doi.org/10.4995/HEAd24.2024.17343

Time management and absenteeism: studying the students through machine learning

Santiago Porras[1], Athénaïs Sauvée[2], Julio César Puche[1], Silvia Casado[1], Paula Antón[1], Joaquín Antonio Pacheco[1]
[1]Department of Applied Economics, University of Burgos, Spain, [2]Department of Private Law, University of Burgos, Spain.

How to cite: Porras, S.; Sauvée, A.; Puche, J. C.; Casado, S.; Antón, P.; Pacheco, J. A. 2024. Time management and absenteeism: studying the students through machine learning. In: 10th International Conference on Higher Education Advances (HEAd'24). Valencia, 18-21 June 2024. https://doi.org/10.4995/HEAd24.2024.17343

Abstract
Absenteeism in higher education is a problem that may involve institutional, economic, social, and individual consequences. The present work aims to analyse whether the way students manage their personal time could be an explanation for absenteeism rates. Authors used machine learning based methodology, combined with explainable artificial intelligence methods. This allowed them to design a two-levels analysis, it is to say from a global, and an individual perspective. Factors such as repeating a course have the most negative impact over class attendance. On the contrary, being able to submit an assignment before the deadline has the most positive impact over class attendance. The kind of academic career, the place of living or the hobbies has also influence over the absenteeism.

Keywords: *Absenteeism; Higher education; Support vector machine, Explainable artificial intelligence, Shapley additive explanation, Time management.*

1. Introduction

Absent is always in the wrong, he who is. Absenteeism is often linked to poor academic performance and high drop-out rates. This has consequences in terms of wasted time and economic resources, as well as the need for continuous review of education policies in both schools and higher education institutions.(Abadzi, 2012; Keppens, 2023; Miller, 2023).

To address absenteeism, it is crucial to identify its underlying factors. Sahin et al. (2016) identify five socio-economic sources: familial factors, administrative and teacher-related factors, factors related to the organization of the school, the student's own factors, and external environmental factors such as transportation. At the individual level, the main causes of absence are common health problems, early class schedules, stress, and depression (Şahín 2023). Other factors that may contribute to absenteeism include the student's organization, teaching and learning methodologies, course characteristics, and external factors (Triado-Ivern et al. 2020). Ramchander (2017) explains that homework load is a factor that can contribute to absenteeism. Students may be absent to prepare for tests, assignments, or group work, particularly on the days leading up to a test or on the day of the test itself.

It should be noted that the above is not an exhaustive list of the factors contributing to absenteeism. Other causes of absenteeism require further investigation. This paper analyses one factor related to students: their time management and how the latter affects their absenteeism. Other variables will also be considered within the study. To do so, we begin with the previous findings of Porras et al. (2023), who examined time management among first-year students of *“Administración y dirección de empresas”* (ADE) and *“Finanzas y contabilidad'* (FICO) degrees at a Spanish university. The study identified four explanatory factors. The current study includes additional academic and personal variables of the students, such as extracurricular activities, maintaining a positive mood, and exercising.

The objective of this work is to utilize machine learning techniques to identify the regression model that best fits the data and analyze it using explainable artificial intelligence (XAI) techniques. This will enable us to quantify the weight of the variables studied in the final model and provide a global explanation of student absenteeism. Then, we can identify the factors that positively influence attendance and those that negatively influence absenteeism. The use of XAI models enables analysis at the individual student level, facilitating personalized guidance.

2. Methodology

2.1. Dataset

Porras et al. (2023) identified four factors related to students' time management: their perception of their own time management, the amount of time spent on unproductive tasks, their ability to complete tasks with a deadline, and their use of tools to help with time management and organization. Variables in Table 1 have been added to the above to expand the set and identify additional factors that may influence absenteeism. The selected variables were adapted from Neill, J.T.'s (2016) original questionnaire to gather background information on students, including their academic background, other activities, living environment, and physical and mental health.

