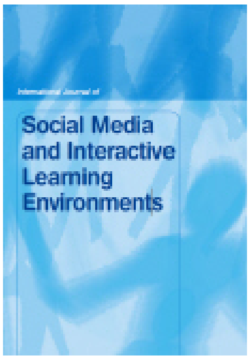
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Social Media and Interactive Learning Environments

ISSN online: 2050-3962 - ISSN print: 2050-3954
https://www.inderscience.com/ijsmile

Detecting students at risk using machine learning: applications to business education

Owen P. Hall Jr.

DOI: 10.1504/IJSMILE.2020.10032203

Article History:

Received:	14 May 2020
Accepted:	28 July 2020
Published online:	08 August 2022

Copyright © Inderscience Enterprises Ltd.

Detecting students at risk using machine learning: applications to business education

Owen P. Hall Jr.

Graziadio Business School,
Pepperdine University,
24255 Pacific Coast Hwy, Malibu, CA 90263, USA
Email: ohall@pepperdine.edu

Abstract: Detecting students at risk continues to challenge the management education community. Traditionally, student examination performance and attendance have been two of the primary metrics used for identifying students at risk. However, waiting until midterm exam results to intervene can often prove problematic. With the advent of cloud-based learning platforms, these traditional factors can now be complemented by a variety of quantitative and qualitative metrics. The results from the current study indicate that machine learning-based classification models can detect struggling students and identify appropriate intervention initiatives. Specifically, student performance on practice quizzes was found to be an effective early warning indicator, which, in conjunction with related student attributes, can be used to identify customised amelioration strategies. The primary purpose of this article is to highlight how machine learning can reduce student dropout rates and improve overall learning outcomes throughout the business education universe.

Keywords: machine learning; business education; student risk detection; practice quizzes; intervention strategies; actionable knowledge discovery.

Reference to this paper should be made as follows: Hall Jr., O.P. (2022) 'Detecting students at risk using machine learning: applications to business education', *Int. J. Social Media and Interactive Learning Environments*, Vol. 6, No. 4, pp.267–289.

Biographical notes: Owen P. Hall Jr. holds the Corwin D. Denney Academic Chair and is a Full Professor of Decision Sciences at Pepperdine Graziadio Business School. He is a Julian Virtue Professor and a Rothschild Applied Research Fellow. He received the Harriet and Charles Luckman Distinguished Teaching Award in 1993, the Howard A. White Teaching Excellence Award in 2009 and 2017, and the Sloan-C Effective Teaching Practice Award in 2013. He is the Vice Chairman of the INFORMS University Analytics Program Committee. He is a registered Professional Engineer, State of California. He received his PhD from the University of Southern California and undertook post-doctoral studies at the Center for Futures Research.

Copyright © 2022 Inderscience Enterprises Ltd.

1 Introduction

Identifying students at risk and formulating appropriate intervention plans has become a hot button topic throughout academe especially in light of the COVID-19 pandemic. Students at risk can be defined as those who are planning to drop out of the institution, are under consideration for termination by the institution, and those that are on academic probation. Current data suggests that nearly 30% of incoming freshman enrolled in a four-year college do not graduate within the standard seven-to-nine-years norm (Eller and DiPrete, 2018). Overlaying these trends is that many schools are now facing increased competition, demanding customers, and a growing aversion toward debt financing. This combination of forces tends to drive up the cost of student acquisition and retention. The emergence of the internet generation as the new student body, who are web savvy and heavily engaged in social media, requires institutions to develop a real-time and interactive approach to enhance student learning opportunities and outcomes, which is heavily reliant on 'big data'. Until recently, the processes used by most business schools have focused on data warehousing, which tends to provide a backward perspective, for example, assessing student performance after the fact. In contrast, a forward-looking emphasis be extremely helpful in detecting students at risk at an early stage and delivering customised content in a timely manner.

More specifically, machine learning can be used to both identify early on students at risk and design specific implementation amelioration strategies. The expression *machine learning* was initially coined by Samuel in the late 1950s and is usually defined as a computer's ability to automatically learn and improve from experience without being explicitly programmed by humans (Samuel, 1959). This 'semi-automated' capability is essential considering the large number of potential students at risk and the large number of factors that are used to pinpoint risk candidates and design intervention strategies. Today, the state-of-the-art in machine learning has advanced significantly compared to even a decade ago (Bakhshinategh et al., 2018; Agrawal and Mavani, 2015). Specific machine learning educational developments include visual analytics, plagiarism detection and virtual and augmented reality (Amidwar, 2017; Endert et al., 2017; Jantjies et al., 2018). For example, the integration of machine learning with visualisation methods can offer the student substantially more advanced feedback in a user-friendly learning environment. Each of these technological developments enhances the process for improving student retention through early detection and intervention (Nespereira et al., 2016).

The use of machine learning in higher education is closely aligned with educational data mining (EDM), which has seen increased usage over the past decade (Depren et al., 2017; Kaur et al., 2019; Paraiso et al., 2015; Sarra et al., 2018). The primary focus of EDM is on collecting, archiving, and analysing educational data with the goal of improving student success (Kumar, 2015). Some specific EDM objectives include (Hernández-Blanco et al., 2019):

- Predicting student performance: the goal is to estimate in advance students' achievement and the accomplishment of learning outcomes.
- Detecting undesirable student behaviours: the focus is on identifying behaviours such as low motivation, limited interactions, cheating, or dropping out.

- Profiling and grouping students: the purpose is to profile students based on different factors and to use this information to group students for various purposes, including study teams.
- Creating alerts for stakeholders: the aim is to serve as an online platform for informing stakeholders (e.g., faculty) or creating alerts in real time regarding various student behaviours.

The application of machine learning in conjunction with the appropriate EDM processes can be used to achieve these objectives. This paper is organised as follows: a review of the students at risk literature, an introduction to machine learning, an example application, and an overview of risk intervention strategies. This article's primary contribution to management education is to illustrate how machine learning can be used to detect students at risk and develop effective intervention plans. In this regard, the article helps fill a gap between the need to identify students at risk and the technical means to accomplish this goal.

> "Poorer academic outcomes and problematic health behaviours are linked to students' distress, and these wider implications also highlight the need for appropriate policies and services to support students during what is clearly a challenging time." (Sharp and Theiler, 2018).

2 Detecting students at risk

The goal of identifying students at academic risk early in the matriculation process is not new (Seidman, 1996). He proposed the following relationship, which links student retention to both early detection and continuous intervention:

$$\text{Student retention} = \text{Early detection} + \text{Continuous intervention}$$

This construct highlights that early detection of students at risk as well as continuous intervention can be key to student retention. The impact of student attrition should be viewed from a holistic perspective, including financial, developmental and future advancement. Two illustrative examples of this viewpoint are the lost revenues to the institution and the economic burden placed on the student as a result of dropping out of school. Furthermore, students can be at risk for a variety of reasons (e.g., behavioural issues); however, for this study, the focus is on academic risk. Additionally, interventions can take on many forms (e.g., faculty counselling), however, for this study, the interventions are based on providing customised learning content in a timely manner via the web.

Today, initiatives to improve student retention are receiving increased attention throughout the higher education universe (Azcona et al., 2019; Febro and Barbosa, 2017; Demetriou and Schmitz-Sciborski, 2011; Moekotte et al., 2017). At the core, the goal is to predict future student performance based on a combination of web-based activities and student characteristics. This capability, in turn, leads to the design and implementation of specific interventions where both customised content and motivations are provided. Furthermore, it has been discovered that early intervention programs can reduce the gap between the lower and higher-performing students (Kent et al., 2020). For example, Purdue University's Signals Project has addressed the problem of enhancing student success, which has resulted in improved retention and graduation rates (Pistilli and Arnold, 2012). This project has led to the development of student success data mining algorithms with intervention messages sent to students based on performance via dashboards. More specifically, these systems can be employed to: select student groups with similar characteristics and reactions to learning strategies, detect student misuse and lurking, recognise students who, in certain types of test formats, are hint-driven, locate students who exhibit low motivation and find alternate means of reaching them, and predict probable student outcomes (Shahiri and Rashid, 2015).

