International Journal of Educational
Technology in Higher Education

Open Access

Framework for automatically suggesting remedial actions to help students at risk based on explainable ML and rule-based models

Balqis Albreiki[1,2,3] , Tetiana Habuza[1,3] and Nazar Zaki[1,3]

*Correspondence:
200907523@uaeu.ac.ae

[1] Department of Computer
Science and Software
Engineering, College
of Information Technology, UAE
University, Al Ain, UAE
[2] Office of Institutional
Effectiveness, UAE University, Al
Ain, UAE
[3] Big Data Analytics Center
(BIDAC), UAE University, Al Ain,
UAE

Abstract

Higher education institutions often struggle with increased dropout rates, academic underachievement, and delayed graduations. One way in which these challenges can potentially be addressed is by better leveraging the student data stored in institutional databases and online learning platforms to predict students' academic performance early using advanced computational techniques. Several research efforts have focused on developing systems that can predict student performance. However, there is a need for a solution that can predict student performance and identify the factors that directly influence it. This paper aims to develop a model that accurately identifies students who are at risk of low performance, while also delineating the factors that contribute to this phenomenon. The model employs explainable machine learning (ML) techniques to delineate the factors that are associated with low performance and integrates rule-based model risk flags with the developed prediction system to improve the accuracy of performance predictions. This helps low-performing students to improve their academic metrics by implementing remedial actions that address the factors of concern. The model suggests proper remedial actions by mapping the students' performance in each identified checkpoint with the course learning outcomes (CLOs) and topics taught in the course. The list of possible actions is mapped to this checkpoint. The developed model can accurately distinguish students at risk (total grade < 70%) from students with good performance. The Area under the ROC Curve (AUC ROC) of binary classification model fed with four checkpoints reached 1.0. Proposed framework may aid the student to perform better, increase the institution's effectiveness and improve their reputations and rankings.

Keywords: Student performance, Explainable ML, Students at risk, Early prediction, Rule-based system, Classification

Introduction

Student performance modeling is crucial in evaluating and enhancing the performance of students' learning discourse. The desired outcome is to determine the most effective educational data mining (EDM) methodologies, in terms of extracting meaningful information for predicting students' academic performance, and identify students who may underperform in their learning courses.

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 2 of 26

There have been many attempts to forecast student performance, including the automatic identification of at-risk students, with the aim of ensuring student retention (Bengio et al., 2021) and allocating appropriate courses and resources. Conventionally, educational institutions use older teaching methods to provide technical and non-technical education (Kuzilek et al., 2015). However, a new form of education based on e-learning must be adopted if an educational institution is to overcome the current challenges (Ha et al., 2020; Hussain et al., 2018). The internet has made it easier for modern educational institutions to compete well in the modern environment. Moreover, students can study at home or learn new skills using various e-learning platforms, such as intelligent tutoring systems (ITSs) (Mousavinasab et al., 2021), learning management systems (LMSs) (Costa et al., 2017; Zhao et al., 2020), and massive open online courses (MOOCs) (Al-Rahmi et al., 2019).

The competitive environment also provides higher education institutions with many ways to sustain long-lasting innovation. Data mining (DM) (Hernáandez-Blanco et al., 2019) is particularly effective when combining ideas from different fields, and has been used to extract important information from raw data. Recent studies (Liao et al., 2019; Iatrellis et al., 2021) have identified new possibilities for technology-enhanced learning systems that can be tailored to each student's needs. The application of EDM can ensure a learning environment that is appropriate for specific students (Romero et al., 2013; Prenkaj et al., 2020).

Predicting students' performance

The performance of individual students can be predicted with great accuracy using educational data (Koprinska et al., 2015). Prediction assists students in making informed decisions about which courses to choose based on their skills (Kuzilek et al., 2015), can be used to develop study plans (Ha et al., 2020), and aids instructors and administrators in ensuring that students obtain the best possible outcomes. This minimizes the number of official warning signals, and consequently the expulsion rate, which may otherwise affect an education institution's reputation (Ha et al., 2020). Early predictions of student performance may allow decision-makers to take appropriate action at the right time. Furthermore, it may allow them to plan proper training schedules in order to increase student success. For instance, dropouts may experience increased risk of poverty or antisocial behavior, as well as difficulties adjusting to society. Thus, failing to increase the retention rate may negatively aect students, parents, academic institutions, and society (Ha et al., 2020). The detection of at-risk students can be used to improve student retention rate and institutional effectiveness.

Monitoring students' performance is a challenging task for several reasons, i.e., identification of at-risk students (e.g., special needs, low performance) (Bengio et al., 2021), restricted access to certain aspects of the curriculum / education / assessment (basic skills versus the whole spectrum of courses) (Koprinska et al., 2015), and difficulty in using limited data to supply and predict instructional techniques, interventions, and supports.

Each course has specific requirements for enrollment based on educational background, skill-set, and hands-on experience. The overall objective of automatically identifying students' performance at the course level can help to modify existing programs.

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 3 of 26

Lowering the dropout rate by assisting students in predicting their chances of success in a course before they enroll is therefore crucial (Goga et al., 2015). Student performance may be improved if the course instructors have a better understanding of their students' capacities, allowing teaching tactics to be modified accordingly (Koprinska et al., 2015).

Machine learning and algorithms

Special forms of education, primarily virtual education, have received considerable attention (Lykourentzou et al., 2009). As a result, many businesses and educational institutions focus on automated performance analysis to measure academic success and determine student requisites (Iatrellis et al., 2019; Liao et al., 2021). Various machine learning (ML) algorithms (Ha et al., 2020; Iatrellis et al., 2021; Liao et al., 2019; Tomasevic et al., 2020), are currently being used to train, analyze, and evaluate the performance of students, aided by data collection techniques that improve the learning platform's usability and interactivity. All of this can be equated with artificial intelligence (AI) (Evangelista, 2021).

ML is very accurate in the early prediction of a student's performance (Buenaño-Fernández et al., 2019; Fahd et al., 2022), and can thus be used to improve education programs (Romero et al., 2013), reduce dropout rates (Goga et al. 2020; 2015), and enhance retention rates (Bengio et al., 2021). Numerous studies (Buenaño-Fernández et al., 2019; Fahd et al., 2022; Ha et al., 2020; Iatrellis et al., 2021; Tomasevic et al., 2020) have proposed ML- and statistical-based techniques for the early prediction of students' performance, but only a few have proposed remedial solutions (Goga et al., 2015; Tomasevic et al., 2020; Zhao et al., 2020). The primary purpose of these research papers was to establish a scale that could be used to assess undergraduate students' impressions of course content and determine which were in danger of failing (Ha et al., 2020). They also examined whether novel teaching approaches lowered dropout rates. A third goal was to understand aspects that may have an impact on perceptions of anxiety and performance in a course setting (Goga et al., 2015). Finally, they determined whether or not the instructors would use the suggested approach to enhance student learning (Koprinska et al., 2015; Tomasevic et al., 2020).

Prospective goals

For many years, educators and legislators have been working to create a reliable system that would aid instructors in identifying students who were in danger of poor performance (Evangelista, 2021). However, the most intricate systems are expensive, heavily reliant on data, and only provide forecasts (Goga et al., 2015). Thus, it is important to develop a reliable warning system that does not require the installation of a complex database or high expenditure, so that all students have equitable access to an education and a brighter long-term future. Towards this goal, the present study aims to use ML- and rule-based models to automatically identify and help students who are at risk of failing a course and suggest remedial actions. The ML- and rule-based models operate by finding important patterns in the students' data through EDM (Romero & Ventura, 2013). The overall aim is to help students to achieve their educational goals and for academic institutions to control their dropout rates.

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 4 of 26

Literature review

Educational institutions are finding it increasingly challenging to evaluate and forecast the performance of at-risk students due to a scarcity of labeled data and appropriate statistical techniques (Alboaneen et al., 2022). This has led to an increasing number of students with poor grades and a rise in student dropout rates (Koprinska et al., 2015). Therefore, techniques based on support vector machines (SVMs), random forests (RF), linear regression (LR), and additive regression (AR) have been proposed (Goga et al., 2015; Koprinska et al., 2015). Big data plays a crucial role in addressing real-life challenges, because different data mining techniques can be used to create value from the enormous volumes of data that are continually being created. Some studies have developed their own datasets, while others have used existing datasets (Prenkaj et al., 2020). Only the development of ML methods has made it possible to provide more reliable predictions about students' performance (Li et al., 2012).

Several studies have presented methodologies for using students' grades and course evaluations to forecast the performance of at-risk students (Albreiki et al., 2021a; Altujjar et al., 2016). Techniques such as Naive Bayes classifiers, K-nearest neighbors (KNN), SVMs, and neural networks have revealed the variety of variables that affect students (Koprinska et al., 2015; Kruck & Lending, 2003). For example, Kruck & Lending (2003) discussed the aspects connected with school, community, and family, all equally contributing to putting students at risk of dropping out.

Predicting students' performance in higher education helps to identify students that may underperform in various subjects (Moonsamy et al., (2021). Recognizing the necessary support required by at-risk students can be extremely helpful because instructors can then take timely and appropriate actions to improve the skills of these students (Purwaningsih & Arief, 2018). Moreover, the capacity to anticipate student achievement in a course or program opens doors to new possibilities, such as improving educational outcomes for all students (Alturki et al., 2016). Compared with past practices, the advent of accurate prediction systems that can successfully determine students' performance allows teachers to better distribute resources and teach according to the students' needs.

