Student Performance Prediction Model Based on Course Description and Student Similarity

David Mäder [a], Maja Spahic-Bogdanovic [a,b,*] and Hans Friedrich Witschel [a]
[a] *FHNW University of Applied Sciences and Arts Northwestern Switzerland, Switzerland*
E-mails: david.maeder@alumni.fhnw.ch, maja.spahic@fhnw.ch, hansfriedrich.witschel@fhnw.ch
[b] *University of Camerino, Italy*
E-mail: maja.spahic@fhnw.ch

Abstract. Choosing courses at the beginning of each semester is a complex decision that affects students' future careers and academic performance, especially when given the freedom to choose. Among other factors, the expected grade at the end of the semester and/or the expected ability to successfully complete a course plays an important role in course selection. This paper introduces a prototype for predicting student performance using state-of-the-art natural language processing techniques. The prototype, designed to assist students in course selection, uses historical course enrollment data and current course descriptions to predict possible grades and warn students of possible negative performance. A large language model, BERT, was used to analyse text and create course description embeddings. For this purpose, descriptions of courses a student has attended were considered and formed the basis for the student knowledge profile. In addition, student performance profiles are created by examining grades from the historical enrolment data. This two-pronged analysis is used to identify patterns that lead to negative study results. Although the idea of creating knowledge profiles based on course descriptions is promising, the evaluation showed room for improvement in terms of accuracy and recall.

Keywords: Performance prediction, Student's performance, Decision making, Educational data mining

1. Introduction

In study programs that offer a freedom of choice (e.g. elective modules) to students, course selection at the start of each semester is a complex decision affecting students' future careers and academic performance. Othman, Mohamad, and Barom [17] identified, based on a survey with 396 students, five key factors influencing students during course selections: class and lecturer, time-space, ease and comfort, course mate, and commitment factors. Each factor contains several items that influence the decision. In four categories, the grades were mentioned as one of the items; for example, lecturers tend to give marks easily or are lenient when grading. The possible grades that students can achieve are important for the course selection. Therefore, helping students in this decision-making process and predicting their grades before enrollment is valuable.

Due to digitisation, curriculum-relevant information becomes available and can be extracted from Academic Information Management Systems (AIMS). The discipline of Educational Data Mining (EDM) [19] uses the data logs retrieved from the AIMS to explore the underlying connections between the educational data and predict students' academic performance. Personal information, student activities, courses attended, grades, and course descriptions can be used to extract important indicators to design a prediction model [19], which, in turn, can help students make well-considered decisions.

*Corresponding author. E-mail: maja.spahic@fhnw.ch.

Various approaches to predicting grades exist, but only a few consider unstructured data as an attribute. Further, while the focus of prior work has been identifying course-taking patterns that positively impact achievement and developing predictive models to forecast student grades, the impact of negative course-taking patterns is largely unexplored. This paper introduces a new approach to assist students in course selection at the start of the semester, utilising historical data and current course descriptions. This method predicts potential grades by analysing factors and patterns that negatively affect course performance and using these insights to recommend or advise against selected courses. We achieve this by generating individual student knowledge profiles and correlating them with course performance. We employ the Large Language Model (LLM) BERT to create averaged vector embeddings from the course descriptions of each student's course history. By only advising against certain courses, our approach otherwise leaves students with maximum freedom of choice.

Enrollment process data from the Business Information Systems (BIS) master's degree program at the FHNW University of Applied Sciences and Arts Northwestern Switzerland (FHNW) was used to develop the approach. The data set, extracted from AIMS, contains information about course registrations, enrollment mutations, and students' final grades per course from the fall semester of 2012 to the fall semester of 2023.

This paper is organised in the following manner: Section 2 provides an overview of the latest research in this field. Section 3 details the procedures and steps undertaken to develop our approach. Section 4 describes the dataset used in this study and the pre-processing steps applied. Section 5 outlines the prototype creation process. Section 6 discusses the metrics employed for assessing the prototypes and presents the findings of this evaluation. Finally, Section 7 reflects on the contributions of this research and explores potential avenues for future investigation.

2. Related Work

Several research studies have focused on developing methods for predicting student grades. This focus benefits students by detecting potential problems early, improving the overall educational experience and student success. Existing studies use various attributes to describe student backgrounds, interests, performance, and living circumstances. Two groups of attributes are very commonly used – with very different success:

- **Demographic attributes** are very widely used [2, 4, 8, 10, 15, 20, 25], but have consistently been found to be less predictive of academic achievements.
- On the other hand, **prior academic performance**, also very widely used [4, 10, 15, 25], often in the form of (Cumulative) Grade Point Averages ((C)GPA), has repeatedly been shown to be the best predictor of future success.

2.1. Predicting Student Grades

A study from Cheng, Liu, and Jia [4] utilises a Portuguese educational dataset with 33 attributes to assess student performance. These attributes, drawn from questionnaires and academic records, include student demographics (sex, age, school, residence type), parental details (education, occupation, cohabitation status), and household characteristics (family size, family relationships). It also covers school-related factors like travel time, study time, past failures, and participation in extra courses, as well as personal aspects like internet access, ambition for higher education, romantic relationships, leisure activities, alcohol consumption, and health status. Grades for three evaluation periods (G1, G2, G3), with G3 as the final grade, and school absences are also included. These grades and absences serve as model outputs and are categorised into four groups based on performance: poor (0–12), medium (12–14), good (14–16), and excellent (16–20). The study compared the performance of five different classifiers: the Random Forest Classifier, the Decision Tree Classifier, the K Neighbors Classifier, the MLP Classifier, and the XG-Boost Classifier. The Enhanced Artificial Ecosystem-Based Optimization + XG-Boost hybrid method provided high accuracy and F1-score values, indicating the potential to effectively combine machine learning techniques with metaheuristic algorithms for predicting and classifying students' performance.