The standard variables for predicting student performance have often included incoming GPA, work experience and admission exam scores (Maldonado and Seehusen, 2018; Pratt; 2015). In today's data rich environment, these parameters can be augmented by threaded discussion posts, chat message, social media and surveys (Wu et al., 2018).

Table 1 extends this classical list of evaluation factors into four general risk assessment categories: admissions, demographic, behaviour, and performance, once the student has enrolled (Duarte et al., 2014; Embse et al., 2017; Fish and Wilson, 2009; Malau-Aduli et al., 2017).

Student persistence, as measured by attendance, for example, was discovered to be a strong indicator for predicting dropout rates. The current literature suggests that quizzes can be used as both a learning vehicle and a mechanism for finding students at risk (Grison et al., 2011; Kwan, 2018; O'Dowd, 2018; Wickline and Spektor, 2011). Additionally, they can serve as a vehicle for engaging students in a proactive learning environment. Modifying the format of the practice quizzes over the course of the term can also enhance the learning experience by minimising boredom.

Table 1 Candidate student risk detection categories

Admissions	*Demographic*	*Behaviour*	*Performance*
GPA[1]	Economic status	LMS engagement[2]	Quizzes
Waiver	Gender	Attendance	Deadlines[3]
Work experience	Race	Team engagement	GPA[4]

Note: [1]Incoming GPA, [2]LMS = learning management system, [3]meets course assignments in a timely manner and [4]current GPA.

In summary, the role of quizzes can be characterised as follows: to improve student learning in a 'low stakes' environment, and to serve as an indicator of potential achievement in summative 'high stakes' evaluations like midterm examinations, final examinations and post-graduation licensed examinations (Azzi et al., 2014; Becerra et al., 2018; Kibble, 2007; Malkemes and Phelan, 2017).

Summary findings of these four studies are highlighted below:

- The first study ($N = 164$) utilised multiple-choice and short answer quizzes associated with each laboratory session over the course of two semesters which yielded a modest correlation between the average quiz scores and the final exam ($r = 0.35$, p-value < 0.05). The results also revealed that midterm exam performance was a better predictor of final exam achievement ($r = 0.61$, p-value < 0.05).

- The second study ($N = 370$), conducted over two trimesters, found a statistically significant correlation between weekly practice quiz score averages and the final examination ($r = 0.58$, p-value < 0.001). The basic conclusion was that practice quizzes could be considered as one of the measurable predictors of academic performance.

- In the third study (N = 339), the results showed that students who elected to use the online quizzes performed better in summative examinations ($r = 0.34$, p-value < 0.05). Offering course credit of between 0.5% and 2% per quiz increased student participation. However, evidence was found of widespread inappropriate use of unsupervised quizzes.
- In the fourth study (N = 54), the students were required to take practice quizzes during their final year. This study utilised an adaptive quiz system where the level of difficulty could vary based on student responses to calibrated standards. The analysis revealed a modest positive correlation between the number of questions a student answered and their overall mastery level examination score ($r = 0.28$, p-value < 0.05).

The reported outcomes of these studies revealed that the degree of making the quizzes mandatory and making them appear more like the summative assessments are two key issues. With respect to the first issue, there are three basic options: completely optional, optional but provide an incentive, and mandatory.

Recent evidence suggests that the use of mandatory quizzes was found to be very unpopular with students-based, in part, on student comments on end-of-semester course evaluations (Brown et al., 2015; Day et al., 2017). Regarding the second issue of increasing the number of practice questions per quiz as a vehicle to better emulate the summative assessments, the obvious consequence will be a reduction in student participation, especially in terms of taking the practice quizzes more than once (Cook and Babon, 2017).

Web-based practice quizzes have several attractive characteristics, including: a low stakes context as mentioned previously, automated delivery and assessment, real-time student feedback, and ability to modify the quizzes based on student performance.

As such, the use of practice quizzes as an early indicator of performance provides a quantitative vehicle, when used in combination with related factors, such as student engagement and student demographics, for identifying students at risk with a potentially high degree of precision. Utilising web access logs represents another medium for detecting high-risk students. A recent study found that students who generally spent more time on online movies and games tended to underperform on examinations compared to students who engaged websites related to document downloading (Zhou et al., 2017). The use of quizzes and access logs for early detection, however, represents only the first component of the Seidman equation. The second and equally important component, continuous intervention, will be addressed in the intervention strategies section of this article.

> "Universities should build an intervention programme that will target specific retention problems including, 1) maintaining high expectations of students, 2) explaining institutional requirements, 3) providing academic, social, and personal support, 4) showing students that they are valued, and 5) offering frequent contact with the staff." (Clark et al., 2010)

3 Machine learning

Machine learning is the science of discovering and communicating meaningful patterns in data and supporting the development of actionable plans. Today, machine learning is seeing increased use throughout academia (Delen, 2010; Mason et al., 2017; Sin and Muthu, 2015; Wanjau et al., 2018). In many previous educational classification studies, logistic regression and ordinal regression have been the methods of choice (Adejumo and Adetunji, 2013; Buskirk and Kolenikov, 2015). Some of the limitations of these models include a decrease in performance as the feature space or the number of categorical features increases and often the need for transformations to address nonlinear effects. Furthermore, these models do not provide discrete outright categories. Instead, they produce probabilities associated with each observation, which then requires a subsequent step to translate probabilities into classifications based on some norm. To address these concerns, the following two machine learning techniques were employed: neural nets (NN) and classification regression trees (CART). Neural networks have seen increased use in EDM studies (Asken and Gokalp, 2011; Aybek and Okur, 2018). Some of the advantages of NN include: address nonlinear relationships, handle outliers, and no prior knowledge regarding the nature of possible relationships.

A lack of understanding regarding the relationship between the inputs and outputs is one of the major operational disadvantages of NN. CART is a non-parametric analytical procedure that generates variable-based structural trees: classification trees for categorical target variables, and regression trees for continuous target variables.

One of the advantages of the basic CART model is that it provides a graphic rendering of the model variables. Trees are formed by a collection of rules based on the values of the predictor variables. CART has also seen considerable application in the educational field (Chiheb et al., 2017; Mesaric and Sebalj, 2016; Migueis et al., 2018). There is a family of CART algorithms which include the basic model, random forest trees (RFT), and extreme gradient boosting (XGB) trees. The XGB algorithm tends to outperform other machine learning models in many classification applications (Dinh et al., 2019).

Table 2 Standard confusion matrix – binary classifier[1]

	Actual condition			
	Positive	*Negative*		
Predict positive	TP	FP	TP / (TP + FP)	PPV[2]
Predict negative	FN	TN	TN / (FN + TN)	NPV[3]
	TP / (TP + FN)	TN / (FP + TN)	(TP + TN) / T	Accuracy
	Sensitivity	Specificity		

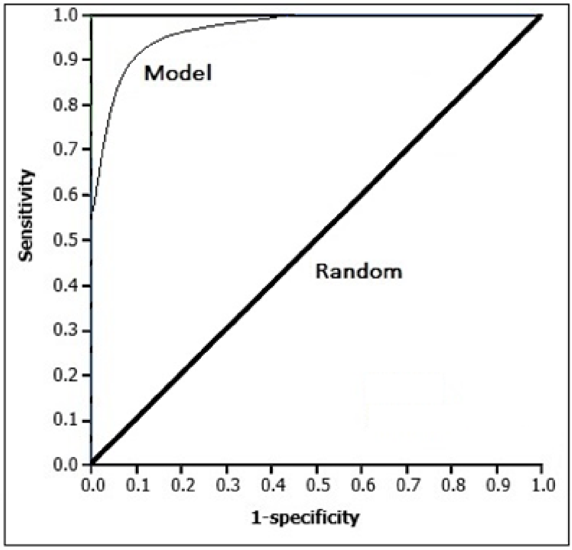
Note: [1]TP = true positive, TN = true negative, FP = false positive, FN = false negative, T = TP + TN + FP + FN, [2]PPV = positive predictive value and [3]NPV = negative predictive value.

The confusion matrix is one of the standard methods used for assessing the performance of the classifier and is illustrated in Table 2. The confusion matrix compares predictions with the actual observed conditions. For example, the statistic sensitivity, also called recall, measures the proportion of actual positives that are correctly identified, while specificity reports the proportion of actual negatives that are correctly characterised.