Learning management systems

One of the more novel ways of assessing student performance is to employ an LMS. The development of e-learning technology has made it simpler for educational institutions to deliver quality learning materials to their students (Hu et al., 2014). These LMSs also give valuable insights into how students interact with the system, their engagement time, and behavior analytics. Parameters such as the number of times a student has interacted with the course content (Zhao et al., 2020), how many times a student has taken quizzes and tests, and how active a student is while viewing an educational video or textual content can easily be recorded and analyzed. However, setting up an LMS requires an enormous amount of time. This is because ensuring that all teachers are comfortable with e-learning demands proper training, which is costly and time-consuming. Moreover, there ongoing administrative expenditures are incurred in ensuring that the interface remains tailored to the requirements. There is also the disadvantage of requiring coding and IT expertise to modify and update the LMS according to the organization's

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 5 of 26

requirements, which places a financial burden on higher education institutions. Finally, several LMSs (Zhao et al., 2020) have adopted a "freemium" model with restricted functionality, with only paid features offering extra support and reporting. This is another challenging issue.

Machine learning algorithms

An effective LMS relies on efficient data processing. This is where ML algorithms (MLAs) come in. The use of MLAs, statistical methodologies, learning analytics, and data mining technologies has enabled researchers to examine and anticipate student performance in higher educational institutions. Different studies have utilized different MLAs, such as regression models (Hasan et al., 2020), to uncover findings related to student performance (Shahiri et al., 2015). For example, one investigation examined how students' programming activity impacts the course results (Watson et al., 2013). In another study, a model was developed to estimate how well students would do in their first college-level course (Kruck & Lending, 2003). A dashboard that allows instructors to monitor students' progress in different courses has also been proposed (Yadav et al., 2012), enabling early intervention when a student is thought to be underperforming in certain courses (Gong et al., 2019). These models have revealed that ML techniques are valuable for the early prediction of students' performance.

A recent study (Alboaneen et al., 2022) used ML and deep learning classifiers to predict student performance. The authors used LR, SVM, KNN, RF, and one neural network-based technique. The mean absolute percentage error was used to evaluate the classifiers predictions. The results showed that the midterm exam score greatly affected students' performance. Finally, the authors concluded that academic factors such as the students' background have a greater impact on performance than demographic factors. Another study (Urkude & Gupta, 2019) proposed a predictive model that outperformed naive Bayes, baggage, boosting, and RF methods in terms of categorizing and predicting students' performance, while a further study (Hu et al., 2014) used a decision tree classifier for early predictions of students' performance. Some recent work (Qazdar et al., 2019) used an ensemble of bagging, boosting, and voting to automatically predict students' performance.

Prediction models for academic success have been established using the ID3 decision tree induction technique (Altujjar et al., 2016). Data relating to students from King Saud University in Riyadh, Saudi Arabia, who were enrolled in the Bachelor of Science degree in information technology were used to train and validate the models. In contrast, a different study (Li et al., 2012) used data from UWF's (University of West Florida) autumn 2008, fall 2009, and fall 2010 semesters to evaluate students' performance in "Elements of Statistics," one of the most popular courses in general education. They summed up the different applicable solutions for different subjects, such as programming courses (Alturki et al., 2016), English language (Purwaningsih & Arief, 2018), and radiology (Cornell-Farrow & Garrard, 2020), as a means of ensuring effective learning and lessening student dropout rates. Likewise, various MLAs have been compared in terms of examining student academic performance (Romero et al., 2013) and enhancing the educational framework (Liao et al., 2019). The accuracy and recall were used to evaluate the robustness of the proposed model. A recent study (Prenkaj et al., 2020) predicted

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 6 of 26

the final exam scores of students in the third week of the term using data collected by instructors using the Peer Instruction methodology.

Most studies (Alapont et al., 2005; Cornell-Farrow & Garrard, 2020) validate the robustness of their models using well-known metrics such as the mean absolute error (MAE), root mean squared error (RMSE) (Costa et al., 2017; Sarker et al., 2013; Wagner et al., 2002), relative absolute error (RAE) (Moonsamy et al., 2021; Zhao et al., 2020), and root relative squared error (RRSE) (Fahd et al., 2022; Ha et al., 2020; Hussain et al., 2018). A literature review of recent studies on predicting students' performance revealed that supervised MLAs, particularly logistic regression, outperformed conventional statistical models in predicting academic performance (Hasan et al., 2020; Shahiri et al., 2015; Namoun & Alshanqiti, 2021) and providing accurate predictions for monitoring student academic progress.

Data mining

MLAs will only work with the data fed into them. This is why they need to be coupled with DM techniques. EDM and ML aid the analysis of classroom settings for students. For example, a case study at Greece's University of Thessaly (Ha et al., 2020) proposed a method for testing student performance. The authors used equivalent educational criteria and measurements to categorize the case study participants. In another study, the authors showed that the demographic data had no impact on classification and regression accuracy (Fahd et al., 2022), and artificial neural networks outperformed traditional MLAs when given student participation and past performance information. Moreover, the authors of a separate study reported that students' final marks might be estimated using ML classifiers based on their prior performance (Buenaño-Fernández et al., 2019). In contrast, other researchers (Zhao et al., 2020) used prediction algorithms and trained them with semester-level performance data provided by course teachers. A forecasting model that predicts the first third of a semester's student learning success has been presented (Dekker et al., 2020), and video learning analytics and DM have been employed to forecast students' overall performance at the start of the semester (Namoun & Alshanqiti, 2021).

Different publications assert the existence of distinctions between data qualities, data complexity, the degree of contribution significance, and the limitations of algorithms used in diverse applications (Zhao et al., 2020). For such purposes, large and complex datasets may be automatically analyzed by ML models, providing accurate results concerning students' performance and minimizing unexpected risks.

At-risk students and dropouts

One of the primary goals of utilizing LMSs, MLAs, and DM is to help at-risk students and prevent dropouts. One study reported that the dropout rate of students in computer programming courses was more than 50%, which was unexpectedly high compared with other courses (Kruck & Lending, 2003). The author reported that students experienced considerable variations in programming courses because of different coding abilities, different teaching methods and materials, and the students' interests, learning styles, and self-discipline. Another study used a supervised naive Bayes classifier to determine student performance in an English language course (Purwaningsih & Arief, 2018).

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 7 of 26

The study revealed that student backgrounds and prior skills at the start of the course could be used as predictors for measuring performance. It is important to note that these previous studies did not determine the possible reasons for students dropping out. Dropouts must be differentiated/segmented depending on student behavior, institutional level, and time. The limited effect of university officials regarding certain reasons for dropping out is another restriction. Finally, the findings of previous studies have revealed that the educational staff of higher education institutions are largely unaware of the dropout problem.

To reduce the dropout rate, we must consider several different perspectives. For example, (Xing et al., 2015) emphasized that student mental health is a crucial factor in determining the likelihood of dropping out. The authors recommended chatbot treatments and a curriculum-wide life-crafting intervention. Recent research (Gupta et al., 2020) employed 12 semi-structured interviews with university staff and LSS (Lean Six Sigma) professionals to better understand student dropout rates and the impact of LSS tools in reducing these rates. The authors suggested that higher education institutions should retain extensive data and educate the appropriate authorities on the effect of student dropout rates so as to establish a student dropout typology. Moreover, the authors emphasized that educational settings should be less punishing. Dropouts can be minimized via consultation and tutoring, because consultation significantly improves the number of students focusing on given activities and reduces the number of inefficient instructors.

Student performance model

One model that takes advantage of all aforementioned strategies is the student performance model (SPM). A recent survey (Albreiki et al., 2021b) highlighted the most promising strategies for predicting students' performance, along with the current limitations and challenges. Different ML and statistical methods have been used to determine the academic and demographic characteristics of those students who are most at risk of failure. Many existing SPMs are based on statistical approaches, using probability and estimation to predict students' performance, and thereby offer a strong basis for decision-making as a means of improving teaching/learning outcomes. Moreover, several studies (Alhassan et al., 2020; Prenkaj et al., 2020) have proposed predictive models and discussed the influence of hidden factors that are peculiar to students, lecturers, the learning environment, and the family, together with their overall effect on student performance, using balanced and unbalanced datasets (Inyang et al., 2019).

Interventions

So how do ML techniques provide remedial interventions for at-risk students? The authors of a recent study (Borrella et al., 2022) used two primary techniques to provide interventions. First, the proposed prediction algorithm identified students at risk of dropping out, and a portion of these students were assigned to an A/B testing experimental environment. Second, the authors employed data analysis to identify target populations of at-risk students. The study recommended that educators assess whether the instruction time is sufficient and students are getting adequate attention, because students need a certain amount of time with appropriate instruction,

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 8 of 26

practice, and feedback. In addition, the study also recommended that educators assess whether the class learning environment promotes opportunities for students to respond and whether the teaching is aligned with students' learning requirements. Instructors should promote one-on-one instruction, which often suits the learning requirements of students who demand more explicit and methodical teaching. As a result, the classroom atmosphere can be improved, and dropouts and suspensions can be reduced. However, the outcomes of this research (Borrella et al., 2022) are subjective due to the diverse range of student backgrounds.

Learning outcomes

The results from the aforementioned interventions can be gauged by the use of learning outcomes. A learning outcome is a statement describing what students should know or do after a class, course, or program, and explains why students should achieve the desired goals. These outcomes assist students in making connections between what they have learned and how they may use it in other situations, such as in their professional lives (Koprinska et al., 2015; Tomasevic et al., 2020; Zhao et al., 2020). The emphasis of learning outcomes is not the quantity of material covered, but how well students can apply what they have learned, both inside the classroom and in the real world (Tomasevic et al., 2020). Moreover, student learning objectives should be obvious, visible, and quantifiable at both the course and program levels, and they should mirror the course and program requirements.