Similarly, Khudhur and Ramaha [10] examined a student performance dataset from Cortez and Silva [5], collected from two Portuguese secondary schools using school reports and questionnaires. This dataset encompasses grades,

Table 1
Frequently used attributes to predict student performance [8]

Attributes	Attribute Domain
Academic performance	GPA, Grade level, High school score, attendance to lessons, number of courses per semester
Demographic	Gender, nationality, place of birth, age
Behavioral	Raised hands, visit resources, school satisfaction, discussion, attend class, answer questions
Psychological	Personality, motivation, learning strategies, approach to learning, contextual influence
Family background	Mother and father Education, family income, location of parents
School environment	School size, medium of instruction, lecturer/teacher behaviour in class

demographic details, and social and school-related features from Mathematics and Portuguese language courses. The relationship between the grades of the different periods is an important aspect of this data set. The final grade at the end of the year, referred to as G3 and issued at the end of the third period, strongly correlates with the grades of the first (G1) and second (G2) periods. This correlation means that it is difficult to predict the final grade (G3) without considering the grades from the earlier periods (G1 and G2). Employing a Generative Adversarial Network (GAN), the dataset was expanded from 1,044 to 46,044 rows. To predict final grades, five machine learning models were used: Decision Tree (DT), Random Forest (RF), Radial Basis Function SVM (RBF SVM), Linear SVM, and K-Nearest Neighbours (kNN). The results showed high accuracy in grade predictions with a range between 99.52% to 99.83%. GAN's ability to create synthetic data, which increased the data set size, helped classifiers to make more accurate predictions. Also, Nachouki et al. [16] developed a model using the random forest algorithm to predict student course performance and identify important predictors of course grades. Compared to previous research, the different course delivery modes, including face-to-face teaching mode, online mode during the pandemic, and hybrid mode after the pandemic, were considered. The data was collected from 650 transcripts of undergraduate computer science students and included four categorical features (high school type, course category, gender, and mode of course delivery) and three numerical features (high school score, student course attendance percentage, and grade point average). The result shows that the most significant predictors were the high school score and grade point average (GPA).

In addition, Tormon et al. [21] have investigated how psychological variables (perceived stress, student engagement, resilience, and growth mindset) affect first-year engineering students' GPA. The results suggest that stress lowers grades, but resilience can lessen this effect while being very engaged in studies can worsen it. A growth mindset did not significantly predict students' GPA. Likewise, Gonzalez-Nucamendi et al [6] examines how intrinsic multiple intelligences (MI) and self-regulation, learning, and affective strategies (SRLAS) can predict engineering students' success. They discovered that logical/mathematical intelligence is crucial for engineering students' academic success. Further, understanding student profiles can help lecturers in implementing effective teaching strategies. Also, McKenzie and Schweitzer [15] examined the predictors of academic performance in Australian first-year students and reviewed previous research on academic, psychosocial, cognitive, and demographic predictors of academic performance. Prior academic achievement was identified as the most significant predictor of academic performance. In addition, academic self-efficacy and job responsibility were found to predict academic performance.

A systematic literature review was performed by Issah et al. [8], analysing 84 publications between 2016 and 2022 relevant to predicting student performance. Classification and decision trees were identified as common machine-learning approaches. Further, they identified frequent attributes for the prediction and mapped them into six categories, as shown in Table 1. Grades and test scores are the main attributes, followed by demographic attributes. Only a few studies considered school environment, family background, and behaviour. A few years earlier, Shahiri, Husain, and Rashid [20] also performed a systematic literature review to identify key methods and attributes for predicting student achievements. At this point in time, cumulative grade point average (CGPA) was a commonly used attribute. Student demographics like gender and age and final exams are also widely used. Other factors, such as extracurricular activities, high school background, and social interactions, were less common, together with psychometric factors like interest and study habits. The latter were rare because of the data quality and difficulty getting reliable data.

A different approach to predicting students' performance has been introduced by Li et al. [12]. Based on students' behaviour features, a Multi-View Hypergraph Neural Network (MVHGNN) was constructed to represent complex relationships among the students. Three different types of behaviours were considered: learning, dining, and sports. It was discovered that certain behaviour patterns, like regular meals and library usage, are associated with better academic performance. The study showed that considering multiple behaviours and high-order relations among students improves prediction accuracy.

Next, student interactions with LMS were also considered as one of the attributes. Waheed et al. [25] developed an approach to predict students' success in online courses by incorporating scores in the first assignment, demographic data, and clickstream interactions of students with the Virtual Learning Environment (VLE). The study uses the Open University Learning Analytics Dataset (OULAD), which includes information from Open UK University courses from 2013-14. The dataset was enhanced with the assessment results and interaction logs of 32,593 students across 22 courses. The study found early student course performance prediction is possible by analysing their first assessment submission and engagement information. Furthermore, they showed that including demographic data only slightly improved the prediction. Additionally, Arifin et al. [2] utilised academic and non-academic data to train a regression model, with the final Cumulative Grade Point Average (CGPA) serving as a numerical predictor value. The data included information about students' current semester CGPA, demographic and economic background, involvement in campus organisations, and learning activities derived from Moodle. The gradient-boosted trees regression model demonstrated the lowest error rate among the examined models. Leelaluk et al. [11] used a neural network model to predict students' performance based on the weekly reading habits for each lecture material. Unlike no-risk students, at-risk students show less engagement in reading and reviewing the learning material. The study suggests that considering content data in the model can help classify students and identify those who are likely to fail. Jović et al.[9] used students' academic performance and learning data collected from an LMS and an AIMS. Academic performance included grades from homework assignments, tests, projects, and class participation, which, together with the final exam score, determined the students' final grades. Students' interactions with online forums, lectures, and assessments like homework, projects, and tests were collected as learning data. The SVM classifier was chosen for model training and evaluation, showing good results in predicting student performance and achieving a 90.3% accuracy. Hasan et al. [7] predicted student semester performance by analysing their online learning activities and including students' interaction with video material provided on Moodle. Features such as grades, registered courses or the number of plagiarism accusations, students' time on Moodle activities on and off campus, and interaction with the video material were considered to train the model. The Random Forest algorithm had the highest accuracy of 88.3%.

