USING MACHINE LEARNING TO SUPPORT STUDENTS' ACADEMIC DECISIONS

AISHA GHAZAL FATEH ALLAH

Lecturer, Higher Colleges of Technology, Department of Computer Information Science, U.A.E

E-mail: aisha.ghazal@gmail.com

ABSTRACT

Making the right decision for students in higher education is vital, as it has a great influence on their study, career, life, and eventually, the whole society. Predicting the future performance of students can inform their choice of majors, concentrations, and courses. It also helps teachers and advisors provide the necessary support to students as needed.

While many studies address the issue of predicting students' performance, they mainly predict student performance at only one particular stage of their study. For example, literature has papers on predicting student's performance at enrollment, or in a particular course, which is not enough to help students throughout their study journey. This work addresses this gap in literature and proposes a holistic framework for assisting students in their decision throughout their entire study journey, and not only at one point of their study - as currently in literature. First, at enrollment, this work predicts a student's GPA in different majors using enrollment data such as high school average, placement test results, and IELTS score. Second, after completing their first year, this work predicts student's GPA in different concentrations using grades of Year-1 courses. Third, at any point of time after the student finishes some courses, a user-based collaborative filtering approach using K-Nearest Neighbor is used to predict a student's grade in a future course. This approach uses other students' grades to make a prediction.

Furthermore, this research tests and compares the performance of Decision Trees, Random Forests, Gradient-Boosted trees, and Deep Learning machine learning regression algorithms to predict student GPA. Gradient Boosted Trees performed the best when predicting student's Major GPA, while Deep Learning performed the best for predicting Concentration's GPA.

Keywords: *Machine Learning, Educational Data Mining, Decision Trees, Random Forests, Gradient-Boosted Trees, Deep Learning, Regression, Predicting Student GPA, Predicting Student Major GPA, Predicting Course Grade,*

1. INTRODUCTION

Students' success continues to be a key concern to individuals, higher education institutions, policymakers, and nations. Students who do not succeed in their study, lose time and effort in their failed pursuits, and they and their families can suffer financially and emotionally. Institutions lose the scarce resources they invested as well.

Last year, three of my students were dismissed from college after reaching year 4, because they could not improve their grades within the given timeframe. This semester, a large group of students moved from one major to another, after struggling to keep good grades at their first major. It is saddening to witness students suffering the consequences of non-optimal academic choices. Students are the future of our nations, and as educators, we hope to see our students successful, in every way, and we are entrusted with the responsibility of providing our students with advice and support to their academic choices, and this how this research idea started.

There is a gap in the literature, as there is no cohesive solution that informs students' academic decision throughout their study journey. The work in this research addresses this gap by utilizing the advances in machine learning to predict students' performance throughout their study years, from the time they enroll in college, till they graduate; in order to help them choose majors, concentrations, and courses.

1.1. Problem Statement

From the time students decide to continue their higher education, they are asked to make decisions concerning their education, many of which can be challenging. When students join college, they choose a major. The main offered majors at the college of study are Business, Engineering, and Information technology. After finishing their first year, students choose specific concentrations in their majors. For example, in Information Technology, students can choose Security, Programming, or Networking concentration. Throughout their study, students decide which courses to take next, which general studies courses to register for, and which upper-level electives to choose.

Wrong academic decisions have a great and direct impact on students' success and future. Choosing a major or a concentration in which they cannot perform well, can result in failure, and perhaps moving to a different major and losing time, or dropping out of college altogether. If a student is on academic probation, choosing which courses to take next becomes a critical decision. If a student continues to have low grades and fails to raise his/her CGPA within a year, the student will be dismissed from the college. In the higher education institution under study, 2,892 students are currently on academic probation, which comprises 22% of the total number of students (Figure 1), and they are not alone. In the USA, around 30% of year-one students do not return for their second year, and more than $9 billion is spent on them (Aulck et al., 2016). Furthermore, the completion rates of 4-year degrees in the US are around 50% (Sweeney et al., 2016). These alarming figures require every possible effort to support students and the higher education institutions in this critical struggle.

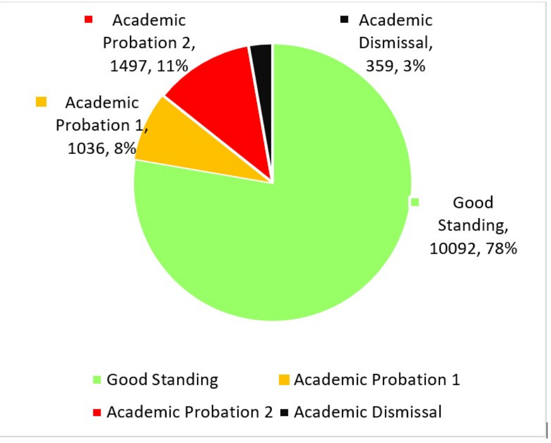


Figure 1: Academic standing of students

1.2 Research Objectives

To help alleviate those problems and to support students' academic decisions throughout their study journey, this work develops a framework that utilises historical data and machine learning algorithms to estimate how well a student will perform in different Majors, Concentrations, and Courses that they have not yet taken. Multiple machine-learning algorithms are tested and compared, to find the best performing amongst them.

The performance prediction is important to students as it can be used by them and their academic advisors to make informed choices. This can also help identify the appropriate action to take and create personalised degree pathways that enable them to successfully and effectively acquire the necessary knowledge to complete their degrees in a timely fashion.

The prediction using machine learning algorithms is done at three stages:

1) At enrollment, this research uses enrollment data to predict student performance (measured in GPA) in each of the three main offered majors: Business, IT, and Engineering. This can reduce the percentage of students changing majors after finding out that the major they selected was too challenging for them.

2) After year 1, this research uses grades of the finished courses to predict student GPA in different concentrations, such as Networking, Security, or Programming. This can inform the selection of the concentration that best matches their capabilities and maximises their chances of success.

3) At any point of time after finishing some courses, this research uses the student's finished course grades in addition to other students' grades to predict future course grades. Providing prediction of student's grades in different courses can assist the student in choosing their courses.

These tools help not only students, but also academic advisors, teachers, and administrators while supporting students at different stages of their study journey. Students on probation can avoid courses with low predicted grades and opt for courses with the highest predicted grade. Alternatively, students, their teachers, and advisors can take preventative measures and actions if a student is predicted to perform poorly in a mandatory course. Stakeholders (mainly top management) can greatly benefit from such prediction. I met with some senior management, and they were very interested in how this research can help reduce the number of students on probation. Furthermore, they are planning to offer custom specialisations and interdisciplinary degrees to students to encourage entrepreneurship and innovation, and predicting performance would be of value for students while making their choices.

To achieve the above objectives, this research will be answering the following research questions:

RQ1: How effectively can student performance in a Major be predicted at enrollment?

1.1. What are the best performing machine-learning algorithms?

RQ2: How effectively can student performance in a Concentration be predicted after year one?

2.1 What are the best performing machine-learning algorithms?

RQ3: How effectively can student performance in a course be predicted?

The remaining sections of the research paper are organised as follows: Section 2 has the literature review. Section 3 covers the overall methodology and process , in which sections 3.1, 3.2, and 3.3 describe in detail the three main research areas: Predicting Major GPA, Predicting Concentration GPA, and Predicting Course Grade. In each of those sections, I describe the data; the required preprocessing, the configuration and performance of algorithms used for the prediction, followed by discussions of the results at each stage. Finally, section 4 has a conclusion and future work.

