Predicting the academic success of architecture students by pre-enrolment requirement: using machine-learning techniques

Ralph Olusola Aluko[1], Olumide Afolarin Adenuga[2], Patricia Omega Kukoyi[3], Aliu Adebayo Soyingbe[2] and Joseph Oyewale Oyedeji[4]

[1]*Olabisi Onabanjo University, Nigeria*
[2]*University of Lagos, Nigeria*
[3]*Nelson Mandela Metropolitan University, South Africa*
[4]*Bells University of Technology, Nigeria*

Abstract

In recent years, there has been an increase in the number of applicants seeking admission into architecture programmes. As expected, prior academic performance (also referred to as pre-enrolment requirement) is a major factor considered during the process of selecting applicants. In the present study, machine learning models were used to predict academic success of architecture students based on information provided in prior academic performance. Two modelling techniques, namely K-nearest neighbour (k-NN) and linear discriminant analysis were applied in the study. It was found that K-nearest neighbour (k-NN) outperforms the linear discriminant analysis model in terms of accuracy. In addition, grades obtained in mathematics (at ordinary level examinations) had a significant impact on the academic success of undergraduate architecture students. This paper makes a modest contribution to the ongoing discussion on the relationship between prior academic performance and academic success of undergraduate students by evaluating this proposition. One of the issues that emerges from these findings is that prior academic performance can be used as a predictor of academic success in undergraduate architecture programmes. Overall, the developed k-NN model can serve as a valuable tool during the process of selecting new intakes into undergraduate architecture programmes in Nigeria.

Keywords: Academic achievement, architecture students, classification, k-NN, prior academic performance, selection criteria.

Submission type: Research article

Introduction

Nigerian universities are experiencing an increase in the number of applicants seeking admission. For admission committees this poses serious yearly problems in terms of identifying, selecting and admitting outstanding students into the various departments. The selection of students by such committees is usually based on prescribed criteria for each programme at the different universities. A robust selection process should ensure that the best students are selected for each programme (Holt *et al.*, 2006; Neame *et al.*, 1992; Young, 1989). Furthermore, the previous academic performance of applicants is crucial information needed to aid the selection process. Grades obtained in secondary school terminal examinations, the Unified Tertiary Matriculation Examination (UTME) and post-UTME test scores (the post-UTME test is conducted by each

university) are sources of key information especially in selecting exceptional students into the architecture undergraduate programme.

Conventionally, admission policies are designed to select applicants with high prior performance. This is because it is generally believed that high prior academic performance may be positively related to high academic success. The main challenge faced by many admission screening committees is the identification of the criteria (i.e. prior academic performance) which influence academic success of undergraduate students. Research into the relationship between the two variables (i.e. prior academic performance and academic success) has a long history (e.g. Young, 1989). Although the findings of previous studies (e.g. Abisuga *et al.*, 2015; Allen and Carter, 2007) have shown that a positive relationship exists between these two variables, a significant portion of those studies are focused on 'explanatory' modelling. Shmueli (2010) points out that 'explaining' and 'prediction' are two distinct terms in statistical modelling. Further, it has been argued that predictive modelling is essential for testing theories which leads to generation of new knowledge (Shumeli, 2010; Runeson, 2011). In addition, Coleman (2007) asserted that prediction is a valid approach to test the validity of a theory against alternative theories.

The objectives of this research are to: (1) develop models for predicting academic success of undergraduate architecture students using prior academic performance as predictors; and (2) identify the most important predictors (i.e. academic performance) which have a significant impact on academic success. Two predictive models, namely: the k-NN algorithm and linear discriminant analysis were applied in this study. The need for this investigation is justifiable for several reasons. First, the number of applicants seeking admission into Nigerian universities have significantly increased in recent years. The developed model could be used as a decision support tool during the process of screen applicants. Second, reduction in attrition rates and effective utilization of resources expended on student's training. Government policies targeted at attracting applicants from educationally disadvantaged areas so as to meet the educational needs of such communities. The developed model could assist in identifying 'weak' students, who will require additional learning support. Finally, modelling provides a platform for evaluating alternatives strategies (Ogunlana, Li and Sukhera, 2003) and theories (Shmueli, 2010). The information provided by the model would be useful for developing strategies for identifying the 'best' applicants, who possess the appropriate knowledge for academic success in the undergraduate architecture programme.

