



Study regarding the influence of a student's personality and an LMS usage profile on learning performance using machine learning techniques

Juan Ramón Rico-Juan¹ · Cristina Cachero¹ · Hermenegilda Macià²

Accepted: 20 April 2024 / Published online: 8 May 2024
© The Author(s) 2024

Abstract

Academic performance (AP) is crucial for lifelong success. Unfortunately, many students fail to meet expected academic benchmarks, leading to altered career paths or university dropouts. This issue is particularly pronounced in the early stages of higher education, highlighting the need for the instructors of these foundational courses to have access to simple yet effective tools for the early identification of students at high risk of academic failure. In this study, we propose a streamlined conceptual model inspired by the Model of Human Behavior (MHB) to which we have incorporated two dimensions: capacity and willingness. These dimensions are assessed through the definition of three variables: Prior Academic Performance (PAP), Personality and Academic Engagement, whose measurements can easily be obtained by the instructors. Furthermore, we outline a Machine Learning (ML) process that higher education instructors can use to create their own tailored models in order to predict AP and identify risk groups with high levels of transparency and interpretability. The application of our approach to a sample of 322 Spanish undergraduates studying two mathematical subjects at a Spanish university demonstrates its potential to detect failure early in the semester with a precision that is comparable with that of more complex models found in literature. Our tailored model identified that capacity was the primary predictor of AP, with a gain-to-baseline improvement of 21%, and the willingness variables increasing this to 27%. This approach is consistent over time. Implications for instructors are discussed and an open prediction and analysis tool is developed.

Keywords Student dropout · Academic performance · Observational study · eXplainable Artificial Intelligence (XAI) · Machine Learning (ML)

1 Introduction

Academic performance (AP) is widely recognized as being crucial for lifelong success. Students who excel academically tend to enjoy better health, overall well-being, and higher

salaries [62]. Moreover, AP is a proven critical indicator of university dropout rates [63, 79], thus posing a significant challenge for modern economies. The demands for individuals with advanced skills is escalating in order to satisfy labor market requirements and foster productive and equitable environments [88], and high dropout rates are a formidable barrier to these objectives. At a worldwide level, a mere 38% of students enrolled on Bachelor's degree programs complete their studies within the designated time frame [62]. Although the situation in Spain is less dire, it is still troubling, with 27.4% of students at public universities discontinuing their studies in order to either change majors (12.5%) or exit the Spanish university system entirely (14.9%). These percentages rise to 37.1% in the fields of Engineering and Architecture, with 18.3% leaving the Spanish university system entirely) [11]. It is particularly noteworthy that the highest dropout rates occur in the first year of study. These

✉ Hermenegilda Macià
hermenegilda.macia@uclm.es

Juan Ramón Rico-Juan
juanramonrico@ua.es

Cristina Cachero
ccachero@dlsi.ua.es

¹ Department of Software and Computing Systems, University of Alicante, 03690 Alicante, Spain

² Higher Technical School of Computer Engineering of Albacete, Universidad de Castilla-La Mancha, Campus Universitario s/n 02071 Albacete, Spain

statistics fully justify the scientific community's considerable interest in two key areas: identifying the factors that can predict academic failure, and developing straightforward strategies based on these factors that will promote the development of flexible and interpretable models that are adaptable to various real-world educational environments [66]. These models should additionally enable the early identification of at-risk student groups so as to facilitate the timely execution of academic interventions [66].

With regard to the factors linked to academic performance, various authors argue that predicting academic dropout necessitates a blend of sociodemographic, economic and academic elements [4, 13, 53, 63, 79]. Achieving this requires cooperation between higher education instructors and institutional data system administrators, a process that can be quite burdensome for instructors.

Some researchers therefore advocate predicting academic performance by evaluating factors that are more readily accessible to academic instructors. The relevant literature emphasizes that cognitive factors such as cognitive abilities and self-regulated learning strategies are crucial [38, 51, 69, 81]. However, recent studies suggest that although cognitive factors account for approximately 25% of the variance in academic performance, the relationship between cognitive abilities and academic success is also shaped by various non-cognitive factors [69]. These include personality, emotional intelligence, motivation, academic engagement, self-efficacy, and grit [1, 3, 6, 28, 46, 58, 72, 92, 93]. A significant challenge regarding some methods in this category is their great dependence on subjective or reported measures [65], along with the use of extensive questionnaires [12, 23], which may hamper the reliability of the measurements. Moreover, certain proposed models incorporate a wide range of variables [74], which not only complicates the measurement process but also heightens the risk of overfitting, particularly when combined with machine learning techniques and relatively small sample sizes (around 100) [40].

With regard to facilitating streamlined processes for the early identification of student groups that are at risk, the use of machine learning (ML) techniques to tackle academic underperformance by considering both cognitive and non-cognitive factors has undergone an increase in popularity in recent years. This trend is attributed to the acknowledgment that the connection between these factors and academic performance (AP) is inherently non-linear [69]. The application of ML has consequently advanced educational research by providing considerable insights into these complex interactions. However, integrating ML models into the university setting can pose significant challenges for instructors (see e.g. [19]), who are required to navigate the application of relatively advanced ML techniques, often within the confines of restricted knowledge, time and resources.

The aim of this paper is twofold. Firstly, we introduce and showcase the potential effectiveness of a concise model that incorporates three key variables: previous academic performance, personality and academic engagement, in order to predict academic success. These variables cover both the cognitive and non-cognitive perspectives, which are, according to current literature, considered essential. The related eight measures (or nine, should the instructor aim to develop a model that is applicable to multiple subjects simultaneously) can be swiftly gathered (within 30 minutes) through the use of a concise questionnaire and a brief examination of the Learning Management System (LMS) log files related to the course. Secondly, we present a comprehensive ML process that will allow instructors to craft their personalized versions of the model, adapted to their unique educational contexts. The straightforwardness of the model, along with a detailed step-by-step guide to the ML process that includes explanatory techniques and cohort analysis, facilitates the detection of at-risk students and the derivation of insightful conclusions, even for instructors with minimal knowledge of ML and in scenarios involving small groups and limited resources. Furthermore, the clarity and interpretability of the resulting ML models equip educators with the means to thoroughly comprehend and trust the predictions generated.

In order to demonstrate the viability of our method, we have implemented it on two mathematical courses at a Spanish university. The findings reveal that, in this setting, our approach successfully identified at-risk groups in both subjects just one month into the semester. Equipped with this information, the educators involved in this case study now have a tool that enables them to implement targeted and early interventions so as to reduce the risk of academic failure on these courses.

The remainder of this paper is organized as follows: Section 2 introduces the conceptual model and outlines its three key constructs, while Section 3 provides a review of key empirical studies examining the effects of the model constructs on academic performance and the methodologies employed to evaluate these effects. Section 4 describes the ML process that instructors must follow so as to develop their own predictive models, and Section 5 describes the case study developed in order to illustrate the proposed approach. Section 6 provides an in-depth explanation of the implementation of the data collection phase of the ML process, while Section 7 illustrates the steps involved in identifying the optimal algorithm and ML model for any given context. Section 8 explains how to apply explainability techniques in order to improve the interpretability of the results. Section 9 explores the implications of our contributions and findings for the contemporary educational context, along with their limitations, and finally, Section 10 encapsulates the primary conclusions and proposes avenues for future research.

2 Conceptual framework

In order to empirically evaluate how differences in students' cognitive and non-cognitive profiles can assist in predicting academic performance early in the semester, we adopted the Model of Human Performance (MHP) [16]. This model categorizes the various individual differences that could potentially impact on work productivity into two different dimensions. The first dimension, Capacity, encompasses variables including work experience, educational level and cognitive abilities, while the second dimension, Willingness, includes psychological and emotional traits such as motivation, academic engagement and personality. The model also incorporates a third dimension, Opportunity, which refers to the contextual factors surrounding the work, such as materials, tools and working conditions. The hypothesis driving this research is that the same three factors that, according to the MHP, impact on productivity in work settings are also bound to affect academic productivity. This hypothesis is partly supported by the presence of these same components, albeit structured and named differently, in other theories that are frequently used for the study of academic performance, such as the theory of Self-Regulated Learning (SRL) [64].

We have identified a minimum set of variables for each dimension that, based on existing literature, show promise as regards predicting academic performance at the university level [41, 69, 74] while simultaneously being easy for instructors at university institutions such as ours to collect. The resulting conceptual model is depicted in Fig. 1, in which

the focus is on the relationship between individual differences - Capacity and Willingness - and Performance. The Capacity dimension is represented by a single component, prior academic performance (PAP), while the Willingness dimension includes two subcomponents: personality and academic engagement. Opportunity, which in the educational context refers to external factors that enable or prevent a student from achieving academic success effectively, is presumed to impact equally on all students within a specific educational setting. It has consequently not been integrated into the model as a dimension with which to explain individual differences in academic performance.

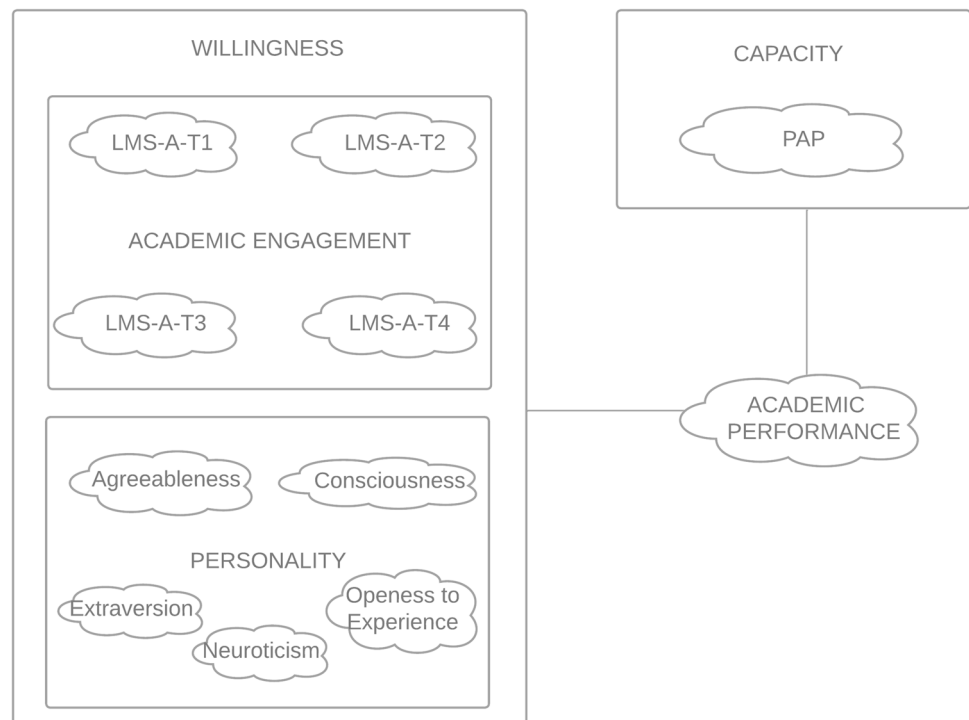
Each of the theoretical constructs involved in this conceptual model is explored in greater depth below.

2.1 Capacity

According to Reed and Jensen [73], cognitive abilities can be defined as "the skills and processes of the mind necessary to perform a task." Cognitive skills assist in the process of information gathering, analysis, understanding, processing, and storing in memory for later use in any activity.

There are four contemporary models of intelligence that have received significant attention in the literature related to assessing cognitive abilities [68]: the theory of Multiple Intelligences (MI), the theory of the three strata or the Cattell-Horn-Carroll (CHC) theory, the Successful Intelligence Theory (SIT), and the Verbal-Perceptual Rotation Theory (VPRT). All of these theories begin with a general

Fig. 1 Conceptual model: overview



concept of intelligence, which is understood to be “the potential that facilitates adaptation, learning, planning, problem-solving, abstract reasoning, decision making, understanding of complex ideas, and the creativity of individuals, which has a biological substrate and is not exclusive to the human species” [39]. While all of these theories have associated measurement instruments, their comprehensiveness makes their application challenging in the university context.

Furthermore, the Spanish University entrance system is based on a composite score that reflects PAP. This score is a combination of the student’s high school grade point average (GPA) and the college admission test score. The PAP score serves as a gatekeeper for Spanish university admissions and is acknowledged to reflect the student’s cognitive abilities beyond that which is captured by an intelligence test, owing to the self-regulatory competencies that influence the high school GPA [69]. Moreover, previous studies have demonstrated that PAP is the strongest predictor of university GPA [69, 74]. In this study we, therefore, use PAP as a proxy for capacity (see Fig. 1).

2.2 Personality

The American Psychological Association (APA) defines personality as “individual differences in characteristic patterns of thinking, feeling, and behaving”. The research community widely accepts that personality traits are useful for predicting intention, achievement and behavior [9].