2. LITERATURE REVIEW

Data mining is the process of discovering useful patterns and trends in large data sets. Educational Data Mining (EDM) is the application of data mining techniques in the field of education, to address important educational questions, and this area of research is growing rapidly (Shahiri and Husain, 2015). Del Rio and Insuasti (2016) surveyed papers that used data mining in predicting academic performance in traditional environments at higher-education institutions. Many of the papers reviewed addressed the issue of predicting academic performance to support student decisions. Majority of the papers reviewed predict course grades, while a few recommended majors or specializations. Also, several studies tried to predict future performance, while others attempted to predict dropouts. To the best of my knowledge, there are no papers that support student academic decisions throughout their journey in higher education, from choosing a major to a concentration, to a course.

The reviewed papers are grouped into four main categories:

1. Papers that predict a course grade,

2. Papers that predict performance in a major or a concentration

3. Papers that predict future performance (such as future GPA)

4. Papers that predict dropout.

The following sections review some of the papers in each category.

2.1 Predicting Course Grades

A large number of studies aimed at predicting course grades and many of them used those grades to recommend courses to students. Elbadrawy and Karypis (2016) in their study attempted to predict Grades of students in courses, and recommend top-n courses to students. The study used multiple sources of data such as course features, student features and academic level. The dataset was comprised of 1,700,000 grades spanning 13 years. The research used collaborative filtering and matrix factorization, in addition to popularity ranking methods. It reported that small sample sizes affect grade predictions accuracy negatively. It used a 0-4 GPA scale and the RMSEs they achieved varied, but the lowest RMSE was 0.65.

Another research that also predicted course grade was done by Iqbal et al. (2017) and used students' pre-university data such as high school grades, entry test scores, course credits, and course grades of 24 courses, for 225 undergraduate students. It also examined Collaborative filtering, Matrix factorization, in addition to Restricted Boltzmann Machines (RBM) techniques. The evaluation metrics used were Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The research concluded that Restricted Boltzmann Machines performed better than other techniques for predicting the students' grades in a particular course. The number of students is relatively small, and the study centralized the data (i.e. subtracted the average GPA of the course from the predicted values), so the RMSE they reported (which is as low as 0.3) is relative.

Ng and Linn (2017) attempted at predicting student rating of the course. The study used data from multiple sources, including course information, professor information, and students preference. Course topics were extracted using machine learning algorithms from a corpus of course descriptions. The authors also performed sentiment analysis on professors' and courses' ratings from RateMyProfessor.com website in an aim to provide a general approach that could be applied in any higher education institution. The research also asked students for their preference of the course type (for example, the quality of the course, how easy it is, etc.). Matrix factorization was used to predict the student rating of the course and make recommendations accordingly. In similar research, Chang, Lin, and Chen (2016) recommended courses to students after predicting student grade and rating of course based on multiple sources of data such as students information (including grades) and professor ratings. They also investigated combining multiple methods including collaborative filtering, clustering, and Artificial immune network.

Polyzou and Karypis (2016) predicted future course grade based on previous courses taken by the student. Their dataset had 76,748 grades of 2,949 students. This study examined both Matrix Factorization and Linear Regression. It used a 0-4 scale for GPA, and their RMSEs ranged between 0.60 and 0.75. The research showed that the accuracy of grade prediction could be improved by focusing on course-specific data, but the degree of improvement depended on the department. Dwivedi and Roshni (2017) also examined collaborative filtering approach to predict students' future grades, but it was item-based, using historical students' grades. The research used Mahout Machine Learning and Hadoop for the recommendation.

Upendran et al. (2016) and Sorour et al. (2015) used different data mining approaches, mainly unsupervised machine learning. Upendran et al. (2016) examined the use of association rule to predict the performance of the student in courses and make recommendations accordingly. The rules with the highest support and confidence were used in the recommendation model. Data used included high school grades (Math, physics, chemistry, biology, English). Sorour et al. (2015) used clustering (namely k-means), in addition to text mining (namely latent semantic analysis (LSM)) to analyse and predict student performance in a course. The text it mined contained free-style comments by students at the end of lessons. The prediction accuracy reached up to 78.5%.

Yang et al. (2018) on the other hand combined Multiple Linear Regression (MLR) and Principal Component Analysis (PCA) to improve the accuracy of predictions. The dataset included data about student's online activity (such as video-viewing), homework, exercises, and quizzes, to predict their final grade in a course.

While the previously reviewed papers provide important work in the area of predicting student performance, they only support students at the course selection stage. Major selection and concentration selection are not addressed in any of the mentioned studies. This work looks at a more comprehensive approach that supports students throughout their journey, and not just at one stage.

2.2 Predicting performance in a Major/Concentration

Not many studies were found to predict performance in a major or specialisations. Bautista et al. (2016) used classification to recommend specialisation for engineering students finishing their general engineering courses. The study used multiple algorithms and found the decision tree to be the best classifier with an accuracy of 80.06%. The study found that students grades of Algebra, Calculus, and physics, in addition to student's gender are the main predictors of success in the engineering specialisations, such as Civil Engineering, Computer Engineering, Mechatronics Engineering, and Manufacturing Engineering, etc. on the other hand, Kusumaningrum et al. (2017) used association rule to recommended majors to students based on their academic history, profile data, and interests.

Mostafa et al. (2014) recommended majors to students transferring from one major to another, using a case-based reasoning system (CBR). The study based the recommendation on the similarity between the previous courses taken by the students and the concepts in the different majors and recommends the major that is nearest to student's learned concepts. Surveys were also given to advisors to evaluate the system, but no results were published.

These studies also focus only on recommending either a major or specialisation and do not support students' academic decisions throughout their study journey.

2.3 Predicting Future performance

Several studies investigated future performance prediction. Naser et al. (2015) used Artificial Neural Network (ANN) to predict senior student performance in the faculty of Engineering. The authors used numerous variables for input such as high school score, Math I, Math II, Electrical Circuit I, and Electronics I scores, number of completed credits, CGPA, high school type, and gender, among others. The data consisted of 150 students only, and they stated that their ANN model was able to correctly predict the performance of more than 80% of prospective students. However, the study only focused on one algorithm only and did not explore other possible algorithms.

Asif et al. (2017) used data mining to predict and understand the performance of students. Firstly, they used classification to predict student graduation performance using socio-economic data for 210 students. Data used included pre-university grades in addition to the first two years. Secondly, the research identified courses that predict good or poor performance. Using decision trees, four courses were found to be the strongest indicators. Lastly, they investigated how students' academic performance progresses over the years of study. The clustering techniques they used revealed that students tend to have the same performance (low, intermediate, or high marks) in all courses, and throughout the years. Tekin (2014) also attempted to predict a student's GPA at graduation. The study investigated the use of Naïve Bayes', Support Vector Machine (SVM) and Extreme Learning Machine (ELM) classifiers. In one scenario, the researcher used students' grades in their first two years to predict their GPA at graduation. In the second scenario, grades for the first three years were used. Their data consisted of the courses taken by 127 students only. Their best-reported accuracy was achieved using SVM and reached up to 93.06% for the first scenario, and 97.98% for the second scenario. Such high accuracy needs further investigation as the model could be overfitting.

Goga, Kuyoro, and Gogan, (2015) addressed predicting student's first-year performance by designing a framework using machine learning. The framework uses background data to make predictions, and utilizes Decision Trees, Neural Networks and Association rules methods. Furthermore, data is fed into a recommender system to suggest the course of action. The study, however, did not provide a detailed evaluation of the work.

Patil, Ganesan, and Kanavalli (2017) developed Feed-Forward Neural Networks and Recurrent Neural Networks to predict students GPA based on the courses they have taken previously. The research RMSE as the evaluation metric to compare the two methods. In a similar study, Al-Barrak and Al-Razgan (2016) used students' grades in previous mandatory courses to predict their future GPA. The dataset comprised of 236 students records and used Decision Tree only for GPA class prediction (A+, A, B+,…, F). The study identified the most important courses for predicting performance in each semester as well.