Review of literature

A review of past studies that focused on the relationship between prior academic performance and the success of undergraduate students was conducted. The information generated from this review served as a basis for comparing the outcome of the present study to what was found in the past.

Academic success

A considerable amount of literature has been published on the academic success of university students (at undergraduate and postgraduate levels). The term 'academic success' refers to a phenomenon that incorporates academic achievement, attainment of learning objectives, acquisition of desired skills and competencies, satisfaction, persistence and post college performance (York *et al.*, 2015). Academic success has also been viewed as completion of academic activities which improves the academic achievement of the student concerned. Academic success is important in achieving the set objectives of knowledge and skill development during the process of learning. Hence, there is a need to understand the underlying factors that affect students' academic success at the university.

The literature on academic success has highlighted that several factors affect academic success of undergraduate students at university level. This has led to development of several theories found in the field of education, such as pedagogical theory, curriculum theory, and learning theory among others. Empirical studies have investigated the impact of a whole range of factors on the academic performance of undergraduate students in universities (Parker *et al.*, 2004; Young, 1989). It is worth noting that the terms 'achievement' and 'academic success' are used interchangeably in the literature. Herminio (2005) classified two factors affecting academic success, namely: internal and external factors. The internal factors are class schedule, class size, classroom environment, role of the lecturers, technology and nature of examination while the external factors include extracurricular activities, family and work activities. Herminio's (2005) findings show that internal factors are much more significant than the external factors. Further, Ling *et al.* (2010) examined the effect of teaching and learning approaches on academic performance. It was found that the growing teaching approach and the achieving motive learning approach are related to improved academic success. Similarly, studies (e.g. McKenzie and Schweitzer, 2001; Win and Miller, 2005) have demonstrated that prior academic performance has significant impact on the academic success of first-year university students. Although Bone and Reid (2011) contend that numerical assessment scores/grades obtained in a subject may not be a true reflection of the knowledge gained by a student on a topic/subject; a numerical score/grade obviously remains the most available proxy for measuring academic success. The aim of the present study is not to investigate all the factors affecting academic success of undergraduate students. For the purpose of the study, the discussion presented in the next section is limited to the impact of prior academic performance on academic success of undergraduate students.

Prior academic performance and success on undergraduate programmes

Recent studies have examined the relationship between previous academic achievement and the academic success of undergraduate students at universities (see Abisuga *et al.*, 2015; Curtis *et al.*, 2007; Whyte *et al.*, 2011; Shahiri and Husain, 2015). Curtis *et al.* (2007) reports that admission criteria are weak predictors of academic performance for first year students in a dentistry programme. Similarly, Whyte *et al.* (2011) determined that the best predictors of academic success for nursing and paramedics students are index scores and student's maturity in age. Furthermore, a number of studies confirmed that academic entry criteria and age maturity are the best predictors of academic success in undergraduate programmes in both nursing and paramedics (Van Rooyen *et al.*, 2006; Whyte *et al.*, 2011). The studies presented thus far provide evidence that a positive relationship exists between prior academic performance and academic success of undergraduate students. However, it should be noted that other factors influence the impact of this relationship.

The generalisability of much published research on this issue (i.e. the positive relationship between academic performance and academic success for undergraduate students) is problematic. Ting's (2001) findings suggest that a combination of prior academic achievement and psychological variables are significant predictors of academic success on undergraduate engineering programmes. However, Poole *et al.* (2007) affirms that admission criteria are significant predictors of academic performance in the first two years of undergraduate study in dentistry but dwindle in later years. In total contrast, studies such as Bone and Reid (2011) report that no relationship exists between prior academic achievement and academic success on undergraduate programmes in medicine and biology respectively, while Kirby and Dempster (2014) attest that providing accommodation and financial support are vital predictors of academic success on a foundation programme in a South African university. Roberts (2007) concludes that cognitive style might be a better predictor of students' ability to attain academic success irrespective of whether a student had previously studied arts or science. Based on the aforementioned, it is evident that there are contrasting results from studies on the relationship

between prior academic achievement and academic success. However, it is important to gain insights into the strength of these relationships in previous studies focused on built environment disciplines.