2.2. Considering Unstructured Data for Performance Prediction

Few have also taken unstructured data into account. Phan, De Caigny, and Coussement [18] proposed an approach to improve dropout prediction using structured and unstructured data. In addition to incorporating student sociodemographic information and course enrollment details, they included vectorised textual data from student feedback into the performance prediction model. This model considered over twenty different attributes. A real-life dataset from a French Higher Education Institution (HEI), including 14,391 students and 62,545 feedback documents, was utilised to test this approach and to identify students at risk of dropping out. The study confirms that incorporating student textual feedback data improves the predictive performance of student dropout models. Further, Liu et al. [13] proposed an explainable Exercise-aware Knowledge Tracing (EKT) framework to predict how well students will perform on future exercises by considering past performance and the actual content of the exercises they have done. The content of the exercises was used to determine the students' knowledge. EKT framework could predict student performance effectively and provide superior interpretability by showing which concepts students understood well.

On the contrary, Priyambada, Usagawa, and ER [19] incorporates the students' course-taking behaviour and domain knowledge to enhance the prediction of students' performance, especially in a vertically coherent curriculum. Each course belongs to a domain like information technology, business management, and general knowledge. Those domains were specified as domain knowledge, and aggregated credits and grades per semester were considered for

the prediction model. Further, in terms of learning behaviour, they categorised students based on how well the proposed curriculum was followed. The proposed framework improved prediction accuracy for student performance and enabled the evaluation of a student who will finish the study program on time.

Another innovative approach from Vertegaal et al. [24] analysed students' participation and voice recordings in student-led tutorials (SLT) to predict their success in an Electromagnetic course. Preparedness for exercises and voice data can predict student performance with reasonable accuracy, with an accuracy of over 65%.

In summary, existing forecasting models focus on analysing student performance in the current semester and predicting the student's performance after final enrolment. Moreover, these models rely primarily on structured data where GPA or CGPA plays a central role. Unstructured data sources are rather overlooked. Additionally, many authors presuppose a vertically integrated curriculum with restricted student choice. Also, the specifics of the courses, particularly the mix of chosen courses, have not been widely considered.

3. Methodology

Design science research (DSR) [23] was selected as the methodological framework for developing a student performance prediction model due to its focus on generating prescriptive knowledge through designing and assessing artifacts to solve specific real-world problems. Problem awareness was raised through a literature review and a case study about the BIS student's course enrollment process at FHNW. In addition, interviews were conducted with BIS students to understand how students select courses and how important a possible grade is in their choice. During the suggestion phase, exported enrollment data was pre-processed and analysed to identify important attributes for grade prediction. Furthermore, artefact requirements were defined based on the findings from the previous phase. The development phase involved the incremental creation of a prototype. The data was prepared to link the course enrollment and corresponding course description data. In the evaluation phase, the prototype's reliability was measured by calculating precision and recall. In addition, the confusion matrix was created, and a cost matrix was calculated to obtain an overall impression of the prototype's reliability. In the conclusion, the contribution to the body of knowledge and further research are discussed.

4. Data Set from Student Enrolment Process

The BIS Master's program starts twice a year and can be taken full-time (1.5 years) or part-time (2.5 years). A large proportion of students are employed and study part-time. Students can design their individual study plans and choose from a pool of four compulsory courses and over 20 elective courses. The grades are awarded on a scale of one to six, rounded to the nearest half-grade. Where one is the lowest grade, six is the highest, and at least a four is expected to pass a course. Students register for offered courses before the beginning of the semester and can de-register from the course in the first two weeks of the semester without penalty. Later, de-registrations will be penalised with a grade of one for the respective course. In addition, students who have failed a course can only repeat it once. The interviews with the students revealed that personal interest in the subject is decisive when choosing a course, with course descriptions playing a crucial role in identifying suitable options. In addition, they are careful to take thematically similar courses within the same semester to get better grades. Nevertheless, there are various course registrations and de-registrations, especially in the first two weeks of the semester. Students often canceled courses after receiving results from the previous semester, particularly if they had already achieved the necessary credit points. Other reasons for cancellations included an unexpectedly high workload, taking a study break, or finding that the course content did not meet their expectations.