Identifying underachieving students and those who are excelling in school may be simplified by ensuring that program learning outcomes (PLOs) and course learning outcomes (CLOs) are fulfilled. Educators and managers may use PLOs and CLOs to design a wide range of educational initiatives. These may help students improve their grades, and may enhance student counseling and tutoring systems (Tomasevic et al., 2020). Moreover, the student solutions for assessment tasks can be submitted online, and the answers are checked against public and concealed tests established by the instructor. This will quickly enable the instructor to identify students' weaknesses and take adequate measures to ensure that the students obtain the necessary expertise and achieve the desired learning outcomes.

The impact of internet usage data on students' academic performance was the subject of recent research (Waheed et al., 2020). The goal of this study was to analyze and report on students' learning processes and contributions to individual achievement, and the proposed model achieved accuracy of 84–93% (Yukselturk et al., 2014). In addition, hierarchical cluster analysis and association rule mining have been used to determine the ideal number of failed course clusters and course grouping (Marbouti et al., 2016). Furthermore, an ML-based framework has been developed for predicting student performance at a high school in Morocco using school data from 2016–2018 (Alboaneen et al., 2022). Finally, an online undergraduate course's learning activities have been used to construct an early warning system using an LMS (Costa et al., 2017), while student learning outcomes have been predicted based on participation in online educational platforms (Wolff et al., 2013).

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 9 of 26

Methodology

Research objectives

The goal of this study is to examine the potential yield of advanced ML strategies to improve the prediction of students' performance at the course level, Fig. 1 summarizes the overview of methodology of this study. Specifically, we investigate the effectiveness of an "Explainable ML" model in conjunction with educational data for predicting students' final performance in programming courses. This research study develops solutions for identifying and predicting students at risk of failure, and suggests appropriate remedial actions to address the significant factors as early as possible. To address the main objective, we formulate the following tasks:

1. ML techniques are used to predict at-risk students as early as possible using course checkpoints.
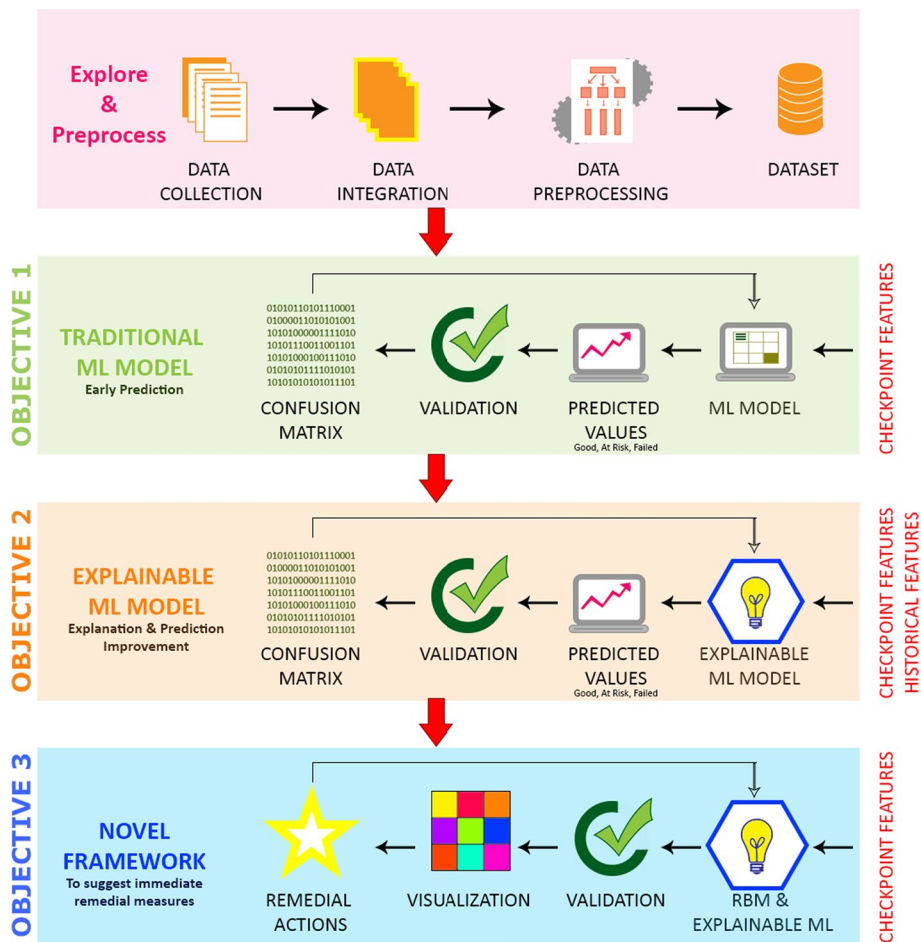2. An Explainable ML model is developed to identify contributing factors that can easily be interpreted by laymen.

Fig. 1 Overview of methodology of this study

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 10 of 26

3 A novel ML-based framework and rule-based models are proposed to improve the identification of students at risk of poor performance during the early stages of the learning process, enabling appropriate interventions to be implemented.

Data collection and dataset description

The educational data used in this study were collected from different sources, such as the Banner system, which contains students' information, instructors that taught programming courses, and documents manually extracted from the Ministry of Education portal. The main data used specifically pertain to programming courses taught to undergraduate students at the College of Information Technology (CIT), United Arab Emirates University (UAEU). The students must take this course to accomplish the university's graduation requirements. Students from other colleges may take the course as an elective as part of their academic study plan. The data represent the performance of students in programming courses over different academic periods from 2016/2017 (fall and spring) until 2020/2021 (fall and spring). General demographics, course registration, and campus details were added to the data. The original dataset contained 730 records with 44 features before data analysis and classification. After removing inconsistent rows and features using univariate feature processing, the final dataset contained 649 samples and 38 features (see Table 1). The courses were not directed or specially designed for the experiments described in this paper. Based on the features of the data, we constructed three nonoverlapping datasets:

- Dataset $D1$ consists of 218 students enrolled in "Algorithms & Problem Solving," a description of which can be found in our previous paper (Albreiki et al., 2021a).
- Dataset $D2$ includes records of 230 students enrolled in "Object-oriented Programming." In addition to the students' performance in this course, we collected some data about their prior performance, demographics, enrolment, etc. (see Historical Features in Table 1).
- Dataset $D3$ consists of 201 students enrolled in "Algorithms & Problem Solving." Along with the students' performance in this course, we added information about the topics and CLOs/PLOs covered in each checkpoint. This allowed us to build a framework for automatically suggesting remedial actions.

Data preprocessing

The data preprocessing was divided into six phases. First, the course assessment files, student data (Banner system), and manually extracted documents were synthesized. Second, the compiled data were cleaned to remove any superfluous entries. Third, because of inconsistencies such as differences in file structures due to courses taught by different instructors, the data were unified to ensure homogeneity (structure unification). Next, missing data values were treated using an imputation technique in which missing entries were assigned the average value of the same coursework components. After data aggregation, standardization was carried out to convert the data from categorical to numerical

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 11 of 26

Table 1 Dataset features

Internal assessment features (checkpoints)		Historical features	
Feature	**Description**	**Feature**	**Description**
ID	Student ID	AGE_ADMITTED	Age of the student when he/she was admitted to university
Quiz1Norm	Score on Quiz 1	AgeCourseStart	Age of the student when he/she took course CSBP219
Quiz2Norm	Score on Quiz 2	Reg. Hrs	Registered Hours
HW1Norm	Score of HW 1	HS_GPA	High School GPA
HW2Norm	Score of HW 2	MATH	Math Grade in High School
HW3Norm	Score of HW 3	PHYS	Physics Grade in High School
HW4Norm	Score of HW 4	Repeated Grade (CSBP119)	How many times the student repeated course CSBP119
MTNorm	Midterm Exam Score	CSBP119-max	Grade of CSBP119 course
Quiz3Norm	Score on Quiz 3	Repeated Grade (CSBP219)	How many times the student repeated course CSBP219
Quiz4Norm	Score on Quiz 4	ESPN	Introduction to Academic English
Quiz5Norm	Score on Quiz 5	CSBP121	Programming Lab I
Quiz6Norm	Score on Quiz 6	MATH105	Calculus I
Qzs	Average Quiz Score	CENG205	Digital Design & Computer Organization
HWs	Average Homework Score	PHYS105	General Physics I
PRJ1Norm	Project Score	CENG202	Discrete Mathematics
FENorm	Final Exam Score	CIT?	Student in College of IT (Yes/No)
TGNorm	Total Grade in CSBP219	AcademicStanding	1 = Good standing, 0 = Probation
		Citizenship	1 = Citizen, 0 = Non-Citizen
		Gender	F/M
		Sponsor	Yes/No
		AlAinResident	0 = Resident outside Al-Ain, 1= Resident of Al-Ain

values, integrate all files into one CSV file, and normalize the marks by employing min-max normalization (rescaling the features to the range [0, 1]).

Finally, before obtaining the final output, we added an additional column based on rules and significant milestones in student performance. We divided the students into three main categories based on their total grade (TG), i.e., Good ($TG \geq 70\%$), AtRisk ($60\% \leq TG < 70\%$), and Failed ($TG < 60\%$) in datasets $D1$ and $D2$; Good ($TG \geq 70\%$) and AtRisk ($TG < 70\%$) in dataset $D3$.