Elbadrawy et al. (2016) aimed at predicting next-term GPA as well as student's performance on in-class assessments (for example, homework). The research used regression-based methods and Matrix Factorization techniques. It reported an RMSE of 0.7381 for next term GPA prediction (GPA is between 0 and 4). The study also concluded that both Personalized Linear Multi-regression and their advanced Matrix factorization techniques could predict next-term grades with lower error rates than traditional methods. The data set included admissions records and grades in courses that were already taken by all the students; in addition to course information and instructors' information.

All the studies in this section addressed predicting student's future performance while they are in their current year, but do not predict their performance in particular courses, majors, or specializations.

2.4 Predicting Dropout

Various studies addressed the concern of student's dropout. Aulck et al. (2016) attempted at predicting students retention using student demographic data, pre-college entry information, and first-year transcript records of 32,500 students at the University of Washington. Regularised logistic provided them with the best predictions, and the study reported math, chemistry, psychology, and English courses to be the strongest predictors of attrition, in addition to birth-year and enrollment-year.

Both Manhaes et al. (2014), Sara et al. (2015) used the classification techniques to predict dropout. Manhaes et al. (2014) used student's grades in each semester to predict dropout. The study examined multiple classification algorithms including Naïve Bayes (NB), Support Vector Machine (SVM), and Multilayer Perceptron (MLP). Naïve Bayes achieved the best accuracy of 80%. On the other hand, Sara et al. (2015) found but Random Forest classifier to achieve the best accuracy of 93.5% and an area under the curve (AUC) of 0.965. The research used a large dataset consisting of 36,299 students.

Wolff et al. (2013) predicted students at risk of failing an online course by analysing their clicks. The research used multiple sources of data such as demographic data, assessments, and virtual learning environment. They had three modules: module A consisted of 4,397 students and 1,570,402 clicks, and Module B consisted of 1,292 students and 2,750,432 clicks. The researchers used classification to predict the final result of a student (Pass or fail). It found the level of activity of students and the number of clicks around the exam times to be predictors of student performance. Better accuracy was reported as a result of combining assessments, demographic, and clicks data.

The studies reviewed in the area of predicting dropout focused mainly on whether a student is at risk of dropping out or not, and highlighted the predictors of dropping out. While this serves as a very useful warning system, it does not offer insight into futures course performance, which could greatly help students who are at risk of dropping out choose courses that could take them out of the risk zone.

The work found in the literature review which focused on one stage of student's academic journey or another inspired the work of this research. The aim of this work is to provide a cohesive and comprehensive framework to support students' choices throughout their academic journey and offer them the largest amount of possible support as they choose courses, majors, and concentration, in the hope of maximizing their chances of better performance. In the coming sections, I go throughout the overall methodology and framework, followed by a detailed description and findings of each stage of support.

3. METHODOLOGY AND RESULTS

The college under study has three main majors: Business, Information Technology (IT), and Engineering. There is a total of 13,750 students in the different majors as per Figure 2, where 4,047 students are studying Business, 3,492 in IT, and 6,211 in Engineering.

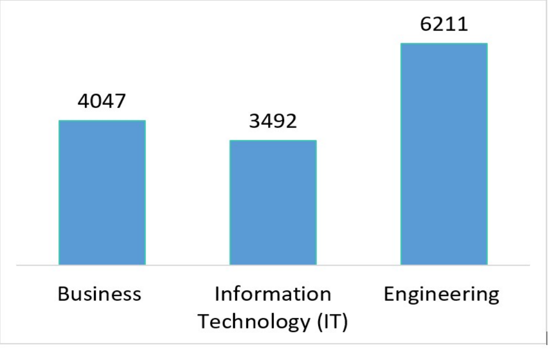


Figure 2: Number of students in the different majors

To supports students in choosing Majors, Concentrations, and Courses, this study develops a framework to predict their GPA at each stage. Figure 3 shows the suggested framework and a summary of the main tasks performed in this research. At each stage: data to be used, the prediction task and the algorithms used are shown.

The main sections of the framework are designed around the following stages of the student's journey:

1) At enrollment: based on their enrollment data, the framework predicts their GPA in different majors to help them choose a major most suitable for their capabilities.
2) At the end of Year 1: based on their grades in Year 1 courses, the framework predicts their GPA in different concentrations to help them in their choice.
3) At any time after year 1 or after finishing some courses: based on their grades in previously finished courses, the framework predicts their grade in any future course to help them choose courses.

The Overall Process:

RapidMiner was used in this study to implement the framework and predict students' performance at the different stages. Below is the general approach used for all the prediction tasks (Major GPA, Concentration GPA, and Course Grade prediction) –also shown in Figure 4:

1. Preprocess data in Excel (basic preprocessing)

2. Retrieve data with a numerical label for regression.

3. Preprocess data as per the requirement of the algorithm and the task at hand

4. Assign "GPA" as the label for regression training and testing

5. Split Data into training data to build the model (70%), and testing data (30%) to test the performance of the model.

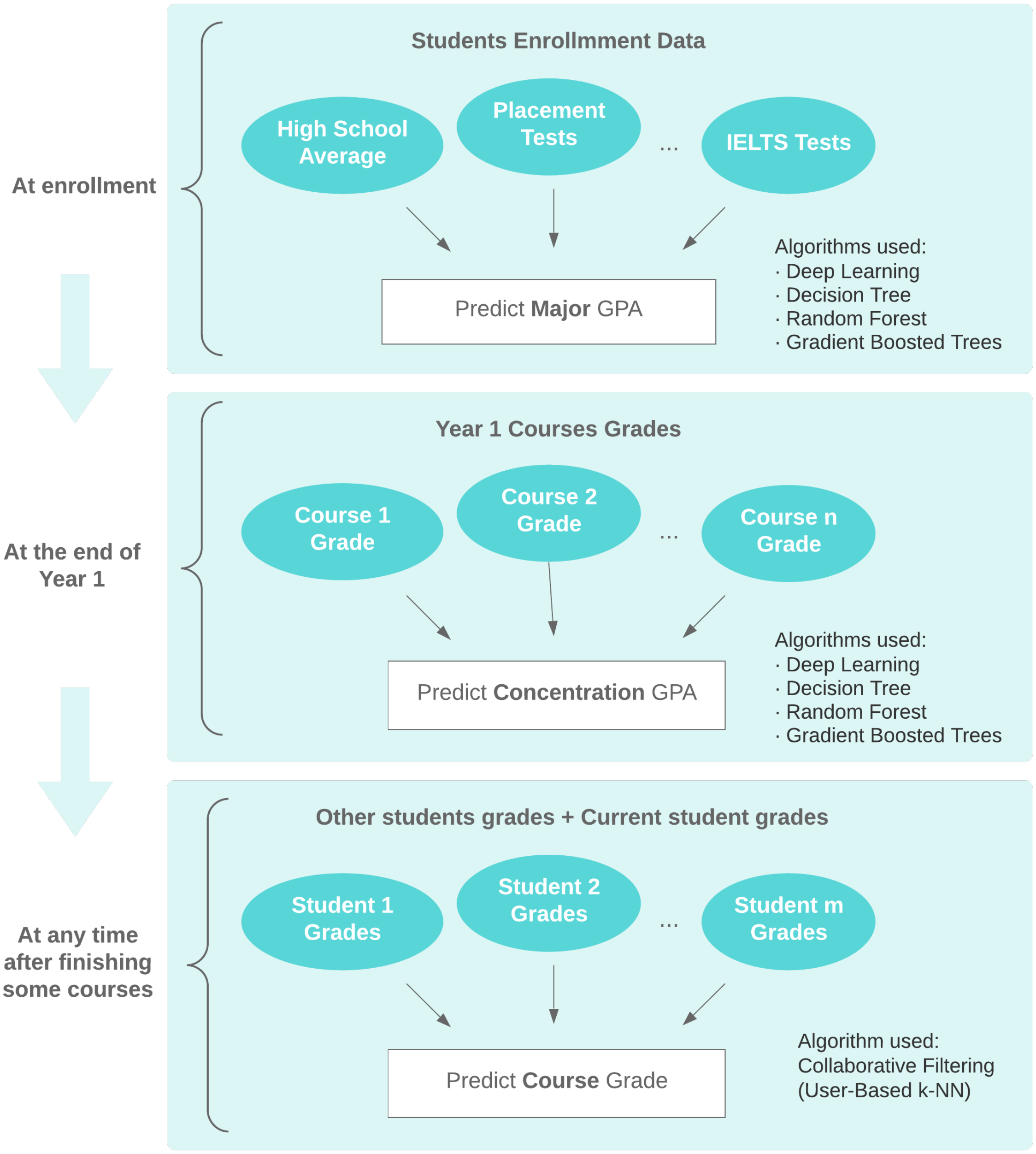


Figure 3: Framework to support students' decisions throughout their study journey