In the context of this study, a positive predictive value is the probability that a student classified at risk is actually at risk. In contrast, a negative predictive value is the probability that a student was classified not at risk when they actually are not at risk. In many studies, the metric accuracy is used to both judge the relative performance of the various candidate models and also serve as a standard for selecting the 'best' model for subsequent usage. Often, there exists a significant imbalance in the database, where, in the case of a binary target variable one of the two categories contains only a few observations (Ohsaki et al., 2017). For example, in an analysis of student plagiarism, most of the cases will be associated with the 'non-plagiarism' category with a very small fraction assigned to the 'plagiarism' category. In these instances, the statistic accuracy may overstate the performance of the model. When this condition occurs, the Gini index is often employed (Bethapudi et al., 2015). The Gini index ranges in value between zero and one, where a value of one indicates that the model is 100% accurate in predicting the outcome. Whereas a Gini value of zero indicates that the model's classification performance is equivalent to coin flipping. In additional to judging model performance, the Gini index is also used by CART to create optimal splits. Basically, the higher the Gini value, the more homogeneity within each leaf. The Gini index is closely associated with the receiver operating characteristic (ROC) curve, which is illustrated in Figure 1. The ROC graph, featuring sensitivity (true positive rate) on the vertical axis and one minus specificity (false positive rate) on the horizontal axis, is designed to illustrate the diagnostic ability of a binary classifier as its discrimination threshold is varied. The random line represents the so-called 'line of no discrimination', which is equivalent to coin flipping. Points above the random line indicated better classification performance compared to the random process. Specifically, the curve at the upper left reports the performance of an example classifier model. Clearly, this model has outperformed the random line in correctly classifying the two categories. The ideal performance case is when the model line rises vertically from the origin to the top of the Y-axis and then horizontally to the end of the X-axis. This graphic can be used to compare the performance of several classification models by plotting their ROC curves on the same graph. A more precise approach is to compare the area under each curve (AUC), which is usually generated by the classifier. Like Gini, AUC ranges in value between zero and one. An AUC value of one suggests perfect discrimination, while a value of 0.5 indicates a random classification process.

> "Adaptive educational systems emphasize that learning processes differ among learners. To ensure that study materials and e-learning services are tailor-made for adaptive learning, a machine learning approach attempts to integrate a capacity to diagnose the specific needs of each individual." (Almohammadi et al., 2017)

Figure 1 Example ROC curve



4 Illustrative example

To illustrate the process outlined above, demographic and assessment data was collected on 266 students engaged in a fully employed MBA business analytics course at a private business school located in the USA with an enrolment of approximately 2,000 students. The primary focus of this core course was on introducing predictive and prescriptive analytics. Some specific operating characteristics of the course were:

- small class sizes with less than 30 students
- close and ongoing student-faculty engagement
- students with significant work experience
- a growing online presence.

Table 3 highlights the various model variable mnemonics and corresponding descriptive statistics for the assembled database. Traditional delivery is defined as primarily in-class instruction. The statistics reported in Table 3 reveal, for example, that 44% of the participants consisted of women and that the average working experience was about eight years. These statistics are consistent with their overall proportions in the MBA program. Practice quizzes were offered each week on a completely optional basis per the finding of the literature review. Each weekly quiz, consisting of ten randomly multiple-choice questions, could be taken more than once since the questions would vary from quiz to quiz. The following three quiz metrics were included in the database: overall quizzes average at the three-week and six-week mark, and number of quizzes taken at the three and six-week mark, and the portion of weekly quizzes taken.

Table 3 Selected variable mnemonics and descriptive statistics (N = 266)

Mnemonic	*Definition*	*Mean*	*SD*
DEL	Delivery method: online = 1, traditional = 0	0.42	-
WRK	Years of working experience	7.79	5.24
UGP	Undergraduate grade point average	3.13	0.41
WAV	Waiver received: yes = 1, no = 0	0.67	-
GND	Gender: female = 1, male = 0	0.48	-
QN3	Number of quizzes taken in 1st three weeks	11.05	9.64
QA3	Average quiz scores in 1st three weeks	68.70	22.56
QNM	Number of quizzes taken prior to midterm	20.82	17.93
QAM	Average quiz scores prior to midterm	67.80	21.20
QP3	Fraction of quizzes taken in 1st three weeks	0.88	-
QPM	Fraction of quizzes taken prior to midterm	0.87	-
MES	Average midterm exam numerical score	77.80	12.81
AMD[1]	Fraction below average on midterm: yes = 1, no = 0	0.42	-
SMD[1]	Fraction below 70 on midterm: yes = 1, no = 0	0.23	

Note: [1]Candidate target variables.

Approximately 88% of the students took all three quizzes at least once during the first three weeks and on average a student took approximately 11 quizzes during this period with an average score of about 69. To be granted an admission waiver, an alternative to the standard GMAT admission requirement, the applicant needed a minimum of three years of working experience, an undergraduate degree, and a B or better in a statistics course. Waivers as an admission criterion substitute for the GMAT in MBA programs have become increasingly popular (Fairfield-Sonn et al., 2010). For this study, a binary classification scheme was employed, which featured two target variables: scoring below the midterm average, and scoring below 70 (the minimum passing grade).

The data shows that approximately 42% scored less than the average and 23% of the students scored less than the minimum passing score of 70. This relatively small proportion for the non-passing class underscores the so-called unbalance problem associated with the current database (Li et al., 2013). There are a variety of different approaches that can be used to help address the unbalanced problem (Longadge et al., 2013). The over-sampling method was adopted for this study, where the proportion of the below passing class was increased to approximately 42%, making it the same proportion as the below average midterm class. Additionally, all but one of the continuous variables were normalised using the min-max procedure since they were highly skewed as measured by the Kolmogorov-Smirnov test (e.g., student age). UGP was the one exception and it was rescaled using the standardisation method, which retains more information.

Figure 2 presents a scatter diagram that highlights the relationship between the average scores for the first three practice quizzes (QA3) and the midterm exam score (MES). The corresponding correlation coefficient is a moderate 0.321, which is statistically significant at the 0.01 level. As can be seen, approximately 5% of the students did not take any of the first three optional practice quizzes. This high participation rate can be attributed, in part, to weekly e-mails that were sent out to encourage students to take the practice quizzes.

Figure 2 MESs versus average quiz scores (1–3) (see online version for colours)

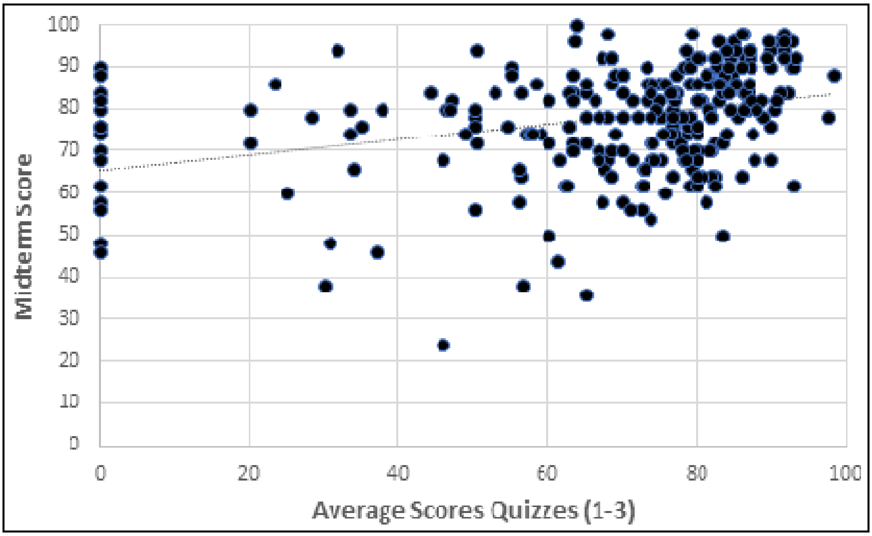


Figure 3 Work experience histogram

Figure 3 shows a relative frequency histogram of student work experience. For example, approximately 45% of the students had five or less years of work experience. Notice that this distribution is highly skewed to the right, which was why this variable was normalised using the max-min procedure discussed above.