A typical data file structure (see Table 2) was employed following that of (Albreiki et al., 2021a). This structure is shown below:

C_i—name of the predefined checkpoint

$g_{i,j}$—grade of the jth student at checkpoint C_i

$max(g_{C_i})$ —maximum possible grade for checkpoint C_i

m—number of students

n—number of checkpoints in the course

i, j—indices, $i = \overline{1, n}, j = \overline{1, m}$

Table 2 Example of student performance file that the model takes as input

Student ID	C_1	C_2	C_3	$\cdots$	C_n
$Student_1$	$g_{1,1}$	$g_{1,2}$	$g_{1,3}$	$\cdots$	$g_{1,n}$
$Student_2$	$g_{2,1}$	$g_{2,2}$	$g_{2,3}$	$\cdots$	$g_{2,n}$
$Student_3$	$g_{3,1}$	$g_{3,2}$	$g_{3,3}$	$\cdots$	$g_{3,n}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
$Student_m$	$g_{m,1}$	$g_{m,2}$	$g_{m,3}$	$\cdots$	$g_{m,n}$
Max grade	$max(g_{C_1})$	$max(g_{C_2})$	$max(g_{C_3})$	$\cdots$	$max(g_{C_n})$

All three datasets included homework components ($HW_i, i = \overline{1, h^D}$, $HW_{mean} = \frac{1}{h^D} \sum_{i=1}^{h^D} HW_i$), quiz scores ($Qz_i, i = \overline{1, q^D}$, $Qz_{mean} = \frac{1}{q^D} \sum_{i=1}^{q^D} Qz_i$), midterm grades MT, final exam grades FE, and the total grade TG, where $\cdot^D$ denotes the dataset used, $h^{D1} = 4, h^{D2} = 1, h^{D3} = 2$, $q^{D1} = 6, q^{D2} = 4, q^{D3} = 5$. All checkpoints were applied cumulatively up to the final exam as input variables to the model.

Explainable ML model

Recent MLAs are very accurate, but are often considered as black box models. When the model is used for decision-making, it is important to explain the reasons for a specific decision. Therefore, insights into the influence/importance of different features are crucial in increasing the confidence in model predictions. For this purpose, an interpretable model must be designed. This model may provide a quantitative relationship between the input variable and the model output. Local fidelity should also be ensured, meaning that the features that are locally important for a prediction can be identified. Finally, the proposed model should be model-agnostic or explain any MLA (Ribeiro et al., 2016). Let us consider a model m that belongs to the class of interpretable models M. We denote an input of model m as $x = \{x_1, x_2, ..., x_n\} \in \mathbb{R}^n$. The corresponding interpretable representation of the vector x is $\overline{x} = \{b_1, b_2, ..., b_k\} \in \mathbb{R}^{\overline{n}}, b_i = \{0|1\}, i = 1, ..., k$. Vector $\overline{x}$ consists of k components that can explain the model output. The complexity of the model plays a crucial role in its "explainability". Let $\Gamma(m)$ be a measure of the model's complexity. For instance, this may be the depth of the tree in a decision tree model.

In classification algorithms, the output of the classification model is the probability that the input vector corresponds to a certain class. In other words, $f(x) = \{p, x \in \mathbb{R}^n, p \in [0, 1], \mathbb{R}^n \mapsto \mathbb{R}\}$. $\Pi_x(s)$ is some local region around the input vector x, where s is a vector located in proximity to vector x, i.e., the distance from x to s is small. As a distance measure, we could use the Euclidian, Manhattan, or cosine distances, among others. For instance, we can use the Gaussian kernel to represent $\Pi_x(z)$ as:

$$\Pi_x(z) = e^{\left(\frac{||x-s||^2}{2\sigma^2}\right)} \tag{1}$$

where $||x - s|| = \sqrt{\sum_{i=1}^{n}(x_i - s_i)^2}$ is a distance norm (i.e., Euclidian norm) and σ is a width parameter.

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 13 of 26

We can now formulate an optimization problem. To ensure that model m approximates function f in proximity of input vector x, we minimize the loss function $L(f, m, \Pi_x)$ while ensuring that $\Gamma(m)$ remains at an appropriate level. We can interpret the model as:

$$L(f, m, \Pi_x) + \Gamma(m) \rightarrow \min_{m \in M} \tag{2}$$

The features that contribute to the final model output can be identified by performing a search using perturbations. In other words, we learn the behavior of function f using input vectors $\bar{x}$ in the proximity of x calculated with Γ_x. For instance, if m is linear, the fidelity function L is as follows:

$$L(f, m, \Pi_x) = \sum_{s, \bar{s} \in S} \Pi_x(s)(f(s) - m(\bar{s}))^2 \tag{3}$$

where S is the set of all perturbed samples used to solve the optimization problem in Eq. (2).

Research design

The proposed method for the early prediction of students at risk of low performance and suggesting appropriate remedial actions is illustrated in Fig. 2. There is an initial preprocessing phase in which the data are collected, integrated, and processed to form a proper dataset (see Sect. "Data preprocessing"). The preprocessed data are then passed through each of the objectives mentioned previously. The basic principle is to add checkpoint features to the ML model cumulatively. The explicit details are given below. Table 3 summarizes the objectives of this research study.
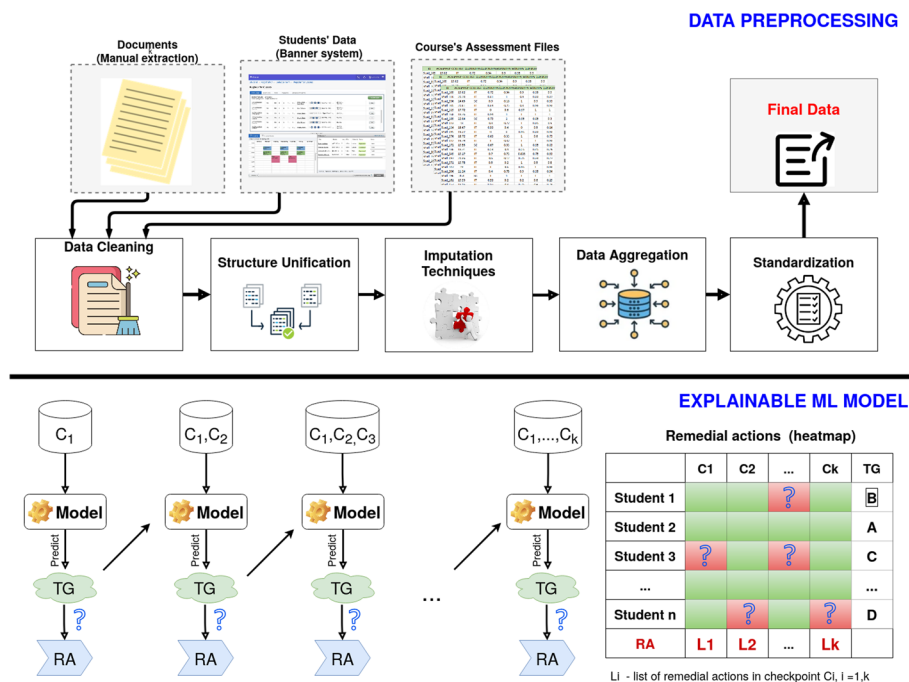

Fig. 2 Pipeline of proposed framework

Table 3 Summary of research objectives

	Objective 1	Objective 2	Objective 3
Dataset (size)	D1 (218), D2 (230)	D1 (218), D2 (230)	D3 (201)
Attributes	Checkpoints	Checkpoints & Historical Features	Checkpoints
Target performance in	TG	MT, TG	MT, TG
Classification	Multiclassification: Good /AtRisk / Failed	Multiclassification:Good / AtRisk / Failed	Binary classification: Good / AtRisk
Remedial Actions	No	No	Yes

MT midterm exam grade, *TG* total grade

For objective 1, we employed advanced ML techniques to identify at-risk students as early as possible using only course checkpoints. Datasets *D*1 and *D*2 were used to classify students into Good, AtRisk, and Failed groups. As the model input, we used all checkpoints obtained prior to the midterm exam (MT). We employed multiclassification prediction models using eight ML techniques, namely the XGB classifier, LightGBM, SVM linear, naive Bayes, ExtraTrees, bagging, RF, and multilayer perceptron. We consistently evaluated whether adding the next checkpoint to the model improved its performance significantly. We also assessed the potential value of historical features in improving the model's accuracy. This allowed us to assess the reliability of the proposed cumulative approach. Five-fold cross-validation was used to generalize the true error rate at the population level.

To address objective 2, we followed the same steps as for objective 1. However, the purpose of this objective was not only to enhance the prediction results, but also to make the model more explainable for non-experts, such as educators and instructors. First, we applied feature selection methods such as information gain, Chi-square test, correlation coefficient, and the mean absolute difference (MAD). This allowed us to find the most informative features with respect to the model output. We then used the local interpretable model-agnostic ML model (see Sect. "Explainable ML model") to provide a qualitative understanding of the relationship between the input variables and the model's response. By explaining a representative set of cases, the user obtains a global understanding of our model. The model provides a generic framework for unraveling black boxes and addressing the "why" behind students' predictions or recommendations for those who are at risk. Finally, we compared the performance of the proposed model using different sets of input features (historical data, checkpoints, historical data and performance in course).

For objective 3, we employed a novel framework using ML- and rule-based models for identifying students at risk of low performance during the early stages of the learning process, enabling appropriate interventions or remedial actions to be taken. We start our analysis by mapping the CLOs to topics and corresponding checkpoints. For this purpose, we worked with three instructors teaching the Algorithms & Problem Solving course. They composed a mapping table and suggested lists of remedial actions for each checkpoint. We applied our rule-based model (Albreiki et al., 2021a) to the checkpoints cumulatively to generate values for the risk flags. Consequently, the ML model was employed to classify students into Good or AtRisk groups. The

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 15 of 26

model inputs were the checkpoints and risk flag values. Finally, based on the model predictions, the corresponding list of remedial actions was invoked for those students identified as being at risk of low performance.