6. Pass the training data to a machine-learning regression algorithm to build a model. The following algorithms were used and compared:

- Deep Learning
- Decision Tree
- Random Forest
- Gradient Boosted Trees
- User-based K-Nearest Neighbors

7. Apply the trained model to the testing data

8. Find Performance of regression using cross-validation: Compare 'label' and 'prediction' to estimate performance.

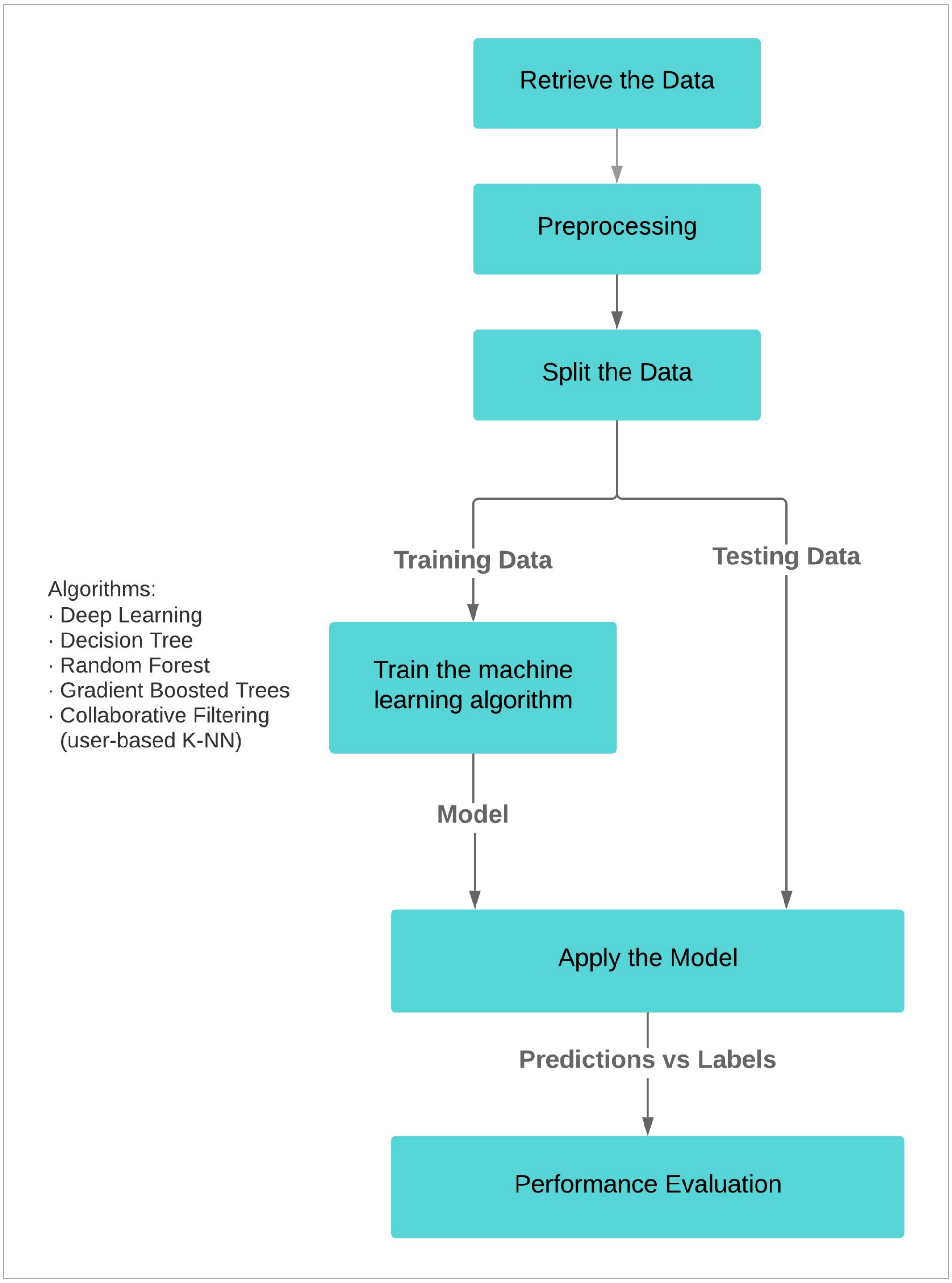


Figure 4: Overall approach for prediction tasks

In the following sections, I describe in detail the three main tasks of the proposed framework which provide answers to the main research questions. Section 3.1 "Predicting Major GPA" answers RQ1 "How effectively can student performance in a Major be predicted at enrollment?" It has the data used for predictions, the preprocessing tasks, the algorithms used, and the performance of algorithms. Section 3.2 "Predicting Concentration GPA" answers RQ2 "How effectively can student performance in a Concentration be predicted after year one?" It also describes the data, the preprocessing, the algorithms, and the results. Finally, Section 3.3 "Predicting Course Grade" answers RQ3 "How effectively can student performance in a course be predicted?"

3.1 Predicting Major GPA (at enrollment)

When students first join the college, they need to choose a major. The only information available is their enrollment data. This data is used to predict student GPA in each of the three main offered majors: Business, IT, and Engineering to help him/her make an informed decision while choosing one of the majors. When a student joins college, we could run his/her enrollment data through each prediction model, and it will give the predicted GPA in each major. For example, if a student's predicted GPA in one major turned to around 2.0 while in another major it was around 3.0, he/she might opt for the major with higher probability of success, and avoid the possibility of wasting time in a major that might not be best suited for his or her capabilities and strengths.

In order to do that, students' historical enrollment data and their achieved GPA (after finishing two years) is used to train the algorithms for GPA prediction. In the coming section, I describe the data, the preprocessing steps, the algorithms trained for predictions, the performance of the different algorithms, a discussion of the results.

3.1.1 Data

I collected enrollment data of 7,230 students studying in the year 2018 in 3 different majors: Information Technology (1,725 students), Business (2,412 students) and Engineering (3,093 students). Table 1 lists the used features obtained.

Table 1: Enrollment data features and range of values

Feature	Values
High school Average	0-100
High school English	0-100
High school Math	0-100
High school Arabic	0-100
IELTS Band	0-9
IELTS Reading	0-9
IELTS Writing	0-9
IELTS Listening	0-9
IELTS Speaking	0-9
College placement tests (CEPA) English	0-210
College placement tests (CEPM) Math	0-210
College placement tests (CEPW) Writing	0-210
Major	Polynomial
Concentration	Polynomial
Employment	Yes/No
Gender	M/F
GPA (**label**):	Continuous value between 0 – 4

3.1.2 Preprocessing

Data collected had to be cleaned and made ready for prediction. Below are the main tasks of preprocessing:

- Anonymized the dataset –removed over 20 features that contained personal details of students such as IDs, names, and contact details.

