I.J. Modern Education and Computer Science, **2022, 6, 1-12**
Published Online on December 8, 2022 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2022.06.01

Machine Learning Algorithms for Quantifying the Role of Prerequisites in University Success

Najat Messaoudi*
Faculty of Sciences A ǹ Chock, University Hassan II of Casablanca, Morocco
Email: najatm2013@gmail.com
ORCID iD: https://orcid.org/0000-0002-2554-2020
*Corresponding Author

Ghizlane Moukhliss
High School of Technology, University Hassan II of Casablanca, Morocco
Email : ghizlane.moukhliss@gmail.com
ORCID iD: https://orcid.org/0000-0003-3699-2008

Jaafar K. Naciri
Faculty of Sciences A ǹ Chock, University Hassan II of Casablanca, Morocco
Email: naciriuh2c@gmail.com
ORCID iD: https://orcid.org/0000-0002- 4903-1604

Bahloul Bensassi
Faculty of Sciences A ǹ Chock, University Hassan II of Casablanca, Morocco
Email: bahloul_bensassi@yahoo.fr
ORCID iD: https://orcid.org/0000-0003-4129-4247

Received: 21 September, 2022; Revised: 23 October, 2022; Accepted: 20 November, 2022; Published: 08 December, 2022

Abstract: The use of machine learning algorithms for higher education performance assessment is an emerging area of research and several works have focused on student performance and related problems. The preliminary goal of this work is to determine and quantify the role of prerequisites in academic success by using machine learning algorithms with the Weka environment. The main objective is the development of a tool based on machine learning algorithms for the prediction of future results for a training program based solely on the previous academic profiles of the students. The interest is to link whether success in previous courses is associated with success in subsequent target courses. This will help to improve the planning of course sequences in a training program on the one hand and the overall academic students' success on the other. The proposed methodology is applied for the analysis of the role of the prerequisites influencing courses success of a training course in Mathematical and Computer Sciences in a Moroccan university. For this purpose, we use several classification algorithms such as Random Forest, J48, and Multilayer Perceptron.

Preliminary results show that the correlation between the prerequisite reliability rates of the courses studied and the accuracy with which the learning algorithms predict the success outcomes of these courses is confirmed.

Also, these results show that the best accuracy and the best Receiver Operator Characteristic ROC area are obtained by using Random Forest algorithm and have reached 86% for the accuracy and 75.6% for the ROC area.

Index Terms: Prerequisites, prediction, academic success, Machine Learning, Random Forest, J48, Multilayer Perceptron, Weka.

1. Introduction

University Courses are generally organized in a hierarchical sequence. The order in which courses are structured is critical to students' expected skills and knowledge acquisition. Identifying the prerequisite relationships between courses is a fundamental step in organizing knowledge for pedagogical purposes. In the learning process context, the simplest concepts that are necessary to understand and address more complex concepts are usually presented first.

Therefore, the identification of the prerequisite relationships between courses is an essential operation for the design of effective courses and is part of a perspective of coherence and relevance for planning training sequences [1]. The goal of this work is to develop a tool for predicting students' performance using machine learning algorithms based solely on students' past academic performance. We will seek to better understand the mechanisms by which prerequisites operate in order to establish the relationships between prerequisites and success in a course or program of study. The question to which an answer is sought can be expressed in the form " Are the prerequisites results sufficient data to predict student success?". Indeed, many factors are involved in the student's success and we can cite among others demographic factors such as age, gender, socioeconomic factors such as parents' occupation, parents' education, parents' income, academic factors such as subject marks, pedagogical methods, previous examination marks, absences and attendance, psychological factors as student interest, stress/anxiety, behavior of study/motivation and the environments factors as class type, semester duration, type of program [2, 3, 4, 5, 6, 7, 8]. However, the data sets for these factors are not easily accessible and their relationship to student success is difficult to quantify. Therefore, it becomes appropriate to examine whether considering only the results of the prerequisites is sufficient to make a reliable prediction. The underlying idea is that all the parameters concerning demographic, socio-economic, pedagogical and environmental factors are to some extent already implicitly incorporated into the results previously obtained by the students during the prerequisites courses.

For this purpose, we will study the reliability of the formal prerequisites of courses and the correlation between this reliability and the accuracy with which Machine Learning models can predict students' success in these courses. The performance of these models can inform about the relevance of the formal prerequisites in predicting the success of these courses. A prerequisite is commonly defined as any knowledge or skill that is crucial for mastering a course or program of study. This prerequisite must be acquired before entry into these studies either as part of previous learning or independently of systematic and explicit training. Mastery of these prerequisites is an important factor in the success of the courses or training considered [9].

In this context, the analysis of the role of formal course prerequisites and their reliability in the students' success in these courses is necessary for the evaluation of the training program's performance. The interest of such an approach is to study whether indeed success in previous courses is associated with success in subsequent target courses. This will help to improve the planning of course sequences in a training program on the one hand and the overall academic students' success on the other. Another objective is to provide assistance and guidance to students in their training programs.

Four main contributions of this work can be summarized as follows: develop a tool to predict the success of students using machine learning algorithms based only on the results of previous years, analyze the reliability of prerequisites results as data for course success prediction, determine the best algorithm for analyzing the reliability of prerequisites, start the implementation of an automatic mechanism that would allow the sequencing of courses in a training program to improve the success of students in these courses and consequently improve the performance of the training as a whole.

The present paper is composed of the following sections. Section 2 summarizes the main results of previous work. Section 3 describes the method for machine learning prediction adopted. Section 4 presents the results obtained by analyzing formal prerequisite reliability using machine learning models and evaluating the performance of machine learning algorithms used in predicting students' success in a given course. Section 5 discusses these results and finally Section 6 which concludes this paper.