The two most representative personality theories are, according to personality psychologists [5, 35], the Big Five (BF) personality model [49] and Eysenck’s Hierarchical Three Factor model (PEN) [33]. Some studies comparing both have concluded that the BF model is more comprehensive and has better measurement reliability than the PEN model [7], thus making it the dominant model in personality research [26, 29, 83] and education [14, 20, 83].

The objective of the BF model is to classify all significant sources of individual personality differences, and it includes five factors: Neuroticism (N), Extraversion (E), Agreeableness (A), Conscientiousness (C), and Openness to experience (O). N reflects the individual’s tendency to experience distress and the cognitive and behavioral styles that result from this. High N scorers experience higher levels of nervous tension, guilt, depression, self-consciousness, frustration, and ineffective coping. E reflects individual differences in positive emotionality, and high E scorers are energetic, optimistic, cheerful, enthusiastic, talkative, dominant, warm, and sociable, while low E scorers are more reserved, quiet, shy, silent, withdrawn, and retiring. A reflects individual differences in friendliness, with high A scorers being modest, caring, nurturing, emotionally supportive, altruistic, and trusting, while low A scorers tend to be more self-centered, spiteful, hostile,

jealous, and indifferent to others. C reflects individual differences in the will to achieve and impulsiveness, with high C scorers being neat, diligent, thorough, achievement-oriented, well-organized, and governed by conscience, while low C scorers are not. Finally, O refers to individual differences in intellect traits: high O scorers have aesthetic sensitivity, imagination, curiosity, intellectual interest, a need for variety, unconventional values, perceptiveness, and openness to values, sensations, feelings, and fantasies. However, individuals that attain low scores for the O factor tend to favor conservative values, repress anxiety, and judge in conventional terms. These five components are illustrated in Fig. 1.

2.3 Academic engagement

Academic engagement is characterized by a favorable, persistent, and comprehensive attitude toward work [78]. Research has consistently shown a positive correlation between academic engagement and academic achievement [69, 77].

With regard to assessing academic engagement, the increasing use of information and computing technologies (ICT) in education, and particularly the integration of Learning Management Systems (LMS), provides instructors with a rich ecosystem of data that can provide a clear picture of the quantity and quality of interaction both inside and outside the classroom, along with a sense of the students’ commitment to the subject. System usage has, in particular, been shown to be significantly related to academic performance, explaining around 20% of the variance in students’ final grades [36].

Of the myriad of LMSs available on the market, the Spanish university system employs Moodle, an open-source e-learning environment that allows instructors to gather a range of metrics with which to assess how students approach the course in blended settings. Some examples of these metrics include total time online, total sessions (accesses to the platform), average inter-session interval, the proportion of time spent on learning resources (relative to total time online), the proportion of learning resources accessed, the proportion of activities completed, number of interventions in fora, and number of messages sent to the facilitator [30].

Although any of these measures (or a combination of them) could have been used to provide a picture of student engagement, given our objective of producing a parsimonious model, we opted for a single variable with which to reflect LMS use. The choice of this variable was guided by the following criteria: (a) The variable had to be easy to collect (standardized by subject) in order to preserve the simplicity of the data gathering process; (b) the variable needed to be sufficiently broad to be applicable to various subjects and instructors, and (c) the variable had to represent platform usage, irrespective of the specific purpose. For

example, some courses may employ active learning methods such as flipped classrooms, necessitating significant access to resources at home without requiring extended interaction with the platform (in terms of time). Conversely, courses that demand student engagement in forums, wikis, or quizzes through the LMS can result in greater platform usage both within and outside the classroom. These two scenarios could yield very different time metrics, although not necessarily divergent counts of platform accesses since they are normalized by subject. We consider this to be an indicator of academic engagement that is less dependent on specific teaching strategies than are other variables, which makes it more adaptable to changes in the academic environment. Moreover, the dynamic nature of academic engagement suggests the suitability of measuring this number of accesses at different time points. Given the 4-month format of Spanish university courses, we have chosen four moments: T1 (week 4), T2 (week 8), T3 (week 12), and T4 (week 15). All of this is reflected in Fig. 1.

3 Related work

3.1 Key factors influencing academic performance: insights from meta-analyses

As mentioned previously, given the acknowledged significance of AP for lifelong success and socio-economic growth, extensive research has been conducted into the variables that may affect it, including several meta-analyses. This section reviews frequently cited studies that examine the effects of the three factors incorporated into the conceptual model illustrated in Fig. 1 (PAP, personality, and academic engagement) on AP.

In terms of capability, a meta-analysis by [74] identified a moderate correlation between high school GPA and American College Testing (ACT). Likewise, the findings of [69] indicated that PAP serves as the most significant predictor of university GPA, a conclusion further supported by other studies [27, 44, 87].

With respect to personality, a meta-analysis by [70] on the Five Factor model of personality and academic performance highlighted conscientiousness (one of the Big Five traits) as a key predictor of tertiary GPA, even after taking high school GPA into consideration. This conclusion is corroborated by [74]. Vedel [86] conducted another meta-analysis, reinforcing the view that conscientiousness is the strongest personality predictor of GPA, with a weighted summary effect of .26. The meta-analysis carried out by [69] similarly recognizes conscientiousness, academic self-efficacy and non-cognitive SRL strategies as powerful predictors of variance in AP.

A more recent meta-analysis was conducted by [60] in order to examine the combined impact of cognitive abilities and personality. This analyzed 267 independent samples ($N = 413,074$) from 228 unique studies. The results indicated that the combined effect of cognitive ability and personality traits can account for 27.8% of the variance in academic performance. Cognitive ability was found to be the most significant predictor, with a relative importance of 64%, while conscientiousness had a relative importance of 28%.

Finally, with regard to academic engagement, data from Learning Management Systems (LMS) have been shown to effectively predict student academic performance with high accuracy. Specifically, metrics such as the frequency and timing of Moodle interactions, along with the types of activities students participate in on Moodle, can be extracted from Moodle logs. These metrics have been identified as being strong predictors of academic performance, accounting for approximately 20% of the variance in student's final grades [2, 32, 36, 54, 71].

It is noteworthy that the three factors included in our conceptual framework (capacity, personality, and academic engagement) appear to maintain their predictive power over academic outcomes across years [42, 69], and also seem to be valid for different cultures [75].

3.2 Academic performance analysis techniques: from traditional statistics to machine learning

With regard to the techniques used to analyze the impact of these factors on academic performance, traditional statistical methods such as correlation analyses and simple/multiple linear regression models have been the norm for many years [36, 79]. However, the emergence of LMSs and their associated data, along with the development of advanced ML techniques, have driven the research community to broaden their understanding of the learning environment by exploring and leveraging a wide range of educational data through the use of these innovative techniques. In particular, recent studies [2, 79] have established the critical role played by ML methods in identifying students at risk and predicting dropout rates. Another recent meta-analysis [34] reports that a significant majority of studies employing ML techniques (approximately 90%) utilize supervised learning (with classification accounting for 78% and regression accounting for 12%), while a smaller fraction (10%) applies unsupervised learning. Within the realm of supervised learning, the meta-analysis identifies the most frequently used ML algorithms for these approaches, including Random Forest, Decision Tree, Support Vector Machine, Naive Bayes, Linear Regression, Logistic Regression, Artificial Neural Network (ANN), K-Nearest Neighbor, Gradient Boosted Trees, and eXtreme Gradient Boosting. Another meta-analysis by [10], which

encompasses 260 studies focused on predicting student performance within the scope of Educational Data Mining (EDM), states that the most popular data-mining algorithms are ANN and Random Forest. Nevertheless, it is crucial to acknowledge that there is no single “best” algorithm with which to predict academic performance, as different ML algorithms excel in various knowledge domains [79]. Furthermore, the effectiveness of these algorithms may vary according to the application of feature selection and whether the evaluation metrics include only precision or both precision and Kappa scores [32].

4 The machine learning process

A step-by-step illustration of our proposed ML process is provided in Fig. 2. This figure shows the three main phases of the approach: data collection, ML model generation and selection, and post-hoc explainability. By following this process, instructors are able to generate their own ML models, tailored to the specific characteristics of their educational contexts, in order to predict academic performance, identify risk groups and implement early interventions effectively.

The instructor starts by collecting data in the three key areas identified by the conceptual model (see Fig. 1). The set of variables (predictors) that must be considered for any instructor who intends to apply our approach for the training of four candidate models (Time 1, Time 2, Time 3 and Time 4, see Fig. 2) includes:

- The five personality variables (N, E, A, C, and O). In order to measure the personality constructs, the Spanish version

of the Big Five Inventory (BFI-44), which has been repeatedly validated by the research community [52], was selected. This questionnaire includes 44 items, divided as follows: N (8 items), E (8 items), A (9 items), C (9 items), and O (10 items).

- PAP: A measure of previous academic performance. A score ranging from 0 to 14, which is obtained by combining the student’s high school grade point average (GPA), the mandatory college admission test score (CATS_M), and the optional college admission test score (CATS_O). The calculation of this score is as follows: a maximum of 10 points is achievable by computing the result of $0.6 * GPA + 0.4 * CATS_M$. The additional 4 points (required in order to attain a total of 14) are obtained from the optional section of the college admission test score.
- Four pairs of LMS-A variables, corresponding to four points in time (T1, T2, T3, and T4) and a distinction between accesses inside the classroom (IA), and outside the classroom (OA). For example, LMS-OA-T3 would correspond to the number of accesses to the Moodle platform made from outside the classroom by each student at the end of week 12.
- Subject code. Moreover, if the instructors’ intention is for the model to predict academic performance across various related subjects, they should include the subject code as an additional variable so as to account for differences in difficulty in different subjects. This inclusion is crucial, because not all subjects, even those within the same domain, share the same level of difficulty, and this variance can significantly influence students’ performance scores.

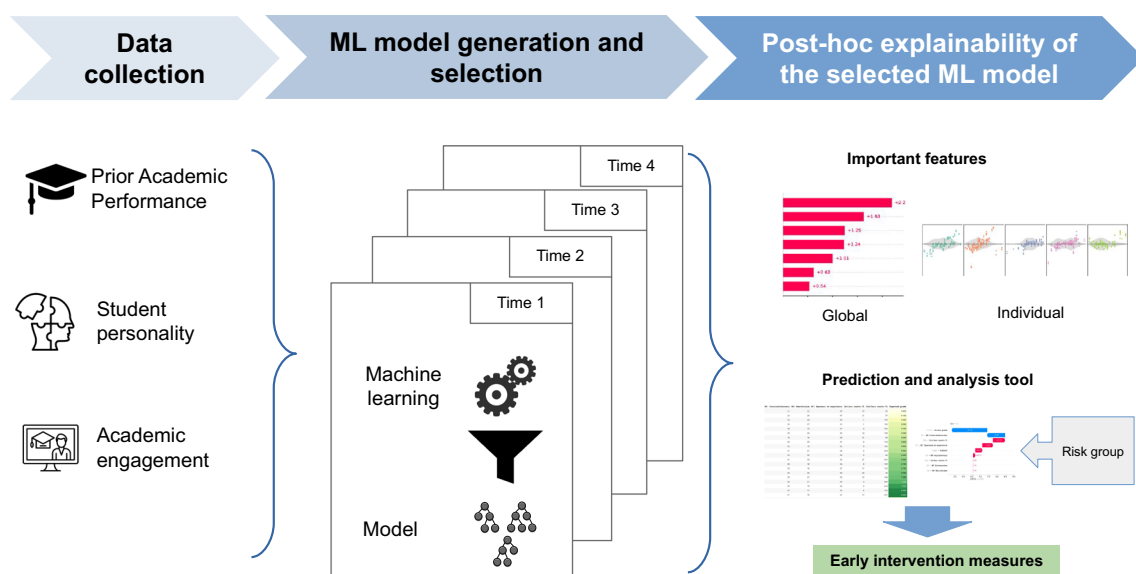


Fig. 2 Proposed ML process: an overview

Four different ML models are then trained, each corresponding to a different point in time (Time 1 through Time 4).

At each point in time, Explainable Artificial Intelligence (XAI) techniques are applied to the corresponding model in order to understand the features that influence the predictions at that point in time. This step helps demystify the decisions made by the ML model and identify which factors are most important as regards predicting academic risk in the instructor's context. The instructor can subsequently perform cohort analysis in order to detect patterns or common characteristics among groups of students at risk. This can involve clustering students on the basis of the predictions made by the model and analyzing the characteristics of these clusters. Finally, using the insights gained from the ML model and the XAI and cohort analyses as a basis, the instructor can devise and implement intervention strategies that are applicable at that point in time with the objective of mitigating the identified risks. The objective of these interventions is to support at-risk students and improve their chances of academic success.

At the end of the course, the instructor will have four potential models (T1, T2, T3, T4) with which to elucidate academic performance. It is then time for the instructor to choose the model that best aligns with the unique characteristics of their educational context. This choice is informed by both the predictive accuracy of the model and the timing of the application, thus favoring the earlier detection of at-risk students whenever feasible. After making this choice, in subsequent years, the instructors will need only to deploy the chosen model at the predetermined time to obtain a prediction of academic success that is customized to the distinctive nature of their particular course.