- Removed noise –removed records that had mistakes such as letters instead of numbers , or wrong data such as 0218 for a year, instead of 2018.

- Removed students with GPA=0, as this is usually to students not showing up and getting a failing grade in all courses due to attendance.

- Removed all newly registered student records, by filtering their catalogue term, since they would not have any GPA,

- Generated a new feature for “Employment” to indicate whether a student is working or not. Original data only had the company name. This feature was generated to find out if employment is a contributing factor to performance prediction.

- Filtered students in Year 3 and Year 4 only. To achieve this, the number of credits completed was checked. Students who have completed more than 60 credits were assumed to have finished year 2.

3.1.3 Algorithms

The following algorithms are popular in literature for regression, and hence are used in this study:

- Deep Learning
- Decision Tree
- Random Forest
- Gradient Boosted Trees

RapidMiner auto model default values are used (unless otherwise stated).

Below is the description and the configurations of the algorithms used to perform the regression task in Rapid miner:

Deep Learning:

- RapidMiner H2O Deep Learning operator is used to predict GPA. Since the label is real, regression is performed.
- The hidden layer sizes parameter is set to 2 layers, each with 50 neurons.

Decision Tree

- RapidMiner H2O Decision Tree operator is used to predict GPA.
- “GPA” is set as a label.
- To use the Decision Tree for regression, 'least_square' is selected as a criterion.

Random Forest

- RapidMiner, Random Forest operator, is used for regression. The model port provides the ensemble of random trees used in combination to obtain a combined prediction. At each leaf of a tree, the average value of GPAs is shown.
- “GPA” is set as a label.
- To use Random Forest for regression, 'least_square' is selected as a criterion.

Gradient Boosted Trees

- RapidMiner H2O Gradient Boosted Trees operator is used to predict the GPA. Since the label is real, regression is performed.
- The operator's distribution parameter is changed to "gamma".
- The algorithm was used to generate 60 Trees to create an ensemble model.

3.1.4 Results

Table 2 shows a summary of the RMSE, standard deviation (STDV), and runtime (in milliseconds), for each algorithm, on each major (Business, Engineering and Information Technology). It also shows the number of records used in each major. Figure 5 shows a graphical comparison of the RMSEs.

Table 2: Summary of algorithms' performance for the Major GPA prediction

	Business			Engineering			Information Technology			
Records	2,412			3,093			1,725			
Algorithm	RMSE	STDV	Run time	RMSE	STDV	Run time	RMSE	STDV	Run time	Avg RMSE
Gradient Boosted Trees	**0.469**	0.012	33805	**0.45**	0.008	16662	0.465	0.025	11867	**0.461**
Random Forest	0.486	0.017	87814	0.462	0.007	84087	**0.458**	0.033	31177	0.469
Deep Learning	0.484	0.016	16375	0.454	0.004	7629	0.481	0.019	3662	0.473
Decision Tree	0.53	0.015	10911	0.512	0.007	2971	0.515	0.026	1503	0.519

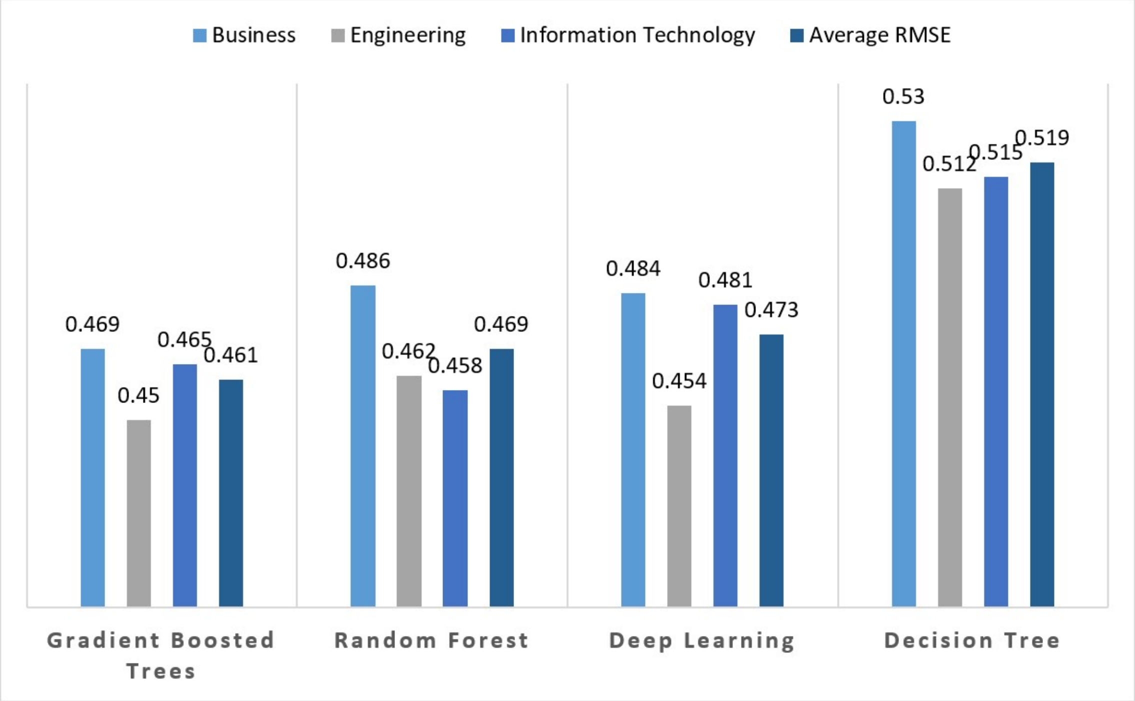


Figure 5: RMSEs of the algorithms used in predicting Major GPA

Elbadrawy et al. (2016) predicted next-term GPA using regression-based methods and Matrix Factorization techniques. Their RMSE was 0.7381 (GPA scale is between 0 and 4). In our research, Gradient Boosted trees algorithm performed the best in predicting Business and Engineering majors GPA (RMSE 0.469 and 0. 45 respectively) while Random forest performed slightly better in predicting Information Technology GPA (RMSE 0.458). Deep closely followed with an average RMSE of 0.473. Decision Tree was the least performing across all data sets with an RMSE average of 0.519. It is interesting to find out that ensemble methods improve the accuracy of predictions. Standard deviations were low in general (the highest was 0.03). Hence, standard deviations are not taken into consideration while comparing the performance.

3.2 Predicting Concentration GPA (after year 1)

After joining a major, and at the end of their first year, students are asked to choose a concentration. For example, the Information Technology major has multiple concentrations, namely: Security, Programming, and Networking. Many students have difficulty choosing between the concentrations and are not sure which of them better matches their strengths and offers them the best chances of success.

To assist students in choosing a concentration by the end of year 1, this work predicts their GPA in the different concentrations, using their marks in five IT-related courses that they take in their first year. The five courses are:

- CIS 1003 - Introduction to Information Systems
- CIS 1103 -Introduction to Networking
- CIS 1203 - Introduction to Web Technologies
- CIS 1403 - Introduction to Programming
- CIS 1303 - Introduction to Database concepts

I built a prediction model for the three concentrations of the IT major. However, the approach applies to any major (which is planned for future research). When a student finishes Year 1 and wants to predict his/her GPA in different concentrations, we can run his/her five courses data through each prediction model, and it will give the predicted GPA in each concentration. This can help students decide on concentrations best suited for their capabilities.

In the coming section, I describe the data, the preprocessing steps, the algorithms trained for predictions, the performance of the different algorithms, a discussion of the results.

3.2.1 Data

The data collected consists of the student's grades in the five courses taken in Year 1, along with their GPA. The total number of grades is 7,740 grades of 1,560 senior students (in Year 3 and Year 4). The number of records in each concentration was as follows: Security (1,715 grades of 343 students), Programming (1,260 grades of 252 students) and Networking (1,160 grades of 232 students). The features are shown in Table 3.