Table 4 reports the Pearson correlation coefficients for the continuous variable set. For example, the data revealed that there is a moderate, positive association between incoming UGP and MES, which was statistically significant (r = 0.231, p = 0.000). Not surprisingly, there is a high degree of multicollinearity between the various quiz-related variables, for example, QP3 and QA3.

Table 4 Correlation matrix (Pearson)

	WRK	*UGP*	*QN3*	*QA3*	*QNM*	*QAM*	*QP3*	*QPM*	*MES*
WRK	1								
UGP	0.033	1							
QN3	0.124*	0.117	1						
QA3	0.059	0.010	0.428**	1					
QNM	0.127*	0.102	0.982**	0.447**	1				
QAM	0.057	0.036	0.456**	0.870**	0.473**	1			
QP3	0.099	0.050	0.431**	0.678**	0.428**	0.669**	1		
QPM	0.061	0.051	0.455**	0.682**	0.481**	0.680**	0.922**	1	
MES	–0.073	0.231**	0.262**	0.321**	0.262**	0.385**	0.259**	0.291**	1

Note: Correlation significant at the **0.01 and *0.05 level.

Table 5 Correlation matrix (Kendall-Tau)

	WRK	*UGP*	*QN3*	*QA3*	*QNM*	*QAM*	*QP3*	*QPM*	*SMD*
WRK	1								
UGP	–0.019	1							
QN3	0.035	0.080	1						
QA3	0.000	0.009	0.365**	1					
QNM	0.035	0.073	0.910**	0.388**	1				
QAM	0.008	0.045	0.411**	0.679**	0.427**	1			
QP3	0.083	0.033	0.509**	0.318**	0.495**	0.359**	1		
QPM	0.044	0.045	0.554**	0.302**	0.580**	0.345**	0.795**	1	
SMD	0.033	–.174**	–.129*	–.180**	–.138*	–.220**	–.150*	–.174	1

Note: Correlation significant at the **0.01 and *0.05 level.

Table 5 reports the Kendall-Tau correlation coefficients for the continuous predictor variable set and the binary target variable SMD. For example, there was a modest, inverse correlation between UGP and SMD, which was statistically significant ($\tau = -0.174$, p = 0.001). These results also show a high degree of multicollinearity between the various quiz-related variables, for example, QN3 and QA3.

The standard approach in applying machine learning models is to divide the database into two sets (training and testing). Typically, a minimum sample size of 50 to 100 observations per predictor variable is required to support this approach. Often, 70% of the data is used for training the model and the remaining 30% to test the model (Korjus et al., 2016). This approach tends to ameliorate the impact of overfitting, which frequently results in over-optimistic model performance. A common variation to this two-step approach is to incorporate a 'validation' step between model training and testing. The objective of validation is to 'optimise' the training model's hyperparameters, those not learned from the data but set prior to the training phase. The splitting criteria are an example of a hyperparameter in a CART analysis. The validation results can also be used in selecting the 'best' model when there is more than one candidate. In this design formulation, the data is often partitioned using the 60/20/20 rule, where 60% is used for training, 20% for validation, and 20% for testing (Rifat et al., 2019).

The relatively small sample size associated with the current study precluded either of the above designs, which would have called for approximately 1,000 observations given the number of candidate variables (Park and Yu, 2018; Zavorka and Perrett, 2014). However, it was also recognised that using the entire database for training could greatly exaggerate model performance due to overfitting. Accordingly, based on the above guidelines, 70% of the data was used for training, while the entire database was used for testing. While not completely satisfactory, this approach provided more realistic performance outcomes than simply using the entire dataset for training.

The current model is designed to detect if a current student is at risk. However, an even more instructive approach is to characterise students at various levels of risk. In this scenario, the confusion matrix would be extended beyond two classifications (Xu et al., 2020). For example, introducing an ordinal target variable with three categories consisting of: 0 = low risk, 1 = moderate risk and 2 = high risk, where again risk could be measured based on quiz performance. The benefit of this approach would permit an even more targeting of specialised content based on the risk level. However, this expanded model design would require an even larger database (Dobbin and Simon, 2006; Indria et al., 2015; Kanaris et al., 2016; Shao et al., 2013).

The two machine learning models discussed above (NN and XGB) were employed to evaluate the database for the following three case scenarios:

Case #1 (DEL, WRK, UGP, WAV, GND)

Case #2 (DEL, WRK, UGP, WAV, GND + QN3 + QA3 + QP3)

Case #3 (DEL, WRK, UGP, WAV, GND + QNM + QAM + QPM)

In Case #1, the focus was on assessing the classification power of only demographic factors. In Case #2, those factors were augmented by practice quiz results for the first three class sessions. For Case #3, the demographic factors were supplemented by practice quiz outcomes during the first six sessions. The midterm examination occurred during session seven.

Table 6 Comparison of XGT and NN classification results (below average) – Case #2

	Actual BA[1]	*Actual NBA*[2]	*Total*	*%*	
Predict BA	99/97	17/17	116/114	89.2/90.8	PPV[3]
Predict NBA	12/14	138/138	150/152	89.0/85.1	NPV[4]
Total	111/111	155/155	266/266	89.1/88.3	Accuracy (%)
%	89.1/87.4	89.0/89.0			
	Sensitivity	Specificity			

Note: [1]Midterm score below average (BA), [2]midterm score average or above (NBA), [3]positive predictive value and [4]negative predictive value.

Table 6 provides a comparison of the performance of the two classifiers for Case #2 using the midterm average as the target variable. The numbers in the body of Table 6 represent frequency count, for example, XGB correctly classified 99 of the cases as below average that actually were below average. The results from Table 6 suggest that the classification performance of the two models was approximately the same. The corresponding AUC values were 0.94 and 0.92, respectively.

Table 7 Comparison of XGT and NN classification results (below 70) – Case #2

	Actual B70[1]	*Actual N70*[2]	*Total*	*%*	
Predict B70	103/107	19/19	122//126	84.4//95.7	PPV[3]
Predict N70	10//6	134//134	144//140	93.1//84.9	NPV[4]
Total	113/113	153/153	266//266	89.1//90.6	Accuracy (%)
%	91.2//94.7	87.6//87.6			
	Sensitivity	Specificity			

Note: [1]Midterm score below 70 (B70), [2]midterm score 70 or above (N70), [3]positive predictive value and [4]negative predictive value.

Table 7 contrasts XGT and NN classification performance, again for Case #2, where the target variable was a midterm score below 70. Recall that in this case the proportion of employees that left the organisation was increased to 42% via the up-sampling method while holding the overall sample size to 266. Again, the classification performance of the XGT and NN was about the same as measured by accuracy.

Table 8 Summary of classification accuracy for below average by model (XGT/NN)

Target	*Case 1 (%)*[1]	*Case 2 (%)*[2]	*Case 3(%)*[3]
Below 70	80.5/74.4	89.1/90.6	91.0/93.6
Below ave.	80.8/73.3	89.1/88.3	88.0/84.6

Note: [1]Pre-quiz, [2]quizzes 1–3 and [3]quizzes 1–6.

Table 8 summarises the two classifiers performance over the three case scenarios for the two target variables based on overall accuracy. The biggest jump in classifier improvement occurred between Case #1 and Case #2, where Case #2 included quiz performance for the first three quizzes. Adding quizzes 4 through 6 does not seem to improve classifier performance, which is a useful outcome since the basic thesis of this article is that students at risk need to be detected early and then engaged with effective and time-sensitive interventions.

Table 9 Summary of classification impurity (Gini) by model (XGT/NN)

Target	*Case 1*[1]	*Case 2*[2]	*Case 3*[3]
Below 70	0.76/0.65	0.88/0.85	0.93/0.88
Below ave.	0.73/0.63	0.88/0.84	0.86/0.79

Note: [1]Pre-quiz, [2]quizzes 1–3 and [3]quizzes 1–6.

As reported earlier, the Gini index is another measure used to assess classifier performance, especially when the data is unbalanced. Table 9 presents a performance assessment of the two classifiers for the three cases and two target variables using Gini.

The general results are very similar to the data given in Table 8. This should not be too surprising since the database was not significantly unbalanced after over-sampling was applied.