5 Design of case study

In order to demonstrate the feasibility of the approach proposed in this paper, we have applied it to our educational context. This case study can be classified as a cross-sectional study within the observational research category. Observational studies are a type of empirical research in which the independent variables cannot be manipulated (as occurs in experiments or quasi-experiments). They are instead observed, and the researcher draws conclusions based on those observations [15]. When compared to other types of empirical studies, cross-sectional studies have one main disadvantage: they allow neither the establishment of cause-effect relationships nor alternative explanations for the results obtained. However, despite this limitation, they are still valuable in education as they can help confirm assumptions and inform educational actions [21].

5.1 Objectives and definition of context

The main objective of this case study was to illustrate the viability of the approach proposed in this paper, and to generate an ML model and an associated tool that would be capable of predicting, with sufficient precision, students' academic performance in the context of first and second-year university students enrolled on two mathematical subjects at a Spanish university.

5.2 Research questions

We additionally sought to address the following research questions based on the case study:

- RQ1: What is the performance improvement ratio of the optimal model, generated using our methodology, when compared to the baseline for student academic performance (AP) in our case study?
- RQ2: At what point during the semester can, according to our case study findings, at-risk student groups be accurately identified in order to prevent academic failures?
- RQ3: How do individual features correlate with AP in our case study, and is this relationship consistent across the four proposed models?

5.3 Variables and measurement instruments

The objective of our case study was to discover a model that would be capable of elucidating academic performance across two related courses. We included both the eight mandatory variables (N, E, A, C, O, PAP, LMS-IA and LMS-OA) and subject code as input variables.

The outcome variable (academic performance), meanwhile, refers to the predicted final score obtained by the students on the course.

5.4 Context of case study

The objective of our case study was to collect data from 322 first- and second-year students enrolled on the Computer Engineering program on Albacete Campus of the University of Castilla-La Mancha in Spain.¹

As discussed in Section 1, this university period is particularly important because the transition from high school to university can be challenging for students in terms of developing independent learning skills, self-assessment

¹ <https://www.esiiaab.uclm.es/plan.php?que=grado&curso=2020-21&idmenup=planestudios>

Table 1 Participants in the study

Subject	#Total Enrolled	#Female Enrolled	#Male Enrolled	#Total Participants	#Female Participants	#Male Participants
CNM	167	17	150	119	12	107
Logic	155	20	135	127	14	113

abilities and meeting expectations regarding academic performance [89, 96]. It is consequently essential to identify students who are at risk of failure during this period, as the majority of dropouts occur at this time [50, 63, 79].

The selection of courses for the study was determined as follows:

- Calculus and Numerical Methods (CNM): first semester of the first year of the degree program. This included 167 students (17 female), with a median age of 18 years and fewer than 16% on their second enrollment.
- Logic: first semester of the second year of the degree program. This course had 155 students (20 female), with a median age of 19 years and fewer than 8% on their second enrollment

This selection was made because these subjects are within the same domain (mathematics) and, along with programming courses, tend to imply the greatest challenges for students on the Computer Engineering degree program, where instructors typically observe higher rates of academic failure.

6 Phase 1: data collection

In this phase, the instructor needs to gather the measures associated with the Personality, PAP and Academic Engagement variables (see Fig. 2).

In our case study, this gathering process started during the first week of the first semester of the course (week 1). During this week, the students completed the BFI-44 questionnaire (Spanish version), together with a question in which they stated their PAP score. Table 1 shows the number of students expected (total enrolled) and the number of students that participated in the study by filling in the BFI-44 questionnaire (total number of participants). We also show the same numbers disaggregated by gender. The students were

Table 2 Final grade of the participants in CNM and Logic subjects

Subject	count	mean	std	min	25%	50%	75%	max
CNM	119	6.1	2.0	1.0	4.6	6.3	7.5	10.0
Logic	127	7.0	1.8	2.3	6.0	7.2	8.2	10.0

not warned in advance that a questionnaire was going to be administered, signifying that we can safely assume that the lack of participation in the study was not related to the study itself.

The Cronbach's Alpha values calculated for the five scales on the Big Five (BF) questionnaire are the following: Extraversion: 0.807 (8 items); Agreeableness: 0.702 (9 items); Conscientiousness: 0.783 (9 items); Neuroticism: 0.829 (8 items); and Openness: 0.753 (10 items). The fact that all values fall between 0.7 and 0.9 indicates good internal consistency for the scales, without the items being redundant [82].

Data regarding the number of accesses to the LMS was then gathered by one of the instructors at four points in time during the semester: after one month (week 4), after two months (week 8), after three months (week 12) and just before the exam took place (week 15).

The students on both courses had to perform a set of evaluative activities that would enable them to obtain 45% of their grade. The remaining 55% was calculated on the basis of several partial exams. In the event of the students failing one or more partial exams, they had the opportunity to retake the course by participating in a final exam during the official examination period established by the university, before the beginning of the next semester. Both the activities and the exams (partial and final) were graded by the instructor responsible for the course on a scale of 0 to 10. In order to pass a course, a weighted final mark of 5 out of 10 is required, with at least a 4 out of 10 in the examination part.

6.1 Descriptive statistics

These data can be analyzed using various open-source software tools and libraries.²

Descriptive statistics of the final grades (the output measure), disaggregated by subject, can be seen in Table 2. Figure 3 comprises the combined violin and box plot graph that represents the distribution of these same grades disaggregated by subject. This graph demonstrates that both subjects have a wide distribution of grades, implying significant vari-

² For our case study, several Python routines were developed by one of the authors. The primary Python libraries utilized include scikit-learn v1.2.2, xgboost v1.7.4, and catboost 1.1.1, which are used in order to implement ML algorithms, the SHAP library v0.41, which are employed in order to analyze XAI-ML aspects, and plotnine v0.10.1 and ggplot2 v4.2.3 (R statistical language), which serve to construct plots.

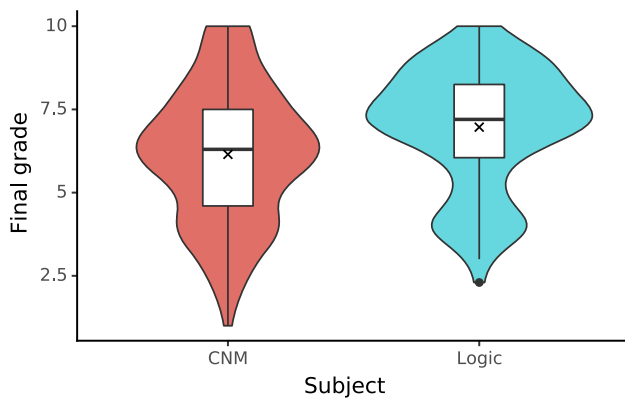


Fig. 3 Combined violin and box plot showing the distribution of final grades in CNM and Logic subjects

ability in student performance. In CNM, the mean is below the median, indicating a distribution that is slightly skewed towards lower grades. In Logic, the mean is very close to the median, suggesting a more symmetrical distribution of grades. CNM also has a longer lower tail, which implies that there are more students with lower grades. In summary, this distribution seems to suggest that the subject of Logic is easier for students, whose academic performance is generally better than in CNM.

Figures 4, 5, and 6 show the combined violin and box plot graph, illustrating the distribution of the input data required by our approach.

The shape of the violins and the position of the mean in Fig. 4 (PAP variable), suggest that there is no significant difference between the distribution of the entrance grades for CNM and Logic, although CNM has a slightly higher median. This indicates that the mean entrance grades for these two subjects are similar.

Figure 5 (Personality) additionally suggests that there are no marked differences between the personality profiles of the CNM and Logic students with respect to the five traits

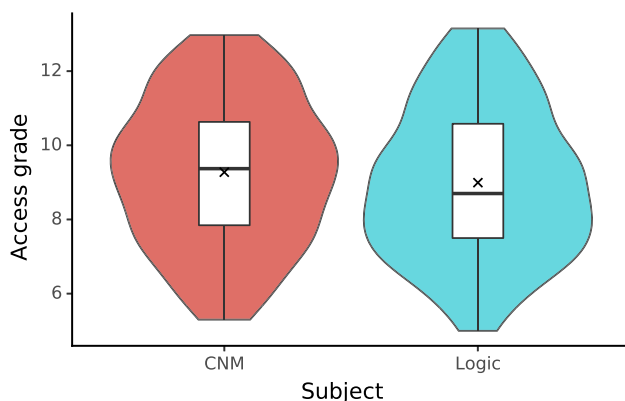


Fig. 4 Combined violin and box plot depicting the distribution of the PAP variable

evaluated. The medians and means are quite similar for each trait, which implies that the groups of students in these two subjects are alike in terms of personality.

Finally, Fig. 6 (Academic Engagement) shows the distribution of the variables measuring (a) number of accesses to the Moodle platform during in-person class hours (LMS-IA) measured at four points in time (T1 to T4) and (b) number of accesses outside these hours (LMS-OA), also recorded at the same four points in time (T1 to T4). A robust normalization (R_n) was applied in order to mitigate the impact of outliers in the distribution of academic engagement variables (Fig. 6) (1). In this formula, the functions q_1 , q_2 , and q_3 correspond to the first, second and third quartiles, respectively, according to the subject under consideration, s .

$$R_n(y, s) = \frac{y_i - q_2(y, s)}{q_3(y, s) - q_1(y, s)} \quad (1)$$

This normalization makes the measures comparable across different time horizons and subjects. As expected, there is a greater dispersion for the LMS-OA variables, especially in the Logic subject.

7 Phase 2: generation and selection of ML model

After collecting the necessary data, the instructors are now ready to employ ML techniques so as to identify the optimal algorithm and ML model tailored to their context of usage (see Fig. 2).

7.1 Selection of ML algorithm

Figure 7 provides a more detailed examination of the activities undertaken in this phase.

7.1.1 Selection of an initial set of machine learning algorithms

The initial suite of ML algorithms incorporated into our methodology, which were obtained from the relevant literature [48], encompasses the array of most widely-used supervised algorithms highlighted in Section 3, along with several new algorithms with which to represent the principal machine learning styles in a comprehensive manner. These include:

- *Baseline*: This algorithm computes the average outcome without considering predictors (independent variables).
- *Linear regression*: This is a common type of regression used for value prediction, which assumes independence between variables and applies a coefficient for each input

Fig. 5 Combined violin and box plot depicting the distribution of the BF personality variables: Agreeableness (A), Conscientiousness (C), Extraversion (E), Openness to Experience (O), and Neuroticism (N)

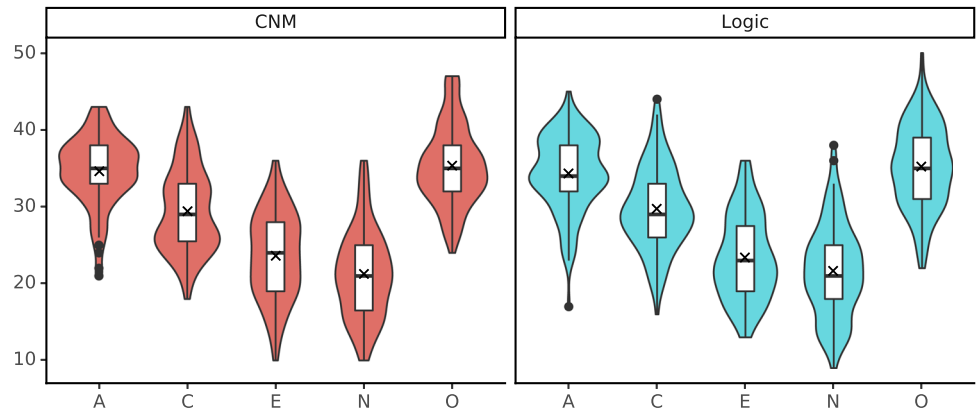


Fig. 6 Combined violin and box plot depicting the distribution of the academic engagement variables (in-class and out-of-class) at different time points, using robust normalization