Table 3: Year 1 data features and the range of values

Feature	Values
CIS 1003 Grade -Introduction to Information Systems (IS) (0-4)	0-4
CIS 1103 Grade -Introduction to Networking (NW)	0-4
CIS 1203 Grade - Introduction to Web Technologies (WEB)	0-4
CIS 1403 Grade - Introduction to Programming (PRG)	0-4
CIS 1303 Grade - Introduction to Database concepts (DB).	0-4
GPA (Label)	0-4

3.2.1 Preprocessing

To filter students in year 3 and year 4 only, the number of credits completed is checked. Students who have completed more than 60 credits are assumed to have finished Year 2.

Data is anonymized, and cleared of errors. Certain columns were combined, as surprisingly enough, the same course grade was stored in different columns for different students because the same course had multiple codes (with a different suffix).

The individual grades obtained were in Letter format (A, A-, B+,…, F), so a new feature was generated to compute Grade Points (GP) in numbers (between 0 and 4), following the college grading system as shown in Table 4.

Table 4: Grade letters and their corresponding grade points

Grade Letter	Grade Points
A	4
A-	3.7
B+	3.3
B	3
B-	2.7
C+	2.3
C	2
C-	1.7
D+	1.3
D	1
F	0

3.2.3 Algorithms

To predict Concentration GPA, the same general approach is applied- as outlined in methodology (section 3) and used the same algorithms used for predicting Major GPA (section 3.1.3), namely, Deep Learning, Decision Tree, Random Forest, Gradient Boosted Trees. In this stage, the algorithms were trained on Year 1 course grades data (as opposed to enrollment data in the previous stage).

3.2.4 Results

Table 5 shows a summary of the RMSEs, and standard deviations (STDV) for each algorithm, and in each concentration (Security, Programming, and Networking). It also shows the number of records used in each concentration.

Table 5: Summary of algorithms' performance for the Concentration GPA prediction

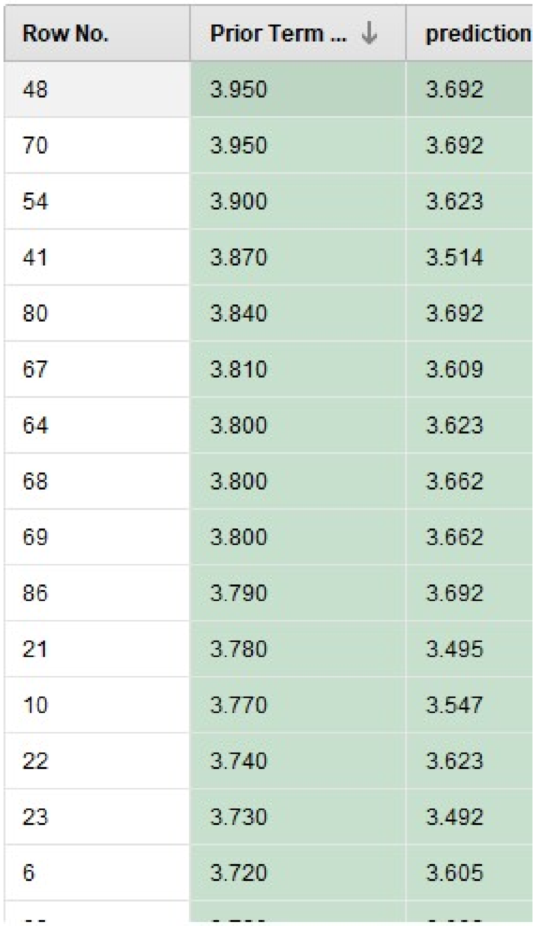
	Security		Programing		Networking		
# of Grades	1,715		1,260		1,160		
# of students	343		252		232		
Algorithm:	RMSE	STDV	RMSE	STDV	RMSE	STDV	Average RMSE
Deep Learning	0.18	0.05	0.24	0.02	0.22	0.03	0.22
Random Forest	0.22	0.03	0.27	0.01	0.26	0.02	0.25
Gradient Boosted Trees	0.22	0.02	0.27	0.03	0.27	0.03	0.25
Decision Tree	0.33	0.02	0.31	0.02	0.35	0.05	0.33

The results are exciting, as the RMSEs for predicting concentration's GPA are relatively low. The lowest average RMSE obtained was 0.21 using Deep Learning, followed by Random Forest and Gradient Boosted Trees (both had an average of 0.25 RMSE). Deep learning performed particularly well for the security concentration, with an RMSE of 0.18, perhaps due to the relatively larger number of records in the security dataset. Decision trees had the highest RMSE with an average of 0.33. It is worth noting that Decision Trees also had the least performance in the previous task of Major GPA prediction (section 3.1).

Overall, the standard deviation is also small (an average of 0.03 across concentrations) which means that this small error in prediction (the RMSE) is relatively consistent. Hence, it was not taken into consideration while comparing the performances of the algorithms.

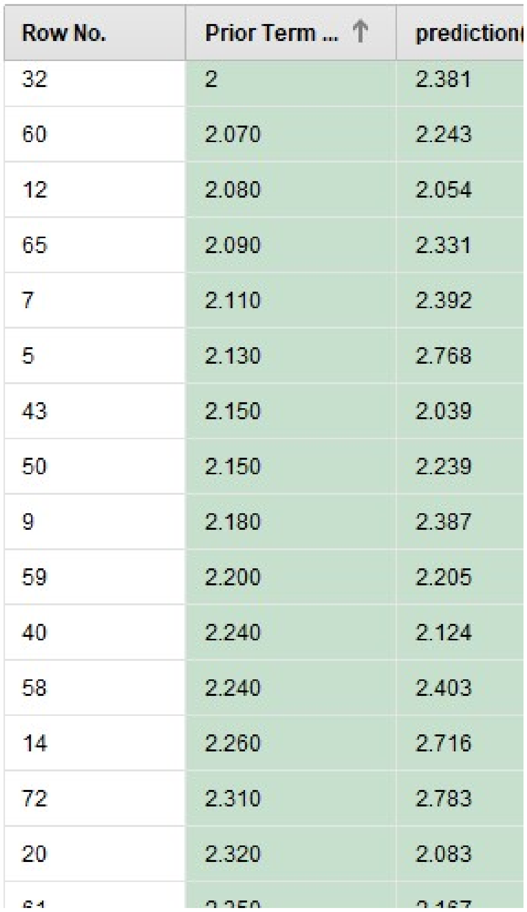
I manually inspected the prediction results, both for high and low GPAs, to see how close the prediction are to the actual values, and they were very close (Please see Figure 6 and Figure 7).

A significant observation is that the error in predicting concentration GPA after one year of study (using Year 1 courses) is considerably smaller than the error in predicting major GPA based on enrollment data (RMSE 0.22, and 0.46 respectively). This means that predicting performance becomes more accurate as students take courses in the college and that enrollment data is not as accurate in predicting GPA.



Row No.	Prior Term ... ↓	prediction
48	3.950	3.692
70	3.950	3.692
54	3.900	3.623
41	3.870	3.514
80	3.840	3.692
67	3.810	3.609
64	3.800	3.623
68	3.800	3.662
69	3.800	3.662
86	3.790	3.692
21	3.780	3.495
10	3.770	3.547
22	3.740	3.623
23	3.730	3.492
6	3.720	3.605

Figure 6: Actual vs. Predicted Concentration GPA- High GPAs



Row No.	Prior Term ... ↑	prediction
32	2	2.381
60	2.070	2.243
12	2.080	2.054
65	2.090	2.331
7	2.110	2.392
5	2.130	2.768
43	2.150	2.039
50	2.150	2.239
9	2.180	2.387
59	2.200	2.205
40	2.240	2.124
58	2.240	2.403
14	2.260	2.716
72	2.310	2.783
20	2.320	2.083

Figure 7: Actual vs Predicted Concentration GPA- Low GPAs

3.3 Predicting Course Grade

After finishing their first year, and throughout their study, many students, and in particular the ones at-risk, struggle to choose the next courses, especially with electives, and general studies courses. Students who are on probation are at risk of academic dismissal due to a low GPA. Choosing a course with the highest probability of success offers them a better chance to move out of probation. Furthermore, the college is promoting entrepreneurship and is moving to flexible degrees where students can customise their study plans and take interdisciplinary certificates. Hence, predicting student's performance in a course can greatly support students' decision in choosing a course for the above reasons.