The exported dataset includes records from 856 students from the spring semester of 2012 to 2023. The data set includes students' demographic (age, sex) and academic performance (grade per course and semester, current semester) attributes. Figure 1 shows the grade distribution before pre-processing. The data set was cleansed during pre-processing, and outliers were removed, like grades from students who de-registered too late and got a grade of one or sets with course registrations but without grades. Afterwards, the data set showed a higher imbalance in the distribution of grades, as positive performance was more common than negative performance. To overcome

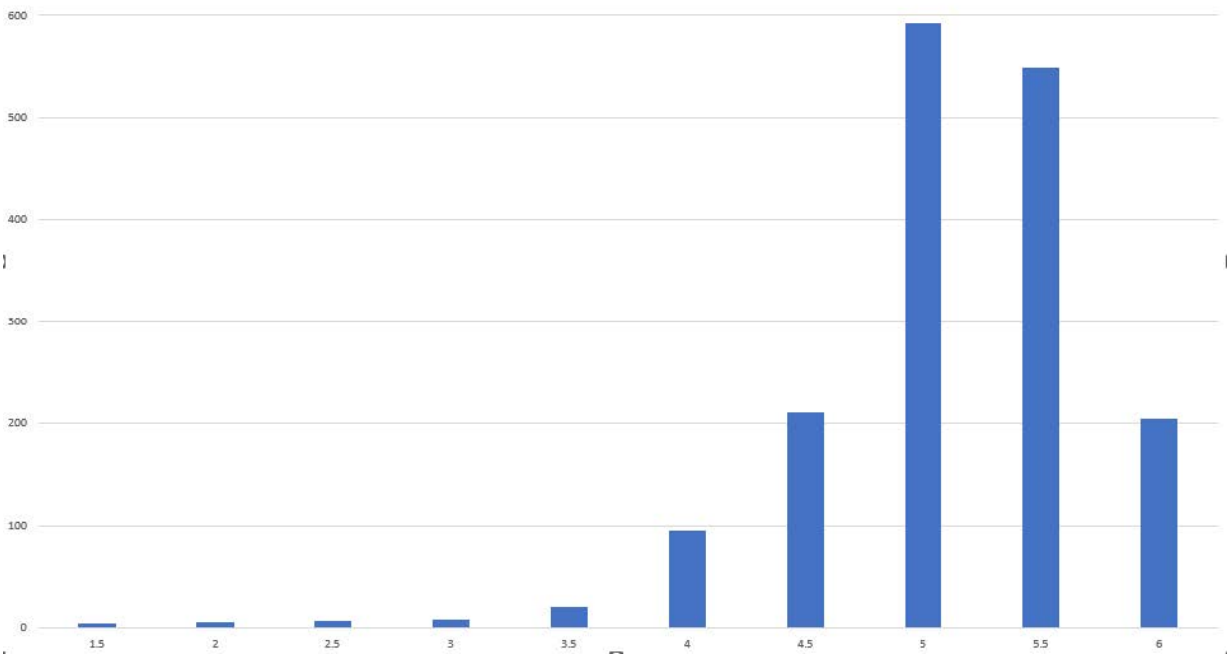

Fig. 1. Grade distribution from the spring semesters of 2012 to 2023 following pre-processing

this imbalance and to make the learned model more sensitive to potential failures, a straightforward oversampling method was employed: students with at least one negative grade during their academic career, along with their semester records, were replicated threefold, thereby augmenting the original dataset.

Table 2 shows an example of the data after pre-processing. Each row in the dataset represents a single semester for an individual student. The semesters for each student are sequentially arranged in the order they were undertaken. A separate column represents each course. For every semester of a student, the dataset includes both the grades obtained in that current semester and the cumulative grades from all previous semesters of the same student.

Table 2
Example of data after pre-processing

ID Person	Semester Nr	Anrede	Age	max_semester	Cloud Computing	Alignment of Business and IT	Research Methods in Information Systems	Business Process Management	Strategic Management in the Digital Economy	Emerging Topics for Business Information Systems	Business Intelligence	Supply Chain Management	Compliance Management and Governance of IT	Knowledge Processing and Decision Making	Digitalization of Business Processes	Master Thesis Proposal	Managers Shadow Project	Master Thesis	...	Competitive Strategy in the Information Age	Information Management	User-centered Design and Design Thinking	Artificial Intelligence for Business Processes	Unique Identifier	Embeddings
708299	1	0	0.30769231	5	5.5	5	0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	708299-1	[-0.02533539 - 0.0201073 -
708299	2	0	0.30769231	5	5.5	5	5	5	5.5	6	0	0	0	0	0	0	0	0	...	0	0	0	0	708299-2	[-0.02170304 - 0.02289835 -
708299	3	0	0.30769231	5	5.5	5	5	5	5.5	6	5	5	5.5	0	0	0	0	0	...	0	0	0	0	708299-3	[-0.02136705 - 0.02271209 -
708299	4	0	0.30769231	5	5.5	5	5	5	5.5	6	5	5	5.5	5.5	5.5	5.5	5.5	0	...	0	0	0	0	708299-4	[-0.02268211 - 0.02359489 -
708299	5	0	0.30769231	5	5.5	5	5	5	5.5	6	5	5	5.5	5.5	5.5	5.5	5.5	6	...	0	0	0	0	708299-5	[-0.02290371 - 0.0235733 -

5. Design and Development of the Prototypes

A hybrid performance prediction system is suggested, utilising similarities among students. Based on the insights from student interviews, we equipped our similarity measure with two components: one to capture students' interests in certain subjects and the other to capture their performance in previous courses. The first component involves comparing students using a 'student knowledge profile,' derived from averaged vector embeddings of the course descriptions from each student's course history. Secondly, the system compares students' course performances through an Euclidean Distance similarity measurement. An overall student similarity score is then computed by integrating these two components, applying a weighted combination of the knowledge profile comparison and the Euclidean Distance measurement.