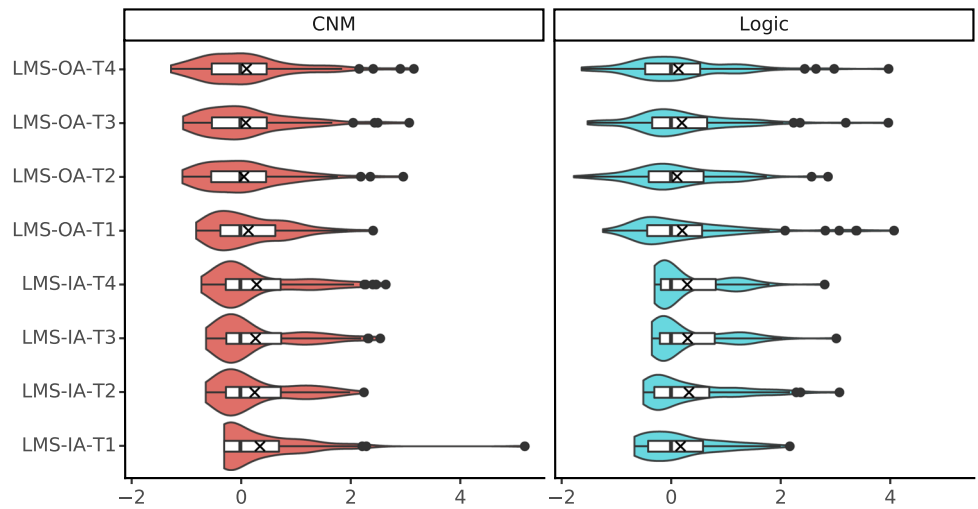
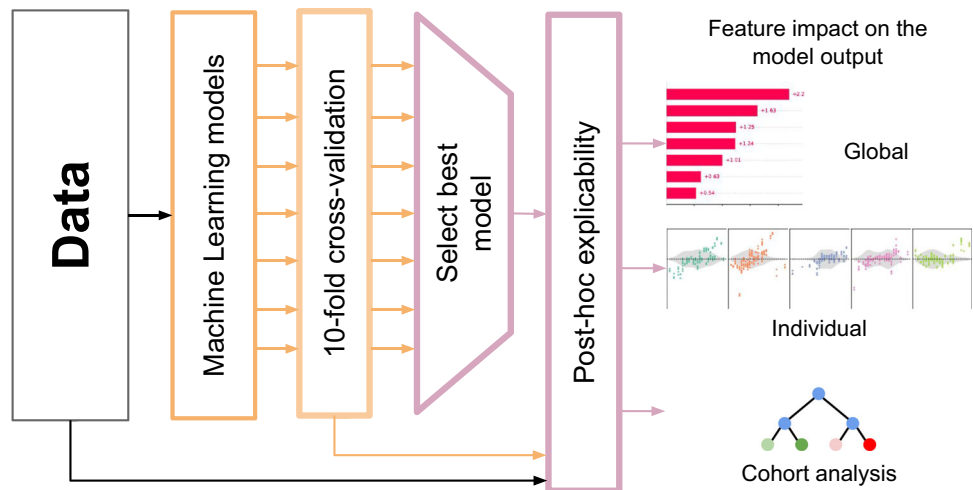


Fig. 7 Diagram illustrating the ML-based process utilized in order to extract pertinent information



variable plus a single overall fit value. In this study, we tested different variants by employing coefficient adjustment methods. These variants are:

- Simple Linear Regression with least squares adjustment [90]
 - LASSO [84] with l1 regularisation
 - Ridge [45] with l2 regularisation
 - Bayesian Ridge [85] with l2 regularisation and noise adjustment
 - Automatic Relevance Determination (ARD) [59] with Gaussian precision and noise adjustment using the evidence maximization technique.
- *Decision tree* [18]: This algorithm predicts the value of a sample in a hierarchical manner on the basis of straightforward learning rules. The tree is constructed using the training samples, and only one feature is taken into account for each rule.
 - *Random Forest* [17]: In order to create a more stable and robust model than a single decision tree, this algorithm builds several decision trees. All individual forecasts are considered in order to calculate the final prediction.
 - *AdaBoost (Adaptive Boosting)* [37]: This algorithm constructs several linear regression models during the first training phase. The algorithm predicts the value of a new sample by taking into account all the predictions of the regressors weighted with weights learned during the training phase.
 - *Tree boosting: XGBoost (eXtreme Gradient Boosting)* [22] and *CatBoost (Category Boosting)* [31] are based on the use of multiple decision trees, and boosting is employed in the training phase. They also use derivable cost functions and gradient descent for weight adjustment, similar to that which occurs with neural networks, but each applies different strategies.
 - *Support vector machine* [24]: This algorithm consists of two phases. In the first phase, the original data space is translated into a usually higher dimensional space. In the second phase, a hyperplane with which to separate the samples in the new space is found.
 - *Neural Network (Multilayer Perceptron)* [43]: All the layers in this basic neural network architecture are fully connected.
 - *Nearest Neighbors* [25]: This algorithm uses a similarity function, commonly the Euclidean distance, to find the k (parameter) closest known samples in the training set. Its prediction is computed on the basis of these closest samples. In our study, we set the values of k to 1, 3, 5, 7, and 9.

For all algorithms, we propose to adhere to the default configurations for all model parameters, with the exception of the case of the neighborhood-based algorithm. With this

algorithm, experimenting with varying numbers of neighbors resulted in improved performance, especially when dealing with noisy data. Table 3 enumerates the algorithms, the corresponding library functions, and their respective parameters in order to facilitate the reproducibility of our experiments,

7.1.2 Selection of the best model using the 10-fold cross-validation technique and pairwise comparisons

The second activity consists of evaluating the set of ML algorithms selected in order to identify that which is optimal, under the assumption that different contexts may yield different optimal algorithms.

We propose to identify the best algorithm by using the k -fold cross-validation technique, a method commonly employed in this field. This technique involves dividing the original sample set into several partitions of a similar size (typically 10) and iterating through them, with one partition serving as the test set and the remaining partitions used as the training set. This generates multiple results (equal to the number of partitions) for each algorithm, which can then be averaged in order to yield a final result or be tested against hypotheses so as to compare algorithms.

The evaluation metric that we propose for the ML models is the root mean square error (RMSE) [67]. This metric, which is defined in (2), utilizes vectors y and \hat{y} of size n , where y represents the true values and \hat{y} represents the predicted values. The RMSE is commonly utilized for regression problems, as it (a) utilizes the square of the difference between the true and predicted values (residuals) to further penalize differences, and (b) maintains the same units of measurement by applying the square root to the average of these residuals.

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (2)$$

To return to our case study, Fig. 8 presents the average RMSE for the outcomes of the 10-fold cross-validation applied separately to the defined predictors: personality (encompassing the five personality traits), academic engagement (represented by the eight LMS-A variables) and capacity (PAP). The findings illustrate that linear approximation algorithms generally yield a superior performance in our context, likely attributable to the linear characteristics of the variables under consideration. Of these, the ARD algorithm stands out as one of the top performers. In this figure, the most favorable RMSE values were observed for PAP (1.54), Personality (1.75), and LMS-A (1.76), with gain-to-baseline ratios of 0.794, 0.903 and 0.906, respectively (the lower the better).

The instructor must now explore the average RMSE of the 10-fold cross-validation outcomes for all the predictors

Table 3 Algorithms, functions and parameters used in the experiments

Algorithm	Function	Parameters
Baseline	DummyRegressor	Mean
Linear Regression	LinearRegression	Intercept=True
ARD	ARDRegressor	tol=0.001, alpha_1=1e-06, alpha_2=1e-06, lambda_1=1e-06, lambda_2=1e-06, threshold_lambda=10000.0, fit_intercept=True
Bayesian Ridge	BayesianRidge	tol=0.001, alpha_1=1e-06, alpha_2=1e-06, lambda_1=1e-06, lambda_2=1e-06, fit_intercept=True
Lasso	Lasso	alpha=1.0, fit_intercept=True, tol=0.0001, max_iter=1000
Ridge	Ridge	fit_intercept=True, tol=0.0001, solver='auto'
NN	KNeighborsRegressor	n_neighbors={1,3,5,7,9}
Linear SVM	SVM	kernel='linear', degree=3, gamma='scale', coef0=0.0, tol=0.001, C=1.0, epsilon=0.1, shrinking=True
RBF SVM	SVM	kernel='rbf', degree=3, gamma='scale', coef0=0.0, tol=0.001, C=1.0, epsilon=0.1, shrinking=True
AdaBoost	AdaBoostRegressor	n_estimators=50, learning_rate=1.0, loss='linear'
Neural Net	MLPRegressor	hidden_layer_sizes=(100,), activation='relu', solver='adam', alpha=0.0001, batch_size='auto', learning_rate='constant', learning_rate_init=0.001, power_t=0.5, max_iter=200, shuffle=True, tol=0.0001, momentum=0.9, early_stopping=False, validation_fraction=0.1, beta_1=0.9, beta_2=0.999, epsilon=1e-08, max_fun=15000
Decision Tree	DecisionTreeRegressor	criterion='squared_error', splitter='best', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, min_impurity_decrease=0.0, ccp_alpha=0.0
Random Forest	RandomForestRegressor	n_estimators=100, criterion='squared_error', min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=1.0, min_impurity_decrease=0.0, bootstrap=True, oob_score=False, ccp_alpha=0.0
XGB	XGBRegressor	learning_rate=0.3, n_estimators=100, max_depth=6, min_child_weight=1, subsample=1, colsample_bytree=1, gamma=0, reg_alpha=0, reg_lambda=1
CatBoost	CatBoostRegressor	learning_rate=0.03, n_estimators=1000, max_depth=6, l2_leaf_reg=3.0, bootstrap_type='MVS', subsample=1.0, colsample_bylevel=1.0

together at each point in time so as to ensure that their combination improves the performance of the model. We propose that this validation method be uniformly applied across all machine learning algorithms utilizing the same sample partitions, thus facilitating the subsequent pairwise comparisons.

In our case study, this exploration is presented in Fig. 9, which combines PAP, personality, and a dynamic evaluation of academic engagement characteristics throughout a four-month period. Once again, the best algorithms in our case study tend to involve linear approximations.

The third step when selecting the optimal model involves conducting a pairwise comparison in order to assess the significance of the RMSE generated by the 10-fold cross-validation of the algorithms under consideration for prediction.

Figure 10 illustrates this pairwise comparison within the context of our case study. The pairwise comparisons were carried out using the Wilcoxon signed-rank test [91], adopting the standard 95% confidence level [80] with Holm's adjustment [47], a non-parametric statistical test that does

not assume a normal distribution of the means. The findings indicate that linear models significantly outperform the others (as denoted by the green bullets).

The instructor can now employ all of these results to select the best model. In our case study, ARD has been selected as the best algorithm because, although the differences between all the linear models are not significant, ARD achieved the best average results.

Furthermore, note that the average RMSE values for ARD at the different time intervals are similar: 1.43, 1.42, 1.43 and 1.42 for T1, T2, T3 and T4, respectively. This means that, in the context of our case study, it is safe to make predictions regarding academic performance at T1, because they are very similar to those that would be made at the end of the academic year, T4.

It is also worth noting that the gain over the baseline using this model and a combination of features is around 0.73, which is better than the 0.79 obtained using the PAP alone.

The results achieved thus far can be employed as the basis on which to address RQ1 by noting that the optimal model

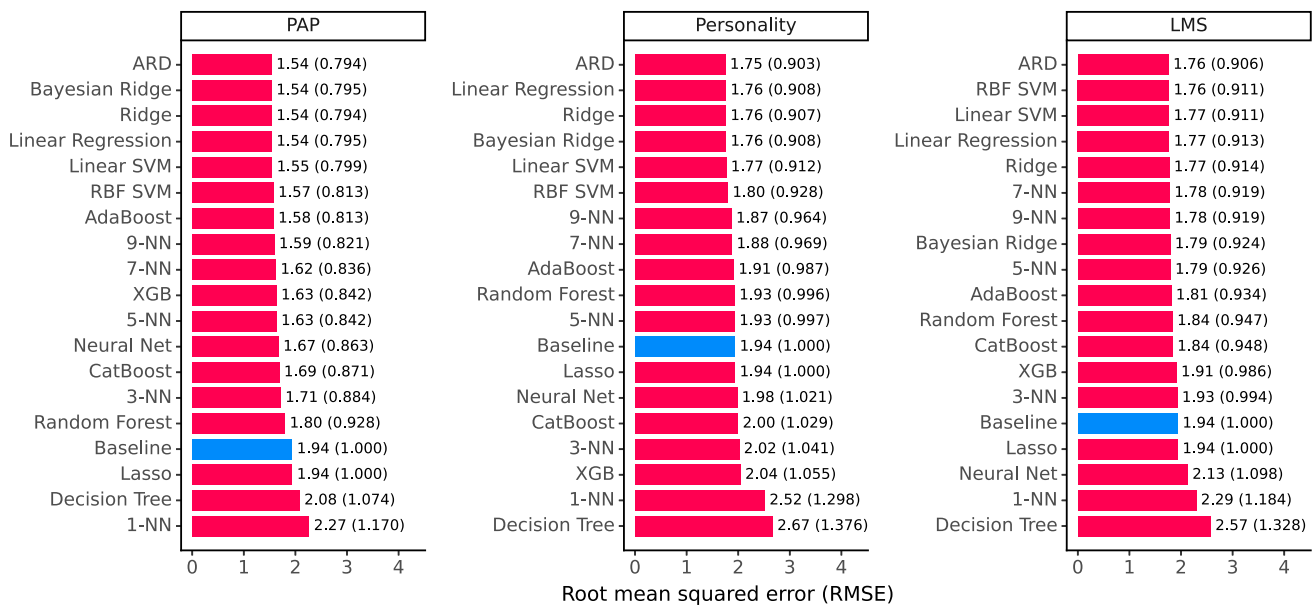


Fig. 8 Average RMSE of the 10-fold cross-validation results for PAP, the five personality predictors (personality), and the eight LMS-A variables (academic engagement) using various machine learning algorithms. Lower error values indicate a better performance. The numbers

at the end of the bars indicate the RMSE error, and the percentage of relative improvement from the baseline is shown in parentheses as $error(algorithm)/error(baseline)$. The baseline results are highlighted in blue

generated using our methodology yields a gain-to-baseline ratio of 0.73, indicating an approximate increase of 27% in precision when compared to the baseline. Moreover, the applicability of this model as early as four weeks into the semester provides a response to RQ2.