2. Litterature Review

The importance and the role that prerequisites play has led to many studies that have examined how well students who have or do not have the necessary prerequisites succeed in the following courses [10, 11, 12, 13]. These studies were conducted using statistical analysis methods such as one-way Anova, SPSS statistics, multiple regression methodologies, and some of them even used surveys and interviews. These studies' results confirm the impact of prerequisites on new learning and affirm that mastery of these prerequisites influences success in higher-level courses. Thus, access to a course is subject to the satisfaction of a list of prerequisites that aim to prepare students for the course and guarantee success in that course. However, these prerequisites vary in importance, and sometimes the lists of prerequisite courses are incomplete or irrelevant [14].

The most recent studies to identify prior relationships between teaching sequences use machine learning models. Indeed, Machine Learning has proven to be of great value in data mining problems where large databases may contain valuable implicit regularities that can be discovered automatically. The application of machine learning approaches to educational data is a recent research trend and is emerging as an important decision support tool in education [15, 16, 17, 18]. Higher education institutions are increasingly called upon to evaluate their own and their students' performance. As a result, there is a need to collect, analyze, and interpret data to create intervention strategies to mitigate factors that may negatively affect student and institutional performance. Machine learning algorithms help educational institutions make predictions of variety of critical educational outcomes such as success [2, 19, 7], performance [20, 21, 22, 23, 24], satisfaction [25] and dropout rate [26]. The goal of predicting educational outcomes is to improve students' success and educational performance as well as facilitate students' orientation based on their previous test performance.

Thus, in order to identify the correlations between a subject and its prior subjects, Das et al. in their works [27] used the association rule mining technique which is applied to the data of computer science and engineering undergraduate students. This data consists of subject grades of eight semesters for 117 students preparing a computer science and engineering bachelor at the university of West Bengal. This study identified 61 associations between prerequisite subjects and corresponding dependent subjects, where the length of each association varied from 2 to 4. These results concern only items that have a uniform quality as 'good' in three subjects or 'bad' in the same three subjects; the combination of both qualities is not considered in this study.

Ref. [1] applied machine learning techniques to student outcomes at a private research university in the United States to study the relationship between completion of preparatory courses and grades in subsequent courses. The study used a logistic regression model to predict whether or not a student would pass a course using previous courses taken as predictors. The determination of which previous courses most strongly predict success in the courses studied is specified by the weights associated with each predictor course. The results showed that including the student's scores in all previously studied courses as possible predictors, better-predicted target course outcomes than including only formal pre and co-requisites as possible predictors. We will discuss this finding later because the results we obtained show the same trend, but not systematically.

Ref. [28] propose a deep learning approach using recurrent neural networks that aims to better represent how the chronological order of courses affects the probability of success. The proposed model also can make recommendations not by course but widely by a combination of courses. This is an ideal feature because the overall academic success for the expected term is not only dictated by the complexity of each individual course, but also by the complexity of the course combination as a whole.

In Ref. [19], a strategy to identify the precedence relation between learning resources based on the concept-based prerequisite relationships using Machine Learning methods is proposed. Different binary classifiers have been evaluated based on a dataset extracted from DBpedia and Core corpus and which are combined with new data that take into account the co-occurrence of concepts in a corpus, the distance between their categories, and the readability of the text on the associated Wikipedia page. The results showed that there are cases where this relationship could not be correctly identified and this is mainly explained by the incomplete annotation of the concepts that a resource addresses. This level of granularity requires detailed information for each course. This information is generally not available in the student databases and therefore a usable tool cannot be easily obtained.

Ref. [29] were interested in studying the extent to which the sequence of courses in a curriculum that a student must take to obtain a degree has an impact on his or her progress. To this end, they propose an analytical framework, using data mining methods, to quantify the importance of course sequences on a student's GPA. This measure is strongly correlated with graduation rates. Specifically, they analyze the orders of course enrolment sequences that best contribute to student performance and success.

Research has been reported in the literature on predicting students' academic performance using machine learning algorithms, but so far there is no clear procedure to help in designing course sequences to improve students' academic performance by taking into account the dependency relationship between courses. Subsequently, these machine learning methods will be used to investigate the possibility of identifying potential prerequisites other than those formally defined that may improve this success rate.

Thus, we can provide an automatic mechanism that can guide the design of training unit sequences as well as a tool to assist students and instructors in determining which prior courses must be mastered to pass the target courses and thus improve the performance of training programs and reduce failure rates in university studies.

3. Data and Methodology

This paragraph presents the machine learning tools used to establish success forecasts for training programs. The adopted environment is WEKA (Waikato Environment for Knowledge Analysis) which is used to evaluate the proposed classification models and to make comparisons. Weka is an environment that provides a comprehensive collection of machine learning algorithms and data preprocessing tools to researchers and practitioners. It allows users to quickly try out and compare different machine learning methods on new data sets. Its modular extensible architecture allows sophisticated data mining processes to be built up from the wide collection of base learning algorithms and tools provided [30]. Weka is an open-source software issued under the General Public License (GNU). It is used in many research works in Educational Data Mining to build prediction models of academic performance using different machine learning algorithms [31, 6, 32, 33, 21].

A. Methodology

To identify the most appropriate prerequisites of a course that are reliable and allow students to succeed in this course, this study will be done in two steps. The first step concerns the analysis of the reliability of the formal prerequisites of the chosen course. During this step, we check whether all students who meet the formal prerequisites succeed in the course. For this purpose, we compare the success rate of students who have met the formal prerequisites with the one of students who have not met these prerequisites. This first indicator will allow us to determine the level of reliability of a prerequisite and its relevance.

