



Developing and Comparing Data Mining Algorithms That Work Best for Predicting Student Performance

Hoda Ahmed Abdelhafez, College of Computer & Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia & Faculty of Computers and Informatics, Suez Canal University, Ismailia, Egypt

 <https://orcid.org/0000-0002-0143-173X>

Hela Elmannai, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia & Laboratoire de Teledetection et Remote Sensing LTSIRS, INSAT, Tunisia *

 <https://orcid.org/0000-0003-2571-1848>

ABSTRACT

Learning data analytics improves the learning field in higher education using educational data for extracting useful patterns and making better decisions. Identifying potential at-risk students may help instructors and academic guidance to improve the students' performance and the achievement of learning outcomes. The aim of this research study is to predict at early phases the student's failure in a particular course using the standards-based grading. Several machine learning techniques were implemented to predict the student failure based on support vector machine, multilayer perceptron, naïve bayes, and decision tree. The results on each technique show the ability of machine learning algorithms to predict the student failure accurately after the third week and before the course dropout week. This study provides a strong knowledge for student performance in all courses. It also provides faculty members the ability to help at-risk students by focusing on them and providing necessary support to improve their performance and avoid failure.

KEYWORDS

Decision Tree, Learning Analytics, Machine Learning, Multilayer Perceptron, Naive Bayes, Predictive Analytics, SVM

INTRODUCTION

The recent Researches predict the student difficulties (Maheshwari, et al., 2020; Chui, et al., 2020), the student achievements (Lerche and Kiel, 2018) and academic performance (Son and Fujita, 2019; Tomasevic, et al., 2020; Waheed et, al., 2020), which aim to empower learning practices design (Galloway, et al., 2020; Goodyear, 2020) and learning environment effectiveness (Quinn and Gray, 2019; Dyrbye, et al., 2020; Asarta and Schmidt, 2020). This emerging landscape have emerged especially in higher education (Viberg, et al., 2018) to improve and support the learning process and provide feedbacks for educators and education decision makers. The main addressed problem is

DOI: 10.4018/IJICTE.293235

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

whether the data analytics sciences can be applied to the learning field and can be deployed widely into educational institutions.

The instructors and decision makers are increasingly interested in modern learning analytics techniques, which provide accurate predictions for student achievements. These learning analytics techniques are based on educational data and data mining to provide a comprehensive analysis and optimization of the learning experience.

In the literature, learning analytics refer to a mean to help educators and learning environments (Christensen, et al., 2018; Christensen et al., 2019) using collection and analysis of learners' data for the main goal of understanding and optimizing the learning process and outcomes (Viberg, et al., 2018). Further goals support the learning design and regulation of learning by offering a dashboard that provides needed feedback for learner and instructors (Sedrakyan, et al., 2020). The aim of advanced learning analytics research studies is to offer a personalized learning approaches (Tsai, et al., 2020) and analyze factors influencing learners' achievements (Moreno-Marcos, et al., 2020).

Especially, developing an educational warning system (Zhang, et al., 2020; Du, et al., 2020) to predict academic performance, dropping-out, failure and advanced education indicators has great interest. Many researchers have addressed the students retention and students at-risk. These two parameters are narrowly related especially in the first year in higher education. Universities reports demonstrated that more than 37% of students leave the educational institution in the first or second year for four-year institutions (ACT Institutional Data File, 2018). This highlighted the importance of prediction student at risk in the enrolled courses during the first year to establish suitable action plan learning practices review and updates.

Few works for student performance predictions were based only on standards-based gradings. Major studies include extra factors including social, demographic, and behavioral factors. This additional information require data collect, learning environment feedbacks and student's information related to age, sex, social conditions, etc. However, standards-based grading is directly mapped to the course leaning objectives and measure the student ability and weakness regarding the course objectives (Selbach-Allen, et al., 2020). Course objectives also reflected the program or academic department objectives and directly translate the intended skills of the program exit students. Therefore, it is crucial that educational early warning systems will be based on standards-based gradings and basic assessments. Matching learning activities (assessments tools) to the course learning objectives is a trustable to build the failure prediction model. Another advantage in using the standards-based learning data is the availability of the data. A few research studies addressed the problem and provided remedial solutions to introduce in-time feedback to instructors, students, and educators (Albreiki, et. al., 2021).

The remainder of this paper is organized as follows. In research purpose section, we will present the purpose of the research study and addressed the research questions. The literature review section discusses related works and background. Research methodology section provides information about different machine learning techniques used for the student failure prediction. The analysis dataset and discussing the results sections are presented respectively and finally, the conclusion section, which highlights the study achievements and limitations as well as the suggestion for future prediction models improvements.

RESEARCH PURPOSE

Identifying student at risk in early stages during the semester could allow the instructors and program committee to provide needed curative action such as reviewing assessment tools et learning practices. This curative action should be able to focus on student at-risk and provide needed support. Furthermore, the curative approach should be able to allow a proactive approach in the next semester based on course review. These iterations should ameliorate the student retention rate by reducing student failure and ameliorating learning and assessment practices. Proactive approach is driven by

each course learning experience to improve the learning outcomes and finding appropriate practices to reduce student failure, dropout and to enhance employability.