2.2. Regression analysis

After combining the variables into a single dataset, the data is split into training (75%) and test (25%) subsets. A regression analysis is then conducted, with the number of classes attended by the student as the dependent variable and the remaining variables as independent variables. The regression is performed using the Caret. package in R.(Kuhn et al., 2020). The following regression methods are employed: Random Forest, Stochastic Gradient Boosting, Neural Network, eXtreme Gradient Boosting, Partial Least Squares, Support Vector Machines with Linear Kernel, The lasso, Bagged CART and Cubist. All methods were cross-validated with 10 folds and default parameters. The root mean square error (RMSE) was chosen as the metric to assess model accuracy.

Table 1 Variables added to the study. Source: Own Elaboration

Variable	Values
Repeater	New student (1) or Repeater student (2)
Genre	Male (1) or Female (2)
Degree	ADE (1) or FICO (2)
Baccalaureate	None (1), General (2), Humanities and social sciences (3), Arts (4) ,Science and tecnology (5)
Scholarship	No (1) or Yes (2)
Residence	Family home (1), On its own (2), Student residence (3) ,Shared accommodation (4)
Academy	Attend an academy (2) or not (1).
Work	Working (2) or not working (1).
Sport	Participate in sports (2) or not (1)
Languages	Study languages (2) or not (1)
Cultural	Participate in cultural activities (2) or not (1)
Other	Perform other activities (2) or not (1)
None	Do any activity (2) or not (1)
Fitness	Very Low (1), Low(2), Normal(3), High (4),Very High (5)
Optimism	Very Low (1), Low(2), Normal(3), High (4),Very High (5)

2.3. Explainable artificial intelligence.

Machine learning and artificial intelligence models are often treated as 'black boxes', where input variables are introduced, and a result is obtained without knowledge of the process in between. This lack of transparency, explanation, and interpretability can be addressed using explainable artificial intelligence models. In recent years, XAI models have been developed rapidly, providing insight into these gaps (Vilone & Longo, 2021). Angelov et al. (2021) identified six XAI methods: feature-oriented models, global models, concept models, surrogate models, local pixel-based methods, and human-centric methods. This work will use the feature oriented SHapley Additive exPlanation model (SHAP), introduced by Lundberg and Lee (2017), which employs coalition game theory to calculate the contribution of each feature to the final model. The main concept behind SHAP is to compute Shapley values for each variable in the set to be interpreted. Each Shapley value represents the impact that the associated variable has on the prediction.

To achieve this, we will use the DALEX library (Law Biecek, 2018) in R, which implements various XAI models, and the shapviz library (Mayer, 2024) to visualise the outcomes. This will enable us to analyze both the complete model and individual instances, in this case, students.

3. Results

Table 2 shows the RMSEs obtained by the different regression methods with the test set. The methods produced similar results, with the Support Vector Machine with Linear Kernel, Random Forest and Bagged CART trees performing the best.

Table 2 RMSE of used methods. Source own elaboration.

Method	RMSE
Random forest	4.917
Stochastic Gradient Boosting	5.144
Neural Network	5.602
eXtreme Gradient Boosting	6.032
Partial Least Squares	5.097
Support Vector Machines with Linear Kernel	4.869
Lasso	5.079
Bagged CART	4.980
Cubist	5.165

Figure 1 illustrates the three variables that carry the most weight in explaining absenteeism across all three methods: whether the student is a repeater or not, factor 3 (which relates to the ability to complete assignments on time), and grade. The remaining variables vary across the methods, but the influence of time management factors and the performance of other activities can be observed.

It is important to determine whether the variables have a positive or negative impact on class attendance. To achieve this, we analyzed the values obtained by applying SHAP in the best model, Support Vector Machine Linear Kernel, as shown in Figure 2. Each point in the figure represents a student, and the color spectrum represents the original value of each variable.