The second step is the use of Machine Learning algorithms to build models to predict student success in a course by introducing only the formal prerequisite courses outcomes. The performance of these models will inform on the predictive relevance of the selected course outcomes based on the formal prerequisites as attributes and hence inform on the reliability of the formal prerequisites in predicting student success in that course. In this sense, the performance of these models is increasing according to the essentiality of the prerequisites. The formal prerequisites of a course are the prerequisites defined by the pedagogical team when designing the training program to which the course belongs. These formal prerequisites are usually explicitly given in the training description document. To achieve this, we will test different Machine Learning algorithms for different years and different courses and compare them. The considered algorithms are Decision Tree (J48), Multilayer Perceptron (MLP) and Random Forest (RF).

Let us recall that the "J48 decision tree" algorithm is a non-parametric supervised learning algorithm that is used for classification and regression problems. It generates models in the form of a tree structure, starting from root attributes and ending with leaf nodes describing the relationship among attributes and the relative importance of attributes. These relationships represent rules that can be easily understood and interpreted by users and do not require complex data preparation [20].

The multilayer perceptron (MLP) is a complement to the feedforward neural network. It consists of three types of layers: the input layer, the output layer, and the hidden layer. The input layer receives the input data to be processed, then the prediction and classification is performed by the output layer. An arbitrary number of hidden layers, placed between the input layer and the output layer, provide the true computational engine of the MLP. MLPs are designed to model complex relationships between inputs and outputs and very often give very good results [34].

Random forest RF is a Supervised Machine Learning Algorithm that is used in Classification and Regression problems. It consists of a large number of individual decision trees that operate as an ensemble and it takes their majority vote for classification and average in case of regression. It performs better results for classification problems [35].

These classification algorithms are selected because they are often used in data mining research [16] to predict students' academic performance [2, 19, 20, 21, 22, 26].

Thus, the proposed methodology allows, by using machine learning algorithms, to make predictions of student success once prerequisites are defined. These predictions can be validated by analyzing the performance metrics of these algorithms. Thus, predictions based on real prerequisites can be made while testing their validity.

Let us note moreover that other prerequisites can be tested by the proposed methodology which allows to confront the relevance of the real prerequisites with other sets of prerequisites.

B. Performance of Machine Learning algorithms

To carry out this study, we evaluate the performance of the algorithms used in order to analyze the importance of the prerequisites of the courses studied which are considered as predictors of these algorithms and also to perform a comparison between the three chosen algorithms.

In general, to evaluate the performance of machine learning algorithms, the confusion matrix based on the following four measures is used:

- True Positive (TP): number of students who passed the course classified correctly as "passed"
- False Positive (FP): number of students who passed the course classified incorrectly as "not passed"
- True Negative (TN): number of students who did not pass the course classified correctly as "not passed"
- False Negative (FN): number of students who did not pass the course classified incorrectly as "passed"

The performance of the algorithms is measured by using several metrics calculated from this confusion matrix. The most frequently used metrics are listed below [2]:

- Accuracy: the number of all correct predictions made by the algorithm overall type of predictions made

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

- Recall (or Sensitivity/TP rate): the proportion of students who passed the course that was classified correctly as "passed" for all students who passed the course

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

- Precision: the proportion of students who passed the course that was classified correctly as "passed" for all students predicted by the algorithm as passed the course

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

- Specificity (or TN rate): the proportion of the students who did not pass the course that was classified correctly as "Not passed" for all students who did not pass the course

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

- F1- Score: is the harmonic mean of precision and sensitivity

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{5}$$

- AUC - ROC curve: the Receiver Operator Characteristic (ROC) curve is a probability curve and is plotted at TPR (True Positive Rate) vs. FPR (False Positive Rate) where the TPR is on the Y axis and the FPR is on the X axis. The Area Under the Curve (AUC) represents the degree of separability: If is closer to the 1, it means the model has a high-class separation capacity and if is closer to the 0, it means the model has no class separation capacity.

In this study, we will focus on two performance metrics, namely the "Accuracy" which allows knowing the proportion of correct predictions compared to all predictions, and the "AUC-ROC" curve which is one of the most important evaluation metrics for checking any classification model's performance. It tells how much the model is capable of distinguishing between the students who passed the course and who didn't. The higher the AUC is, the better the model is at predicting students who passed the course as passed and the one who did not pass the course as not passed.

C. Data Description and Pre-Processing

Data preprocessing is an essential step to use machine learning models. Data preprocessing deals with data preparation and transformation of the dataset and seeks at the same time to make knowledge discovery more efficient. Preprocessing includes several techniques like cleaning, integration, transformation, and reduction [36].

Thus, to eliminate noise and outliers, pre-processing is applied that includes data cleaning, data transformation and selection of appropriate features. The data are usually received in a spreadsheet format and need to be converted into a WEKA compatible format. Weka can work with comma-separated values (CSV) or attribute-relation format files (arff), so the spreadsheets are converted into CSV files.

For the preparation of the dataset for this study, a group of data from University Hassan II of Casablanca (Morocco) database for the organization and management of courses and students (Apogée) is used. The collection of the data for this study is done following the ethics of exploitation of institutional data and the respect of the confidentiality of the student's personal data. The data generated in this study will not be made available to the public. The Faculty of Sciences Aïn Chock grants permission to use the data and publish the results, but not to publish the data. The used DATA concerns courses taught at the Faculty of Sciences Aïn Chock of the University Hassan II of Casablanca in Morocco for the graduation of the license in Applied Mathematical Sciences.