Evaluation measures

To assess the quality of the outcomes given by the classification methods, we calculated the sensitivity, specificity, area under the receiver operating characteristic (ROC) curve (AUC), accuracy, and balanced accuracy metrics. A confusion or error matrix was constructed for each predictive model to show how well it distinguished between classes. The ROC curve and its AUC were used to evaluate the performance of the classifiers and summarize the trade-off between the true positive rate (TPR) and false positive rate (FPR) using different probability thresholds. We define:

$$TPR(sensitivity) = \frac{TP}{TP + FN} \tag{4}$$

$$TNR(specificity) = \frac{TN}{TN + FP} \tag{5}$$

The overall accuracy of the model is defined as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where *TP*, *TN*, *FP*, *FN* are the true positive, true negative, false positive, and false negative values representing the confusion matrix of the classification model, respectively. All models were trained using k-fold cross-validation. The metrics were calculated for each fold separately, and then the averaged values were used as the final measure.

Experimental results

In this section, we present the main results from the experiments outlined in Sect. 3. We show that the advanced and explainable ML- and rule-based models can improve the identification of students at risk of low performance during the early stages of the learning process, so that appropriate interventions can be implemented.

Exploratory data analysis

First, we inspected the attributes in datasets $D1$, $D2$, $D3$ for Gaussianity. A Shapiro–Wilk test revealed the non-normal distribution of all attributes. Therefore, we utilized non-parametric statistical tests for further analysis. To check whether the data from the studied categories came from a common distribution, we applied the Kruskal–Wallis test to continuous features and the Chi-square test to quantitative features.

> **D1:** Of the 218 students in the dataset, 60.09/16.97/22.94% were identified as being in the Good/AtRisk/Failed groups, respectively. All groups were significantly different in terms of students' performance for all checkpoints ($p < 0.05$). A statistical test revealed no significant differences between genders ($p = 0.458727$) for the observed groups.

D2: Of the 230 students, 57.39/17.39/25.22% were identified as being in the Good/ AtRisk/Failed groups, respectively. The observed groups were significantly different in terms of all course checkpoints ($p < 0.05$). This trend was also evident when we compared grades in previously taken courses. For instance, the grades in high school math ($p = 3.82081 \times 10^{-5}$), high school physics ($p = 4.43989 \times 10^{-7}$), Calculus I ($p = 9.47833 \times 10^{-5}$), and Algorithms & Problem Solving ($p = 5.76147 \times 10^{-18}$) differ significantly between the Failed, AtRisk, and Good groups. The number of times the course was repeated also contributes to the segregation ($p = 1.55162 \times 10^{-08}$). Historical features revealed that the admitted age, college, and gender had no effect on the total course scores.

D3: Of the 1 students from eight different sections, 81.59/18.41% were identified as being in the Good/AtRisk groups, respectively. A statistical test revealed significant differences between the performance for all checkpoints, except homework assignments. No influence of term or year on performance was evident ($p > 0.05$). There were significant differences between groups in terms of sections ($p = 0.00278$). This may be related to teaching style as well as gender differences. Due to the gender segregation policy in UAEU, each section is offered for either male or female students. We applied our previously proposed model (Albreiki et al., 2021a) to D3 to identify at-risk students at early stages. All risk flag values differed significantly between the Good and AtRisk groups. The number of remedial actions invoked was also significantly different ($p = 5.93659 \times 10^{-15}$).

Correlation analysis shows that students' performance for all checkpoints is positively correlates with MT, and TG.

ML techniques for predicting at-risk students using course checkpoints

We divided the students into three main classes based on their total grades for the course (Good, AtRisk, and Failed). We applied eight MLAs (XGB classifier, Light-GBM, SVM linear, naive Bayes, ExtraTrees, bagging, RF, and multilayer perceptron) to D1 and D2 to predict the groups of students based on their TG performance. We considered only those checkpoints before the midterm exam, which are Quiz1Norm, HW1Norm, Quiz2Norm, and HW2Norm for D1 and Quiz1Norm, HW1Norm, and Quiz2Norm for D2. Finally, we calculated the precision, recall, F1-score, and AUC for all of the algorithms. Table 4 summarizes the results for objective 1.

Table 4 Performance of classification models in predicting students' groups from checkpoints before the midterm exam (datasets D1 and D2)

Dataset	D1	D2
Predictors	Qz1, HW1, Qz2,HW2	Qz1, HW1, Qz2
Correlation	Positive 0.66%	Positive 0.54%
Best ML classifier	ExtraTrees	ExtraTrees
Accuracy	0.86	0.87
ROC AUC	0.96	0.95

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 17 of 26

For D1, the ExtraTrees classifier achieved the best performance for this objective. It outperformed the other seven state-of-the-art algorithms with an AUC score of 0.96 and an accuracy score of 0.86. For D2, the ExtraTrees classifier outperformed the other algorithms, achieving an accuracy score of 0.87 and an AUC score of about 0.95, as shown in Table 5.

Advanced and explainable ML Model for enhancing prediction results by adding prior knowledge

Even though the traditional ML model successfully predicts at-risk students, it cannot identify the factors that contribute to students falling into this category. Thus, we conducted a series of experiments to identify at-risk students at a sufficiently early stage and predict their MT and TG performance during the course period. The predictions were obtained in three experiments using different features, as described below:

- Experiment 1: Using only historical features. We used 18 features from the dataset in this experiment. These features cover historical student data, such as the student's age, registered hours, high school GPA, math grade, physics grade, number of repeated programming courses, citizenship, gender, sponsor, residency, and so on (see Table 1).
- Experiment 2: Using only course checkpoints. We used three features from the dataset, namely Quiz1Norm, Quiz2Norm, HW1Norm.
- Experiment 3: In this mixed experiment, we combined all of the 21 features used in experiments 1 and 2.

We used the same eight ML classifiers. The attribute feature_importances in Python were used as a feature selection method to improve the efficiency and effectiveness of the predictive model. Figure 3 shows the most important features in the dataset. Based on the ten most important features identified by each classifier, we then attempted to predict which of the students would fall into the three groups of Good, AtRisk, and Failed.

Table 6 presents the prediction results using the combined features. Based on the ten most important features, we were able to predict the groups of students based on their MT and TG performance with AUC scores of 0.95 and 0.97, respectively. By incorporating prior knowledge and selecting the most important data points, we were able to improve the prediction results. Table 6 also shows that there are overlapping features/predictors (such as HS_GPA, Qz1Norm, CENG205, HW1Norm, MATH, and PHYS) that affect the performance of the students in this course.

Table 5 Performance of ML models (AUC ROC) classifying students into Good, AtRisk, and Failed groups

	XGB	LightGBM	SVM	GaussianNB	ExtraTrees	Bagging	RandomForest	MLP
D1	0.9274	0.9452	0.8460	0.8115	**0.9627**	0.8907	0.9620	0.8089
D2	0.9230	0.9195	0.8430	0.8499	**0.9489**	0.8498	0.905	0.8509

⁻Checkpoints before the MT were used as model inputs

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49
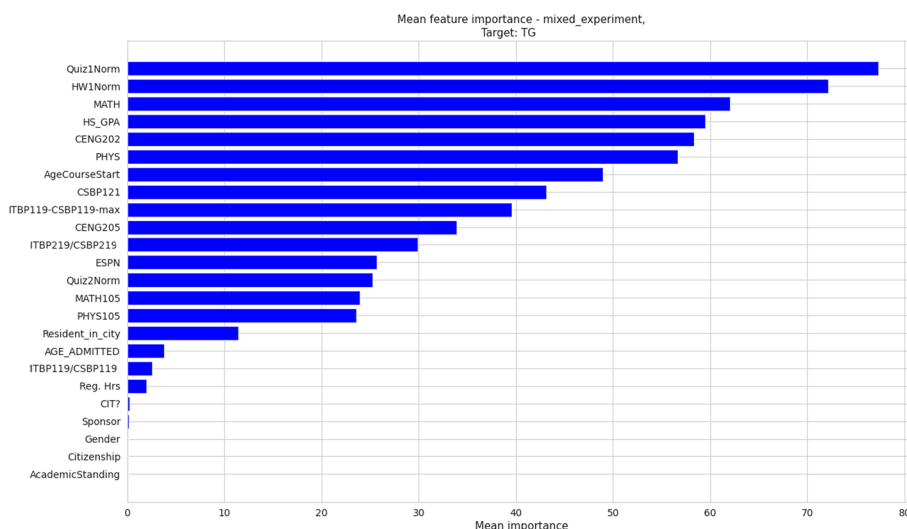
Page 18 of 26

Fig. 3 Important features for predicting students' groups (Failed/AtRisk/Good) from D2 prior knowledge and course checkpoints

Table 6 Performance of the classification model in predicting students' groups from prior knowledge and course checkpoints in dataset D2

Target performance in	MT	TG
Predictors	HS_GPA, Qz1Norm, AgeCourseStart, PHYS, Qz2Norm, HW1Norm, MATH, CENG202, CSBP119, CSBP119-max	Qz1Norm, CENG202, HW1Norm, MATH, HS_GPA, AgeCourseStart, PHYS, CENG205, CSBP119-max, Qz2Norm
Correlation	Positive 0.56% (CSBP119)	Positive 0.79% (MATH, PHYS)
Best ML classifier	XGB	ExtraTrees
Accuracy	0.83	0.84
ROC AUC	0.95	0.96

After predicting the students' performance successfully, our objective was to generate trust in our model. For this, it is important to explain the model to ML experts and domain experts such as instructors and educators. As such, Fig. 4 presents the results after Explainable ML was run for experiment 3. There are three sections in Fig. 4: Failed students are displayed in blue, AtRisk students are indicated in orange, and Good students are shown in green. All three sections consist of three columns.