Previous studies into the relationship between prior academic performance and academic success for undergraduate programmes in built environment disciplines are relatively few. Allen and Carter (2007) revealed that academic success is largely dependent on grades obtained by students in core-knowledge courses taken at earlier stages of study in a real estate programme. Abisuga *et al.* (2015) reported that admission criteria are weak predictors of academic success for building technology students. In a similar vein, Newell and Mallik (2011) found that prior academic performance in mathematics is significantly related to academic success for property undergraduate degree programme. In view of all the studies that have been reviewed so far, certain inferences can be drawn: (1) a relationship exists between prior academic achievement and academic success across all disciplines. However, the strength of this relationship weakens as students progress to later years; (2) studies have largely focused on 'explaining' the relationship rather than 'prediction' (see Shmueli, 2010 for detailed information on the difference between 'explain' and 'predict'); (3) the preference for linear modelling techniques was evident in previous built environment disciplines (Abisuga *et al.* 2015; Allen and Carter, 2007; Newell and Mallik, 2011). Therefore, the present study investigates the relationship between prior academic performance and student's academic success on an undergraduate architecture programme using machine-learning techniques. The study addresses a gap in the literature by applying machine-learning techniques which possess the ability to capture non-linear relationships present in real life data. Also, the estimated relationship is used for prediction.

Method

Many researchers have examined the determinants of academic success in built environment programmes using various methods. In literature, the research methods utilized include: questionnaire survey (Ling *et al.*, 2010), simulation (Long *et al.*, 2009) and longitudinal survey (Guillermo *et al.*, 2014). Although several research methods have been applied in literature, it is important to note that the suitability of a particular approach to addressing a research problem is a principal factor considered in selecting a research method. The modelling research method offers an effective way of examining the relationship between dependent and independent variables (Fellows and Liu, 2015). The underlying pattern that is uncovered can then be used for prediction. In addition, prediction modelling is useful in theory/hypothesis testing (Shmueli, 2010). Therefore, two machine learning modelling techniques (i.e. k-NN algorithm and linear discriminant analysis) were used to predict academic success of undergraduate students on an architecture programme in a Nigerian university. Linear discriminant analysis was applied because this technique was employed in early studies (e.g. Young, 1989). Also, the k-NN modelling technique has provided useful predictions in different fields of academic endeavours, such as ergonomics (Sánchez *et al.*, 2016); traffic engineering (Yoon and Chang, 2014) and medicine (Zhu *et al.,* 2007) among others. The predictive accuracy (i.e. generalisation) of the developed models were compared. Due to space constraints, readers interested in the application of machine-learning models are referred to Cortez (2015).

Ethical consideration

Permission was obtained from the concerned academic department prior to data collection. In order to comply with ethical requirements, the academic records of each student was anonymised by the administrators in charge of the database. Subsequently, the anonymised data was transferred to the first author. This ensured that no connection could be made with any individual. It should be noted that issues relating to admission criteria and access to student's

entry information are handled by senior academics and top university administrators. In addition, there is an internal process for cross-checking marks assigned for each course taken by each student. This serves as quality assurance in order to preserve objectivity in the metrics used to evaluate students' academic success.

Data

The architecture undergraduate programme at the Olabisi Onabanjo University began in 2003. The curriculum includes studio-based learning strategies and integration of relevant courses from other built environment disciplines. Currently, the admission criteria are designed to provide entry opportunities into the architecture programme for high academic achievers (based on prior academic achievement) and applicants from educationally disadvantaged areas. Further, the undergraduate architecture programme is accredited by the National Universities Commission.

The developed models were trained using data collected from the Department of Design and Architecture, Olabisi Onabanjo University, Ogun State, Nigeria. The collected data contained information on 102 students that completed the undergraduate programme between 2011 and 2014. This provided information relating to prior academic achievement and academic success for each student. Before fitting the collected data to the models, data cleaning was carried out. Owing to some missing information, the data for only 101 students were used during the model development phase of the current study.