Future course Grade Point (0-4) in this task is predicted using a collaborative filtering approach. In this approach, the grades of any courses finished by the student, in addition the grades of other similar students, are used to predict future grade.

In the coming section, I describe the data, the preprocessing steps, the algorithms trained for predictions, the performance of the different algorithms and a discussion of the results.

3.3.1 Data

The obtained data consists of 227,507 grades in all offered courses across all the majors. There are 80,324 grades for the Business students, 60,440 grades for IT students, and 86,743 grades for Engineering students for the year of 2018. Table 6 shows the used features.

Table 6: Course grade prediction features and the range of values

Feature	values
Student ID	Polynomial
Course Code	Polynomial
Major	Polynomial
Grade in course	0-4

3.3.2 Preprocessing

The following preprocessing steps are performed on the data:

- Integrate data from multiple files
- Un-pivot the data to follow the format (user, item, rating) that is necessary for prediction. The matching format for this research would be (student, course, grade) as shown in Table 7.

Table 7: Un-pivoted format of (student, course, grade) data

Student_ID	Course_Code	Grade Point
1	TEC-112	4
1	GEN-453	3.3
1	TEC-001	2.3
2	TEC-112	1
2	GGB-100	2

- Remove records that have no grades.
- Remove records that have missing values
- Clean course codes from duplicates
- Remove records that have unwanted grades (such as “W” for Withdrawn courses as opposed to A, B, C, D, F)
- The individual grades obtained were in Letter format (A, A-, B+, …, F), so a new feature is generated to compute Grade Points (GP). The Grade Point is between 0 and 4 using the college grading system

3.3.3 Algorithms

To predict course Grade point (0-4) this work uses the Collaborative Filtering approach; a recommender system approach (commonly applied in recommender systems to predict ratings, but here it is used to predict grades). This approach predicts one student’s grade on non-graded courses based on similarity with other students. Recent studies started using this approach for predicting students grades such as Elbadrawy and Karypis (2016), Iqbal et al. (2017), Polyzou and Karypis (2016), Ng and Linn (2017), Chang, Lin, and Chen (2016), and Dwivedi and Roshni (2017).

I tested the algorithm on the courses of the three majors: Business, IT, and Engineering. If a student is a Business student, his record could be compared to business students’ records to speed up the process. I also tested the algorithm on all records combined, in case a student wants to take courses from different majors.

The algorithm used is Weighted User-Based K-Nearest Neighbor with Pearson Similarity (available through RapidMiner recommender system extension > item rating prediction > user k-NN). It executes a Collaborative Filtering recommender based on (student, course, grade) matrix. The algorithm compares student grades to other students’ grades, and similar students are found. By similar, we mean students who took the same courses and achieved close results. The K-Nearest Neighbor in Collaborative Filtering works as follows:

1. The algorithm looks for students who share the same grades patterns with the current student (the student whom the prediction is for).
2. The algorithm measures how similar each student in the database to the current student using K-Nearest Neighbors with Pearson’s correlation coefficient as a similarity measure.
3. The resulting similarity is used as a weight while calculating the weighted average of the grades of similar students.
4. The resulted grade is used as a prediction for the current student’s grade.

An advantage of using collaborative filtering is that there is no need to build a profile of features for each course. The approach has limitations such as cold start and data sparsity, however, these limitations are at their minimal in our case, since we have ample data, with many students taking the same courses. This method is most useful after year one as per our plan. Hence, the student would have finished some courses, and this avoids the cold start issue and allows for better predictions.

Unlike regression algorithms where we only need to specify the "Label" column to be predicted, in the collaborative filtering we need to specify both the "Label" and the "Item identification" columns. Table 8 shows the feature and the target role assignment in RapidMiner. The k value chosen was 20 (it was found to have the best prediction). The minimum rating is set to 0, and maximum is set to 4 since course grades fall in this range.

Table 8: Target role assignment for user-based k-NN in RapidMiner

Feature name	Target role
Mark (or Grade)	Label
Short_Course (shortened Course Code)	Item identification

3.3.4 Results

Table 9 shows the number of records and the RMSE for each major. RMSEs for Business, IT, and Engineering were 0.69, 0.46, and 0.66 respectively. When all records were combined, RMSE was 0.66. The least error that could be achieved was 0.457 for IT grades prediction.

Table 9: Summary of the performance for the Course grade prediction

	Business	IT	Engineering	All	Avg RMSE
Records	80,324	60,440	86,743	227,507	
RMSE	0.69	0.46	0.67	0.66	0.62

Even though the error of predicting a Course Grade at this stage (with an average RMSE of 0.62) is larger than previous stages (average Major GPA prediction was 0.46, and Concentration GPA prediction was 0.22 - Figure 8), I believe this is still acceptable, for multiple reasons.

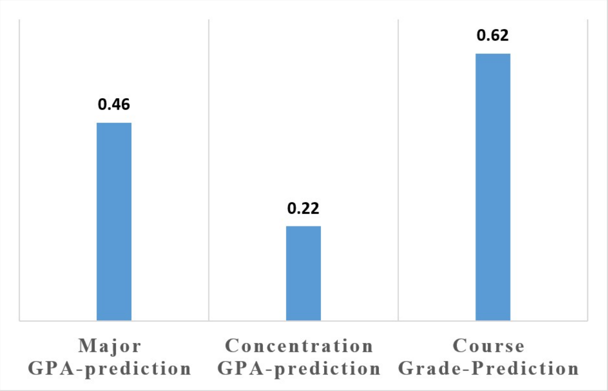


Figure 8: Average RMSEs of the main stages in this research

First, unlike GPA prediction where the GPA is the average of many courses, and most GPAs will be within a smaller range, namely between 1.75 and 4.0 (because students must maintain a GPA above 2.0 to proceed), here we are trying to predict a single course grade. This can take any value in the range between 0-4, and not necessarily in the upper range. This can make the prediction problem harder and the chance of getting a higher error is bigger, because the range of the data is larger.

Second, let us say the actual course grade was 4.0 (A), and the algorithm predicted 3.5 (B+), or even 3.0 (B). This is an error of 0.5 and 1.0 respectively. I would argue that this would not be considered a very far prediction since the grade is still relatively high. It's very unlikely that the algorithm would predict 0(F) or 1 (D) for that student.

Third, the measure used here is RMSE, which penalizes larger errors, hence it is considered stricter as compared to other measures of error such as MAE.

Lastly, the results obtained in this research (average RMSE of 0.62) are comparable (and sometimes better) than other published research predicting course grades on a scale 0-4. For example, Elbadrawy and Karypis (2016) reported an RMSE of 0.65 using collaborative filtering for predicting course grades, and Polyzou and Karypis (2016) reported RMSEs between 0.60 and 0.75 using both linear regression and matrix factorization techniques.

Having said that, I am still very interested in finding ways to improve prediction accuracy to maximize the value of these predictions. I have tried different approaches, such as filtering for only a specific type of courses and concentrations for example, but none of them improved the performance. I would like to investigate more ways such as incorporating hybrid approaches or improving the prediction algorithm itself.