The dataset, not particularly extensive, encompasses students in the Master's program who possess Bachelor's degrees in closely related fields, such as business informatics, computer science, or business administration. However, their professional experiences vary significantly. Post-Bachelor's, these students have pursued diverse career paths. Consequently, their academic and professional histories were not considered in the analysis. Given the dataset's limited size, constructing student profiles based on their academic and professional backgrounds for predictive purposes presents a considerable challenge. Interviews with students indicated that course selections are primarily interest-driven. Therefore, the Large Language Model BERT was employed using course descriptions to identify students with similar course interests.

5.1. Course Description Similarity

Course descriptions were collected from a publicly accessible course description website[1] and from the university administrations. The course descriptions contain information about the competencies to be achieved, content, teaching and learning methods, and assessment. As far as possible, the course descriptions from the respective semesters reflected the reality at the time. Course descriptions were transformed into a vector space using BERT, a transformer-based language model. For each student, the courses they had enrolled in up to the semester preceding the target semester for prediction are analysed. The embeddings of these course descriptions are processed and normalised, and their average is calculated. This average embedding represents the student's knowledge profile, reflecting the content of courses they have already completed. L2 Normalization was applied to normalise the course description embeddings. This process guarantees that each course description embedding vector has a length of one, an essential step for averaging them effectively. Normalising the embeddings ensures that each course contributes equally to the average calculation, thereby maintaining uniformity in weight. Figure 2 shows the procedure.

Once the knowledge profiles for each student were computed, cosine similarity was employed to identify the most similar students. This was done by comparing the averaged vector embeddings constituting the student knowledge profiles. The cosine similarity formula $cos(\theta) = \frac{\sum_{i=1}^{n} E1_i E2_i}{\sqrt{\sum_{i=1}^{n} E1_i^2}\sqrt{\sum_{i=1}^{n} E2_i^2}}$, utilised in this context, facilitates the calculation of similarities between various student knowledge profiles.

5.2. Course Performance Similarity

To succinctly represent each student's historical cumulative semester performance, their course performance data was pre-processed into a single row per student in the dataset. The course performance similarity was then calculated by transforming each student's grades to the semester before the one being predicted into a vector space. Each course and its corresponding grade are represented in this space by a distinct vector dimension. Subsequently, the Euclidean Distance was computed between student grade vectors for courses attended by both students, and a normalised similarity measure was derived. The Euclidean Distance calculation underwent a slight modification to accommodate scenarios where courses were attended by only one of the compared students. In these instances, a smoothing factor of 0.1 was added to the Euclidean Distance calculation. Figure 3 illustrates the method for calculating individual grade differences as part of the Adjusted Euclidean Distance Calculation.

The methods involve several key formulas:

[1] See https://modulbeschreibungen.webapps.fhnw.ch/

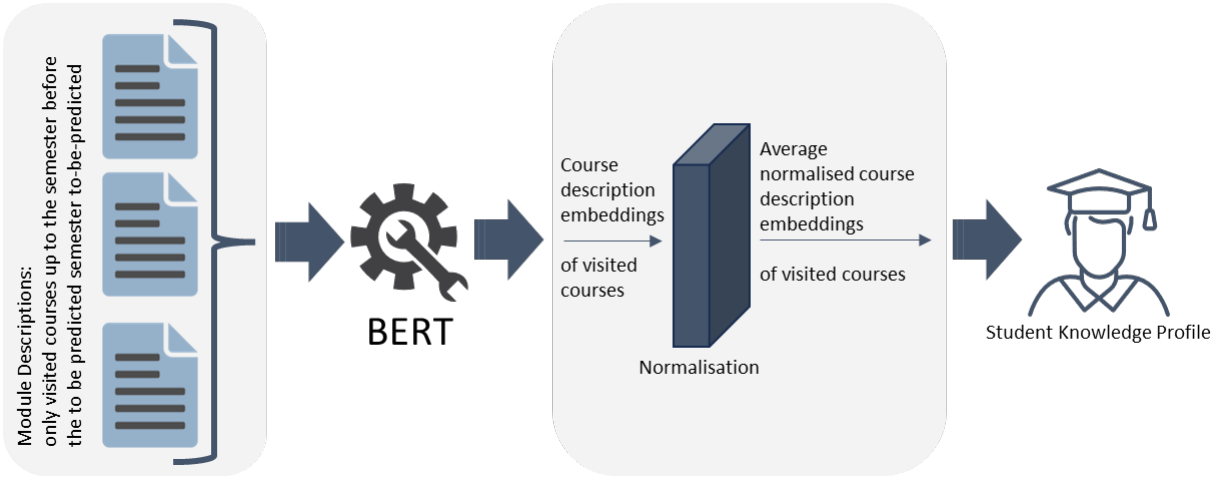

Fig. 2. Calculating students knowledge profile

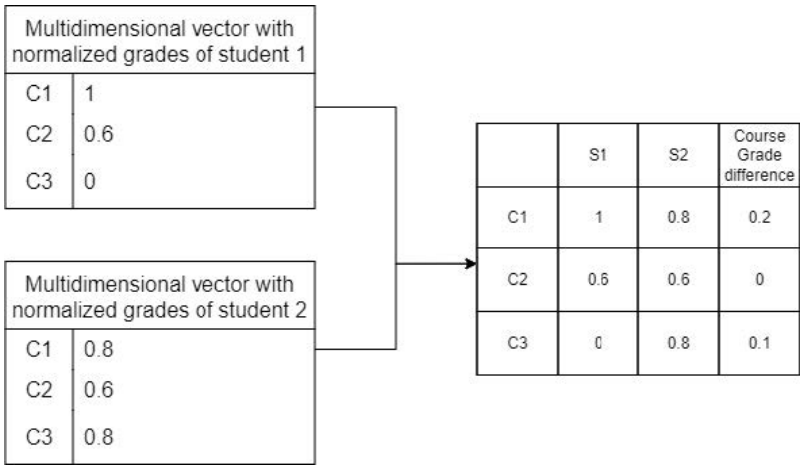

Fig. 3. Calculation of course grade differences

- Min-Max Normalization: This process converted each student's grades into a range between 0 and 1 ($normalized_i = \frac{grade_i - mingrade}{maxgrade - mingrade}$).
- Adjusted Euclidean Distance: The normalised grades were inputted into a multi-dimensional vector, encompassing all courses attended and unattended up to the semester preceding the forecasted one