The left-hand side of the visualization (blue section) presents the predictive probability distribution per class. This student will fail with 90% confidence. Based on the LGB model results, the features with the most influence on the "Failed" class are presented on the right-hand side. In the center of the plot, we see a condition per influential feature and its strength (i.e., contribution/influence to the model). We find that 45% of this score can be attributed to the "Repeated Grade (ITBP219/CSBP219) " value, 20% of this score comes from Quiz1Norm being less than or equal to 0.42 (normalized value), and the remainder is attributable to the values of HW1m CENG205, CENG202, CSBP121, PHYS, MT, MATH, and so on.

The AtRisk student falls into the orange section with 97% confidence. Based on the LGB model results, the center of the plot gives a condition per influential feature and
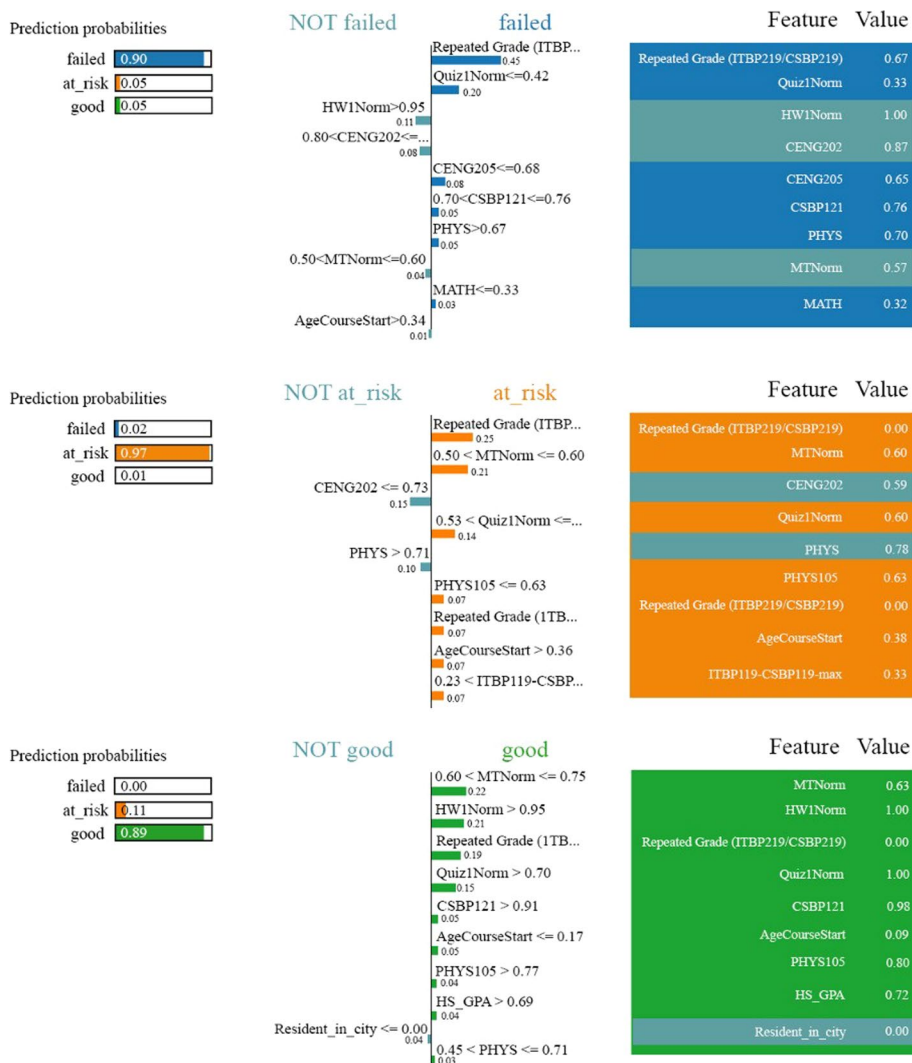
Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 19 of 26

Fig. 4 Results of explainable ML model trained on prior knowledge and course checkpoint data from dataset D2

its strength. In this case, 25% of the score can be attributed to the "Repeated Grade (ITBP219/CSBP219) " value, 21% of the score comes from MT being greater than 0.50 and less than or equal to 0.60, and the remainder is attributable to the values of CENG202, Qz1, PHYS, PHYS105, and so on. Finally, the green section gives the predictive probability distribution per class for a student classified as "Good" with 89% confidence. Some 22% of this score comes from the MT value being between 0.6 and 0.75, 21% can be attributed to the HW1 value being greater than 0.95, and the remainder comes from the other values.

Novel framework and immediate remedial actions for improving students' performance

We first created a mapping table to link the CLOs with the topics and assessment checkpoints. Three instructors teaching the D3 course composed Table 7. From this table, we can see that some assessments address two or more topics. For each checkpoint, a list

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 20 of 26

Table 7 Contribution of assessments to topics and CLOs

Topic(s)	Topic description	CLO1	CLO2	CLO3	CLO4	RA
Topic 1(T1)	Computer fundamentals and evolution of programming languages (Chap. 1)	Qz1, FQ1				RA_1
	Concepts and properties of algorithms. Problem-solving process. Problem solving using pseudocode (Chap. 1)					
Topic 2 (T2)	Introduction to Java (Chap. 1)		Qz1, FQ1			RA_2
Topic 3 (T3)	Identifiers, literals, operators, variables, expressions, and data types (Chap. 2)		Qz2, FQ1			RA_3
Topic 4 (T4)	Object definition of predefined classes with string or math class as an example (Chap. 3)	HW1, MQ3, FQ1	HW1, MQ3, FQ1			RA_4
Topic 5 (T5)	Reading and writing from keyboard and files (I/O events) (Chap. 3)	HW1, MQ4, FQ1	HW1, MQ4, FQ1			RA_5
Topic 6 (T6)	Control structure—selection (Chap. 4) • If switch • Develop simple algorithms that use decision-making constructs	Qz3, MQ1	Qz3, MQ2, FQ2			RA_6
Topic 7 (T7)	Control structure—repetition (Chap. 5) • While, For, Do-While • Implement searching and selecting algorithms	Qz4	Qz4		Qz4	RA_7
Topic 8 (T8)	Use of predefined functions (Chap. 7) • User-defined functions (Chap. 7) • Develop simple algorithms to solve problems using methods	HW2, FQ3	HW2, FQ3		HW2	RA_8
Topic 9 (T9)	One-dimensional array declaration, definition, initialization, and use (Chap. 9)			Qz5, FQ4	Qz5, FQ4	RA_9
Topic 10 (T10)	Passing arrays to methods (Chap. 9) • Two-dimensional arrays (Chap. 9). • Implement searching (sequential/ binary) • Sort an array using bubble/ insertion sort			FQ4	FQ4	RA_{10}

$RA_i, i = 1, 10$ are the list of RAs invoked for each topic/checkpoint

of remedial actions $(RA_i, i = 1, 10)$ was proposed. Before the beginning of the course, instructor should compose such table. Once it is done, the proposed framework can be used by feeding the model with formative or summative assessments.

We now propose a novel framework that uses ML- and rule-based models to identify students at risk of low performance during the early stages of the learning process, enabling appropriate interventions to be implemented. Our model combines a rule-based model (Albreiki et al. 2021a) with binary ML classification to predict each student class based on the students' cumulative grades, i.e., Good ($TG \geq 70\%$) and AtRisk ($TG < 70\%$) in D3. Using the rule-based model, we can generate risk flag (*RF*) features every time

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 21 of 26

new checkpoint values are inserted into the model. When student performance drops below a certain threshold (less than 70%), the cumulative *RF* value is updated (Albreiki et al., 2021a). We then add the checkpoints and *RF* features to the model cumulatively to predict the performance of students based on their groups. In addition, we compare the output of the proposed model with two sets of input features (course checkpoints only and checkpoints with *RF* features). Based on the mean AUC value, the ExtraTrees classifier performed best with both sets of input features (see Table 8), outperforming the other seven classifiers.

Table 9 presents the mean AUC values of the best classifier for both sets of input features. The prediction results clearly improved as the features were cumulatively added. Table 9 also shows that, by adding risk flags from the rule-based model, the performance improved by 2.31%. As a result, we can predict the students' performance at the first checkpoint of the course with a reasonable level of accuracy, which will benefit both students and instructors. Finally, proper remedial actions can be taken during the course by mapping the predicted risk probability to a list of actions associated with each checkpoint, as shown in Fig. 5.

To validate the usage of the proposed framework, we assessed the distribution of the total grade values with respect to the number of remedial actions. Figure 6 shows that a greater number of remedial actions corresponds to a lower total grade. The linear relationship between the number of remedial actions and the total grade was also assessed using Pearson's correlation coefficient. The calculated value of -0.735 is statistically significant ($p = 9.25 \times 10^{-36}$). Therefore, the proposed customized model can be considered and used as an effective warning system to identify at-risk students in the early stages of a course.