Let us recall that access to studies in Moroccan universities is conditioned by obtaining the baccalaureate. The baccalaureate is the diploma that completes the 12 years of primary and secondary education. The teaching provided by the universities in Morocco is organized in training cycles according to the LMD system (License-Master-Doctorate). The first cycle of training is the License (L) which is a diploma which corresponds to the first 3 years of university studies and is organized in 6 semesters. The Master's degree (M) is awarded at the end of a two-year program (4 semesters) after the "License" cycle. Finally, the post-master Doctoral cycle (D) of usually three years duration.

The data used in this study is divided into two files. One for the 2020/2021 academic year and another for the 2018/2019 academic year. It contains results of the students enrolled in semester 3 of the Applied Mathematical Sciences program. The year 2019/2020 impacted by covid19 and the implementation of distance learning is not considered in this study as it needs a specific ulterior study for which the present results and developed tools will be useful.

Each of these files includes 20 attributes composed of 6 attributes related to the courses taken during semester 3 by the students in the academic year concerned, and 14 attributes related to the courses taken before semester 3 by these students (Semester 1 and 2). The attributes that will be used will be selected based on their relevance to our goals and objectives. Table 1 shows the list of courses taught in semester 3 and taken by students enrolled in the Applied Mathematical Sciences program, as well as the courses taken in semester 1 and 2.

Table 1. List of courses by semester in the Applied Mathematical Sciences program

Semester	Courses
Semester 3 (S3)	Analysis 4
	Analysis 5
	Algebra 4
	Probability - statistics
	Electricity 3
	Computer Science 3: Algorithms and programming)
Semester 2 (S2)	Analysis 2
	Analysis 3
	Algebra 3
	Electricity
	Optics
	Computer Science 2
	Language and terminology 2
Semester 1 (S1)	Analysis 1
	Algebra 1
	Algebra 2
	Mechanics
	Thermodynamics
	Computer Science 1
	Language and terminology 1

The dataset contains the students enrolled in semester 3 of the Applied Mathematical Sciences program, and their results in each course taken during the three semesters S1, S2, and S3. The student numbers were deleted to protect the student's identity. The results of a course are passed (P) or not passed (NP). In this study, we will focus on the analysis of the prerequisites of the courses “Computer Science 3: Algorithms and programming” and “Analysis 5” which are taught in semester 3 of the Applied Mathematical Sciences program.

4. Results

A. Study of the Prerequisites for the Course "Computer Science 3" for the 2020/2021 Academic Year

The course in consideration is "Computer Science 3: Algorithms and Programming" which is a major course taught in semester 3 of the Applied Mathematical Sciences program. For this course, we have complete verifiable data defined by a single prerequisite, which corresponds to the simplest situation. We will give later examples with modules having several prerequisites.

The number of students enrolled in the course "Computer Science 3" for the year 2020/2021 is 82 students of which 60 students have passed this course. The success rate of this course is then 73.2%. According to the description of this program, this course has only one formal prerequisite which is the course "Computer Science 2: Algorithms 1" taught in semester 2 in this same program.

Step 1: Analyze of the Reliability of the Formal Prerequisite

A prerequisite is considered reliable if all the students satisfying the required prerequisite pass the target course. Thus, we can define the reliability rate as the ratio of the number of students who passed the target course and met the required prerequisite to the total number of students satisfying the required prerequisite. For the course studied and as shown in Figure 1 below, among the 82 students enrolled in "Computer Science 3", 74 students have the required prerequisite course and 8 students do not have the required prerequisite.

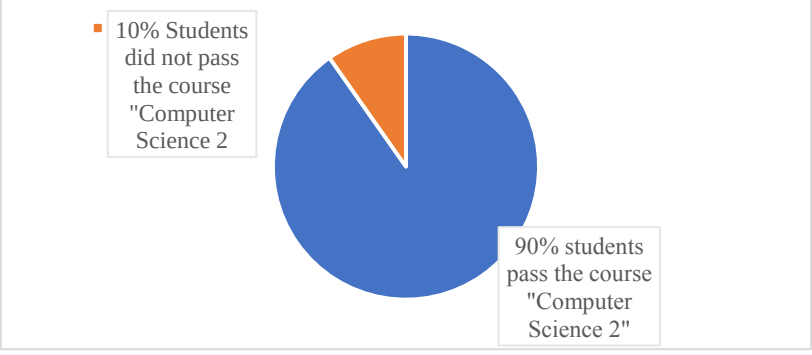

Fig. 1. Composition of students enrolled in “Computer sciences3” in 2020/2021

Of the 74 students satisfying the required prerequisite, 58 students passed the "Computer Science 3" course and 16 students did not pass the course. Thus, the reliability rate of this prerequisite is 78.38% as shown in Table 2. We also note in this table that among the 8 students who did not pass the required prerequisite course, two students did pass the course “Computer Science 3”.

Table 2. Distribution of students enrolled in the course “Computer Science 3” in 2020/2021 and who have the required prerequisite

	Number of students enrolled in “Computer Science 3” and met the required prerequisite	Number of students enrolled in “Computer Science 3” and did not meet the required prerequisite
Number of students who passed the course “Computer Science 3”	58	2
Number of students who did not pass the course “Computer Science 3”	16	6
Total	74	8
Reliability rate	**78.38%**	

The reliability rate calculated above, which is 79%, allows us to conclude that 79% of the students who meet the prerequisite of the course studied were able to pass this course, whereas 21% of these students did not pass this course, while two students of the 10 students who did not meet the prerequisite were able to pass the “Computer Science 3” course.

We can then conclude that the officially defined prerequisite is necessary and has an impact on the success in this course but does not allow a good success rate (not exceeding 73%) for this course.

Step 2: Analyze the Reliability of the Formal Prerequisite by Using Machine Learning Algorithms

To analyze the reliability of a course prerequisite for success, we use machine learning algorithms to build models which enables to predict student success in this course based on their results in the prerequisite course and analyze the performance of these models in predicting success in this course. The performance of these models is expected to inform on the predictive relevance of the results of the chosen course based on the formal prerequisites as predictors.