However, RQ3 can be addressed only by carrying out a further analysis.

8 Phase 3: Post-hoc explainability of the ML model selected

Having selected the best model, it is now time for the instructor to apply ML post-hoc explainability techniques [8] (see the last activity in Fig. 7). These techniques can uniformly explain the results, regardless of the particular algorithm, the linear/nonlinear relationships between variables and the learning styles. Their application assists as regards interpreting the results produced by the model and enhances the instructor's confidence in its predictions.

Two main approaches can be employed in order to achieve a certain level of interpretability and determine the importance of a variable input in the model. The first approach involves a statistical technique based on performing permutations on the input variables [17], establishing a relationship between these modifications and the predictions, and thus estimating the importance of the variable.

The second approach involves building an additional linear model on top of the original model. This approach is led by the Shapley values [76] and is based on game theory. In essence, the Shapley values quantify the average incremental contribution of a variable when combined with other variables. The Shapley approximation is particularly useful in those scenarios in which each variable contributes unequally to the final outcome (target variable), thus ensuring local accuracy, omission handling and consistency. It is for these reasons that the Shapley approach has been used in this study.

One limitation of the Shapley approach is that its values are general and its calculation applies to the total values studied. However, recent advances [56, 57] make it possible to extend its application to different contexts, such as groups of samples or individual samples. The SHAP (SHapley Additive exPlanations) [55] tool makes it possible to leverage these advancements.

A scheme of the integration of the SHAP post-hoc explainability process with the ML training process that we advocate in our approach is presented in Fig. 11. This shows how an explainer model is built (using a linear approach) on top of the trained ML model in order to provide an explanation for the prediction.

The outcomes of applying this SHAP explainability process to the ARD model chosen for our case study are shown below, and we also show how the results have allowed us to address RQ3.

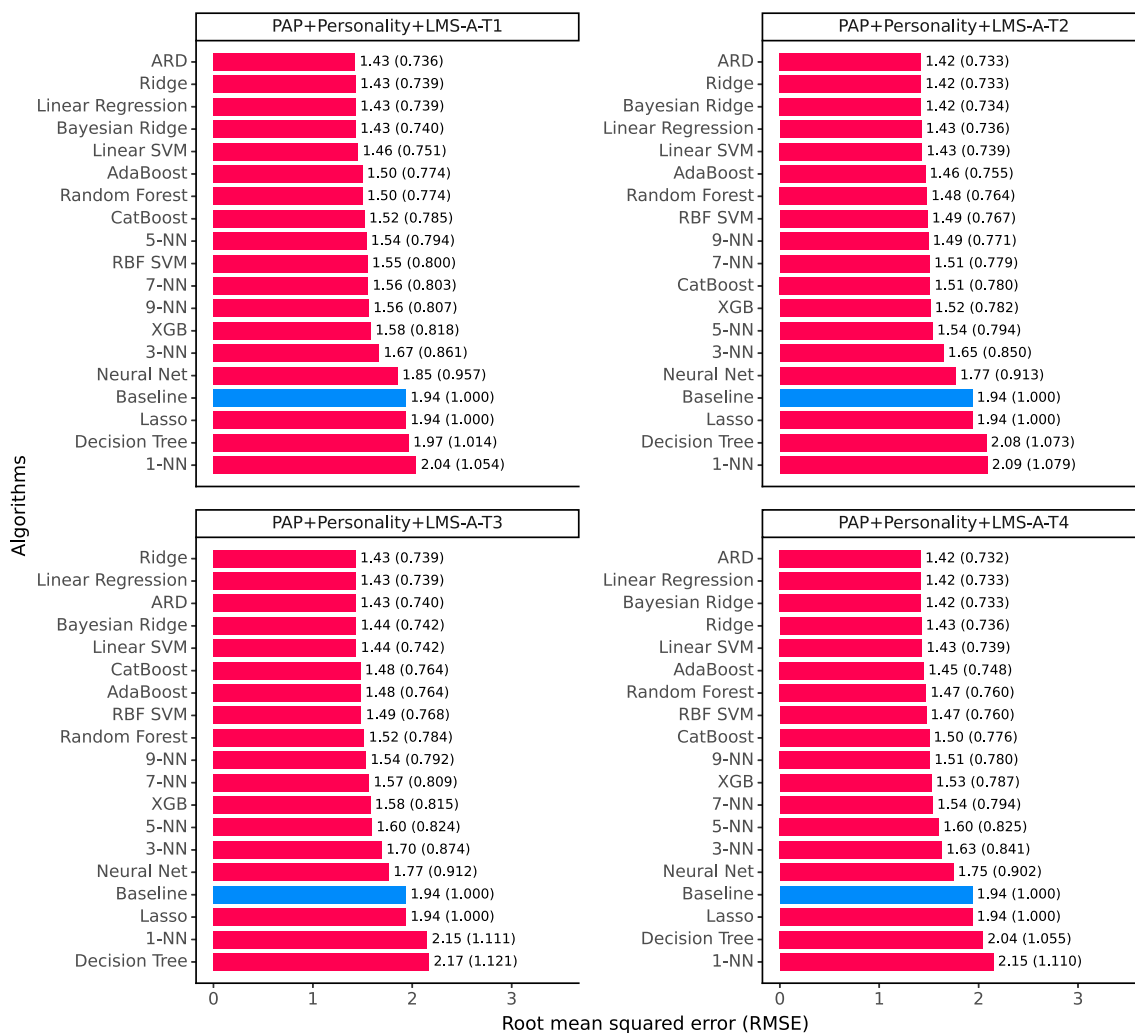
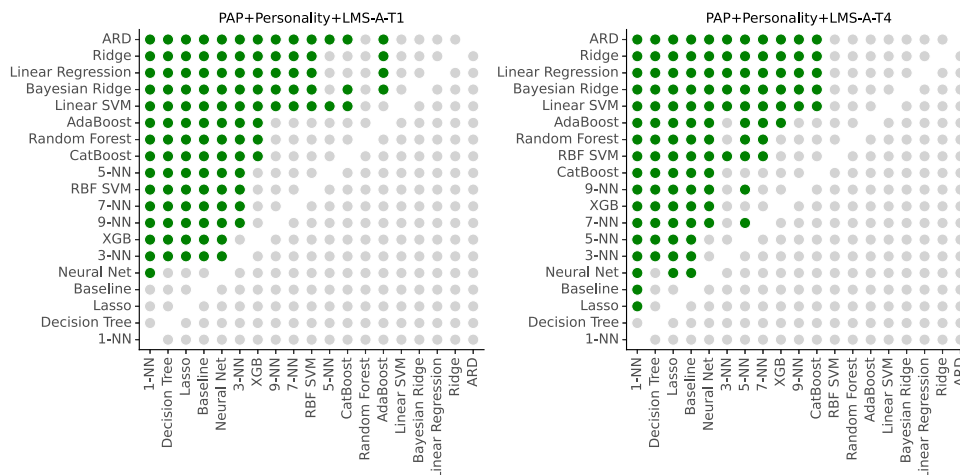


Fig. 9 Average RMSE of the 10-fold cross-validation outcomes for the personality and academic engagement predictors at different time intervals (spaced one month apart) using several machine learning algorithms together. Lower error values indicate a better performance. The

numbers at the end of the bars indicate the RMSE error, and the percentage of relative improvement from the baseline is shown in parentheses as $error(algorithm)/error(baseline)$. The baseline results are highlighted in blue

Fig. 10 Pairwise comparison of significance between the RMSE values obtained from the 10-fold cross-validation after applying the Wilcoxon signed-rank test. Green bullets indicate that the row algorithm is significantly better than the column algorithm



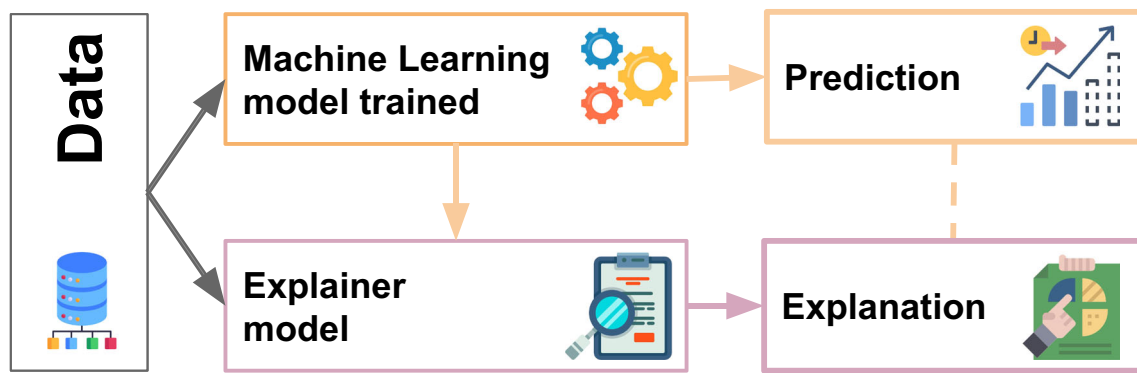


Fig. 11 General scheme of the post-hoc explainability process that is performed on top of the trained ML model

8.1 RQ3: consistency of features correlation with AP

For the sake of comparability, Fig. 12 shows the average impact of the chosen set of input variables on AP at T1 and T4.

Our results indicate that, in our case study, the same variables are relevant as regards predicting AP at both points in time (PAP, Subject code, Conscientiousness, Openness to experience, Agreeableness, and LMS outclass counts). The fact that the same predictors operate at T1 and T4 suggests the consistency of the model over time.

In order to further explore the individual relationships between relevant features and AP, Fig. 13 presents a grid plot depicting the shape of the relationship for each combination. The figure highlights that PAP is the most relevant variable with a positive relationship on both courses. The Openness to Experience personality variable also has a negative relationship on both courses, meaning that higher Openness to Experience correlates with lower final grades. Moreover, the student's access to the LMS outside class time (LMS-OA-T1 and LMS-OA-T4) has a positive relationship, with slight differences between CNM and Logic. Agreeableness also has a slight influence, with a negative relationship. Finally, the course itself also has an impact on the grades, with a tendency toward better grades in Logic, which is a second-year subject, than in CNM (a first-year subject), indicating a greater difficulty of the CNM subject.

8.2 Cohort analysis

The analyses conducted thus far have resulted in a customized model with which to predict a student's final grade at a specific point in time (T1 to T4). In order to enable instructors to attain a greater understanding of the factors that most influence performance in their analytical context, our proposal includes a cohort analysis as the final step. This analysis allows the division of the students for whom the ML model has been customized into n performance groups (n cohorts)

on the basis of a set of specific rules, conditions or criteria that define each group.

Our proposal suggests that this group formation should be automated in a manner similar to the way in which decision trees are created (decision tree method). In this automatic cohort formation, the most relevant variables for the samples analyzed are evaluated on the basis of the target variable, and a split is applied to one of them in each step so as to form two partitions. The use of the SHAP tool and the application of the Gini metric (see (3)) to the samples of the partitions being created are proposed for this purpose.

$$Gini(y) = 1 - \sum_{i=0}^n P(y_i)^2 \quad (3)$$

where y is a vector with n elements and $P(y_i)$ represents the probability of the element y_i being misclassified.

The process starts with the total set of samples and divides it into subsets until the maximum number of sets is attained.

The main advantage of this additional step is that the algorithm automatically identifies the optimal cohorts on the basis of logical criteria, and helps uncover rules that might not be obvious a priori.

Returning to our case study, using the ML model at T1 as a basis, we decided to create five cohorts, which, in our educational context, were considered sufficient to differentiate between various profiles while still ensuring that the subsets contained an adequate number of samples.