Discussion and future work

Several studies using ML classifiers to predict student performance have obtained varying degrees of accuracy—56.25% (Yadav et al., 2012), 65% (Romero et al., 2013), 80% (Muñoz-Carpio et al., 2021), 85% (Iatrellis et al., 2021), 93% (Evangelista, 2021), and an AUC score of 0.79 (Liao et al., 2019). However, the present study has proposed a method that obtained 96% accuracy in terms of predicting the total grade as early as possible before the midterm exam. Furthermore, ML- and statistical-based techniques for early prediction of students' performance have been utilized in a variety of studies (Buenaño-Fernández et al., 2019; Fahd et al., 2022; Ha et al., 2020; Iatrellis et al., 2021; Tomasevic et al., 2020). Nonetheless, previous works (Ha et al., 2020; Iatrellis et al., 2021; Tomasevic et al., 2020) primarily focused on detecting at-risk students, and only a few explainable ML and rule-based models have been discussed. These studies did not examine the features that are most influential in predicting

Table 8 Performance of ML models (AUC ROC) classifying students into Good and AtRisk groups

	XGB	LightGBM	SVM	GaussianNB	ExtraTrees	Bagging	RandomForest	MLP
CP	0.9447	0.9690	0.9220	0.9349	**0.9918**	0.9422	0.9601	0.9231
CP + RF	0.9842	0.9781	0.9390	0.9606	**1.000**	0.9777	0.9980	0.9455

⁻*CP* checkpoints; *RF* risk flag

⁻Checkpoints before MT from D3 dataset are used as model inputs

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49
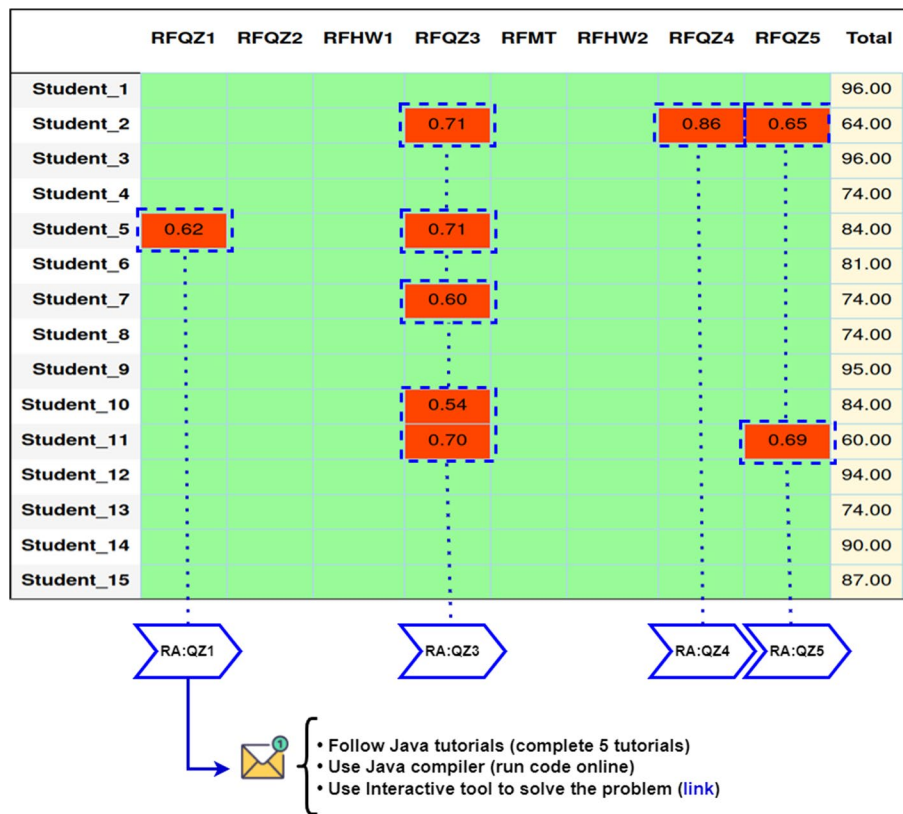
Page 22 of 26

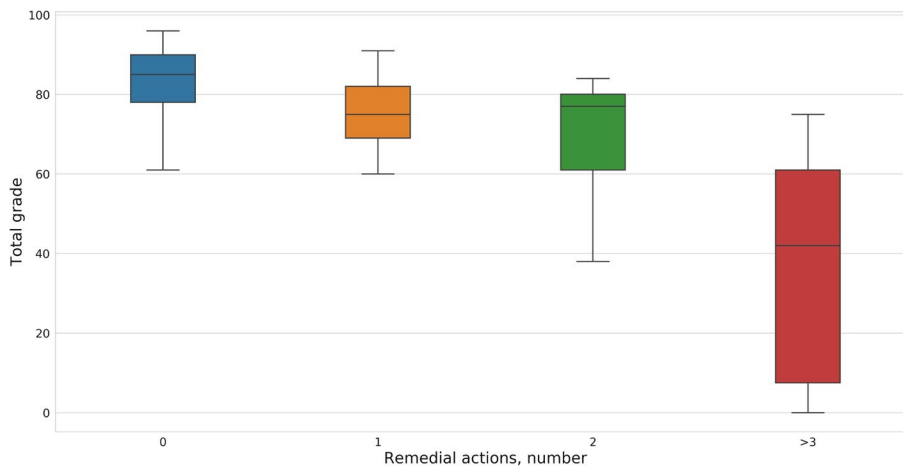**Fig. 5** Visualization of at-risk students using the heat-map technique

Fig. 6 Relationship between total grade in the course and the number of remedial actions invoked

students' performance or identify what factors put a student at risk. In contrast, the proposed method has not only predicted the performance in the total grade with high accuracy, but also produced explainable ML outputs, providing insightful and useful information to non-experts about the features that affect the students' total grade, either the course checkpoints (e.g., Qz1, HW2) or student-based factors (e.g., high school GPA, high school grade, pre-requisite courses, age).

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 23 of 26

Table 9 AUC performance of ExtraTrees model classifying students into Good and AtRisk groups

Features	C1	C1,C2	C1,C2,C3	C1, C2, C3, C4
C	0.664	0.930	0.946	0.992
C + RF	0.691	0.958	0.959	1.000
Gain	+ 4.07%	+ 3.01%	+ 1.37%	+ 0.8%

C_i checkpoints added cumulatively

RF risk flag added to each checkpoint

Early predictions of at-risk students' performance are crucial. Providing relevant and appropriate remedial solutions to these students is another important problem. Several studies (Gupta et al., 2020; Koprinska et al., 2015; Tomasevic et al., 2020) have provided a list of remedial interventions for students considered to be at risk of poor performance, such as an intense academic program (Tomasevic et al., 2020), mental health support (Xing et al., 2015), less punishing educational settings (Gupta et al., 2020), and timely feedback (Borrella et al., 2022). However, the available research does not provide any clear suggestions on how these remedial solutions assist at-risk students. The present research has suggested proper remedial actions by mapping the students' performance in each checkpoint with the CLOs and topics taught in the course. For example, if the proposed model predicts that a student will not perform well in quiz 2, the student will be directly notified that he/she needs to take specific remedial actions. The list of possible actions is mapped to this checkpoint. This will help the student to perform better and increase the institution's effectiveness.

In future research, the authors aim to implement this framework as an automated solution for academic institutions and test it in a real settings. For example, if a student is at risk, an automatic notification will be sent to the student, and the instructor will be notified with a list of suggested remedial actions. Moreover, the authors hope to improve the prediction results by tuning the hyperparameters and designing more sophisticated features using deep learning models. Finally, the authors will seek to apply this model to other datasets (courses) to validate the model output, and will liaise with instructors to obtain further feedback and inputs.

Conclusions

Early predictions of students' academic performance can play a significant role in planning suitable interventions, such as student counselling, intelligent tutoring systems, continuous progress monitoring, and policymaking. In particular, such interventions can improve academic performance during the learning process and reduce the number of students who drop out or graduate late. As such, effective prediction models directly help educational institutions improve their reputations and rankings. Despite recent technological advances, educational institutions continue to face issues obtaining early and accurate predictions of students' performance due to the non-incorporation of performance modules in most online and offline learning platforms. Therefore, an accurate prediction model of student performance is an urgent requirement for educational institutions. Furthermore, assessing students' performance in the early stages of the learning process helps facilitate the implementation of suitable strategies to mitigate the factors leading to dropouts or low performance at both the student and instructor levels.

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 24 of 26

This research study has developed a model that accurately identifies students who are at risk of low performance, while also delineating the factors that contribute to this phenomenon. The model employs explainable ML techniques to delineate the factors that are associated with low performance and integrates rule-based model risk flags with the developed prediction system to improve the accuracy of performance predictions. This may help low-performing students to improve their academic metrics by implementing remedial actions that address the factors of concern.

Abbreviations

Acc	Accuracy
AI	Articial intelligence
AUC	Area under the curve
AR	Additive regression
CLOs	Course learning outcomes
CSV	Comma-separated value
EDM	Educational data mining
MAE	Mean absolute error
ML	Machine learning
MLAs	Machine learning algorithms
PLOs	Program learning outcomes
ROC	Receiver operating characteristic curve
RA	Remedial action
RF	Random forest
RBM	Rule based model
Spec	Specificity
SVM	Support vector machine
HW	Homework assignment
ITS	Intelligent tutoring systems
LMS	Learning management systems
LGB	Light gradient boosting
LR	linearregression
MOOC	Massive open online courses
MT	Mid-term exam
Qz	Quiz
XGB	eXtreme gradient boosting

Availability of data and materials
The data of this study"EduRisk" is available from the corresponding author upon reasonable request through bi-dac.com/download.

Declarations

Competing interests
The authors declare that they have no competing interests.

References

Al-Rahmi, W., Aldraiweesh, A., Yahaya, N., Kamin, Y. B., & Zeki, A. M. (2019). Massive open online courses (moocs): Data on higher education. *Data in Brief, 22,* 118–125.