To verify the reliability of the formal prerequisite course studied, we analyze the performance of the machine learning algorithms in predicting the students’ success in the "Computer Science 3" course based only on the results of the prerequisite required course "Computer Science 2". This performance of the algorithms informs on the predictability of the "Computer Science 3" outcomes from the "Computer Science 2" outcomes. This analysis is performed using Random Forest (RF), J48, and Multilayer Perceptron (MLP).

Thus, we load in Weka a file that contains only the results obtained by the students in the course studied "Computer Science 3" and the results they had obtained in the formal prerequisite course which is "Computer Science 2".

To build the models and as our dataset contains a limited amount of data to divide it into training and test data, we use a statistical technique called Tenfold cross-validation. With this method, the data are randomly split into 10 subsets. Each time, nine of them are used as training data and the tenth as test data. This procedure is repeated 10 times on the different subsets. Finally, the 10 success estimates are averaged to obtain an overall success estimate. Figure 2 shows as an example the output of a classifier given by Weka.

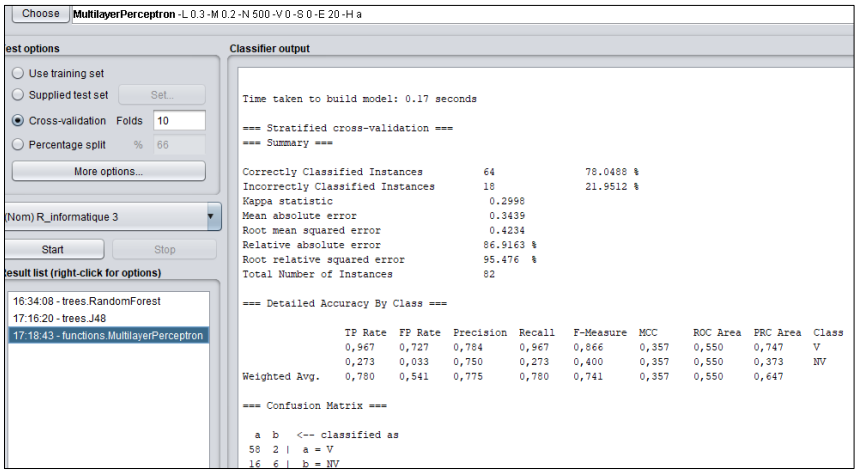

Fig. 2. MLP output in Weka

The results of the performance metrics obtained by the three classifiers on Weka are summarized in Table 3 below.

Table 3. Results of the performance metrics of the algorithms used (Prerequisite of "Computer Science 3" for 2020/2021)

Algorithm	Accuracy	Precision	Recall	F1 - score	ROC Area
RF	78%	77.5%	78%	74.1%	52.3%
J48	78%	77.5%	78%	74.1%	52.1%
MLP	78%	77.5%	78%	74.1%	55%

Figure 3 presents the design of the knowledge flow for the comparison of the ROC curves of the three algorithms used.

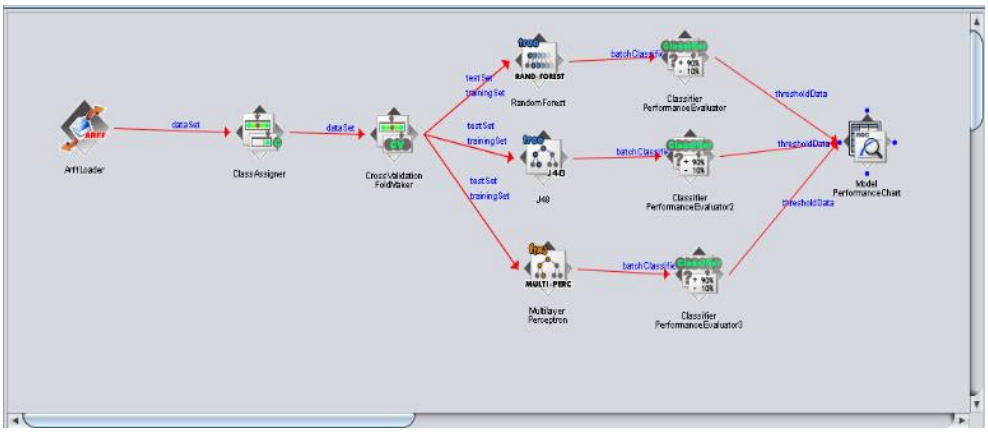

Fig. 3. Model performance chart for comparison of ROC Curves

The result of this model is shown in Figure 4 below.