Table 4 shows the rules automatically generated, defining the five specified cohorts for our case study. Similarly, Fig. 14 illustrates these rules as a binary tree. In both representations, there are some descriptors of the groups formed, such as their size, median of the final grades, and the significance of differences between them determined by applying the U Mann-Whitney U test [61]. In the first group, individuals with a PAP score exceeding 10.85 achieve the highest final grades, averaging 8.6. In the case of those who do not meet this criterion, their subject performance is determined

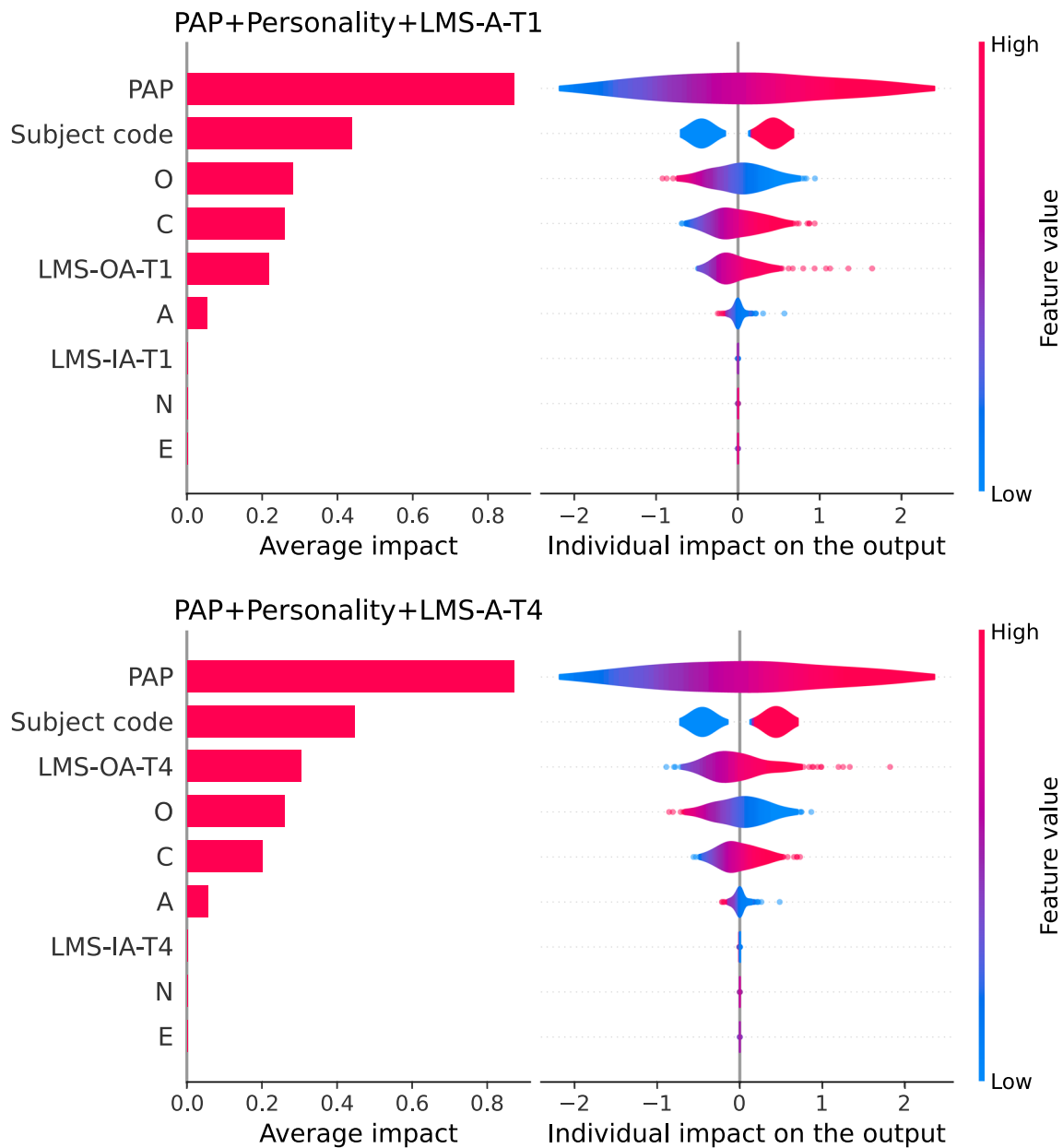


Fig. 12 Average impact of input factors on AP. The graph at the top represents T1 (week 4), while the graph at the bottom represents T4 (end of the course)

by the number of accesses they make outside class time during period T1. If the number of accesses is above 104, they are assigned to groups 2 (average 7.2) or 3 (average 6.0) on the basis of their PAP score (greater or lower than 7.4), whereas if it is below 104, they are assigned to groups 4 (average 5.8) or 5 (average 3.7) on the basis of whether they have a PAP score that is greater or lower than 6.58.

Using the cohorts generated as a starting point, it is possible to identify at-risk groups by selecting those cohorts whose mean predictor value falls below a certain threshold. In our case, and based on the criteria outlined in Table 4 and

in Fig. 14, cohorts 4 ($PAP \leq 10.85$) & ($LMS-OA-T1 \leq 104$) & ($PAP > 6.58$) and 5 ($LMS-OA-T1 \leq 104$) & ($PAP \leq 6.58$) indicate student groups of a potentially low academic performance, with the latter group being at extreme risk, with a median AP score of 3.7 over 10.

8.3 Prediction and analysis tool

Finally, in order to facilitate both individual-level academic performance prediction with explainability and group-level estimation and analysis in the context of our case study, we

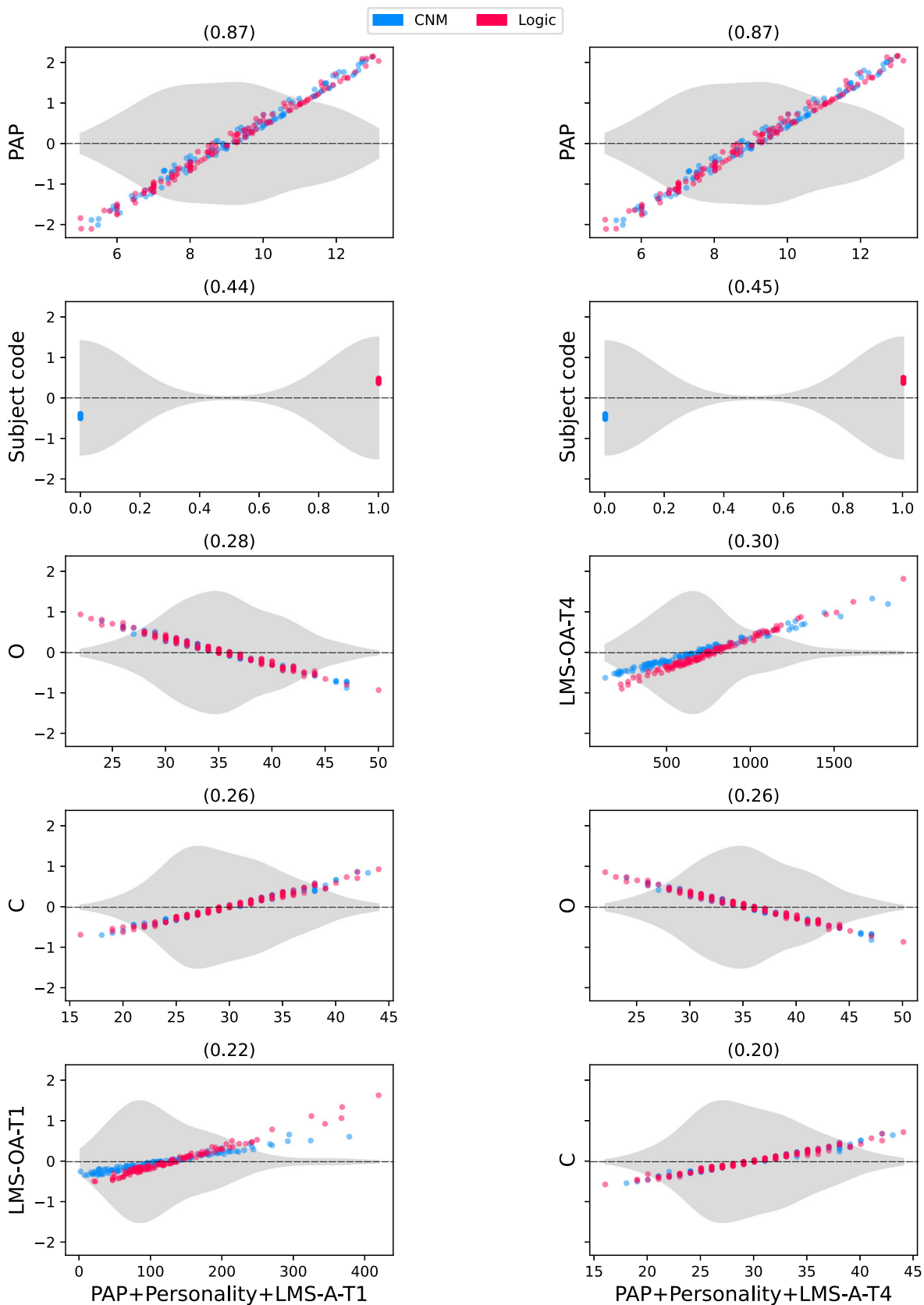


Fig. 13 Individual impact of the most relevant variables on AP. The sub-figures are ordered from top to bottom on the basis of their average impact value indicated in brackets. The horizontal axis of each sub-figure shows the density of samples in each corresponding area

Table 4 Number of students and median value of the output predictor using cohort rules at T1

Id	Cohort	Size	Median	
(1)	('PAP' > 10.85)	55	8.6	** (2,3,4,5)
(2)	('PAP' <= 10.85) & ('LMS-OA-T1' > 104) & ('PAP' > 7.4)	65	7.2	** (3,4,5)
(3)	('PAP' <= 10.85) & ('LMS-OA-T1' > 104) & ('PAP' <= 7.4)	33	6.0	* (4), ** (5)
(4)	('PAP' <= 10.85) & ('LMS-OA-T1' <= 104) & ('PAP' > 6.58)	81	5.8	** (5)
(5)	('PAP' <= 10.85) & ('LMS-OA-T1' <= 104) & ('PAP' <= 6.58)	12	3.7	

Asterisks (**) indicate significant differences (at 95% confidence level) between cohorts (1), (2), (3), (4) and (5) based on the output variable

have created a simple yet functional and effective tool. We have utilized Python Jupyter notebook technology with the Google Colab online service, which was chosen for its free access and the requirement of only a web browser for use. It is available at [this link](#). This tool is already being used by the instructors of the subjects included in our case of study, and could also serve as an example and a basis for other developments, as its source code is accessible.

9 Discussion

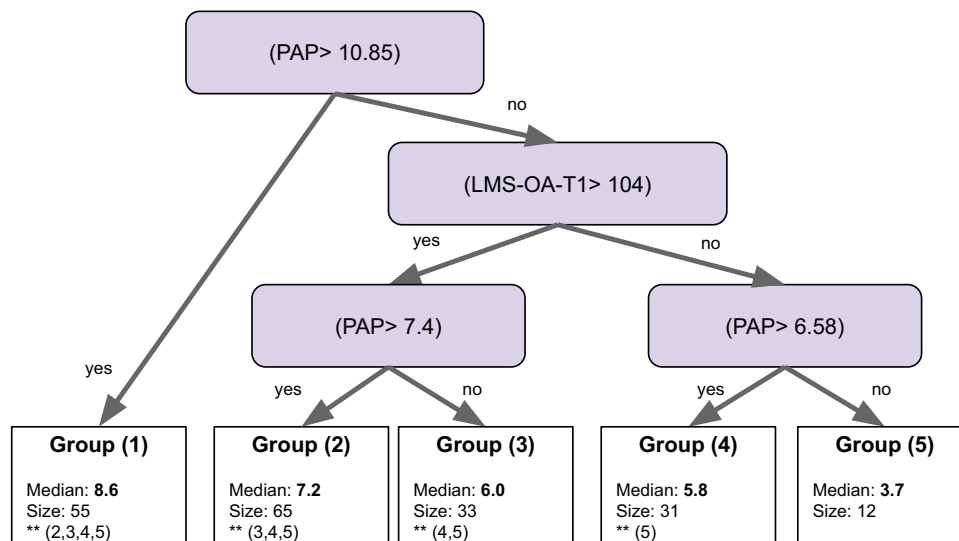
This paper makes two main contributions. Firstly, it proposes a conceptual model that is based on the MHP and includes three factors: personality, academic engagement and capacity. This model is sufficiently simple to allow instructors to collect data quickly and without having to resort to academic authorities, thus increasing its practical utility.

Secondly, it proposes an ML methodology with which to generate tailored models for the early prediction of academic outcomes and the detection of at-risk groups in single or related academic subjects. The process involves three phases: (a) a data gathering phase, (b) a model generation and selection phase and (c) a post-hoc explainability phase.

The use of XAI techniques increases the transparency and interpretability of the model selected, thus enabling educators to better understand and trust in their predictions, which can inform the proposal of more effective intervention plans. Moreover, the inclusion of cohort analysis techniques allows group-level estimation and analysis.