Table 1: Class discretisation

Class	Possible value (CGPA Range)
Pass	5.00 – 2.40
Fail	2.39 – 0.00

Table 2: The pre-processed input and output variables

Attribute	Description (Domain)
Grades obtained in 'O' level examinations	
Mathematics (MATH)	student's grade (numeric: from 9 - A1 to 0 - no grade)
English (ENG)	student's grade (numeric: from 9 - A1 to 0 - no grade)
Physics (PHY)	student's grade (numeric: from 9 - A1 to 0 - no grade)
Biology (BIO)	student's grade (numeric: from 9 - A1 to 0 - no grade)
Chemistry (CHEM)	student's grade (numeric: from 9 - A1 to 0 - no grade)
Local language (e.g. Yoruba- YOR)	student's grade (numeric: from 9 - A1 to 0 - no grade)
Geography (GEO)	student's grade (numeric: from 9 - A1 to 0 - no grade)
Technical Drawing/Fine Arts (TD)	student's grade (numeric: from 9 - A1 to 0 - no grade)
Economics (ECON)	student's grade (numeric: from 9 - A1 to 0 - no grade)
Further Mathematics (FM)	student's grade (numeric: from 9 - A1 to 0 - no grade)
Agricultural science (AGRIC)	student's grade (numeric: from 9 - A1 to 0 - no grade)
Total UTME score[a] (JAMB)	Total score (numeric: from 1 to 0)
Direct entry (DE)	Mode of entry (binary: yes-1 or no)
Academic Success	CGPA at graduation (Binary: 1- Pass or 0-Fail)

[a]Total UTME score = student's UTME score divided by 400

The collected data include 13 input (independent) variables, which are measures of prior academic achievement, and one output (dependent) variable. The input variables are the grades obtained in the Ordinary ('O') level examination, total score in the University Matriculation Examination (also called JAMB or UTME) and mode of entry (which is a dummy variable 1= Direct Entry and 0 = JAMB). The collected data are similar to those used in an earlier study (see Young, 1989). For the output variable, the cumulative grade point average (CGPA) obtained by each student upon completion of the architecture undergraduate programme was used as a

measure of academic success. The numerical values of CGPA were transformed into categorical classes. The classification places each student into either of the two classes 'Pass' and 'Fail'. The detail of each class is presented in Table 1. It is important to note that achieving a minimum CGPA of 2.4 is a criterion used in selecting those students that will proceed to the Master of Science (MSc) Architecture programme. Therefore, this criterion is used as a measure of academic success. The details of the input variables and output variables used in developing the models are presented on Table 2.

Machine-learning models

In the present study, two modelling techniques were used to predict students' academic success: linear discriminant analysis and k-nearest neighbour. The machine-learning algorithms (i.e. linear discriminant analysis and k-nearest neighbour) were implemented in *R* programming software (R Core Team, 2015) and *Rminer* R-package (Cortez, 2010). Classic linear regression models are used to examine the relationship between a set of independent variables and a dependent variable. It is imperative to note that the linear regression model is not suitable for predicting a categorical dependent variable (i.e. a classification problem). Hence, the two machine-learning techniques were applied in the present study.

Linear discriminant analysis is a useful method for classifying cases (each student) into one of two groups based on a set of features (i.e. admission criteria for architecture undergraduate students). Each student is assigned to one of the predetermined groups based on CGPA at the end of their undergraduate programme (see Table 1). Discriminant analysis has been shown to be useful for classification problems in several academic disciplines. Typical examples can be found in classification of citrus fruits (Iqbal *et al.*, 2016), screening for elderly drivers (Ferreira *et al.,* 2012) and academic performance of first-year students (Young, 1989). (For additional details on the linear discriminant analysis model, Izenman (2008) may be consulted.)

K-nearest neighbour (k-NN) is a machine-learning technique that has been applied to classification and regression tasks. In this research k-NN is used for classification. According to Parsian (2015), the underlying principle behind the k-NN algorithm is that no prior assumption is made about the function *f*:

$$y = f\left(x_1 + x_2 + ... + x_n\right) \tag{1}$$

Where *y* is a dependent variable and x_i are the independent variables. The function *f* is non-parametric because no parameter is estimated. Given new data sets (i.e. test data), the algorithm dynamically identifies *k* observations in the training data that are similar to *p* (the k nearest neighbour). The neighbours are determined by a similarity measure that is computed between the observations based on independent variables. The Euclidean distance between the independent variables in the training set $\left(x_1, x_2, ..., x_n\right)$ and test set $\left(p_1; p_2, ..., p_n\right)$ can be expressed as:

$$\sqrt{\left(x_1 - p_1\right)^2 + \left(x_2 - p_2\right)^2 + ... + \left(x_n - p_n\right)^2} \tag{2}$$

Parsian (2015) provides a detailed explanation of k-NN.