Table 10 Predictor variables relative importance by model (Case #2 – below average)

XGT		*NN*	
Factor	*Weight*	*Factor*	*Weight*
QA3	100.0	WRK	100.0
UGP	97.5	QA3	81.8
QN3	82.0	QN3	33.6
WRK	68.0	GND	26.2
DEL	52.5	UGP	17.2
GND	50.9	DEL	13.7
WAV	46.7	QP3	6.6
QP3	30.3	WAV	5.1

Table 10 identifies the relative importance of the predictor variables for the two classification models for Case #2 using the below average binary score as the target variable. A Spearman test of the two variable ranks (XGT and NN) yielded borderline results ($r = 0.69$, $p = 0.058$). It should be noted that the sample size ($N = 8$) is at the low end of the desired minimum sample size. Also notice that the relative weights for the NN drop-off at a considerable higher rate. QA3 and QN3 seem to be the two most influential flexible attributes while WRK appears to be the most dominant stable attribute.

Table 11 Predictor variables relative importance by model (Case #2 – below 70)

XGT		*NN*	
Factor	*Weight*	*Factor*	*Weight*
QA3	100.0	WRK	100.0
UGP	92.7	WAV	90.5
WRK	74.3	QN3	70.4
QN3	70.8	QP3	37.8
WAV	51.3	GND	20.7
GND	48.5	DEL	12.9
DEL	48.4	UGP	4.8
QP3	25.9	QA3	3.1

Table 11 identifies the relative importance of the predictor variables for the two classification models for Case #2 using the below passing binary score as the target variable. A second Spearman test found that the two variable rankings in Table 10 were significantly different ($r = –0.2619$, $p =0.53$).

The classification accuracy results reported above are consisted with similar studies involving student risk assessment (Bayer et al., 2015; Khasanah, 2017; Marbouti, 2016). The first study found that the addition of student social behavioural data (e.g., explicitly expressed friendship and mutual e-mail conversation) to standard student demographics (e.g., age and entrance examination scores) significantly improved the classification performance. The second study, which employed Bayesian networks (BN) and CART, discovered that first semester GPA and attendance were two key factors in identifying which students were likely to drop out of the program. The performance of the BN and CART models, as measured by accuracy, was approximately the same. It should be noted that the overall sample size for this study was less than 200. In the third study, which featured standards-based grading (N = 780), the target variable was passing/failing a first-year engineering course. With a course failing rate of less than 10%, the problems associated with an unbalanced database were present. Gini scores were not reported. Recognising the potential of using machine learning for detecting students at risk represents only the first step in the process. Of equal importance is the design of intervention strategies, which is the subject of the next section.

> "Increasing numbers of at-risk students are going to college with multiple risk factors, including being first-generation college students. This requires educators to think and act differently in achieving their educational mission, to identify high-risk factors, delineate models to address them, and document effective strategies that challenge students in their thinking." (Horton, 2015)

5 Intervention strategies

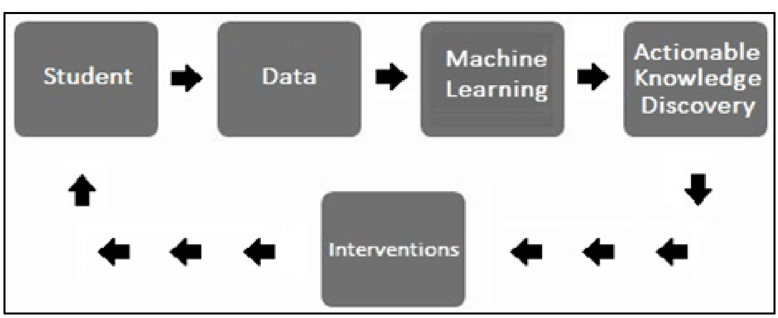
The above example illustrates how machine learning can be used in concert with a variety of data metrics to discover students at academic risk. The next step is to design early-on intervention strategies that can assist these struggling students. To that end, machine learning can also be used to specify customised learning resources and specialised counselling for students that require early interventions (Cox et al., 2017; Sneyers and Witte, 2017). Specifically, machine leaning models can specify an array of interventions based on the characteristics and situation of a given student. For example, if the model has determined that one of the major factors in classifying the student at risk is a lack of engagement, then preventive counselling via either a human or intelligent tutor could be initiated. The evidence suggests that early and frequent interaction with faculty, staff, and peers, clearly communicated academic expectations and requirements, learning opportunities that increase involvement with other students, and social networking can all contribute to enhancing student retention and academic success (Almahaireh et al., 2018; Harvey et al., 2016; Salem et al., 2017; Szlyk, 2018). Specifically, avatar-based counselling on a 24/7 basis was discovered to improve retention, progression and student achievement. Personalised pre-term orientation is another potentially effective intervention strategy, especially for online programs. For first-term students, only admissions and demographic data would be available to pinpoint risk candidates and deploy implementation strategies. However, over subsequent terms, performance data can be added to enhance the precision of the pre-term intervention (Howard and Flora, 2015). Group interventions, which are clearly more manageable for faculty, also have been found to be effective in improving student engagement and performance (Collins et al., 2017). Student choice of instructional activities also increased on-task behaviour compared to the teacher-choice interventions (May, 2018). For example, the student selects a business simulation instead of a MOOC. The basic goal of intervention should be to increase on-task engagement, which in turn should have a positive impact on academic performance. Some reasons these interventions appear so promising include: enables quick engagement of the student in a self-directed way, and facilitates customised content and delivery modalities, for example, an interactive case, which can lead to quicker mastery of the subject material.

Another, perhaps more drastic strategy is to counsel the student to withdraw from one or more of their classes. The data reveals, not surprisingly, that course withdrawal negatively impacts both the time required to graduate and subsequent term retention levels (Akos, 2017; Nicholls and Gaede, 2014).

Machine learning-based tutors represent one promising technology for providing customised learning opportunities for students at risk (Dani, 2016; Sales and Pane, 2019). These systems are designed to provide learners with both customised content and feedback at a performance-driven pace. To be effective, a machine learning-based tutoring system requires a pedagogical framework that is based on the latest learning constructs and that can justify its choice of activities, presentation formats, and decisions. For example, Bayesian knowledge tracing (BKT) models that incorporates partial credit scores, reasoning about multiple attempts to solve problems, and integrating item difficulty were found to improve student performance and engagement compared to more static learning techniques (David et al., 2016). These automated systems are not without their critics (Baker, 2016). Specifically, an automated system cannot recognise when there is a change in context or learning environment. Moreover, automated interventions tend to be brittle and may not be responsive enough to the exact needs and requirements of the student. Nevertheless, machine-driven automated interventions hold considerable potential for providing effective interventions (Anderson and Anderson, 2017).

Figure 4 presents an overview of the proposed machine learning-driven intervention paradigm. The process consists of five integrated elements: student, data (demographic, perspective, performance, organisational and external), machine learning algorithms, post-processing actionable knowledge discovery via decision trees, and implementation of interventions along with follow-up.

Figure 4 Machine learning-driven intervention paradigm



Often, many machine learning exercises end after the development of the detection model with the associated predictor variables. However, as outlined above, specific operational decisions are also needed. This process is technically known as actionable knowledge discovery, which has seen increased use in the customer relationship management field to identify specific actions that would transform a fickle customer to one loyal to the organisation (Subramani and Balasubramaniam, 2016; Yang et al., 2008). The approach, as applied to students, is to develop a decision tree that reports the risk levels, i.e., percentages, for the various relevant categories, e.g., gender, quizzes, work experience. These trees can be used to identify a set of interventions (e.g., specific content) that maximises the chances of transforming a student at risk into the mainstream student body. The actionable knowledge discovery decision tree process requires at least one flexible attribute, i.e., one that can be changed at reasonable cost.

The integrated design, as highlighted in Figure 4, must consist of more than simply ‘attaching’ a series of standalone interventions. Generally, the system must provide a seamless and dynamic interface between the student and the intervention process. Some specific administrative implementation tasks include: training faculty for successful system deployment and usage, providing high-quality and consistent system access, setting specific performance goals and metrics, preparing students for entry and ongoing use, sustaining system operation, and establishing and maintaining the overall learning culture (Jiménez et al., 2018).

The use of intervention technology, which is receiving increased acceptance throughout higher education, represents a key pillar in the deployment of the risk mitigation paradigm (Scherer et al., 2019). The student’s attitude toward and experience with the learning technology represents an important element in the eventual success of an effective intervention (Nakamura et al., 2019).