In order to illustrate the feasibility of the approach, the paper shows the application of the proposed methodology to a case study in which an ML model was selected and trained with data concerning students enrolled in two different mathematics-related subjects taught in the first and second year of a Computer Engineering degree program. While recent meta-analyses have reported a prediction capacity of up to 25% for cognitive ability [69], the proxy of academic capacity (PAP) of our model has a slightly lower gain-to-baseline (20.6%), using just one easily-collected measure. This gain-to-baseline increases to 26.4% when all the variables are considered together. The context of the case study (mathematics and first and second year subjects) maximizes the practical impact of the model obtained, as they are the years and subjects in which the highest academic failure rates occur in the Spanish system. In this case study, significant insights were obtained from relatively small sample sizes, demonstrating a uniform pattern across two different

Fig. 14 Cohort rules at T1, represented as a decision tree. Asterisks (**) indicate significant differences (at 95% confidence level) between cohorts (1), (2), (3), (4) and (5) based on the output variable



subjects, despite variations in teaching methods and instructors, within the same academic discipline.

Our case study reveals that, while non-cognitive variables in related literature can explain up to an additional 25% of the variability in performance, this percentage greatly varies depending on the particular context. In our case study, the non-cognitive components included (academic engagement and personality) explain only an additional 6%, although when analyzed separately, they explain 9.4% and 9.7%, respectively (see Fig. 8). One possible explanation for this is that the variables of willingness, and particularly personality, which is measured in a manner similar to that shown in related literature, may not have a significant predictive capability in mathematics subjects when compared to other fields. Whatever the reason, this fact underscores the need for instructors to develop customized models that are finely tuned to the unique characteristics of their educational environment. The three personality variables that are most strongly associated with academic achievement in our case study are consciousness, openness to experience, and agreeableness. These variables have also been identified as being the most relevant in related meta-analyses [69, 74, 86]. The positive impact of consciousness is expected and easily justifiable, as thoroughness and diligence are widely recognized as having a positive impact on academic outcomes. However, in contrast to other studies, our findings indicate that openness to experience has an inverse relationship with success in our context. Students with greater openness to experience perform worse, as shown in Fig. 13. This negative impact of openness suggests that students with higher levels of imagination, aesthetic sensitivity and need for variety may not perform as well in mathematical subjects, despite the fact that these traits are considered positive in other areas and at lower levels of education [70]. Our case study therefore supports the approach of analyzing subjects belonging to different areas of knowledge separately. Finally, agreeableness is the third most important influential personality variable in our context, which is consistent with existing meta-analyses [86].

Our cohort analysis shows that academic engagement is the most important willingness-related factor as regards classifying students at risk in the context of our case study. Despite its dynamic nature [2], our case study demonstrates that this variable can be useful in certain contexts when measured as soon as even one month into the semester (T1). This can be justified on the grounds that procrastination is a major impediment to academic success at university [74], and students engaging early in the class dynamics consequently have greater possibilities of success. Also related to this analysis, it is worth noting that the decision tree derived during the cohort analysis relies solely on the PAP and the LMS-OA-T1 variables to determine the groups, indicating that academic engagement and capacity overshadow the impact

of personality in the context of this study, which simplifies the identification of students at risk.

Finally, and similar to that which occurs in other contexts [71], the case study shows that combining the capacity and willingness variables increases the global classification accuracy of our model when compared to relying exclusively on one type of variable. Our gain-to-baseline result (26%) is comparable to the gains shown in a recent meta-analysis that included studies that predicted academic performance on the basis of a mixture of cognitive and non-cognitive measures [60]. Moreover, our results highlight the superior predictive power of PAP, with a relative importance of 50.24% (slightly lower than the relative importance of 64% reported for a similar measure in [60]), while personality has a relative importance of 21.24% (also slightly lower than the 28% reported in [60]). These differences can be attributed to the inclusion of academic engagement, which was not considered in the meta-analysis carried out by [60].

9.1 Main limitations of the approach

The case study presented in this paper highlights that the proposed approach has the potential to generate customized ML models with which to predict academic performance. However, it is vital to acknowledge its main limitations. Firstly, the creation of an optimal model necessitates a full academic term for data gathering and analysis in order to identify the most suitable model for the specific context. This model can then assist in identifying at-risk students in future terms, provided that the context of the subject remains relatively unchanged. Secondly, instructors need a fundamental understanding of ML in order to apply this process effectively. Lastly, the dependence on LMS data as a surrogate for academic engagement could restrict its use for those instructors who do not employ these platforms.

9.2 Practical implications

The proposed methodology, which involves integrating a conceptual framework with a Machine Learning process so as to generate personalized models with which to predict academic performance in higher education, provides significant practical benefits for instructors:

- Early Identification of Needs: It helps identify students who may require additional support, thus enabling early interventions.
- Learning Personalization: It allows instructors to tailor educational strategies to the individual needs and characteristics of students, thus improving learning effectiveness.
- Improvement of Pedagogical Strategies: It provides insights into how different aspects of personality and

academic engagement affect performance, thus guiding instructors as regards optimizing their teaching methods.

- Encouragement of Academic Engagement: By better understanding the influence of engagement on academic performance, instructors can design activities that increase student participation and motivation.

These implications highlight the potential of the proposal to transform higher education, emphasizing personalization and the continuous improvement of the teaching-learning process.

10 Conclusions and future work

In this paper, we have presented a method that includes both a conceptual model and a ML process with which to predict academic performance. We believe that the parsimony of the conceptual model and the clear step-by-step description of the process significantly increase the likelihood of our proposal being adopted in university settings.

Our approach has the objective of assisting instructors in the detection of at-risk groups and implement targeted interventions with which to prevent academic failure as soon as possible in the semester, ultimately leading to improved academic outcomes. The use of XAI techniques and cohort analysis allows a more comprehensive understanding of the factors that influence performance in each particular context, leading to more effective risk mitigation strategies.

The application of the approach to a case study involving two first and second-year mathematical subjects demonstrates that the accuracy of the resulting model is consistent with the findings reported in literature. The model and the associated [Colab notebook](#) are currently aiding instructors of these subjects to devise targeted strategies aimed at lowering the rates of academic failure, thereby enhancing educational outcomes.

As part of our future work, we plan to apply this method in various new contexts so as to assess the ability of the approach to generate similarly valuable models for other courses and universities. Moreover, we are exploring the incorporation of other easily collectible variables in order to enrich the willingness dimension of the conceptual model, such as the admission option (the order in which students chose their current degree program) [79] and the grades they aspire to achieve [74] so as to determine whether these additions enhance the predictive accuracy of the models. Finally, we aim to investigate how our approach compares (in terms of both effectiveness and complexity) with other new approaches, such as those that use expanded user and item features by employing the latent factor space of auxiliary domains. These approaches enrich the feature set with underlying patterns in user-item interactions that may not be directly observable in the current context [94, 95]. One

potential advantage of these methods is that, once these latent vectors have been obtained for the students, the feature set of any target domain in which these students participate can be expanded by incorporating these vectors without the need for any further data collection or analysis. This expanded feature set provides a richer representation of the students, thus enabling more nuanced predictions about student performance or preferences.

Acknowledgements We would like to thank the students who kindly agreed to participate in our study.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This work has been partially funded by the Instituto de Ciencias de la Educación (ICE) of the University of Alicante through their ‘Programa de redes de investigación en docencia universitaria’ in the 2023/2024 edition (Red 5942), and the UCLM group, cofinanced with ERDF funds (research grant 2022-GRIN-34113). This paper reflects the views of only the authors, and the funding bodies cannot be held responsible for any use which may be made of the information contained therein.

Data Availability Dataset is available for research purposes from the first author.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Affuso G, Zannone A, Esposito C, Pannone M, Miranda MC, De Angelis G, Aquilar S, Dragone M, Bacchini D (2023) The effects of teacher support, parental monitoring, motivation and self-efficacy on academic performance over time. *Eur J Psychol Educ* 38(1):1–23. <https://doi.org/10.1007/s10212-021-00594-6>
2. Albreiki B, Zaki N, Alashwal H (2021) A systematic literature review of student performance prediction using machine learning techniques. *Educ Sci* 11(9):552. <https://doi.org/10.3390/educsci11090552>
3. Alhadabi A, Karpinski AC (2020) Grit, self-efficacy, achievement orientation goals, and academic performance in University students. *Int J Adolesc Youth* 25(1):519–535. <https://doi.org/10.1080/02673843.2019.1679202>

4. Alshamaila Y, Alsawalqah H, Aljarah I, Habib M, Faris H, Alshraideh M, Salih BA (2024) An automatic prediction of students' performance to support the university education system: a deep learning approach. *Multimedia Tools and Applications*, pp 1–28
5. Avia M, Sanz J, Sánchez-Bernardos M, Martínez-Arias M, Silva F, Graña J (1995) The five-factor model—II. Relations of the NEO-PI with other personality variables. *Pers Individ Differ* 19(1):81–97 (1995). [https://doi.org/10.1016/0191-8869\(95\)00007-S](https://doi.org/10.1016/0191-8869(95)00007-S)
6. Azevedo R (2020) Reflections on the field of metacognition: Issues, challenges, and opportunities. *Metacognition Learn* 15:91–98. <https://doi.org/10.1007/s11409-020-09231-x>
7. Balijepally V, Mahapatra R, Nerur SP (2006) Assessing personality profiles of software developers in agile development teams. *Commun Assoc Inf Syst* 18(1):4. <https://doi.org/10.17705/1CAIS.01804>
8. Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, García S, Gil-Lopez S, Molina D, Benjamins R, Chatila R, Herrera F (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion* 58:82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
9. Barrick MR (2005) Yes, personality matters: Moving on to more important matters. *Hum Perform* 18(4):359–372 (2005). https://doi.org/10.1207/s15327043hup1804_3
10. Batool S, Rashid J, Nisar MW, Kim J, Kwon HY, Hussain A (2023) Educational data mining to predict students' academic performance: A survey study. *Educ Inf Technol* 28(1):905–971. <https://doi.org/10.1007/s10639-022-11152-y>
11. BBVA F (2019) U-ranking 2019. indicadores sintéticos de las universidades españolas. <https://www.fbbva.es/wp-content/uploads/2019/04/Informe-U-Ranking-FBBVA-Ivie-2019.pdf>
12. Beckham NR, Akeh LJ, Mitaart GNP, Moniaga JV (2023) Determining factors that affect student performance using various machine learning methods. *Procedia Comput Sci* 216:597–603
13. Behr A, Giese M, Tegum Kamdjou HD, Theune K (2020) Dropping out of university: a literature review. *Rev Educ* 8(2):614–652. <https://doi.org/10.1002/rev3.3202>
14. Bergold S, Steinmayr R (2018) Personality and intelligence interact in the prediction of academic achievement. *J Intell* 6(2):27. <https://doi.org/10.3390/jintelligence6020027>
15. Bluman AG (2012) Elementary statistics: A step by step approach. McGraw-Hill
16. Blumberg M, Pringle CD (1982) The missing opportunity in organizational research: Some implications for a theory of work performance. *Acad Manag Rev* 7(4):560–569. <https://doi.org/10.5465/amr.1982.4285240>
17. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
18. Breiman L (2017) Classification and regression trees. Routledge. <https://doi.org/10.1201/9781315139470>
19. Bressane A, Zwirn D, Essiptchouk A, Saraiva ACV, de Campos Carvalho FL, Formiga JKS, de Castro Medeiros LC, Negri RG (2024) Understanding the role of study strategies and learning disabilities on student academic performance to enhance educational approaches: A proposal using artificial intelligence. *Comput Educ Artif Intell* 6:100,196
20. Caprara GV, Vecchione M, Alessandri G, Gerbino M, Barbaranelli C (2011) The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. *Br J Educ Psychol* 81(1):78–96. <https://doi.org/10.1348/2044-8279.002004>
21. Carlson MD, Morrison RS (2009) Study design, precision, and validity in observational studies. *J Palliat Med* 12(1):77–82. <https://doi.org/10.1089/jpm.2008.9690>
22. Chen, T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System. *CoRR* [arXiv:1603.02754](https://arxiv.org/abs/1603.02754)
23. Cheng B, Liu Y, Jia Y (2024) Evaluation of students' performance during the academic period using the XG-Boost classifier-enhanced AEO hybrid model. *Expert Syst Appl* 238:122,136
24. Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297. <https://doi.org/10.1023/A:1022627411411>
25. Cover TM, Hart PE (1967) Nearest neighbor pattern classification. *IEEE Trans Inf Theory* 13(1):21–27. <https://doi.org/10.1109/TIT.1967.1053964>
26. Cruz S, da Silva FQ, Capretz LF (2015) Forty years of research on personality in software engineering: A mapping study. *Comput Hum Behav* 46:94–113. <https://doi.org/10.1016/j.chb.2014.12.008>
27. Cyrenne P, Chan A (2012) High school grades and university performance: A case study. *Econ Educ Rev* 31(5):524–542. <https://doi.org/10.1016/j.econedurev.2012.03.005>
28. De Feyter T, Caers R, Vigna C, Berings D (2012) Unraveling the impact of the Big Five personality traits on academic performance: The moderating and mediating effects of self-efficacy and academic motivation. *Learn Individ Differ* 22(4):439–448. <https://doi.org/10.1016/j.lindif.2012.03.013>
29. De Raad B, Schouwenburg HC (1996) Personality in learning and education: A review. *European J Personal* 10(5):303–336
30. Del Valle R, Duffy TM (2009) Online learning: Learner characteristics and their approaches to managing learning. *Instr Sci* 37:129–149. <https://doi.org/10.1007/s11251-007-9039-0>
31. Dorogush AV, Ershov V, Gulin A (2018) Catboost: gradient boosting with categorical features support. [arXiv:1810.11363](https://arxiv.org/abs/1810.11363)
32. Dunatov SM, Kasalo K, Lovrinčević A, Maljković J, Prnjak A (2022) Predicting students' final exam grades based on learning material usage extracted from moodle logs. In: 2022 International conference on software, telecommunications and computer networks (SoftCOM), pp 1–6. IEEE. <https://doi.org/10.23919/SoftCOM55329.2022.9911477>
33. Eysenck HJ (1994) The big five or giant three: criteria for a paradigm. Lawrence Erlbaum Associates, Inc
34. Fahd K, Venkatraman S, Miah SJ, Ahmed K (2022) Application of machine learning in higher education to assess student academic performance, at-risk, and attrition: A meta-analysis of literature. *Education and Information Technologies*, pp 1–33. <https://doi.org/10.1007/s10639-021-10741-7>
35. Feldt R, Angelis L, Torkar R, Samuelsson M (2010) Links between the personalities, views and attitudes of software engineers. *Inf Software Technol* 52(6):611–624. <https://doi.org/10.1016/j.infsof.2010.01.001>
36. Filippidi A, Tselios N, Komis V (2010) Impact of Moodle usage practices on students' performance in the context of a blended learning environment. *Proceedings of Social Applications for Life Long Learning*, pp 2–7
37. Freund Y, Schapire RE (1997) A decision-theoretic generalization of on-line learning and an application to boosting. *J Comput Syst Sci* 55(1):119–139. <https://doi.org/10.1006/jcss.1997.1504>
38. Galla BM, Shulman EP, Plummer BD, Gardner M, Hutt SJ, Goyer JP, D'Mello SK, Finn AS, Duckworth AL (2019) Why high school grades are better predictors of on-time college graduation than are admissions test scores: the roles of self-regulation and cognitive ability. *Am Educ Res J* 56(6):2077–2115. <https://doi.org/10.3102/0002831219843292>
39. Guerra S (2016) Prueba Neuropsicológica para medir la inteligencia: utilización del test de Retención Visual de Benton. Ph.D. thesis, Universidad de Valladolid
40. Hastie T, Tibshirani R, Friedman JH, Friedman JH (2009) The elements of statistical learning: data mining, inference, and prediction, vol 2. Springer
41. Heckman JJ, Kautz T (2012) Hard evidence on soft skills. *Labour Econ* 19(4):451–464