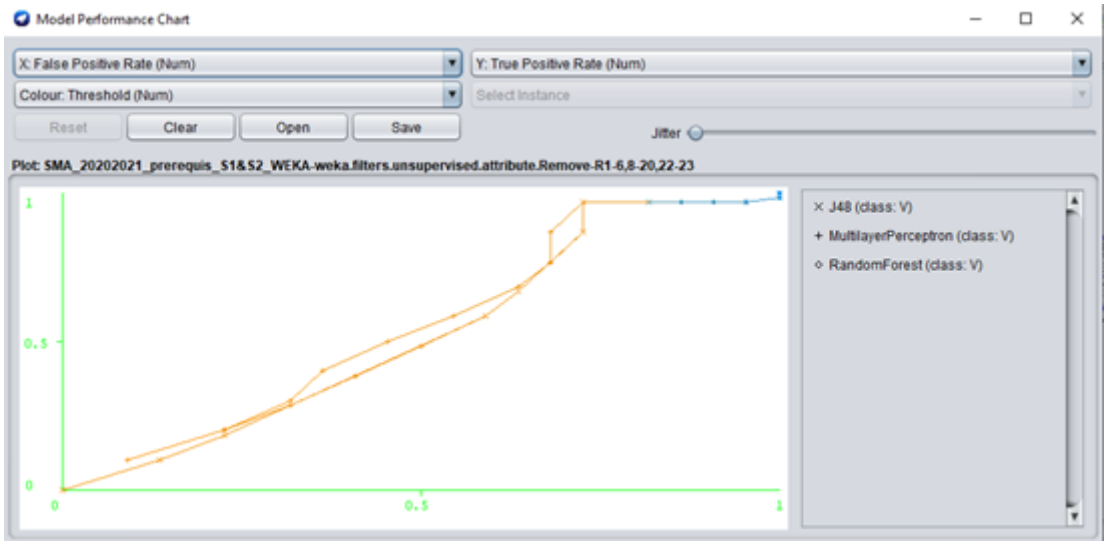

Fig. 4. Comparison of ROC Curves of the algorithms used

We notice that the accuracy of the three algorithms Random Forest, J48, and Multilayer Perceptron is the same that is 78.78% but the best ROC Area is given by Multilayer Perceptron (55%). We can deduce, thus, that there is a correlation between the accuracy of the prediction algorithms used (78.78%) and the reliability rate which is 78.38%. We can therefore assume that the prerequisite thus determined (Computer Science 2) has an impact on the success of the course studied (Computer Science 3) and allows an adequate accuracy of prediction but not enough to build a relevant prediction model since the ROC Area is only 55%.

To test if this correlation between the accuracy of the prediction algorithms and the reliability rate, we propose to test this approach on other academic years and courses.

B. Analysis of the correlation between algorithms' accuracy and the reliability of prerequisites from other courses and academic years

In this section, steps 1 and 2 above are applied for the same course for the 2018/2019 year and for one other course "Analysis 5" for the 2020/2021 and 2018/2019 years.

The course "Analysis 5" is a course that is taught in semester 3 of the Applied Mathematical Sciences program. According to the description of this program, this course has three formal prerequisites which are: the course "Analysis 1" taught in semester 1 in this same program and the courses "Analysis 2" and "Analysis 3" taught in semester 2 in this same program. The choice of the course "Analysis 5" is due to the fact that it requires three courses as prerequisites instead of one as is the case with the course initially studied "Computer Science 3". This also allows us to analyze the interest in having multiple prerequisites for the success of a course.

Table 4 below shows the results of step 1 of the approach conducted on the different courses and different academic years.

Table 4. Success rate and Reliability rate for the different courses and different academic years

	Course «Computer Science 3 »		Course «Analysis 5 »	
	2020/2021	2018/2019	2020/2021	2018/2019
Number of students	82	101	82	101
Success rate	73.2%	58.4%	73%	74.25%
Formal prerequisites	«Computer Science 2 »		«Analysis 1 », «Analysis 2 »and «Analysis3 »	
Number of students who have met the formal prerequisites	74	89	65	75
Reliability rate of the formal prerequisites	78.38%	64%	89.2%	86.67%

We report in table 5 below the results of Step 2 of the simulations conducted for each course (CS3: the course “Computer Science 3”, A5: the course “Analysis 5”) and each year academic.

Table 5. Results of the simulations of the different courses and different years

		Accuracy		Precision		Recall		F1-Score		ROC Area	
		2020/21	2018/19	2020/21	2018/19	2020/21	2018/19	2020/21	2018/19	2020/21	2018/19
CS3	RF	78%	66.3%	77.5%	72.1%	78%	66.3%	74.1%	66.4%	52.3%	55.7%
	J48	78%	66.3%	77.5%	72.1%	78%	66.3%	74.1%	60.4%	52.1%	55.3%
	MLP	78%	66.3%	77.5%	72.1%	78%	66.3%	74.1%	60.4%	55%	56.9%
A5	RF	86,58%	78.2%	86.6%	76.3%	86.6%	78.2%	85.7%	75.8%	75.6%	63.8%
	J48	80.48%	74.2%	80.4%	70.9%	80.5%	74.3%	77.7%	71.3%	63.6%	55.3%
	MLP	85.36%	74.2%	85.4%	71.4%	85.4%	74.3%	84.2%	72%	76.5%	65.7%

Comparing the results of the performance of the models used, we notice that for the course “Computer Science 3”, the accuracy of the three models is practically the same since we introduced only one attribute which is the prerequisite of this course to predict the results of the studied course. On the other hand, we notice a slight difference between the ROC areas of the three algorithms, and the best performance is given by Multilayer Perceptron for the two academic years studied. The ROC area obtained remains poor and does not allow a good prediction of the results of the course studied.

For the course “Analysis 5”, since we have three courses as prerequisites and which constituted the prediction attributes for these models, the accuracy of the three models is not the same, and the best accuracy is given by Random Forest which is 86.5%. On the other hand, the best ROC area is given by Multilayer Perceptron (76.5%), which is considered the best model for this study and the two academic years studied. Despite this, this performance remains average and can be improved by introducing other attributes or by acting on the parameters of the algorithm.

5. Discussion

According to these results, we notice that for the course "Computer Science 3" which has only one prerequisite, the reliability rate of this prerequisite and the accuracy of the models were moderately important (between 66.3% and 78%) and therefore this prerequisite is considered not sufficient to ensure success in this course. On the other hand, for the course "Analysis 5" which has three prerequisites, the reliability rate of these prerequisites and the accuracy of the models are very important (between 86% and 92% for the reliability rate and between 74% and 86.5% for the accuracy) which allows us to conclude that these prerequisites have a role in the success in the target course but according to the Roc area, the success rate can be improved by introducing other attributes as prerequisite courses. The results also show that the more prerequisites a course has, the higher the reliability of the prerequisites. Thus, the machine learning models can be used to identify the relevant prerequisites that can give better results.