The data shows that being a new student increases the likelihood of attending class, while repeating a year drastically decreases attendance. Additionally, students who are better able to complete assignments on time (Factor 3) are more likely to attend. The grade level indicates that FICO students are more likely to attend class than ADE students. The type of residence appears to have an impact on attendance rates, with students living at home more likely to attend classes than those living in halls of residence, and those more likely to attend than those in shared accommodation. Additionally, there is a gender disparity in attendance rates, with males attending more frequently than females. In the sample, it is evident that only a small number of

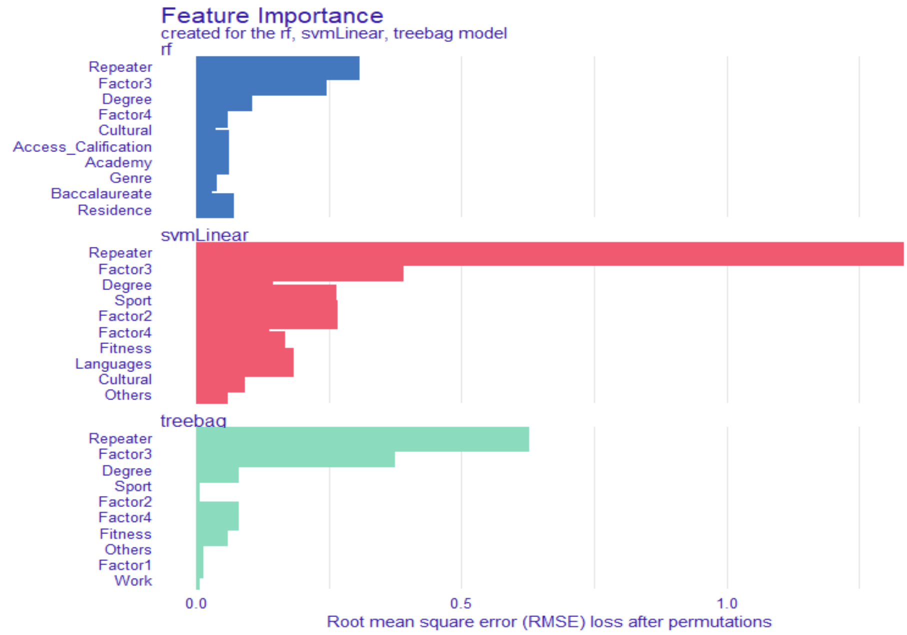

Figure 1 Top 10 variables for each model. Source own elaboration

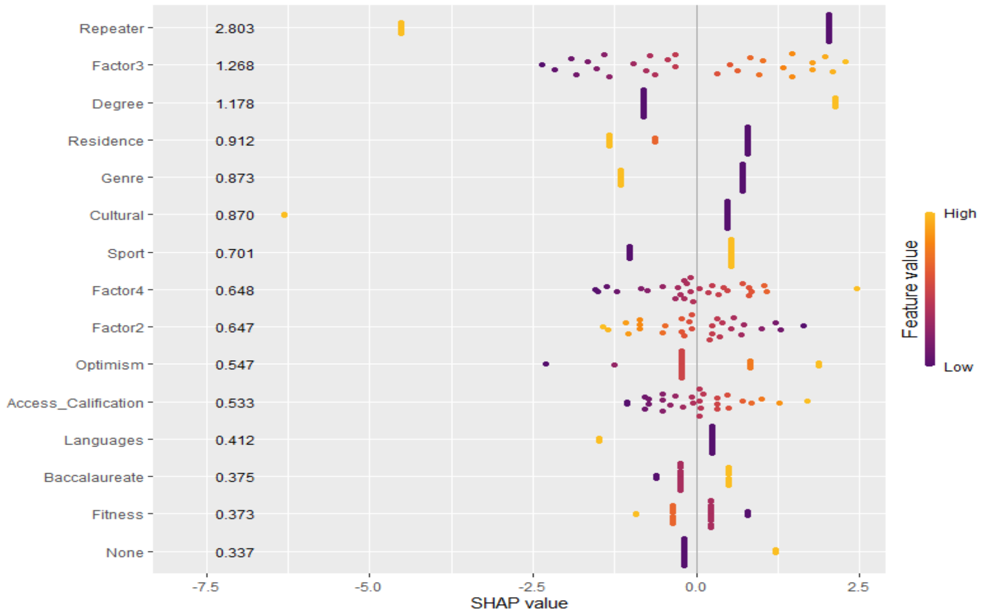

Figure 2 SHAP values for SVMLinear. Source own elaboration.

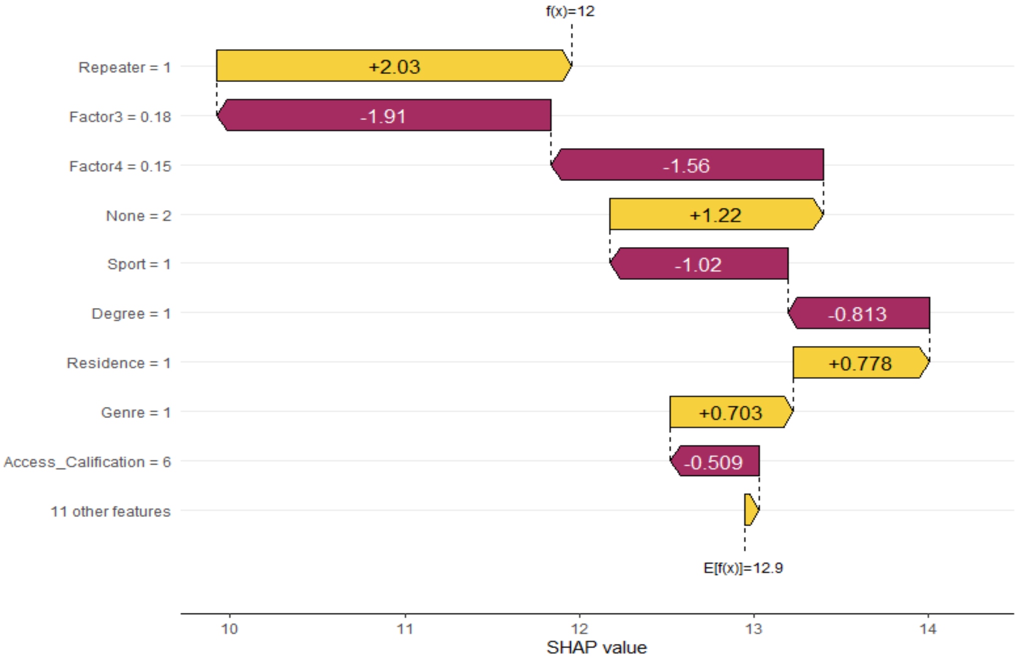

Figure 3 SHAP values for one student. Source Own Elaboration.

students participate in cultural or language activities. However, these activities do have an impact on attendance, as those who do not participate are more likely to miss classes. Unlike other activities, practicing sports has a positive effect on attendance. However, this contradicts the fact that students who attend classes are more likely to consider themselves in poor or normal physical shape. Regarding time management, it can be concluded that using time management tools (Factor 4) and avoiding unproductive activities (Factor2) positively affects attendance. Students who have achieved higher grades in previous stages are more likely to attend class, while those with lower grades are less likely to attend. Attendance is also influenced by mood, with higher levels of optimism increasing the likelihood of attendance.

The practical implications of the proposed methodology's ability to analyse individual factors affecting a student's attendance are significant. It allows for focused work on the student's weak points, optimising available resources and achieving results in a shorter period. This results in economic savings and enables students to better manage their time in the short term. This leads to a more efficient use of academic resources and improved academic performance, potentially reducing the dropout rate. Figure 3 displays the profile of a new student in ADE with an entry mark of 6. The student does not use time management tools and struggles to complete homework on time. Additionally, they live in the parental household and do not participate in any extracurricular activities, resulting in an attendance range of 10 to 14 classes.