Providing trustable prediction method based on student grades is therefore necessary to achieve the institution goals such as increasing student performance and retention (Wiley, et al., 2020). The aim of this work is to predict the student at-risk in three stages. The first and second stages are early stages to allow needed curative approaches for both the instructor and the learners. The third stage is to avoid failure by dropping out.

Our challenge is to use relatively limited data variables based on in-semester grades to predict students at-risk. These data have the advantages to be available and easy to collect without any need of specific data.

LITERATURE REVIEW

Related work

Various research studies have been focused on predicting students at-risk using different machine learning algorithms. Polyzou and Karypis (2019) applied decision tree and linear support vector machine as base classifiers, and random forest (RF) and gradient boosting as ensemble classifiers in their research. They used the data that represented the students' grades from the University of Minnesota, an undergraduate public institution. They predicted students' performance using binary classification, based on data provided before the start of the semester. The factors that can lead to student's poor performance were derived from the original data different features. These features are (1) the student-specific that are independent from course, (2) course-specific features that are independent from student and (3) course-specific features dependent on student. The features related to course's difficulty and popularity as well as courses specific features, manage to achieve the same accuracy as when using all the features. This indicated that the reasons of a student drop a course are related to the course itself, rather than to the students themselves. Therefore, the student-course specific features have better performance, while the student-specific features have worst performance (Selbach-Allen, et al., 2020).

Wakelam et al., (2020) applied and compared three machine learning techniques, Decision Tree, K-Nearest Neighbors and Random Forest to analyze and predict student performance, applied at appropriate points during module delivery. These machine learning methods were selected to analyze the student performance. The methods are used to create a model for predicting target variable which is the students' assessment marks. The data collected from the final year of undergraduate students in computer Science university, and the student data included lecture or tutorial attendance, accessing virtual learning environment and intermediate assessments for 15 weeks. The Experimental results showed some potential for analyzing and predicting student assessment marks on courses using small student cohorts, with very limited student datasets. The analyses demonstrated three opportunities for module leadership to identify at risk students and taking into consideration appropriate timely interventions.

Another study accumulated year-on-year data to examine the effects of using multi-year data on prediction accuracy. Meier et al., proposed an algorithm to predict the final grade of each student in a class based on multi-year data. Their experiments were based on a dataset from an undergraduate digital signal processing course taught at UCLA over the past 7 years. The course consists of 7 homework assignments, one in-class midterm exam (taking place after the third homework assignment), one course project and the final exam. The duration of this course is 10 weeks in each year. The developed algorithm predicted the final grades of students based on their scores in early performance assessment such as homework assignments or midterm exams. This algorithm was able to learn from past data, the predictions were robust even when the course was taught by different instructors. They demonstrated that in-class exams were better predictors of the overall performance of a student than the homework

assignments. They showed that for 85% of the students they could predict with 76% whether they were going to do well or poorly in the class after the fourth course week (Meier, et al., 2016).

Baarsa et al., (2017) developed a model for the early and reliable prediction of students who fail to pass the first year of Erasmus medical school. The aim was to refer these students to another program. Five consecutive cohorts of students from year 2001 to year 2005) were included in this study. These cohorts comprised 272, 332, 405, 403 and 407 students, respectively. Prediction for failure in the first-year medical curriculum were made at 0, 4, 6, 8, 10 and 12 months. Predictive variables included pre-admission variables, age, gender, pre-university education GPA, the way students were selected, and post-admission variables. The model was logistic regression analyses using SPSS for predicting the students fail. The experimentations identify two categories: optimal category of students that have strong chance to pass the first year and non-optimal category. The results demonstrated the earliest moment with the highest specificity to predict student failure in the first-year curriculum was at 6 months in medical school. However, they concluded that additional factors are needed to improve the power of prediction operation.

A study conducted by Marbouti et al., (2015) built three models to identify at-risk students in a specific large first-year engineering course. The dataset included 3400 first-year engineering (FYE) students at a large midwestern university. They used the course performance data from Spring 2011 as the training data and tested the models with Spring 2012 data. The performance data were selected according to the course syllabus. The data included grades on weekly attendance, homework, and quizzes, as well as exams. They created three predictive models for week 2, week 4, and week 9. First, they implemented logistic regression on data from Spring 2011. Second, they refined the model to keep only the most significant factors and minimize the complexity of the model. Third, they tested the refined model on the Spring 2012 dataset. These models were able to identify 79% of at-risk students at week 2, 90% at week 4, and 98% at week 9. This high accuracy illustrated the value of creating course specific prediction models instead of developing generic ones.

Marbouti, et al., (2016) proposed later another study to identify student at-risk in one course using only standards-based grading for the prediction. The study variables included learning objectives. The authors used seven models (logistic regression, Support Vector machine, Decision Tree, Multi-Layer Perceptron, Naïve Bayes Classifier, K-Nearest Neighbor and Ensemble Model based on Majority vote) and a feature selection method. Results show that Naïve Bayes and Ensemble model achieved best results.