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 25 of 26

Alapont, J., Bella-Sanjuán, A., Ferri, C., Hernández-Orallo, J., Llopis-Llopis, J., & Ramírez-Quintana, M. (2005). Specialised tools for automating data mining for hospital management. In: Proceedings of First East European Conference on Health Care Modelling and Computation, pp 7–19.

Alboaneen, D., Almelihi, M., Alsubaie, R., Alghamdi, R., Alshehri, L., & Alharthi, R. (2022). Development of a web-based prediction system for students' academic performance. *Data, 7*(2), 21.

Albreiki, B., Habuza, T., Shuqfa, Z., Serhani, M. A., Zaki, N., & Harous, S. (2021). Customized rule-based model to identify at-risk students and propose rational remedial actions. *Big Data and Cognitive Computing, 5*(4), 71.

Albreiki, B., Zaki, N., & Alashwal, H. (2021). A systematic literature review of student'performance prediction using machine learning techniques. *Education Sciences, 11*(9), 552.

Alhassan, A., Zafar, B., & Mueen, A. (2020). Predict students' academic performance based on their assessment grades and online activity data. *International Journal of Advanced Computer Science and Applications (IJACSA) 11*(4), 185–194.

Altujjar, Y., Altamimi, W., Al-Turaiki, I., & Al-Razgan, M. (2016). Predicting critical courses affecting students performance: A case study. *Procedia Computer Science, 82*, 65–71.

Alturki, R. A., et al. (2016). Measuring and improving student performance in an introductory programming course. *Informatics in Education-An International Journal, 15*(2), 183–204.

Bengio, Y., Lecun, Y., & Hinton, G. (2021). Deep learning for AI. *Communications of the ACM, 64*(7), 58–65.

Borrella, I., Caballero-Caballero, S., & Ponce-Cueto, E. (2022). Taking action to reduce dropout in moocs: Tested interventions. *Computers & Education, 179*(104), 412.

Buenaño-Fernández, D., Gil, D., & Luján-Mora, S. (2019). Application of machine learning in predicting performance for computer engineering students: A case study. *Sustainability, 11*(10), 2833.

Cornell-Farrow, S., & Garrard, R. (2020). Machine learning classifiers do not improve the prediction of academic risk: Evidence from australia. *Communications in Statistics: Case Studies, Data Analysis and Applications, 6*(2), 228–246.

Costa, E. B., Fonseca, B., Santana, M. A., de Araújo, F. F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior, 73*, 247–256.

Dekker, I., De Jong, E. M., Schippers, M. C., Bruijn-Smolders, D., Alexiou, A., Giesbers, B., et al. (2020). Optimizing students' mental health and academic performance: AI-enhanced life crafting. *Frontiers in Psychology, 11*, 1063.

Evangelista, E. (2021). A hybrid machine learning framework for predicting students' performance in virtual learning environment. *International Journal of Emerging Technologies in Learning (iJET) 16*(24), 255–272.

Fahd, K., Venkatraman, S., Miah, S. J., & 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, 27*, 3743–3775.

Goga, M., Kuyoro, S., & Goga, N. (2015). A recommender for improving the student academic performance. *Procedia-Social and Behavioral Sciences, 180*, 1481–1488.

Gong, B., Nugent, J. P., Guest, W., Parker, W., Chang, P. J., Khosa, F., & Nicolaou, S. (2019). Influence of artificial intelligence on canadian medical students' preference for radiology specialty: Anational survey study. *Academic Radiology, 26*(4), 566–577.

Gupta, S. K., Antony, J., Lacher, F., & Douglas, J. (2020). Lean six sigma for reducing student dropouts in higher education-an exploratory study. *Total Quality Management & Business Excellence, 31*(1–2), 178–193.

Ha, D. T., Loan, P. T. T., Giap, C. N., & Huong, N. . TL. (2020). An empirical study for student academic performance prediction using machine learning techniques. *International Journal of Computer Science and Information Security (IJCSIS) 18*(3), 21–28.

Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U., & Sattar, M. U. (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences, 10*(11), 3894.

Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., & Navarro-Colorado, B. (2019). A systematic review of deep learning approaches to educational data mining. *Complexity, 2019*(1):1–22.

Hu, Y. H., Lo, C. L., & Shih, S. P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior, 36*, 469–478.

Hussain, M., Zhu, W., Zhang, W., & Abidi, SMR. (2018). Student engagement predictions in an e-learning system and their impact on student course assessment scores. *Computational Intelligence and Neuroscience*. https://doi.org/10.1155/2018/6347186

Iatrellis, O., Savvas, I. K., Fitsilis, P., & Gerogiannis, V. C. (2021). A two-phase machine learning approach for predicting student outcomes. *Education and Information Technologies, 26*(1), 69–88.

Inyang, U. G., Eyoh, I. J., Robinson, S. A., & Udo, E. N. (2019). Visual association analytics approach to predictive modelling of students' academic performance. *International Journal of Modern Education & Computer Science 11*(12), 1–13.

Koprinska, I., Stretton, J., & Yacef, K. (2015). Students at risk: Detection and remediation. In: EDM, pp 512–515.

Kruck, S., & Lending, D. L. D. (2003). Predicting academic performance in an introductory college introductory college-level is course level is course. *Information Technology, Learning, and Performance Journal, 21*(2), 9.

Kuzilek, J., Hlosta, M., Herrmannova, D., Zdrahal, Z., Vaclavek, J., & Wolff, A. (2015). Ou analyse: Analysing at-risk students at the open university. *Learning Analytics Review*, 1–16.

Li, K., Uvah, J., & Amin, R. (2012). Predicting students' performance in elements of statistics. *Online Submission, 10*, 875–884.

Liao, S. N., Zingaro, D., Thai, K., Alvarado, C., Griswold, W. G., & Porter, L. (2019). A robust machine learning technique to predict low-performing students. *ACM Transactions on Computing education (TOCE), 19*(3), 1–19.

Lykourentzou, I., Giannoukos, I., Mpardis, G., Nikolopoulos, V., & Loumos, V. (2009). Early and dynamic student achievement prediction in e-learning courses using neural networks. *Journal of the American Society for Information Science and Technology, 60*(2), 372–380.

Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education, 103*, 1–15.

Albreiki *et al. Int J Educ Technol High Educ* 2022, **19**(1):49

Page 26 of 26

Moonsamy, D., Naicker, N., Adeliyi, TT., & Ogunsakin, R. E. (2021). A meta-analysis of educational data mining for predicting students performance in programming. *International Journal of Advanced Computer Science and Applications, 12*(2), 97–104.

Mousavinasab, E., Zarifsanaiey, N., Niakan Kalhori, R. S., Rakhshan M., Keikha L., & Ghazi Saeedi M. (2021). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments, 29*(1), 142–163.

Muñoz-Carpio, J. C., Jan, Z., & Saavedra, A. (2021). Machine learning for learning personalization to enhance student academic performance. In: *LALA*, pp 88–99.

Namoun, A., & Alshanqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences, 11*(1), 237.

Prenkaj, B., Velardi, P., Stilo, G., Distante, D., & Faralli, S. (2020). A survey of machine learning approaches for student drop-out prediction in online courses. *ACM Computing Surveys (CSUR), 53*(3), 1–34.

Purwaningsih, N., & Arief, D. R. (2018). Predicting students' performance in english class. In: AIP Conference Proceedings, AIP Publishing LLC, vol 1977, p 020020.

Qazdar, A., Er-Raha, B., Cherkaoui, C., & Mammass, D. (2019). A machine learning algorithm framework for predicting students performance: a case study of baccalaureate students in morocco. *Education and Information Technologies, 24*(6), 3577–3589.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016)."why should i trust you?" explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp 1135–1144.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), 12–27.

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use moodle courses. *Computer Applications in Engineering Education, 21*(1), 135–146.

Sarker, F., Tiropanis, T., & Davis, HC. (2013). Students' performance prediction by using institutional internal and external open data sources. eprintssotonacuk.

Shahiri, A. M., Husain, W., et al. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science, 72*, 414–422.

Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & education, 143*(103), 676.

Urkude, S., & Gupta, K. (2019). Student intervention system using machine learning techniques. *International Journal of Engineering and Advanced Technology, 8*(6), 21–29.

Wagner, E. P., Sasser, H., & DiBiase, W. J. (2002). Predicting students at risk in general chemistry using pre-semester assessments and demographic information. *Journal of Chemical Education, 79*(6), 749.

Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from vle big data using deep learning models. *Computers in Human behavior, 104*(106), 189.

Watson, C., Li, F. W., & Godwin, J. L. (2013). Predicting performance in an introductory programming course by logging and analyzing student programming behavior. In: 13 IEEE 13th international conference on advanced learning technologies, IEEE, pp 319–323.

Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2013). Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In: Proceedings of the third international conference on learning analytics and knowledge, pp 145–149.

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior, 47*, 168–181.

Yadav, S. K., Bharadwaj, B., & Pal, S. (2012). Data mining applications: A comparative study for predicting student's performance. arXiv preprint arXiv:1202.4815

Yukselturk, E., Ozekes, S., & Türel, Y. K. (2014). Predicting dropout student: An application of data mining methods in an online education program. *European Journal of Open, Distance and e-learning, 17*(1), 118–133.

Zhao, Q., Wang, J. L., Pao, T. L., & Wang, L. Y. (2020). Modified fuzzy rule-based classification system for early warning of student learning. *Journal of Educational Technology Systems, 48*(3), 385–406.

Publisher's Note

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