In the four simulations conducted, the correlation between the prerequisite reliability rates of the courses studied and the accuracy with which the learning algorithms predict the success outcomes of these courses is confirmed.

This allows us to conclude that the accuracy of the prediction models increases according to reliability rate of the prerequisites. Then, to improve this accuracy, it would be necessary to identify further prerequisites that will improve this reliability rate and therefore the success of the course studied.

We also notice from these simulations that if we raise the number of prerequisites for a course, the reliability rate and learning accuracy of the algorithms rise as well.

6. Conclusion

In this study, tool for predicting student performance using machine learning algorithms based only on students' past academic performance is developed. The interest was shown in analyzing the role of formal course prerequisites that are determined during course design in course success by studying the reliability rates of these prerequisites.

Using Machine Learning algorithms, a correlation is established between the prediction accuracy of these algorithms and the reliability rates of the prerequisites which will initiate future studies that will be interested in using these algorithms to determine the most appropriate course prerequisites that would improve student success in these courses. The interest of this approach is the implementation of an automatic mechanism that would allow the sequencing of courses in a training program to improve the success of students in these courses and consequently improve the performance of the training as a whole.

This mechanism would also allow for a better definition of the prerequisites for access to more selective or more specialized training and would also allow for a better orientation of students in their training pathway according to their training achievements.

References

[1] G. M. Davis, A. A. Abu Hashem, D. Lang et M. L. Stevens, "Identifying Preparatory Courses that Predict Student Success in Quantitative Subjects," chez Proceedings of the Seventh ACM Conference on Learning, 2020.

[2] E. Alyahyan et D. Düştegör, "Predicting academic success in higher education: literature review and best practices, " International Journal of Educational Technology in Higher Education, pp. 1-21, 2020.

[3] S. Lakhal et H. Khechine, "Technological factors of students' persistence in online courses in higher education: The moderating role of gender, age and prior online course experience, " Education and Information Technologies, vol. 26, n °%13, pp. 3347-3373, 2021.

[4] M. Martins, V. Miguéis, D. Fonseca et A. Alves, «A Data Mining Approach for Predicting Academic Success – A Case Study », » Advances in Intelligent Systems and Computing, vol. 918, pp. 45-56, 2019.

[5] A. Mueen, B. Zafar et U. Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques, " International Journal of Modern Education and Computer Science, vol. 11, pp. 36-42, 2016.

[6] G. Matafeni et R. Ajoodha, "Using Big Data Analytics to Predict Learner Attrition based on First Year Marks at a South African University, " Advances in Science, Technology and Engineering Systems Journal, vol. 5, pp. 920-926, 2020.

[7] J. G. Perez et E. S. Perez, "Predicting Student Program Completion Using Naïve Bayes Classification Algorithm, " International Journal of Modern Education and Computer Science, vol. 3, pp. 57-67, 2021.

[8] L. Mansangu, A. Jadhav et R. Ajoodha, "Predicting Student Academic Performance Using Data Mining Techniques, " Advances in Science, Technology and Engineering Systems Journal, vol. 6, n °%11, pp. 153-163, 2021.

[9] M.-C. Hublet, S. Lontie, M. Arras et S. Remacle, "Test diagnostique à l'entrée de l'enseignement supérieur: Validation et usages, " Revue internationale de pédagogie de l'enseignement supérieur, vol. 37, n °%13, 2021.

[10] S. Krause-Levy, S. Valstar, L. Porter et W. G. Griswold, "Exploring the Link Between Prerequisites and Performance in Advanced Data Structures, " chez SIGCSE '20: Proceedings of the 51st ACM Technical Symposium on Computer Science Education, Portland, 2020.

[11] B. K. Sato, A. K. Lee, U. Alam, J. V. Dang, S. J. Dacanay, P. Morgado, G. Pirino, J. E. Brunner, V. W. Chan et J. H. Sandhothz, "What's in a Prerequisite? A Mixed-Methods Approach to Identifying the Impact of a Prerequisite Course, " Life Sciences Education, vol. 16, n °%11, pp. 1-9, 2017.

[12] E. S. Krol, R. T. Dobson et K. Adesina, "Association Between Prerequisites and Academic Success at a Canadian University's Pharmacy Program, " American Journal of Pharmaceutical Education, vol. 83, n °%11, pp. 83-92, 2019.

[13] F. Islam, S. Khan, I. Wilson et R. Gooch, "The value of prerequisite courses for statistics, " The Journal of Business Inquiry, vol. 7, n °%11, pp. 61-67, 2008.

[14] W. Jiang, Z. A. Pardos et Q. Wei, "Goal-based Course Recommendation, " chez Proceedings of the 9th International Conference on Learning Analytics & Knowledge, 2019.

[15] K. Shade O., N. Goga, O. Awodele et S. Okolie, "Machine Learning Approach to Determining the Influence of Family Background Factors on Students' Academic Performance, " International Journal of Information Sciences and Application, vol. 4, n °%11, pp. 57-76, 2012.

[16] C. Romero et S. Ventura, "Educational data mining and learning analytics: An updated survey, " WIREs Data Mining Knowledge Discovery, pp. 1-21, 2020.

[17] P. Balaji, S. Alelyani, A. Qahmash et M. Mohana, "Contributions of Machine Learning Models towards Student Academic Performance Prediction: A Systematic Review, " applied, vol. 11, n °%121, 2021.