Online learning platforms also provide information that are useful in predicting outcomes, such as pass/fail and final grade prediction. In addition, a huge amount of behavioral data describing student interaction and engagement in the learning process (Costa, et al., 2017) is generated by online platforms. These features may be used as additional variables such as variables related to the participation in online discussion (Osmanbegović and Suljić, 2012), online platform usage criteria (Romero, et al., 2008). Akçapınar et al., (2019) presented a failure prediction for one course using online learning environment reports and student interaction behavior (regular logging, participation in forum, doing homework, interaction with others, ..). Preprocessing methods were applied for data transform and feature selection is carried to determine most relevant features. Seven classification algorithms were experimented: Naive Bayes, Random Forest, Support Vector Machine, Classification Tree, Neural Networks, CN2 Rules and KNN. The results using selected preprocessing techniques and selected algorithms, show that feature selection allows higher performance. Best accuracies were obtained using KNN and CN2 Rules. The early prediction of student at-risk in the third week allow to predict 74% of failure.

Keshkar et al., (2016) developed a model for predicting the students that were at-risk of attaining a negative learning outcome or at-risk of failure in the semester. They applied classification machine learning algorithms for prediction the pass or fail by predicting the final grade. They used a dataset that was a mix of online-only courses and offline only courses from Moodle. The experiments applied Logistics Model Trees for offline course as well as Decision Tree and Naive Bayes for online courses.

The results showed that using Logistic Model Tree could help in predicting the negative learning performance for the offline dataset. The same result did not occur in the case of the online courses. Therefore, their future work is to investigate an online model that incorporates some of other metadata features such as scholarship, part-time and full-time that may improve the results.

Sarra et al., (2019) identified students at-risk and more likely to dropout using Bayesian Profile Regression. The model is based on students' performance, motivation, and resilience. Feature were extracted from an online questionnaire. The model estimated a joint distribution of dropout tendency for nine profile clusters and demonstrated that similar student profiles have similar tendency. The risk of dropping out from university varies according to these cluster profiles. Students at-risk are determined by sorting the clusters according to the dropout risk. Authors concluded that the student's engagement and retention are conducting factors in the student success.

Background

Many advanced machine learning techniques have been explored for learning data analytics. In this study, we aim to investigate these techniques and compare their abilities to predict the student failure. The five commonly used approaches to acquire adequate background are presented as follows.

Decision Tree (DT) is based on partitioning the data iteratively. Each partition is based on a data feature or attributes until reaching final leaves. Inner nodes hold feature test and leaves are class labels. A branch is therefore a test result. Root-leaf branch provides the classification rules (McGonagle, et al., 2018). The partitioning is leaded by a score function that selects the feature providing the maximum gain (Horning, 2013). Decision tree can handle small and large data but present overfitting problem.

Support Vector Machines (SVM) is a supervised machine learning that was proposed by Vapnik (1995). The classification decision is based on the principle of structural risk minimization. The main idea is to represent the data in a wider feature space and divide the dataset based on a hyperplane separator determined by the support vectors. The mapping to the higher dimension space is provided by a kernel-based function. Among the kernel functions the Radial Basis Function RBF has been widely used for non-linear classification problem.

Multi-Layer perceptron (MLP) is an Artificial Neural Network (ANN). The main concept of ANN is to simulate the human brain by a set of connected neurons that processes the input data for classification and prediction tasks (Gerven M. and Bohte S., 2017). Neurons are aggregated into layers and perform a transform to their input signals by specific activation function, weights and biases. The weights are updated during the training process by backpropagation learning algorithm (Yang, et al., 2009).

Naïve Bayes (NB) is a probabilistic classifier that predicted the data labels as the highest conditional probability. The prediction is based on the Bayes' theorem and the assumption of independency between predicted variables (Nikam S. (2015). The classifier has the advantages of short training time but requires large training set.

Logistic regression is used for classification and regression problems. This technique used a regularization method to avoid overfitting. The models compute the a-posteriori probability of a label from predicting variables by minimizing a regularized cost function. Kernel Logistic Regression (KLR) maps the original data in a high-dimension feature space meaning a Mercer kernel function (Şenkal and Kuleli, 2009). This mapping allows a nonlinear classification and can process few training data.

RESEACH METHODOLOGY

Identifying students at-risk in the early weeks of the academic semester is very critical. In this research, the prediction model was developed to identify at-risk students in a course. The prediction depends

on three critical academic times of the semester. We used students' grades as academic assessment factors during the semester. The course assessment components were identified in the course to determine the student's success or fail. Five machine learning algorithms were used to build the prediction modeling based on the academic assessment factors that represented by scores or grades during the semester. The aim is to compare different prediction models to find best model that could predict at-risk students through evaluating the accuracy of each model.

Prediction Techniques

There are five machine learning algorithms that are popular data mining in education field. WEKA is a platform developed to be more oriented towards classification and regression problems (Abdelhafez, 2016). WEKA was used to apply these algorithms and building the prediction models.

Kernel Logistic Regression (KLR) is a non-linear classification technique. It is a kernel version of logistic regression used in statistical learning and handles multi class classification problems. KLR provides probabilities outputs using maximum likelihood arguments (Elbashir and Wang, 2015; Karsmakers, et al., 2007).

LibSVM is a library for Support Vector Machines uses multi-class classification. It is one of the most widely used SVM software. LIBSVM provides many statistics features such as confusion matrix, recall and precision (Chang and Lin, 2019; Elmannai, et al., 2013).