$$(d = \sqrt{\sum_{i=1}^{n} (if(grade1_i \neq 0 \; and \; grade2_i \neq 0) \; then \; (grade1_i - grade2_i)^2 \; else \; if \; (grade1_i \neq 0 \; or \; grade2_i \neq 0) \; then \; penalty^2 \; else \; 0}).$$

  Here, a value of 0 indicates courses not attended. This Adjusted Euclidean Distance was employed to factor in performance in courses not accounted for by the course description similarity. Additionally, it introduced a penalty for courses exclusively attended by one of the two compared students.
- Course Performance Similarity: This metric is based on the calculated Adjusted Euclidean Distance ($CoursePerformanceSimilarity = \frac{1}{1 + AdjustedEuclideanDisntace}$). It aims to transform the Adjusted Euclidean Distance into a similarity score ranging from 0 to 1, thus quantifying the likeness in course performance between students.

The final student similarity score was derived by integrating the previously computed similarities from course descriptions and performance ($FinalStudentSimilarityScore = w_1 * courseDescSim + w_2 * coursePerfSim$). As these similarities produce values ranging from 0 to 1, a weighted approach was employed for their combination, allowing for a balanced integration of both metrics.

5.3. Performance Prediction

The prototype considers two cases when predicting the grades: (1) Grade prediction for students beyond the first semester and (2) grade prediction for first-semester students.

Grade prediction for students beyond the first semester: The prototype uses a k-nearest neighbour approach and calculates grade predictions by averaging the grades of the five most similar students who have previously taken the same course in the same semester. The number of similar students to be considered for this calculation can be set as a parameter. If fewer students than the parameter have taken the course in the same semester, the prediction uses the actual number of students for averaging. In cases where no student has taken the course in the same semester, the grade is predicted by averaging the grades of students who completed the program within the same number of semesters. If no students meet this criterion, the average is then based on the grades of all students, regardless of the semester they took the course.

Grade prediction for first-semester students: A different approach is adopted for students who lack an academic record. The predicted grade is determined by averaging students who completed the program within the same number of semesters and took the course in their first semester. If no students meet these criteria, the prediction is based on the average grades of students who finished the program within the same number of semesters. Should this also not be applicable, the grade is predicted by averaging all students' grades across all semesters.

The objective of the prototype is to proactively alert students about potential underperformance while aiding them in choosing their courses with minimal restrictions. To achieve this, the output from the performance prediction Prototype has been simplified into a binary response. This response either warns about the possibility of failing or endorses the student's course selection. A predefined threshold determines the classification of grades as positive or negative. A positive grade is considered when the predicted grade is higher or equal to 4. A warning is issued if the grade is below 4.

Figure 4 illustrates an example prediction for a course, where the first seven columns represent the parameters input by students, and the last two columns display the predicted grade and classification.

ID Person	Gender	Age	Course	Performance	Semester Nr.	max_semester	current semester	Evaluation 5 most similar students	Predicted Output
123	1	0	Business Analytics: Quantitative Methods	0	3	5	2	5.4	Module choice is alright

Fig. 4. Example of a performance prediction

6. Evaluation

The Performance Prediction System's effectiveness is assessed by setting a classification threshold: grades below 4 are deemed negative, while those 4 and above are considered positive. This system was evaluated using the dataset prepared in the pre-processing stage, employing Leave-One-Out cross-validation. The evaluation concentrated on comparing two configurations of the prototype, one considering the performance prediction based on five similar students and the other on seven. Various metrics were used to gauge the model's performance:

- Precision: This metric is computed by dividing the number of true positives (correctly predicted negative performances) by the total predicted positives (both correctly and incorrectly predicted negative performances). Precision thus reflects the accuracy of correctly predicted negative performances.
- Recall: Recall is calculated by dividing the number of true positives (correctly predicted negative performances) by the sum of true positives and false negatives (incorrectly predicted positive performances). It indicates the proportion of actual negative performances that were correctly identified.
- Confusion Matrix: This matrix provides a visual and numerical representation of the model's predictions, contrasting them against the actual outcomes. It is segmented into four quadrants: true positives, true negatives, false positives, and false negatives. True positives and negatives represent correctly predicted negative and positive performances, respectively. False positives are instances where negative outcomes are incorrectly predicted, and false negatives are where positive outcomes are incorrectly predicted.
- Cost Matrix: Unlike the confusion matrix, which counts the number of correct and incorrect predictions, the cost matrix assigns a value to the impact of incorrect predictions. For the prototype's evaluation, false positives are assigned a cost of 1 and false negatives a cost of 3, acknowledging that not all incorrect predictions have the same severity. This matrix is crucial for understanding and quantifying the actual impact of these incorrect predictions.

The prototype's main aim is to assist students in course selection, offering freedom in their choices except in situations where performance warnings are raised. A true positive (a correctly predicted negative performance) is valued higher than a true negative, as the cost of over-warning students is less severe than the risk of under-warning, which could lead students into a false sense of security.