Implementation of the intervention strategy requires a broader view than merely assessing student performance. Faculty, administrators, and students need to take ownership of the intervention process. Faculty must understand the importance of ‘classroom’ level calibration and the methodologies to achieve this objective. To support this process, faculty should continuously measure student performance and implement customised intervention strategies in a timely fashion. This approach is consistent with Seidman’s paradigm: Student retention = Early identification + Continuous intervention.

> “Early intervention is referring to a broad range of efforts aimed at helping students improve their performance. The early intervention actions range from simple educational programs to systematic social integration strategies, and the performance measurements are also numerous, the most frequently used ones being school dropouts and test results.” (Zhang et al., 2014)

6 Conclusions

This article outlines how machine learning can be used to identify students at risk and to develop student-specific intervention strategies. The proposed risk assessment template, based on the Seidman formula, combines the following three components: dynamic database, machine learning algorithms and specific intervention strategies. The Seidman formula suggests that early identification of students at risk as well as continuous intervention is at the heart of increasing student retention. The acquisition and maintenance of detailed student data (demographic and performance) is a key ingredient to achieve this goal. Among other things, these relationships can be used to detect students at risk so that specific interventions can be implemented, and adjustments made over time. Furthermore, post-processing actionable knowledge discovery via decision trees can identify actional plans that transform a student at risk into one that meets institutional standards. Not only does this process build learning capacity in the student, it is also highly measurable, and feedback is immediate, which is essential when dealing with students at risk.

The results of a machine learning analysis of data from an ongoing MBA program suggest that practice quiz performance provide useful insights into which students are struggling early in the term. Additional predictive factors for identifying students at risk included work experience and undergraduate GPA. The design of effective quizzes is essential since they play a vital role in the proposed risk assessment system. Changing up quiz formats on a regular basis (e.g., from multiple choice to short answers) offers the student a variety of learning contexts, which should both mitigate boredom and enhance overall engagement. The issue as to whether the quizzes should be optional, optional with incentives, or mandatory remains an open question and should be studied further. The assessment database should be expanded to incorporate additional factors, such as student web engagement, group texting, and deadline performance. Machine learning can also be used for developing and deploying specialised content designed around a student's specific needs and challenges. One of the limitations of this study was the relatively small sample size. The acquisition of sample sizes large enough to properly train and test the developed models will continue to be a challenge, even though an institution might have tens of thousands of enrolled students. This situation is due, in part, to the fact that some performance metrics such as texting and meeting deadlines may not be universally available. Nevertheless, additional research should be given on the acquisition of non-quantitative measurements (e.g., tweets, blogs), which could provide additional insights into student risk detection.

Engaging faculty, educational researchers, and administration in the risk mitigation paradigm is essential for ensuring student success. Faculty-driven collaboration networks can help facilitate the adoption of the proposed strategy through access to community best practices. Collaboration networks provide the business education community with the opportunity to converge, share, and exchange ideas to drive innovation in student risk mitigation. Specifically, these networks facilitate the use of machine learning to help ascertain and fill the dynamically changing gaps between student skills and academic standards. The methodology outlined in this article is also applicable to other aspects associated with students at risk, for example, student behavioural and medical issues. The overall goal is to appreciate that every student is unique and that no single learning approach is optimal for every student. To that end, university leadership can also take concrete steps, such as creating student success centres and developing outreach initiatives, to improve retention and on-time graduation.

References

Adejumo, A. and Adetunji, A. (2013) 'Application of ordinal logistic regression in the study of students' performance', *Mathematical Theory and Modeling*, Vol. 3, No. 11, pp.10–19.

Agrawal, H. and Mavani, H. (2015) 'Student performance prediction using machine learning', *International Journal of Engineering Research and Technology*, Vol. 4, No. 3, pp.111–113.

Akos, P. (2017) 'Are course withdrawals a useful student success strategy?', *National Symposium for Student Retention*, Destin, November.

Almahaireh, A., Aldalaeen, A. and Takhaineh, S. (2018) 'Efficacy of a preventive counseling program for improving psychological hardiness and the positive use of social network sites among students at risk', *International Journal for the Advancement of Counselling*, Vol. 40, No. 2, pp.173–186.

Almohammadi, K., Hargra, H., Alagazzawi, D. and Aldabbagh, G. (2017) 'A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms', *Journal of Artificial Intelligence and Soft Computing Research*, Vol. 7, No. 1, pp.47–64.

Amidwar, S. (2017) 'Plagiarism detection using supervised machine learning algorithm', *Journal of Emerging Technologies and Innovative Research*, Vol. 4, No. 6, pp.200–202.

Anderson, T. and Anderson, R. (2017) 'Applications of machine learning to student grade prediction in quantitative business courses', *Global Journal of Business Pedagogy*, Vol. 1, No. 3, pp.13–22.

Asken, O. and Gokalp, F. (2011) 'Comparing the predictive and classification performance of logistic regression and neural networks: a case study on TIMSS', *4th Internal Conference on New Horizons in Education, Procedia – Social and Behavioral Sciences*, Vol. 106, pp.667–676.

Aybek, H. and Okur, M. (2018) 'Predicting achievement with artificial neural networks: the case of Anadolu University open education system', *International Journal of Assessment Tools in Education*, Vol. 5, No. 3, pp.474–490.

Azcona, D., Hsiao, I. and Smelton, A. (2019) 'Detecting students-at-risk in computer programming classes with learning analytics from students' digital footprints', *User Modeling and User-Adapted Interaction*, pp.1–30 [online] https://doi.org/10.1007/s11257-019-09234-7.

Azzi, A., Ramnanan, C., Smith, J., Dionne, É. and Jalali, A. (2014) 'To quiz or not to quiz: formative tests help detect students at risk of failing the clinical anatomy course', *Anatomical Sciences Education*, Vol. 8, No. 5, pp.413–420.

Baker, R. (2016) 'Stupid tutoring systems, intelligent humans', *International Journal Artificial Intelligent Education*, Vol. 26, No. 2, pp.600–614.

Bakhshinategh, B., Zaiane, O., Ealtia, S. and Ipperciel, D. (2018) 'Educational data mining applications and tasks: a survey of the last 10 years', *Education and Information Technologies*, Vol. 23, No. 1, pp.537–553.

Bayer, J., Bydzovsk, H., Geryk, J., Obsıva, T. and Popelinsky, L. (2015) 'Predicting drop-out from social behaviour of students', *Proceedings of the 5th International Conference on Educational Data Mining*, pp.103–109.

Becerra, D., Grob, M., Soto, R., Trico, J. and Sabag, N. (2018) 'Weekly quizzes as a predictive factor of final academic performance in morphology', *International Journal of Morphology*, Vol. 37, No. 1, pp.296–301.

Bethapudi, P., Reddy, B. and Vama, K. (2015) 'Classification of breast cancer using Gini index based fuzzy supervised learning in quest decision tree', *International Journal of Computer Applications*, Vol. 111, No. 14, pp.50–57.

Brown, G., Bice, M., Shaw, B. and Shaw, I. (2015) 'Online quizzes promote inconsistent improvements on in-class test performance in introductory anatomy and physiology', *Advances in Physiology Education*, Vol. 39, No. 2, pp.63–66.

Buskirk, T. and Kolenikov, S. (2015) 'Finding respondents in the forest: a comparison of logistic regression and random forest models for response propensity weighting and stratification', *Survey Insights: Methods from the Field, Weighting: Practical Issues and 'How to' Approach*, pp.1–17.

Chiheb, F., Boumahdi, F. Bouarfa, H. and Boukraa, D. (2017) 'Predicting students' performance using decision trees: case of an Algerian university', *International Conference on Mathematics and Information Technology*, Adrar, December, pp.113–121.

Clark, T., Zhang, Y., Oussena, S. and Kim, H. (2010) 'Use data mining to improve student retention in higher education: a case study', in *ICEIS 2010: Proceedings of the 12th International Conference on Enterprise Information Systems, Databases and Information Systems Integration*, Vol. 1, pp.190–197.

Collins, T., Hawkins, R., Flowers, M., Kalra, H., Richard, J. and Haas, L. (2017) 'Behavior bingo: the effects of a culturally relevant group contingency intervention for students with EBD', *Psychology in the Schools*, Vol. 55, No. 1, pp.63–75.