Multilayer Perceptron is the type of neural network. It is used for supervised learning or binary classification. The architecture of neural network depends on hidden layers or neural nodes that are stacked between the inputs and the output. These hidden layers assist in learning many complicated features (Singh J. and Banerjee R., 2019; Popescu, et al., 2009).

Naïve Bayes classifier is based on applying Bayes' theorem for probabilistic classification. It provides effective results in many applications. It also shows excellent performance even when the dataset attributes are not independent (Berrar, 2018).

Decision Tree-J48 is C4.5 algorithm for creating C4.5 decision tree. J48 is a binary tree used in WEKA. It takes into consideration the standardized data gain at each node to choose the best split for that node. J48 decision tree can handle specific characteristics and missing attribute estimations of the data (Venkatesan, 2015; Tina, 2013).

The Evaluation of Prediction Model

To evaluation classification problems, the confusion matrix is used as shown in Table 1. The equations derived from the confusion matrix used to evaluate the prediction models (Hossin and Sulaiman, 2015).

Table 1. Confusion matrix for classification

	Actual Positive Class	Actual Negative Class
Predicted Positive Class	True positive	False negative
Predicted Negative Class	False positive	True negative

$$\text{Accuracy} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (1)$$

$$\text{Precision / Specificity} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (2)$$

$$\text{Recall / Sensitivity} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (3)$$

$$\text{F-Measure} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (4)$$

These equations show the accuracy of the model, the precision, the recall, and F-measure for harmonic mean between recall and precision values.

ANALYSIS DATASET AND IMPLEMENTATION

Our analysis is based on real data for 2000 students in the first semester as shown in Figure 1. The dataset was extracted from computing and information technology course that is given to freshman in different universities. This choice is guided by the importance of student failure prediction specially in the first enrolled year in the higher education. Most of students' dropout occur during the first year (Callender and Feldman, 2009) as students may be demotivated and have not enough experience and support. The data contains the assessments details and scheduling for fifteen weeks. Table 2 details the assessment tools and schedule over the semester. The course includes several assessment components which are homework, quizzes, midterm exams and project as well as final exam. The student's grades are collected on a weekly basis.

Figure 1: Students scores and assessment components

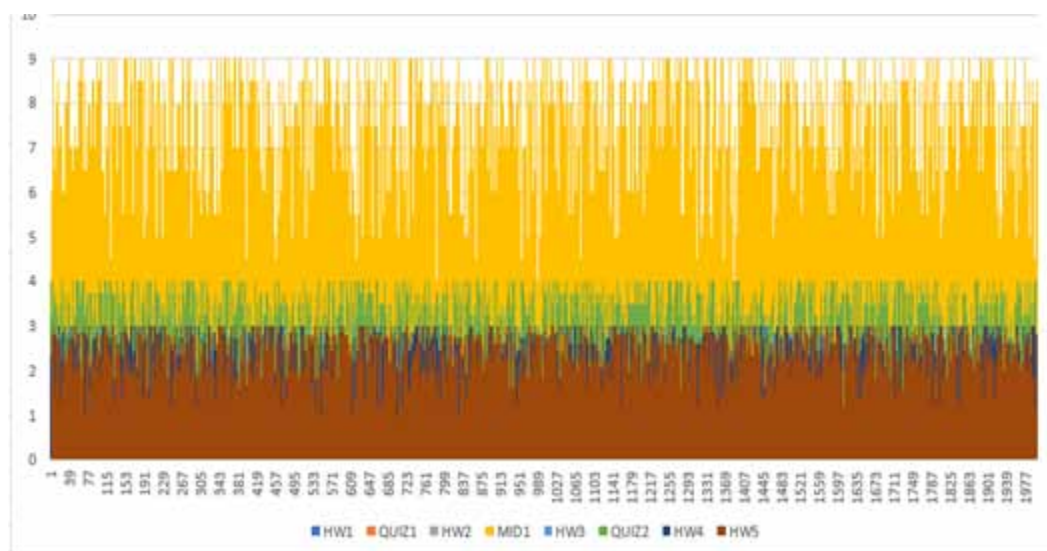


Table 2. Assessment plan and Distribution of grades

Assessment tool	Percentage of final grade	Due week	Dropout weeks
Homework 1 (HW1)	4%	2	
Quiz 1	5%	3	
Homework 2 (HW2)	4%	4	Early dropout
Midterm exam 1	10%	5	
Homework 3 (HW3)	4%	6	
Quiz 2	5%	7	
Homework 4 (HW4)	4%	8	
Homework 5 (HW5)	4%	9	
Midterm exam 2	10%	10	Late dropout
Project	10%	13	
Final exam	40%	15	

The dataset was divided into three sets training, testing, and validation. The 10-fold cross-validation was used according to the academic calendar, we implemented prediction models based on three important dates week 3, week 6 and week 9. For each week, five prediction models were applied and then comparing the results.

EXPERIMENTAL RESULTS AND DISCUSSION

The models were trained on 1200 students which consists of 60% from the cohort. The validation dataset and the test dataset consisted each of 400 students which represent 20% of the student number.