The comparative analysis of the performance prediction system, focusing on the impact of considering five versus seven most similar students, revealed that an increase in the number of similar students doesn't necessarily enhance the prototype's effectiveness. Notably:

- Precision Decrease: The precision dropped from 0.36734 (with five similar students) to 0.24 (with seven similar students), indicating that the prototype's accuracy in correctly identifying negative performances diminished as more similar students were included.
- Recall Decrease: There was also a decline in recall, from 0.45 (with five similar students) to 0.2926 (with seven similar students). This suggests a reduced sensitivity in detecting negative cases, leading to a lower rate of correctly identifying actual negative performances.
- True Positives and False Negatives: The number of true positives declined from 18 to 12 when the count of similar students considered rose from 5 to 7, signifying a reduction in the prototype's effectiveness in correctly identifying negative performances. Concurrently, false negatives increased from 23 to 29, indicating that more negative performances were overlooked.
- False Positives Increase: There was a rise in false positives from 31 to 38, suggesting a higher likelihood of the prototype erroneously categorising positive performances as negative with seven similar students.
- Slight Decline in True Negatives: The number of true negatives marginally decreased from 1622 to 1615, implying a slight drop in the prototype's ability to accurately recognise positive performances.
- Higher Overall Cost: The total cost of the prototype's predictions rose from 100 to 125. This increment reflects the increased inaccuracy in performance predictions, particularly due to the rise in incorrectly predicted negative performances and a decline in accurately identifying negative performances when a larger group of similar students was considered.

Table 3 summarises the comparative analysis:

These findings suggest that while expanding the number of similar students considered for predictions, the prototype may lose precision and recall, potentially leading to higher costs due to increased inaccuracies. The results indicate that expanding the number of similar students factored into the performance prediction adversely affects the prototype's accuracy. Specifically, the increase in the number of students considered similar leads to a decline in the prototype's ability to predict performance outcomes accurately. This finding suggests that the current prototype construction approach requires further refinement, as it fails to identify many negative performances. The implication is that the method of determining similarity among students and how it influences prediction accuracy needs to be re-evaluated and optimised to enhance the prototype's efficacy in predicting student performance.

Table 3
Comparison of the results

	5 Students	7 Students
Precision	0.36734	0.24
Recall	0.45	0.2926
True Positive	18	12
False Negative	23	29
False Positive	31	38
True Negative	1622	1615
Cost	100	125
F1	0.40	0.2637

7. Conclusion

The introduced approach offers promising future development potential. However, the evaluation indicates room for improvement, particularly in precision and recall. A key area for enhancement is the optimisation of the LLM approach to differentiate course descriptions more effectively.

Several potential improvements have been identified. Adding a feature that allows students to input their interests could personalise the systems. Further, the prototype could be improved by considering known positive and negative course sequences reported by students in its performance prediction calculations. Additionally, the current method of addressing data imbalance involves oversampling the minority class by duplicating performance data. Future research could explore more sophisticated sampling methods, like the Synthetic Minority Over-sampling Technique (SMOTE) or Generative Adversarial Networks (GAN), which create synthetic samples rather than duplicating existing ones. This approach could offer a more robust solution to the imbalance issue and reduce overfitting risks. Lastly, the current use of the BERT model for embedding course descriptions faces challenges in differentiating them effectively. Future research efforts could focus on optimising these descriptions to improve embeddings, possibly by applying feature extraction techniques or altering the structure of the course descriptions to better align with the model's learning algorithms. Enhanced differentiation of course description embeddings would significantly improve the prototype's ability to match students based on their interests accurately.

In conclusion, while the prototype marks a significant advancement, the identified improvements offer a roadmap for evolving this tool into a more refined and practical aid for students in their academic journey.

References

[1] H. Altabrawee, O.A.J. Ali and S. Qaisar Ajmi, Predicting Students' Performance Using Machine Learning Techniques, *Journal of University of Babylon, Pure and Applied Sciences* **27**(1) (2019), 194–205.

[2] M. Arifin, W. Widowati, F. Farikhin and G. Gudnanto, A Regression Model and a Combination of Academic and Non-Academic Features to Predict Student Academic Performance, *TEM Journal* **12**(2) (2023), 855–864. doi:10.18421/TEM122-31.

[3] A. Bogarín, R. Cerezo and C. Romero, A survey on educational process mining, *WIREs Data Mining and Knowledge Discovery* **8**(1) (2018), 1–17. doi:10.1002/widm.1230.

[4] B. Cheng, Y. Liu and Y. Jia, Evaluation of students' performance during the academic period using the XG-Boost Classifier-Enhanced AEO hybrid model, *Expert Systems with Applications* **238** (2024), 122136. doi:10.1016/j.eswa.2023.122136.

[5] P. Cortez and A. Silva, Using data mining to predict secondary school student performance, in: *Proceedings of 5th Annual Future Business Technology Conference*, Brito. A and J. Teixeira, eds, EUROSIS-ETI, Porto, 2008, pp. 5–12. ISBN 9789077381397.

[6] A. Gonzalez-Nucamendi, J. Noguez, L. Neri, V. Robledo-Rella, R.M.G. García-Castelán and D. Escobar-Castillejos, The prediction of academic performance using engineering student's profiles, *Computers Electrical Engineering* **93**(August 2020) (2021), 107288. doi:10.1016/j.compeleceng.2021.107288.

[7] R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K.U. Sarker and M.U. Sattar, Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques, *Applied Sciences* **10**(11) (2020), 3894. doi:10.3390/app10113894.