4. Conclusions

This paper analyses absenteeism in higher education from the perspective of time management and other variables.

A machine learning methodology has been presented that allows for the analysis of the factors involved in absenteeism at two levels. At a global level, this enables high-level decision-makers such as center directors or deans to make informed decisions about measures to fight absenteeism. At the student level, personalized monitoring can identify areas for improvement to prevent absenteeism.

The analysis of data using this methodology concludes that absenteeism cannot be solely explained by time management. This study showed that the three most influential factors affecting attendance are repetition of a course, the ability to complete assignments on time, and the field of study. Family factors, such as place of residence, practice of other activities, or optimism, have a relatively minor influence on the decision to attend classes.

Speaking of time management, all factors except for the student's perception of their own time management have been shown to affect attendance. The ability to complete tasks on time is the most significant factor in time management.

Future research will analyze the impact of time management and attendance on academic performance.

Acknowledgements

This work was supported by research grants provided by the University of Burgos and IFIE in the project “Análisis de la gestión del tiempo en estudiantes universitarios: mejora en asistencia a clase y calificaciones a través de machine learning y gamificación.” under the convocatory: “Convocatoria de Ayudas a Grupos de Innovación Docente reconocidos para la elaboración de materiales docentes para los años 2023 y 2024”.

References

Abadzi, H. (2012). *Absenteeism and beyond: Instructional time loss and consequences* (Vol. 4376). World Bank Publications.

Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(5), e1424. https://doi.org/10.1002/WIDM.1424

Keppens, G. (2023). *School absenteeism and academic achievement: Does the timing of the absence matter?* https://doi.org/10.1016/j.learninstruc.2023.101769

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., & Team, R. C. (2020). Package ‘caret.’ *The R Journal*, *223*(7).

Law Biecek, P. (2018). DALEX: Explainers for Complex Predictive Models in R. *Journal of Machine Learning Research*, *19*, 1–5. https://pbiecek.github.io/DALEX.

Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Neill, J. T. (2016). The University Student Satisfaction and Time Management Questionnaire v.9.Retrievedfrom https://en.wikiversity.org/wiki/Survey_research_and_design_in_psychology/Assessment/Survey/TUSSTMQ9

Mayer, M. (2024). *shapviz: SHAP Visualizations*. R R package version 0.9.4, https://modeloriented.github.io/shapviz/.

Miller, A. P. (2023). *Attending to Student Success: Understanding the Antecedents, Realities, and Consequences of Absenteeism in Higher Education*. Rowman & Littlefield.

Porras, S., Puche, J. C., Casado, S., Sauvée, A., Antón, P., & Pacheco, J. A. (2023). Relationship between time management and class attendance in university students: clustering techniques for detection of profiles. *HEAd 2023 - 9th International Conference on Higher Education Advances*, 851–859.

Ramchander, M. (2017). Contextual factors influencing student absenteeism at a higher education institution in South Africa. *Africa Education Review*, *14*(2), 1–14.

Şahín, M. (2023). A Study on Absenteeism of University Students. *International E-Journal of Educational Studies*, *7*(13), 75–88. https://doi.org/10.31458/iejes.1223043

Triado-Ivern, X., Aparicio-Chueca, P., Elasri-Ejjaberi, A., Maestro-Yarza, I., Bernardo, M., & Presas Maynegre, P. (2020). A factorial structure of university absenteeism in higher education: A student perspective. *Innovations in Education and Teaching International*, *57*(2), 136–147. https://doi.org/10.1080/14703297.2018.1538896

Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, *76*, 89–106. https://doi.org/10.1016/j.inffus.2021.05.009