Predictive Models Assessment

The classification problem is binary. The discrimination evaluation of the optimal predictive classifier for student at-risk can be based on the confusion matrix presented in Table 1. The rows represent the predicted class and columns represented the effective class. Following the definitions of the determinant's variables in our model assessment, the number of students who failed and that are predicted as at-risk is called True Positives. The number of students who passed and are not predicted as at-risk is called True Negative. The number of students who failed but were not predicted as at-risk represents the False Negatives. In contrary, False Positives is the number of students who passed but were predicted as at-risk. Therefore, true positive and true negative represent the correctly classified instances and false positive and false negative are the misclassified instances

To assess the model prediction effectiveness, we will consider the accuracy of the student at-risk. As the goal of our study is to identify at-risk students, it is crucial to achieve high predictive accuracy for the students who failed the course. Figure 2 shows the prediction accuracy for the five models in week 3, week 6 and week 9. The best predictive accuracy in identifying at-risk students is obtained by the KLR prediction model in week 3, week 6 and week 9. KLR accuracy was higher to other models. Overall, while the KLR and MLP models performed similarly, the KLB was slightly higher than the MLP model in the three prediction weeks.

As the accuracy metrics has many limitations and produces less discriminative values, we also considerate the F-Measure that combines precision and recall to provide the harmonic mean. the F-Measure performed better than accuracy in discriminating classifiers for binary classification. The

Figure 2. Prediction accuracy for Week1, Week 2 and Week 3

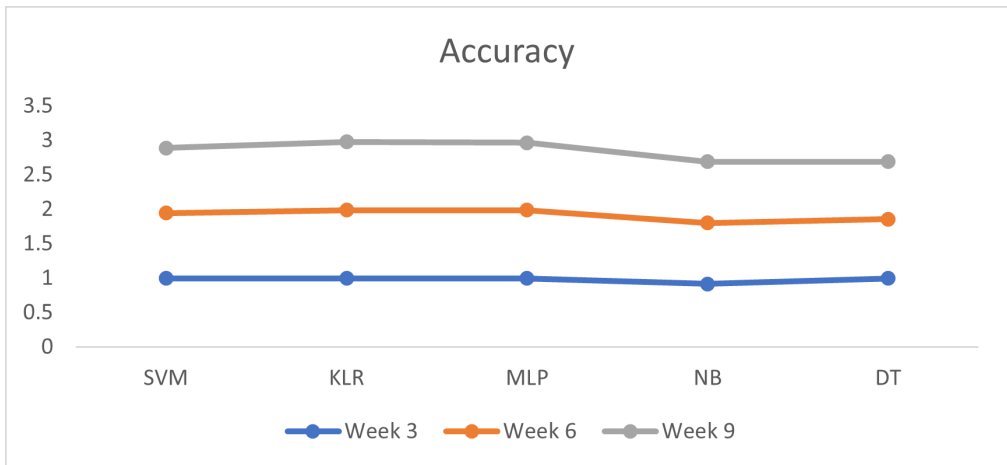


Table 3. The harmonic mean F-Measure for different prediction models and prediction weeks

	SVM	KLR	MLP	NB	DT
Week 3	0.994	0.994	0.994	0.888	0.994
Week 6	0.924	0.983	0.978	0.821	0.764
Week 9	0.894	0.975	0.962	0.829	0.696

Figure 3. Prediction root mean squared error by model for week 3, week 6 and week 9

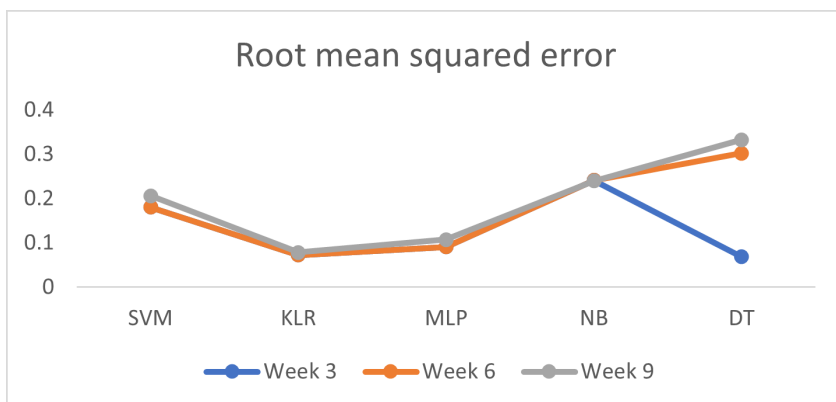


Table 4. Informativeness Analysis with KLR and MLP

Classifier	TP	FP	TN	FN	Total	Accuracy
KLR	155	1	242	2	400	0.9925
MLP	154	1	242	3	400	0.99

KLR has higher F-Measure as shown in Table 3. KLR is more efficient in predicting at-risk student and it outperforms other models especially in week 6 and week 9. KLR model was able to predict at-risk students' success and fail with the minimal error as shown in Figure 3. The NB model accuracy was lower than other models especially in week 6.

As KLR and MLP are equivalent and none of student at-risk class member (minority class) is correctly predicted by the classifiers, thus we focused on the informativeness analysis for the two classifiers as presented in Table 4 in week 6. KLR classifier performs better and was able to predict correctly the two classes more than MLP.