Cook, B. and Babon, A. (2017) 'Active learning through online quizzes: better learning and less (busy) work', *Journal of Geography in Higher Education*, Vol. 41, No. 1, pp.24–38.

Cox, S., McKinley, J. and Hansen, G. (2017) 'Retention and recruitment: using a predictive analytic model to build and implement a strategic graduation and retention action plan', *Proceedings 13th Annual National Symposium on Student Retention*, pp.325–336.

Dani, A. (2016) 'Students' patterns of interaction with a mathematics intelligent tutor: learning analytics application', *International Journal on Integrating Technology in Education*, Vol. 5, No. 2, pp.1–18.

David, Y., Segal, A. and Gal, Y. (2016) 'Sequencing educational content in classrooms using Bayesian knowledge tracing', *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, Association for Computing Machinery, Inc., pp.354–363.

Day, I., Blankenstein, F., Westenberg, M. and Admiraal, W. (2017) 'A review of the characteristics of intermediate assessment and their relationship with student grades', *Assessment & Evaluation in Higher Education*, Vol. 43, No. 6, pp.908–929.

Delen, D. (2010) 'A comparative analysis of machine learning techniques for student retention management', *Decision Support Systems*, Vol. 49, No. 4, pp.498–506.

Demetriou, C. and Schmitz-Sciborski, A. (2011) 'Integration, motivation, strengths and optimism: Retention theories past, present and future', in Hayes, R. (Ed.): *Proceedings of the 7th National Symposium on Student Retention, 2011*, Charleston, October, pp.300–312.

Depren, S., Askin, O. and Oz, E. (2017) 'Identifying the classification performances of educational data mining methods: a case study for TIMSS', *Educational Sciences: Theory & Practice*, Vol. 17, No. 5, pp.1605–1623.

Dinh, A., Miertschin, S., Young, A. and Mohanty, S. (2019) 'A data-driven approach to predicting diabetes and cardiovascular disease with machine learning', *BMC Med. Informatics and Decision Making*, Vol. 19, No. 211, pp.1–18.

Dobbin, K. and Simon, R. (2006) 'Sample size planning for developing classifiers using high-dimensional DNA microarray data', *Biostatistics*, Vol. 8, No. 1, pp.101–117.

Duarte, E., Ramos, A. and Goncalves, H. (2014) 'Identifying at-risk students in higher education', *Total Quality Management*, Vol. 25, No. 8, pp.944–952.

Eller, C. and DiPrete, T. (2018) 'The paradox of persistence: explaining the black-white gap in bachelor's degree completion', *American Sociological Review*, Vol. 83, No. 6, pp.1171–1214.

Embse, N., Kilgus, S., Iaccarino, S. and Levi-Nielsen, S. (2017) 'Screening for student mental health risk: diagnostic accuracy, measurement invariance, and predictive validity of the social, academic, and emotional behavior risk screener-student rating scale', *School Mental Health*, Vol. 9, No. 3, pp.273–283.

Endert, A., Ribarsky, W., Turkay, C., Wong, W., Nabney, I., Blanco, D. and Rossi, F. (2017) 'The state of the art in integrating machine learning into visual analytics', *Computer Graphics Forum*, Vol. 36, No. 8, pp.458–486.

Fairfield-Sonn, J., Singamsetti, R. and Wahab, M. (2010) 'GMAT and other determinants of GPA in an MBA program', *American Journal of Business Education*, Vol. 3, No. 12, pp.77–86.

Febro, J. and Barbosa, J. (2017) 'Mining student at risk in higher education using predictive models', *Journal of Advances in Technology and Engineering Research*, Vol. 3, No. 4, pp.117–132.

Fish, L. and Wilson, F. (2009) 'Predicting performance of MBA students: comparing the part-time MBA program and the one-year program', *College Student Journal*, Vol. 43, No. 1, pp.145–160.

Grison, S., Luke, S., Shigeto, A. and Watson, P. (2011) 'Using low-stakes repeated testing can improve student learning: how (some) practice makes perfect', *Psychonomic Society Conference*, Seattle, November.

Harvey P., Currie, E. and Augusto, J. (2016) 'Enhancing student support with a virtual assistant', in Vincenti, G., Bucciero, A. and Vaz de Carvalho, C. (Eds.): *E-learning, E-education, and Online Training. eLEOT 2015. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, Springer, Vol. 160, pp.101–109.

Hernández-Blanco, A., Herrera-Flores, B., Tomás, D. and Navarro-Colorado, B. (2019) 'A systematic review of deep learning approaches to educational data mining', *Complexity*, Vol. 1306039, pp.1–22 [online] https://doi.org/10.1155/2019/1306039.

Horton, J. (2015) 'Identifying at-risk factors that affect college student success', *International Journal of Process Education*, Vol. 7, No. 1, pp.83–102.

Howard, J. and Flora, B. (2015) 'A comparison of student retention and first year programs among liberal arts colleges in the mountain south', *Journal of Learning in Higher Education*, Vol. 11, No. 1, pp.67–75.

Indria, V., Vasanthakumari, R., Jegadeeshwaran, R. and Sugumaran, V. (2015) 'Determination of minimum sample size for fault diagnosis of automobile hydraulic brake system using power analysis', *Journal of Engineering Science and Technology*, Vol. 18, No. 1, pp.59–69.

Jantjies, M., Moodley, T. and Maart, R. (2018) 'Experiential learning through virtual and augmented reality in higher education', *Proceedings of the 2018 International Conference on Education Technology Management*, Barcelona, December, pp.42–45.

Jiménez, S., Juárez-Ramírez, R., Castillo, V. and Ramírez-Noriega, A. (2018) 'Integrating affective learning into intelligent tutoring systems', *Universal Access in the Information Society*, Vol. 17, No. 4, pp.679–692.

Kanaris, L., Zeng, O., Kandula, S. and Ngo, L. (2016) 'Sample size determination algorithm for fingerprint-based indoor localization systems', *Computer Networks*, Vol. 101, pp.169–177.

Kaur, B., Gupta, A. and Singla, R. (2019) 'An insight into educational data mining', *International Journal of Computer Sciences and Engineering*, Vol. 7, No. 2, pp.83–90.

Kent, G., Traynor, A., Darmody, K. and O'Leary, T. (2020) 'Enhancing the administration of an early intervention educational programme through technology, exploring usage, acceptance and attitudes', *Irish Journal of Technology Enhanced Learning*, Vol. 5, No. 1, pp.1–13.

Khasanah, A. (2017) 'A comparative study to predict student's performance using educational data mining techniques', *IOP Conference Series: Materials Science and Engineering*, Vol. 215, No. 1, pp.1–7.

Kibble, J. (2007) 'Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance', *Advances in Physiology Education*, Vol. 31, No. 3, pp.253–260.

Korjus, K., Herbart, M. and Vicente, R. (2016) 'An efficient data partitioning to improve classification performance while keeping parameters interpretable', *PLOS One*, Vol. 11, No. 8, pp.1–16.

Kumar, J. (2015) 'A comprehensive study of educational data mining', *International Journal of Electrical Electronics & Computer Science Engineering*, pp.58–63.

Kwan, F. (2018) 'Daily quiz – for engagement … and learning', *Journal of Instructional Pedagogies*, Vol. 21, pp.1–10.

Li, S., Xiaohui, F., Ningtao, C. and Leihong, W. (2013) 'Determination of minimum training sample size for microarray-based cancer outcome prediction–an empirical assessment', *PLOS One*, Vol. 8, No. 7, pp.1–9.

Longadge, R., Dongre, S. and Malik, L. (2013) 'Class imbalance problem in data mining: review', *International Journal of Computer Science and Network*, Vol. 2, No. 1, pp.1–6.

Malau-Aduli, B., O'Connor, T. and Teague, P. (2017) 'Risk factors associated with academic difficulty in an Australian regionally located medical school', *BMC Medical Education*, Vol. 18, No. 1, pp.25–37.

Maldonado, E. and Seehusen, V. (2018) 'Data mining student choices: a new approach to business curriculum planning', *Journal of Education for Business*, Vol. 93, No. 5, pp.196–203.