4. CONCLUSION AND FUTURE WORK

Student success is of great importance to students, their families, higher education institutions, society, and nations. Predicting students' future performance can help students, teachers, and advisors make informed choices. This research developed a framework to predict student performance (as measured by GPA or grade) in different Majors, Concentrations, and Courses they are yet to take, using machine-learning. Literature has covered one area or another, but this research fills in the gap, and offers comprehensive support to students' decisions throughout their study journey starting from enrollment (when performance in different majors is predicted), followed by another prediction after one year (when performance in different concentrations is predicted), in addition to perdictions at any point of time after that, (when performance in any course can be predicted). Furthermore, multiple machine-learning algorithms were used, and their performance is compared (summary in presented in the coming paragraphs).

Below are the research questions of this study and a summary of the findings detailed in previous sections:

RQ1: How effectively can student performance in a Major be predicted at enrollment?

1.1. What are the best performing machine-learning algorithms?

At enrollment, student enrollment data (such as high school grades, IELTS scores, college placement tests in English and math) is used to predict student's GPA (scale 0-4) in different majors. Deep learning, Decision Tree, Random Forest, and Gradient Boosted Trees algorithms are used. Gradient Boosted trees algorithm performed the best in predicting Business and Engineering majors GPA (RMSE 0.469 and 0. 45 respectively) while Random forest performed slightly better in predicting Information Technology GPA (RMSE 0.458).

RQ2: How effectively can student performance in a Concentration be predicted after year one?

2.1 What are the best performing machine learning algorithms?

After year 1, students are asked to choose concentrations within their majors. This study uses year one courses to predict student's GPA at different concentrations in the IT major. The error in predicting Concentration's GPA was considerably smaller than the error in predicting Major GPA (RMSE 0.22 vs 0.46). Deep Learning algorithm achieved the least average RMSE of 0.21, followed by Random Forest and Gradient Boosted Trees (both had an average of 0.25 RMSE). Decision Tree had the least performance with an average RMSE of 0.33.

RQ3: How effectively can student performance in a course be predicted?

At any point after finishing some courses, student's grades of previously finished courses can be used to predict their grade in future courses. A Collaborative Filtering approach using K-Nearest Neighbor is used to predict student grade point (0-4). An average RMSE of 0.62 was achieved. Improving the accuracy of prediction is an area for further exploration.

This study, however, did not attempt to optimize the performance of the used algorithms, and the algorithms used are only a few in number. In future, I plan on investigating ways to optimize the performance of the algorithms, and investigate other algorithms and methods hoping they would provide improved results. Furthermore, this study did not study the effect of other data, such as attendance data, detailed coursework grades, and student's course feedback, to list a few, which could possibly contribute to a better prediction of future performance. I will attempt to obtain more data to include other factors in predictions.

In order to fully utilize the power of machine learning, it would be of most value to operationalize this work and integrate it within the business solutions currently available for planning and selecting courses and for advising. It would be exciting to show students the predicted performance in the majors, concentrations, or courses they are interested in. Another significant addition to this work would be to include explanations of predictions, possible interventions, and guidance on how to improve student's chances of success in case they opt for a choice with less predicted GPA. This can be of support to all stakeholders.

REFERENCES:

[1] Al-Barrak, M.A. and Al-Razgan, M., 2016. Predicting students final GPA using decision trees: A case study. International Journal of Information and Education Technology, 6(7), p.528.

[2] Asif, R., Merceron, A., Ali, S.A. and Haider, N.G., 2017. Analyzing undergraduate students' performance using educational data mining. Computers & Education, 113, pp.177-194.

[3] Aulck, L., Velagapudi, N., Blumenstock, J. and West, J., 2016. Predicting student dropout in higher education. arXiv preprint arXiv:1606.06364.

[4] Chang, P.C., Lin, C.H. and Chen, M.H., 2016. A hybrid course recommendation system by integrating collaborative filtering and artificial immune systems. Algorithms, 9(3), p.47.

[5] Del Río, C.A. and Insuasti, J.A.P., 2016. Predicting academic performance in traditional environments at higher-education institutions using data mining: A review. Ecos de la Academia, 2016(7).

[6] Dwivedi, S. and Roshni, V.K., 2017, August. Recommender system for big data in education. In E-Learning & E-Learning Technologies (ELELTECH), 2017 5th National Conference on(pp. 1-4). IEEE.

[7] Elbadrawy, A. and Karypis, G., 2016, September. Domain-aware grade prediction and top-n course recommendation. In Proceedings of the 10th ACM Conference on Recommender Systems (pp. 183-190). ACM.

[8] Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G. and Rangwala, H., 2016. Predicting student performance using personalized analytics. Computer, 49(4), pp.61-69.

[9] Goga, M., Kuyoro, S. and Goga, N., 2015. A recommender for improving the student academic performance. Procedia-Social and Behavioral Sciences, 180, pp.1481-1488.

[10] Iqbal, Z., Qadir, J., Mian, A.N. and Kamiran, F., 2017. Machine Learning Based Student Grade Prediction: A Case Study. arXiv preprint arXiv:1708.08744.

[11] Manhães, L.M.B., da Cruz, S.M.S. and Zimbrão, G., 2014, March. WAVE: an architecture for predicting dropout in undergraduate courses using EDM. In Proceedings of the 29th Annual ACM Symposium on Applied Computing (pp. 243-247). ACM.

[12] Mostafa, L., Oately, G., Khalifa, N. and Rabie, W., 2014, March. A case based reasoning system for academic advising in egyptian educational institutions. In 2nd International Conference on Research in Science, Engineering and Technology (ICRSET'2014) March (pp. 21-22).

[13] Naser, S.A., Zaqout, I., Ghosh, M.A., Atallah, R. and Alajrami, E., 2015. Predicting student performance using artificial neural network: In the faculty of engineering and information technology. International Journal of Hybrid Information Technology, 8(2), pp.221-228.

[14] Patil, A.P., Ganesan, K. and Kanavalli, A., 2017, December. Effective Deep Learning Model to Predict Student Grade Point Averages. In 2017 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)(pp. 1-6). IEEE.

[15] Polyzou, A. and Karypis, G., 2016. Grade prediction with models specific to students and courses. International Journal of Data Science and Analytics, 2(3-4), pp.159-171.

[16] Sara, N.B., Halland, R., Igel, C. and Alstrup, S., 2015. High-school dropout prediction using machine learning: A danish large-scale study. In ESANN 2015 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence (pp. 319-24).

[17] Shahiri, A.M. and Husain, W., 2015. A review on predicting student's performance using data mining techniques. Procedia Computer Science, 72, pp.414-422.

[18] Sorour, S.E., Mine, T., Goda, K. and Hirokawa, S., 2015. A predictive model to evaluate student performance. Journal of Information Processing, 23(2), pp.192-201.

[19] Sweeney, M., Rangwala, H., Lester, J. and Johri, A., 2016. Next-term student performance prediction: a recommender systems approach. arXiv preprint arXiv:1604.01840.

[20] Tekin, A., 2014. Early Prediction of Students' Grade Point Averages at Graduation: A Data

Mining Approach. Eurasian Journal of Educational Research, 54, pp.207-226.

[21] Upendran, D., Chatterjee, S., Sindhumol, S. and Bijlani, K., 2016. Application of predictive analytics in intelligent course recommendation. Procedia Computer Science, 93, pp.917-923.

[22] Wolff, A., Zdrahal, Z., Nikolov, A. and Pantucek, M., 2013, April. 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). ACM.

[23] Yang, S.J., Lu, O.H., Huang, A.Y., Huang, J.C., Ogata, H. and Lin, A.J., 2018. Predicting Students' Academic Performance Using Multiple Linear Regression and Principal Component Analysis. Journal of Information Processing, 26, pp.170-176.