Attributes Analysis

Analyzing the prediction error showed that the prediction accuracy increased in week 9, except for DT model where the error becomes important since week 6. Analyzing in-depth the assessments statistics showed that the MID 1 and Quiz 2 had higher standard deviation for the two classes as shown in Table 5 and Table 6: at-risk students and success students. Therefore, the cumulative impact of the higher standard deviation appears in week 9 for two classes. This should explain that the assessment planification has great impact on student performance and that a notable variability in students' grades should impact the performance prediction. Week 6 and week 9 present therefore the most important key points for predicting at-risk students due to the importance of the scheduled assessments. The DT model implemented throw the Tress-J48 shows that the pruned decision tree has the nodes presented in Figure 4. The most determinant nodes are Mid1, Quiz1 and QUIZ2. This encourages the instructors to modify the assessment plan to balance the tree and reinforce the remain nodes effectiveness.

Table 5. Attributes analysis for at-risk class

	HW1	QUIZ1	HW2	MID1	HW3	QUIZ2	HW4	HW5
mean	1.81	2.3467	1.788	5.023	1.74	2.4444	1.836	1.7916
std. dev.	0.5579	0.6982	0.5794	0.9577	0.5828	0.7446	0.5767	0.6007

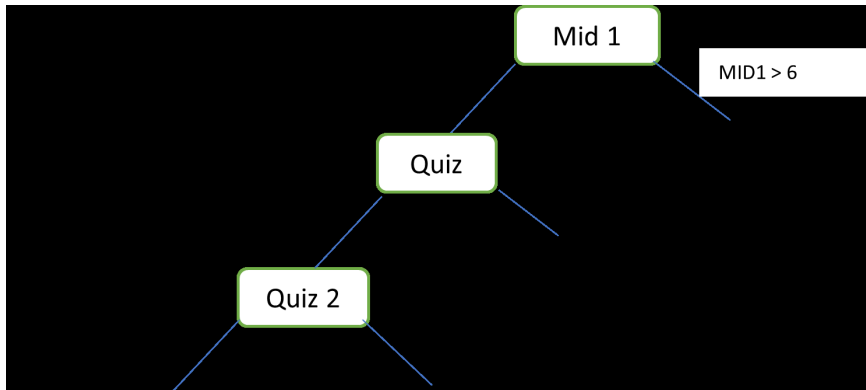
Table 6. Attributes analysis for success class

	Week 3 assessments		Week 6 assessments			Week 9 assessments		
	HW1	QUIZ1	HW2	MID1	HW3	QUIZ2	HW4	HW5
mean	2.0669	2.8853	2.055	6.8088	2.0349	2.89	2.0828	2.076
Standard deviation	0.5972	0.7762	0.6124	1.4745	0.6003	0.7887	0.5975	0.6405
Cumulative Percentage of final grade	4%	9%	13%	23%	27%	32%	36%	40%

Prediction Week Analysis

The cumulative percentage of assessments shows that 45% of the assessment are scheduled before

Figure 4. Attributes in Decision Tree



week 9 (27% among 60%) and 66.67% of the assessment are scheduled before week 9 (40% among 60%). The week 6 and week 9 are the most relevant prediction time for the at-risk students. Week 6 is the most convenient for an early prediction and it should be considered by instructors to prevent students who have been predicted as at-risk. If there were no or minor changes in the course syllabus and assessment exams between the cohort taken as study case and the next cohort, the predictive model based on KLR prediction in week 6 may be able to identify at-risk students in the next cohorts with high accuracy.

CONCLUSION

This research study examines different data mining algorithms and compares between them to accurately predict at-risk students. The prediction methods were implemented based on selecting three significant weeks of the semester. We used performance data of 2000 students for one semester. The five machine learning algorithms were used for creating prediction models for each week to determine at-risk students. The best prediction model was the Kernel Logistic Regression. It represents the best predictive accuracy based on the accuracy and harmonic mean (F-measure) in the three prediction weeks. The results also demonstrated that both week 6 and week 9 are the most relevant prediction time for the at-risk students. For early prediction, week 6 should be taken into consideration to determine the students who succeeded or failed. For future research, we expect to investigate performance data that cover more than one academic semester. In addition, future research might focus on the learning objectives of the course through studying the impact of learning objectives on the students' success in the course.