Malkemes, S. and Phelan, J. (2017) 'Impact of adaptive quizzing as a practice and remediation strategy to prepare for the NCLEX-RN', *Open Journal of Nursing*, Vol. 7, No. 11, pp.1289–1306.

Marbouti, F. (2016) 'Models for early prediction of at-risk students in a course using standards-based grading', *Computers & Education*, Vol. 103, pp.1–15.

Mason, C., Twomey, J., Wright, D. and Whitman, L. (2017) 'Predicting engineering student attrition risk using a probabilistic neural network and comparing results with a backpropagation neural network and logistic regression', *Research in Higher Education*, Vol. 59, No. 3, pp.382–400.

May, M. (2018) 'Effects of differential consequences on choice making in students at risk for academic failure', *Behavior Analysis in Practice*, pp.1–8, Springer-Verlag, New York.

Mesaric, J. and Sebalj, D. (2016) 'Decision trees for predicting the academic success of students', *Croatian Operational Research Review*, Vol. 7, pp.367–388.

Migueis, V., Freita, A., Garcia, P. and Silva, A. (2018) 'Early segmentation of students according to their academic performance: a predictive modelling approach', *Decision Support Systems*, Vol. 115, pp.36–51.

Moekotte, P., Brand-Gruwel, S. and Ritzen, H. (2017) 'Teachers' beliefs about using technology to enhance the learning process of at-risk students', *International Journal of Social Media and Interactive Learning Environments*, Vol. 5, No. 2, pp.108–136.

Nakamura, W., Marques, L., Rivero, L. and Oliverira, E. (2019) 'Are scale-based techniques enough for learners to convey their UX when using a learning management system?', *Brazilian Journal of Computers in Education*, Vol. 27, No. 1, pp.104–131.

Nespereira, C., El-hariri, E., El-Bendary, N. and Vilas, A. (2016) 'Machine learning based classification approach for predicting students performance in blended learning', *Proceedings of the 1st International Conference on Advanced Intelligent System and Informatics*, Beni Suef, November, pp.47–56.

Nicholls, G. and Gaede, R. (2014) 'Exploring the effects of student course withdrawals on time to graduation', *121 ASEE Annual Conference, Proceedings*, Indianapolis, June, No. 9392, pp.1–12.

O'Dowd, I. (2018) 'Using learning analytics to improve online formative quiz engagement', *Irish Journal of Technology Enhanced Learning*, Vol. 3, No. 1, pp.28–41.

Ohsaki, M., Wang, P., Matsuda, K., Katagiri, S., Watanabe, H. and Ralescu, A. (2017) 'Confusion-matrix-based kernel logistic regression for imbalanced data classification', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 29, No. 9, pp.1806–1819.

Paraiso, M., Torres, H., Follosco, R., Curioso, J. and Vinluan, A. (2015) 'Data mining approach for analyzing graduating students' academic performance of new era university bachelor science in computer science', *International Journal of Conceptions on Computing and Information Technology*, Vol. 3, No. 3, pp.308–314.

Park, J. and Yu, H. (2018) 'Recommendations on the sample sizes for multilevel latent class models', *Educational and Psychological Measurement*, Vol. 78, No. 5, pp.737–761.

Pistilli, M. and Arnold, K. (2012) 'Course signals at Purdue: using learning analytics to increase student success', *2nd International Conference on Learning Analytics and Knowledge*, Vancouver, April.

Pratt, W. (2015) 'Predicting MBA Student success and streamlining the admissions process', *Journal of Education for Business*, Vol. 90, No. 5, pp.247–254.

Rifat, R., Imran, A. and Badrudduza, A. (2019) 'Educational performance analytics of undergraduate business students', *International Journal of Modern Education and Computer Science*, Vol. 7, pp.44–53.

Salem, M., Samara, K. and Aldhaheri, M. (2017) 'A secure heuristic framework for online instructor-student advising', *4th Annual Conference on HCT Information Technology Trends*, pp.39–44.

Sales, A. and Pane, J. (2019) 'The role of mastery learning in an intelligent tutoring system: principal stratification on a latent variable', *Annals of Applied Statistics*, Vol. 13, No. 1, pp.420–443.

Samuel, A. (1959) 'Some studies in machine learning using the game of checkers', *IBM Journal of Research and Development*. Vol. 3, No. 3, pp.210–229.

Sarra, A., Fontanella, L. and DiZio, S. (2018) 'Identifying students at risk of academic failure within the educational data mining framework', *Social Indicators Research*, pp.1–20 [online] https://doi.org/10.1007/s11205-018-1901-8.

Scherer, R., Siddiqb, F. and Tondeurc, J. (2019) 'The technology acceptance model (TAM): a meta-analytic structural equation modeling approach to explaining teachers' adoption of digital technology in education', *Computers & Education*, Vol. 128, pp.13–35.

Seidman, A. (1996) 'Retention revisited: RET = E Id + (E + I + C)Iv', *College and University*, Vol. 71, No. 4, pp.18–20.

Shahiri, A. and Rashid, N. (2015) 'A review on predicting student's performance using data mining techniques', *Procedia Computer Science*, Vol. 72, pp.414–422.

Shao, L, Fan, X., Cheng, N., Wu, L. and Cheng, Y. (2013) 'Determination of minimum training sample size for microarray-based cancer outcome prediction – an empirical assessment', *PLOS One*, Vol. 8, No. 7, p.e68579 [online] https://doi.org/10.1371/journal.pone.0068579.

Sharp, J. and Theiler, S. (2018) 'A review of psychological distress among university students: pervasiveness, implications and potential points of intervention', *International Journal for the Advancement of Counselling*, Vol. 40, No. 3, pp.193–212.

Sin, K. and Muthu, L. (2015) 'Application of big data in education data mining and learning analytics – a literature review', *ICTACT Journal on Soft Computing*, special issue on Soft computing models for big data, Vol. 5, No. 4, pp.1035–1049.

Sneyers, E. and Witte, K. (2017) 'Interventions in higher education and their effect on student success: a meta-analysis', *Educational Review*, Vol. 70, No. 2, pp.208–228.

Subramani, S. and Balasubramaniam, S. (2016) 'Post mining of diversified multiple decision trees for actionable knowledge discovery', *17th International Conference on Web Information Systems Engineering*, Shanghai, November, Vol. 10041, pp.553–560.

Szlyk, H. (2018) 'Fostering independence through an academic culture of social responsibility: a grounded theory for engaging at-risk students', *Learning Environments Research*, Vol. 21, No. 2, pp.195–209.

Wanjau, S., Okeyo, G. and Rimiru, R. (2018) 'Data mining model for predicting student enrolment in STEM courses in higher education institutions', *International Journal of Computer Applications Technology and Research*, Vol. 5, No. 11, pp.698–704.

Wickline, V. and Spektor, V. (2011) 'Practice (rather than graded) quizzes, with answers, may increase introductory psychology exam performance', *Teaching of Psychology*, Vol. 38, No. 2, pp.98–101.

Wu, J., Hsiao, Y. and Nian, M. (2018) 'Using supervised machine learning on large-scale online forums to classify course-related Facebook messages in predicting learning achievement within the personal learning environment', *Interactive Learning Environments*,Vol. 28, No. 1, pp.65–80.

Xu, J., Zhang, Y. and Miao, D. (2020) 'Three-way confusion matrix for classification: a measure driven view', *Information Sciences*, Vol. 507, pp.772–794.

Yang, Q., Yin, J., Ling, C. and Pan, R. (2008) 'Extracting actionable knowledge from decision trees', *IEEE Transactions on Knowledge & Data Engineering*, Vol. 19, No. 1, pp.43–56.

Zavorka, S. and Perrett, J. (2014) 'Minimum sample size considerations for two-group linear and quadratic discriminant analysis with rare populations', *Communications in Statistics – Simulation and Computation*, Vol. 43, No. 7, pp.1726–1739.

Zhang, Y., Fei, Q., Quddus, M. and Davis, C. (2014) 'An examination of the impact of early intervention on learning outcomes of at-risk students', *Research in Higher Education Journal*, Vol. 26, pp.1–12.

Zhou, Q., Quan, W., Zhong, Y., Xiao, W., Mou, C. and Wang, Y. (2017) 'Predicting high-risk students using internet access logs', *Knowledge and Information Systems*, Vol. 55, No. 2, pp.393–413.