[8] I. Issah, O. Appiah, P. Appiahene and F. Inusah, A systematic review of the literature on machine learning application of determining the attributes influencing academic performance, *Decision Analytics Journal* **7**(October 2022) (2023), 100204. doi:10.1016/j.dajour.2023.100204.

[9] J. Jović, E. Kisić, M.R. Milić, D. Domazet and K. Chandra, Prediction of student academic performance using machine learning algorithms, in: *Proceedings 13th International Conference on eLearning*, Vol. 3454, M. Saqr, S. López-Pernas, M.Á. Conde and M. Raspopović Milić, eds, CEUR Workshop Proceedings, 2022, pp. 31–39.

[10] A. Khudhur and N.T.A. Ramaha, Students' Performance Prediction Using Machine Learning Based on Generative Adversarial Network, in: *2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, IEEE, 2023, pp. 1–6. doi:10.1109/HORA58378.2023.10156733.

[11] S. Leelaluk, T. Minematsu, Y. Taniguchi, F. Okubo and A. Shimada, Predicting student performance based on Lecture Materials data using Neural Network Models, in: *Proceedings of the 4th Workshop on Predicting Performance Based on the Analysis of Reading Behavior*, Vol. 3120, B. Flanagan, R. Majumdar, H. Li, A. Shimada, F. Okubo and H. Ogata, eds, CEUR Workshop Proceedings, 2022, pp. 11–20.

[12] M. Li, Y. Zhang, X. Li, L. Cai and B. Yin, Multi-view hypergraph neural networks for student academic performance prediction, *Engineering Applications of Artificial Intelligence* **114**(June) (2022), 105174. doi:10.1016/j.engappai.2022.105174.

[13] Q. Liu, Z. Huang, Y. Yin, E. Chen, H. Xiong, Y. Su and G. Hu, EKT: Exercise-aware knowledge tracing for student performance prediction, *IEEE Transactions on Knowledge and Data Engineering* **33**(1) (2021), 100–115. doi:10.1109/TKDE.2019.2924374.

[14] F. Marbouti, H.A. Diefes-Dux and K. Madhavan, Models for early prediction of at-risk students in a course using standards-based grading, *Computers Education* **103** (2016), 1–15. doi:10.1016/j.compedu.2016.09.005.

[15] K. McKenzie and R. Schweitzer, Who Succeeds at University? Factors predicting academic performance in first year Australian university students, *Higher Education Research Development* **20**(1) (2001), 21–33. doi:10.1080/07924360120043621.

[16] M. Nachouki, E.A. Mohamed, R. Mehdi and M. Abou Naaj, Student course grade prediction using the random forest algorithm: Analysis of predictors' importance, *Trends in Neuroscience and Education* **33** (2023), 100214. doi:10.1016/j.tine.2023.100214.

[17] M.H. Othman, N. Mohamad and M.N. Barom, Students' decision making in class selection and enrolment, *International Journal of Educational Management* **33**(4) (2019), 587–603. doi:10.1108/IJEM-06-2017-0143.

[18] M. Phan, A. De Caigny and K. Coussement, A decision support framework to incorporate textual data for early student dropout prediction in higher education, *Decision Support Systems* **168** (2023), 113940. doi:10.1016/j.dss.2023.113940.

[19] S.A. Priyambada, T. Usagawa and M. ER, Two-layer ensemble prediction of students' performance using learning behavior and domain knowledge, *Computers and Education: Artificial Intelligence* **5**(January) (2023), 100149. doi:10.1016/j.caeai.2023.100149.

[20] A.M. Shahiri, W. Husain and N.A. Rashid, A Review on Predicting Student's Performance Using Data Mining Techniques, *Procedia Computer Science* **72** (2015), 414–422. doi:10.1016/j.procs.2015.12.157.

[21] R. Tormon, B.L. Lindsay, R.M. Paul, M.A. Boyce and K. Johnston, Predicting academic performance in first-year engineering students: The role of stress, resiliency, student engagement, and growth mindset, *Learning and Individual Differences* **108**(October) (2023), 102383. doi:10.1016/j.lindif.2023.102383.

[22] N. Třcka and M. Pechenizkiy, From local patterns to global models: Towards domain driven educational process mining, in: *ISDA 2009 - 9th International Conference on Intelligent Systems Design and Applications*, IEEE, 2009, pp. 1114–1119. doi:10.1109/ISDA.2009.159.

[23] V.K. Vaishnavi and W. Kuechler, *Design Science Research Methods and Patterns*, 2nd edn, CRC Press, 2015. doi:10.1201/b18448.

[24] C.J.C. Vertegaal, P. Sundaramoorthy, C. Martinez, R. Serra and M.J. Bentum, Exploring the Potential of Machine Learning to Predict Student Performance in an EM Course, in: *2023 32nd Annual Conference of the European Association for Education in Electrical and Information Engineering (EAEEIE)*, IEEE, 2023, pp. 1–6. doi:10.23919/EAEEIE55804.2023.10181928.

[25] H. Waheed, I. Nisar, M.-u.-N. Khalid, A. Shahid, N.R. Aljohani, S.-U. Hassan and R. Nawaz, Predicting Academic Performance of Students from the Assessment Submission in Virtual Learning Environment, in: *Research and Innovation Forum 2022. RIIFORUM 2022. Springer Proceedings in Complexity*, A. Visvizi, O. Troisi and M. Grimaldi, eds, Springer, 2023, pp. 417–424. doi:10.1007/978-3-031-19560-0$_3$3.