REFERENCES

- Abdelhafez, H. (2016). Mining Big Data in Telecommunications Industry: Challenges, Techniques, and Revenue Opportunity. In *Proceedings of 18th International Conference on Computer Science and Information Technology*. World Academy of Science, Engineering and Technology (WASET).
- ACT Institutional Data File. (2018). *National collegiate retention and persistence to degree rates*. Retrieved from <https://www.act.org/content/dam/act/unsecured/documents/MS2807rev1-retention-persistence-2018-07.pdf>
- Akçapınar, G., Altun, A., & Aşkar, P. (2019) Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 16(40), 1-20.
- Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques. *Education Sciences*, 11, 1-27.
- Asarta, C., & Schmidt, J. (2020). The effects of online and blended experience on outcomes in a blended learning environment. *The Internet and Higher Education*, 44, 100708.
- Baarsa, G., Stijnenb, T., & Splinterc, T. (2017). A Model to Predict Student Failure in the First Year of the Undergraduate Medical Curriculum. *Health Professions Education*, 3(1), 5-14.
- Berrar, D. (2018). *Bayes' theorem and naive Bayes classifier*. In *Encyclopedia of Bioinformatics and Computational Biology* (Vol. 1). Elsevier.
- Callender, C., & Feldman, R. (2009). Part-time undergraduates in higher education: A literature review. Prepared for HECSU to inform Future track: Part-time students, University of London.
- Chang, C., & Lin, C. (2019). *LIBSVM - A Library for Support Vector Machines*. <https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>
- Christensen, B., Bemman, B., Knoche, H., & Gade, R. (2018). Identifying students struggling in courses by analyzing exam grades, self-reported measures and study activities. *Proc. SLERD*, 167-176.
- Christensen, B., Bemman, B., Knoche, H., & Gade, R. (2019). Pass or fail? Prediction of students' exam outcomes from self-reported measures and study activities. *Interact. Des. Archit.*, 39(17), 44-60.
- Chui, K., Fung, D., Lytras, M., & Lam, T. (2020). Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. *Computers in Human Behavior*, 107, 105584.
- Costa, B., Fonseca, B., Santana, M., de Araújo, F., & Rego, J. (2017). Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, 73(supplement C), 247-256. <https://doi.org/10.1016/j.chb.2017.01.047>
- Du, X., Yang, J., & Hung, J. (2020). An Integrated Framework Based on Latent Variational Autoencoder for Providing Early Warning of At-Risk Students. *IEEE Access*, 8, 10110-10122. doi: 10.1109/ACCESS.2020.2964845
- Dyrbye, L., Lipscomb, W., & Thibault, G. (2020). Redesigning the Learning Environment to Promote Learner Well-Being and Professional Development. *Academic Medicine*, 95(5), 674-678. doi: 10.1097/ACM.0000000000003094
- Elbashir, M., & Wang, J. (2015). Kernel Logistic Regression Algorithm for Large- Scale Data Classification. *The International Arab Journal of Information Technology*, 12(5), 465-472.
- Elmannai, H., Loghmari, M. A., & Naceur, M. S. (2013). Support Vector Machine for Remote Sensing image classifications. *Proceedings of the International the International Conference on Control, Engineering & Information Technology*, 2, 68-72.
- Galloway, R., Reynolds, B., & Williamson, J. (2020). Strengths-based teaching and learning approaches for children: Perceptions and practices. *Journal of Pedagogical Research*, 4(1), 31-45.
- Gerven, M., & Bohte, S. (2017). Editorial: Artificial neural networks as models of neural information processing. *Frontiers in Computational Neuroscience*. Advance online publication. doi:10.3389/fncom.2017.00114
- Goodyear, P. (2020). Design and co-configuration for hybrid learning: Theorising the practices of learning space design. *British Journal of Educational Technology*, 51(4), 1045-1060.

- Horning, N. (2013). *Introduction to decision trees and random forests*. American Museum of Natural History.
- Hossin, M., & Sulaiman, M. (2015). A Review on Evaluation Metrics for Data Classification Evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), 1-11.
- Karsmakers, P., Pelckmans, K., & Suykens, J. (2007). Multi-class kernel logistic regression: a fixed-size implementation. *Proceedings of the International Joint Conference on Neural Networks*, 1756-1761. doi:10.1109/IJCNN.2007.4371223
- Keshtkar, F., Cowart, J., & Crutcher, A. (2016). Predicting Risk of Failure in Online Learning Platforms Using Machine Learning Algorithms for Modeling Students' Academic Performance. *Proceedings of the 33rd International Conference on Machine Learning*, 48.
- Lerche, T., & Kiel, E. (2018). Predicting student achievement in learning management systems by log data analysis. *Computers in Human Behavior*, 89, 367-372.
- Maheshwari, E., Roy, C., Pandey, M., & Rautray, S. S. (2020). Prediction of Factors Associated with the Dropout Rates of Primary to High School Students in India Using Data Mining Tools. In S. Satapathy, V. Bhateja, B. Nguyen, N. Nguyen, & D. N. Le (Eds.), *Frontiers in Intelligent Computing: Theory and Applications. Advances in Intelligent Systems and Computing* (Vol. 1013). Springer. doi:10.1007/978-981-32-9186-7_26
- Marbouti, F., Diefes-Dux, H., & Strobel, J. (2015). *Building course-specific regression-based models to identify at-risk students*. American Society for Engineering Education Annual Conference, Seattle, WA.
- Marbouti, F., Diefes-Dux, H., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, 103, 1-15.
- McGonagle, J., George, S., Hsu, A., & Williams, K. (2018). *Backpropagation*. Brilliant.org
- Meier, Y., Xu, J., Atan, O., & Schaar, M. (2016). Predicting Grades. *IEEE Transactions on Signal Processing*, 64(4), 959-972.
- Moreno-Marcos, P., Pong, T., & Kloos, D. (2020). Analysis of the Factors Influencing Learners' Performance Prediction with Learning Analytics. *IEEE Access*, 8, 5264-5282. doi: .10.1109/ACCESS.2019.2963503
- Nikam, S. (2015). A comparative Study of Classification Techniques in Data Mining Algorithms. *Orient. J. Comput. Sci. Technol.*, 8(1), 13-19.
- Osmanbegović, E., & Suljić, M. (2012). Data mining approach for predicting student performance. *Economic Review*, 10(1), 3-13.
- Polyzou, A., & Karypis, G. (2019). Feature Extraction for Next-Term Prediction of Poor Student Performance. *IEEE Transactions on Learning Technologies*, 12(2), 237-248.
- Popescu, M., Balas, V., Popescu, L., & Mastorakis, N. (2009). Multilayer Perceptron and Neural Networks. *WSEAS Transactions on Circuits and Systems*, 8(7), 576-588.
- Quinn, R., & Gray, G. (2019). Prediction of student academic performance using Moodle data from a Further Education setting. *Irish Journal of Technology Enhanced Learning*, 5(1). Advance online publication. doi:10.22554/ijtel.v5i1.57
- Romero, C., Ventura, S., Hervás, C., & Gonzales, P. (2008). Data mining algorithms to classify students. *Proc. Data Mining, Montreal: Int. Conf. Educ.*
- Sarra, A., Fontanella, L., & Di Zio, S. (2019). Identifying Students at-risk of Academic Failure Within the Educational Data Mining Framework. *Social Indicators Research*, 146(1-2), 41-60. doi:10.1007/s11205-018-1901-8
- Sedrakyan, G., Malmberg, J., Verbert, K., Järvelä, S., & Kirschner, P. (2020). Linking learning behavior analytics and learning science concepts: Designing a learning analytics dashboard for feedback to support learning regulation. *Computers in Human Behavior*, 107, 105512.
- Selbach-Allen, M., Greenwald, S., Ksir, A., & Thomley, J. (2020). Raising the Bar with Standards-Based Grading. *Primus*, 30(8-10), 1110-1126. DOI: 10.1080/10511970.2019.1695237