[18] H. Almarebeh, "Analysis of Students' performance by Using Different Data Mining Xlassifiers, " International Journal of Modern Education and Computer Science, vol. 8, pp. 9-15, 2017.

[19] R. Manrique, J. Sosa, O. Marino, B. P. Nunes et N. Cardozo, "Investigating learning resources precedence relations via concept prerequisite learning, " chez International Conference on Web Intelligence (WI), 2018.

[20] K. P. Ajay et P. Saurabh, "Data Mining Techniques in EDM for Predicting the Performance of Students, " International Journal of Computer and Information Technology (ISSN: 2279 – 0764), pp. 1110-1116, 2013.

[21] K. Shade O., N. Goga, O. Awodele et S. Okolie, "Optimal Algorithm for Predicting Sstudents' Academic Performance, " International Journal of Computers & Technology, pp. 63-75, 2013.

[22] H. Altabrawee, O. Ali et S. Qaisar, "Predicting Students' Performance Using Machine Learning Techniques, " Journal of University of Babylon for Pure and Applied Sciences, vol. 27, pp. 194-205, 2019.
[23] A. R. Iyanda, O. D. Ninan, A. O. Ajayi et O. G. Anyabolu, "Predicting Student Academic Performance in Computer Science Courses: A Comparison of Neural Network Models, " International Journal of Modern Education and Computer Science, vol. 6, pp. 1-9, 2018.
[24] M. T. Sathe et A. C. Adamuthe, "Comparative Study of Supervised Algorithms for Prediction of Students' Performance, " International Journal of Modern Education and Computer Science, vol. 1, pp. 1-21, 2021.
[25] E. Alqurashi, "Predicting student satisfaction and perceived learning within online learning environments, " Distance Education, vol. 40, n °%11, pp. 133-148, 2019.
[26] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka et P. C. Nshimyumukiza, "Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization, " Computers and Education: Artificial Intelligence, vol. 3, 2022.
[27] C. Das, S. Bose, A. Chanda, S. Singh, S. Das et K. Ghosh, "Impact of Prerequisite Subjects on Academic Performance Using Association Rule Mining, " chez Advances in Intelligent Systems and Computing, 2021.
[28] N. Araque, V. a. G. Rojas et M. V. Vitali, "UniNet: Next Term Course Recommendation using Deep Learning," chez International Conference on Advanced Computer Science and Information Systems, 2020.
[29] A. Slim, G. L. Heileman, W. Al-Doroubi et C. T. Abdallah, "The Impact of Course Enrollment Sequences on Student Success, " chez International Conference on Advanced Information Networking and Applications, 2016.
[30] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann et I. H. Witten, "The WEKA Data Mining Software: An Update, " SIGKDD Explorations, vol. 11, n °%11, pp. 10-18, 2009.
[31] A. Siddique, A. Jan, F. Majeed, A. I. Qahmash, N. N. Quadri et M. O. Abdul Wahab, "Predicting Academic Performance Using an Efficient Model Based on Fusion of Classifiers, " Applied Sciences, vol. 11, pp. 1-19, 2021.
[32] A. Fazal, I. Fakhund, A. Rahmani, R. Azhar et A. Masood Khattak, "A Predictive Model for Predicting Students Academic Performance, " chez 10th International Conference on Information, Intelligence, Systems and Applications (IISA), 2019.
[33] Sana, I. F. Siddiqui et Q. A. Arain, "Analyzing Students' Academic Performance through Educational Data Mining, " 3C Technologia. Glosas de innovacion aplicadas a la pyme, pp. 402-421, 2019.
[34] S. Abirami et P. Chitra, "Energy-efficient edge based real-time healthcare support system, " Advances in Computers, vol. 117, n °%11, pp. 339-368, 2020.
[35] M. Utari, B. Warsito et R. Kusumaningrum , «Implementation of Data Mining for Drop-Out Prediction using Random Forest Method, » chez 8th International Conference on Information and Communication Technology (ICoICT), 2020.
[36] W. S. Bhaya, "Review of Data Preprocessing Techniques in Data Mining, " Journal of Engineering and Applied Sciences, vol. 12, n °%116, pp. 4102-4107, 2017.

Authors' Profiles

Najat Messaoudi received her Ph.D. degree in automatic and industrial computer from Hassan II University, Casablanca, Morocco in 2018.
She is a Professor at the Faculty of sciences, Hassan II University, Casablanca, Morocco. Her research interests include modelling and simulation, modeling logistic systems, industrial computing, Machine Learning and performance of higher education.

Ghizlane Moukhliss is currently working as digital library manager of Hassan II University of Casablanca. She received her Ph.D. in Computer Science at the High School of Technology, Hassan II University of Casablanca. Her research interests include software engineering, Machine Learning, Network Security and Information Systems.

Jaafar K. Naciri received the Ph.D. degree in mechanical sciences from University Paul Sabatier, Toulouse, France. He is professor emeritus at the faculty of sciences, Hassan II University, Casablanca, Morocco. His research interest includes aeronautical engineering, mechanical engineering, quality assurance in higher education, modelling and simulation

Bahloul Bensassi is a professor in physics department at faculty of sciences, Hassan II University, Casablanca, Morocco. He is responsible of logistics engineering master degree and the Electronics, Electrical, Automatic and Industrial computing master degree. His main research interests are modeling logistic systems, electronics, automatic and industrial computing.

How to cite this paper: Najat Messaoudi, Ghizlane Moukhliss, Jaafar K. Naciri, Bahloul Bensassi, "Machine Learning Algorithms for Quantifying the Role of Prerequisites in University Success", International Journal of Modern Education and Computer Science(IJMECS), Vol.14, No.6, pp. 1-12, 2022. DOI:10.5815/ijmecs.2022.06.01