42. Higgins N, Rathner JA, Frankland S (2021) Development of self-regulated learning: a longitudinal study on academic performance in undergraduate science. *Research in Science and Technological Education*. <https://doi.org/10.1080/02635143.2021.1997978>
43. Hinton GE (1990) Connectionist learning procedures. In: *Machine learning*, Volume III, pp 555–610. Elsevier
44. Hodara M, Lewis K (2017) How well does high school grade point average predict college performance by student urbanicity and timing of college entry. *National Center for Education Evaluation and Regional Assistance, REL*, p 250
45. Hoerl AE, Kennard RW (1970) Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12(1):55–67. <https://doi.org/10.1080/00401706.1970.10488634>
46. Hoffman D (2021) The APS and undergraduate performance in construction economics in south africa. *Lect Notes Netw Syst* 269:232–240. https://doi.org/10.1007/978-3-030-80000-0_28
47. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, pp 65–70
48. 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*, pp 100204
49. John OP, Naumann LP, Soto CJ (2008) Paradigm shift to the integrative big five trait taxonomy. *Handbook of Personality: Theory and Research* 3:114–158
50. Kori K, Pedaste M, Tõnisson E, Palts T, Altin, H, Rantsus R, Sell R, Murtazin K, Rüttemann T (2015) First-year dropout in ict studies, pp 437 – 445. <https://doi.org/10.1109/EDUCON.2015.7096008>
51. Laidra K, Pullmann H, Allik J (2007) Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Pers Individ Differ* 42(3):441–451. <https://doi.org/10.1016/j.paid.2006.08.001>
52. Li H, XU J, CHEN J, FAN Y, (2015) A reliability meta-analysis for 44 items big five inventory: Based on the reliability generalization methodology. *Adv Psychol Sci* 23(5):755. <https://doi.org/10.3724/SPJ.1042.2015.00755>
53. López-García, A., Blasco-Blasco, O., Liern-García, M., Parada-Rico, S.E.: Early detection of students' failure using Machine Learning techniques. *Oper Res Perspect* 11:100,292 (2023)
54. López-Zambrano J, Lara JA, Romero C (2020) Towards portability of models for predicting students' final performance in university courses starting from moodle logs. *Appl Sci* 10(1):354. <https://doi.org/10.3390/app10010354>
55. Lundberg S (2019) Shap (shapley additive explanations). <https://github.com/slundberg/shap>
56. Lundberg SM, Lee SI (2017) Consistent feature attribution for tree ensembles. [arXiv:1706.06060](https://arxiv.org/abs/1706.06060)
57. Lundberg SM, Lee SI (2017) A unified approach to interpreting model predictions. In: *Advances in neural information processing systems*, pp 4765–4774
58. MacCann C, Jiang Y, Brown LE, Double KS, Bucich M, Minbashian A (2020) Emotional intelligence predicts academic performance: A meta-analysis. *Psychol Bull* 146(2):150. <https://doi.org/10.1037/bul0000219>
59. MacKay DJ et al (1994) Bayesian nonlinear modeling for the prediction competition. *ASHRAE Trans* 100(2):1053–1062
60. Mammadov S (2022) Big Five personality traits and academic performance: A meta-analysis. *J Personal* 90(2):222–255. <https://doi.org/10.1111/jopy.12663>
61. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *The annals of mathematical statistics*, pp 50–60. <https://doi.org/10.1214/aoms/1177730491>
62. OECD (2022) Education at a glance 2022. <https://www.oecd.org/education/education-at-a-glance/>
63. Ortiz-Lozano JM, Rua-Vieites A, Bilbao-Calabuig P, Casadesús-Fa M (2018) University student retention: Best time and data to identify undergraduate students at risk of dropout. *Innov Educ Teach Int* 57(1):74–85. <https://doi.org/10.1080/14703297.2018.1502090>
64. Panadero E (2017) A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology*, pp 422. <https://doi.org/10.3389/fpsyg.2017.00422>
65. Parkavi R, Karthikeyan P et al (2023) Predicting academic performance of learners with the three domains of learning data using neuro-fuzzy model and machine learning algorithms. *Journal of Engineering Research*
66. Pelima LR, Sukmana Y, Rosmansyah Y (2024) Predicting university student graduation using academic performance and machine learning: a systematic literature review. *IEEE Access*
67. Pentreath N (2015) *Machine learning with spark*. Packt Publishing Ltd
68. Pérez E, Medrano LA (2013) Teorías contemporáneas de la inteligencia: Una revisión crítica de la literatura. *PSIENCIA: Revista Latinoamericana de Ciencia Psicológica* 5(2):6
69. Pérez-González JC, Filella G, Soldevila A, Faiad Y, Sanchez-Ruiz MJ (2022) Integrating self-regulated learning and individual differences in the prediction of university academic achievement across a three-year-long degree. *Metacognition Learn* 17(3):1141–1165. <https://doi.org/10.1007/s11409-022-09315-w>
70. Poropat AE (2009) A meta-analysis of the five-factor model of personality and academic performance. *Psychol Bull* 135(2):322. <https://doi.org/10.1037/a0014996>
71. Quinn RJ, Gray G (2020) Prediction of student academic performance using Moodle data from a Further Education setting. *Irish Journal of Technology Enhanced Learning* 5(1). <https://doi.org/10.22554/ijtel.v5i1.57>
72. Rafiola R, Setyosari P, Radjah C, Ramli M (2020) The effect of learning motivation, self-efficacy, and blended learning on students' achievement in the industrial revolution 4.0. *Int J Emerg Technol Learn (iJET)* 15(8):71–82. <https://doi.org/10.3991/ijet.v15i08.12525>
73. Reed SK (2012) *Cognition: Theories and applications*. CENGAGE learning
74. Richardson M, Abraham C, Bond R (2012) Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychol Bull* 138(2):353. <https://doi.org/10.1037/a0026838>
75. Rivers DJ (2021) The role of personality traits and online academic self-efficacy in acceptance, actual use and achievement in Moodle. *Education and Information Technologies*, pp 1–26. <https://doi.org/10.1007/s10639-021-10478-3>
76. Roth AE (1988) *The Shapley value: essays in honor of Lloyd S. Cambridge University Press, Shapley*. <https://doi.org/10.1017/CBO9780511528446>
77. Schaufeli WB, Martinez IM, Pinto AM, Salanova M, Bakker AB (2002) Burnout and engagement in university students: A cross-national study. *Journal of cross-cultural psychology* 33(5):464–481. <https://doi.org/10.1177/0022022102033005003>
78. Schaufeli WB, Salanova M, González-Romá V, Bakker AB (2022) The measurement of engagement and burnout: A two sample confirmatory factor analytic approach. *J Happiness Stud* 3:71–92. <https://doi.org/10.1023/A:1015630930326>
79. Segura M, Mello J, Hernández A (2022) machine learning prediction of university student dropout: does preference play a key role. *Mathematics* 10(18):3359. <https://doi.org/10.3390/math10183359>
80. Stapor K, Ksieniewicz, P, Garcia S, Woźniak M (2021) How to design the fair experimental classifier evaluation. *Applied Soft Computing*, pp 107219. <https://doi.org/10.1016/j.asoc.2021.107219>
81. Sternberg RJ (1990) *Metaphors of mind: Conceptions of the nature of intelligence*. Cambridge University Press

82. Tavakol M, Dennick R (2011) Making sense of Cronbach's alpha. *Int J Med Educ* 2:53
83. Terzis V, Moridis CN, Economides AA (2012) How student's personality traits affect Computer Based Assessment Acceptance: Integrating BFI with CBAAM. *Comput Hum Behav* 28(5):1985–1996. <https://doi.org/10.1016/j.chb.2012.05.019>
84. Tibshirani R (1996) Regression Shrinkage and Selection via the Lasso. *J R Stat Soc: Series B (Methodol)* 58(1):267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
85. Tipping ME (2001) Sparse bayesian learning and the relevance vector machine. *J Mach Learn Res* 1(Jun):211–244. <https://doi.org/10.1162/15324430152748236>
86. Vedel A (2014) The Big Five and tertiary academic performance: A systematic review and meta-analysis. *Pers Individ Differ* 71:66–76. <https://doi.org/10.1016/J.PAID.2014.07.011>
87. Vera M, Cortés JA (2021) Emotional and cognitive aptitudes and successful academic performance: using the ecct. *Int J Environ Res Public Health* 18(24). <https://doi.org/10.3390/ijerph182413184>
88. Vossensteyn JJ, Kottmann A, Jongbloed BW, Kaiser F, Cremonini L, Stensaker B, Hovdhaugen E, Wollscheid S (2015). Dropout and completion in higher education in Europe: Main report. <https://doi.org/10.2766/826962>
89. Webster BJ, Yang M (2012) Transition, induction and goal achievement: first-year experiences of hong kong undergraduates. *Asia Pac Educ Rev* 13(2):359–368. <https://doi.org/10.1007/s12564-011-9185-6>
90. Weisberg S (2005) *Applied linear regression*, vol 528. John Wiley & Sons. <https://doi.org/10.1002/0471704091>
91. Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics Bull* 1(6):80–83. <https://doi.org/10.2307/3001968>
92. Wolters CA, Hussain M (2015) Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacognition Learn* 10:293–311. <https://doi.org/10.1007/s11409-014-9128-9>
93. Wu H, Li S, Zheng J, Guo J (2020) Medical students' motivation and academic performance: the mediating roles of self-efficacy and learning engagement. *Med Educ Online* 25(1):1742,964. <https://doi.org/10.1080/10872981.2020.1742964>
94. Yu X, Jiang F, Du J, Gong D (2019) A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domains. *Pattern Recognit* 94:96–109
95. Yu X, Peng Q, Xu L, Jiang F, Du J, Gong D (2021) A selective ensemble learning based two-sided cross-domain collaborative filtering algorithm. *Inf Process & Manage* 58(6):102,691
96. Yucel R, Bird FL, Young J, Blanksby T (2014) The road to self-assessment: exemplar marking before peer review develops first-year students' capacity to judge the quality of a scientific report. *Assess & Eval High Educ* 39(8):971–986. <https://doi.org/10.1080/02602938.2014.880400>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.