- Şenkal, O., & Kuleli, T. (2009). Estimation of solar radiation over turkey using artificial neural network and satellite data. *Appl Energy*, 86(7-8), 1222–1228.
- Singh, J., & Banerjee, R. (2019). A Study on Single and Multi-layer Perceptron Neural Network. *Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019)*. doi:10.1109/ICCMC.2019.8819775
- Son, L. H., & Fujita, H. (2019). Neural-fuzzy with representative sets for prediction of student performance. *Applied Intelligence*, 49(1), 172–187. doi:10.1007/s10489-018-1262-7
- Tina, R. P. (2013). Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification. *International Journal of Computer Science and Applications*, 6(2), 256–261.
- Tomasevic, N., Gvozdenovic, N., & Vranes, S. (2020). An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers & Education*, 143, 103676.
- Tsai, Y.-S., Rates, D., Moreno-Marcos, P. M., Muñoz-Merino, P. J., Jivet, I., Scheffel, M., Drachsler, H., Delgado Kloos, C., & Gašević, D. Tsai, Y. (2020). Learning analytics in European higher education Trends and barriers. *Computers & Education*, 155, 103933. doi:10.1016/j.compedu.2020.103933
- Vapnik, V. (1995). *The nature of statistical learning theory*. Springer-Verlag. doi:10.1007/978-1-4757-2440-0
- Venkatesan, E. V. (2015). Performance Analysis of Decision Tree Algorithms for Breast Cancer Classification. *Indian Journal of Science and Technology*, 8(1), 1–8. doi:10.17485/ijst/2015/v8i29/84646
- Viberg, O., Hatakka, M., Bälter, O., & Mavroudia, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior*, 89, 98–110.
- Waheed, H., Hassan, S., Aljohani, N., Hardman, J., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, 104, 106189.
- Wakelam, E., Jefferies, A., Davey, N., & Sun, Y. (2020). The potential for student performance prediction in small cohorts with minimal available attributes. *British Journal of Educational Technology*, 51(2), 347–370.
- Wiley, K., Dimitriadis, Y., Bradford, A., & Linn, M. (2020). From theory to action: developing and evaluating learning analytics for learning design. *LAK '20: Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 569–578.
- Yang, J., Shen, K., Ong, C., & Li, X. (2009). Feature selection for MLP neural network: The use of random permutation of probabilistic outputs. *IEEE Transactions on Neural Networks*, 20(12), 1911–1922. doi:10.1109/TNN.2009.2032543
- Zhang, J., You, C., Huang, J., Li, S., & Wen, Y. (2020). Research on Application of Frequent Pattern Growth Algorithm in Academic Early Warning. *Proceedings of the 2020 8th International Conference on Information and Education Technology*, 116–121.

Hoda Ahmed Abdelhafez is an associate professor, and she holds a PhD in Information Technology from Alexandria University. She has over 20 publications in big data, data warehouse, decision support, data mining, e-government, and e-learning. She has three book chapters, two of them published in the Encyclopedia of Business Analytics and Optimization and the third one in the Encyclopedia of Information Science and Technology.

Hela Elmannai holds the PhD in Information and Communication Technology from the High School of Communication Technology-University of Carthage in 2017. From 2007 to 2017, she is a lecturer in the Higher Institution of Technological Studies in Tunisia. Since 2018, she is an Assistant Professor in Princess Nourah bint Abdulrahman University in Saudi Arabia. Her research interests are data processing, image classification and artificial intelligence. Corresponding Author: hselmannai@pnu.edu.sa.