To validate the developed models, the collected data were divided into two: training set (70%) and test set (30%). The training data set was initially fitted to the model. Subsequently, the trained model was used to predict previously unseen data, i.e. test data set. This is done to evaluate the predictive performance of the model (generalisation capability). The percentage of correctly classified (degree of accuracy) and Cohen's kappa coefficient (Kappa) are computed as measures of predictive performance. In addition, a sensitivity analysis as explained in Cortex (2010) was applied to evaluate the relative importance of the input variable in the developed model.

Results

The processes carried out prior to fitting the model are described in the preceding section. The training data sets were fitted to linear discriminant analysis and the k-nearest neighbour model. The predictive accuracy of all the models are presented in this section.

Linear discriminant analysis

Admission criteria were used in the linear discriminant analysis to predict the academic success of architecture undergraduate students ('Pass' or 'Fail'). The purpose of linear discriminant analysis is to identify the best linear combination of independent variables to produce the best prediction of classifications (Verma, 2013). The linear discriminant analysis generated the coefficients presented in Table 3, which shows that the variables which impact the most on students' academic success are JAMB scores (with an absolute of 4.01), direct entry applicants (with an absolute of 3.36), and grades obtained in mathematics at 'O' level examinations (with an absolute of 0.49).

Table 3: Coefficients of linear discriminants

Variable criteria	Discriminant function coefficients
Mathematics	-0.486
English	-0.253
Physics	0.215
Biology	0.307
Chemistry	0.328
Local language (e.g. Yoruba)	0.015
Geography	-0.091
Technical Drawing	0.107
Economics	-0.074
Further Mathematics	-0.060
Agricultural Science	-0.044
Total UTME score	-4.009
Direct Entry	-3.357

The predictions generated from the linear discriminant analysis are presented in Table 4. The measure is evaluated by comparing the observed misclassification rate to that expected by chance alone. The percentage of cases correctly classified can be regarded as the benchmark for assessing the effectiveness of the discriminant function. In this study, 50.0% of all cases were correctly classified (see Table 3). However, the kappa statistic is -0.230. This indicates poor agreement between the predictive and actual class in the test data set.

Table 4: Linear discriminant analysis classification

Observed	Predicted Fail	Pass
Fail	1	7
Pass	8	14
Overall =50.00%		
Kappa = -0.230		

K-nearest neighbour (k-NN)

The k parameter in the k-NN algorithm is a user-defined parameter. In this study, an initial experiment was carried out by setting k at values of 1, 3, 4, 5, 6, 7 and 9. Subsequently, k was set at 1. This is because no significant improvement in the accuracy of the k-NN model was observed due to changes in the value of k. The results of out-of-sample prediction (i.e. the test

set) for the k-NN model are presented on Table 5. Overall, 73.33% of the instances (cases) were correctly classified. Kappa statistic is calculated as 0.318.

Table 5: *k*-NN classification

Observed	Predicted	
	Fail	Pass
Fail	4	4
Pass	4	18
Overall = 73.33%		
Kappa = 0.318		

Comparison of various models

This section provides a comparative analysis of the predictive performance of the different models developed in this study for predicting the academic success of architecture graduates in a Nigerian university. The predictive accuracy of the developed models served as a basis for comparison. The results of the evaluation test are presented in the previous section (see Tables 4 and 5). It is evident that the k-NN model outperforms the linear discriminant analysis model, in terms of percentage of correctly classified cases and kappa statistic.

Sensitivity analysis

Compared to k-NN, linear models (such as regression) are easy to interpret; hence, machine-learning techniques are also described as 'black box' models. After using the developed k-NN model for prediction, a sensitivity analysis was carried out. Sensitivity analysis is a technique used to extract additional information on the importance of each independent variable in predicting the dependent variable in machine-learning models (see Cortez *et al.*, 2009; Tinoco *et al.*, 2011). Figure 1 shows the importance attributed by k-NN to each input variable (i.e. prior academic performance) based on sensitivity analysis. As can be seen from the figure, it is evident that the most influential input variables are MATH, PHY, and CHEM.
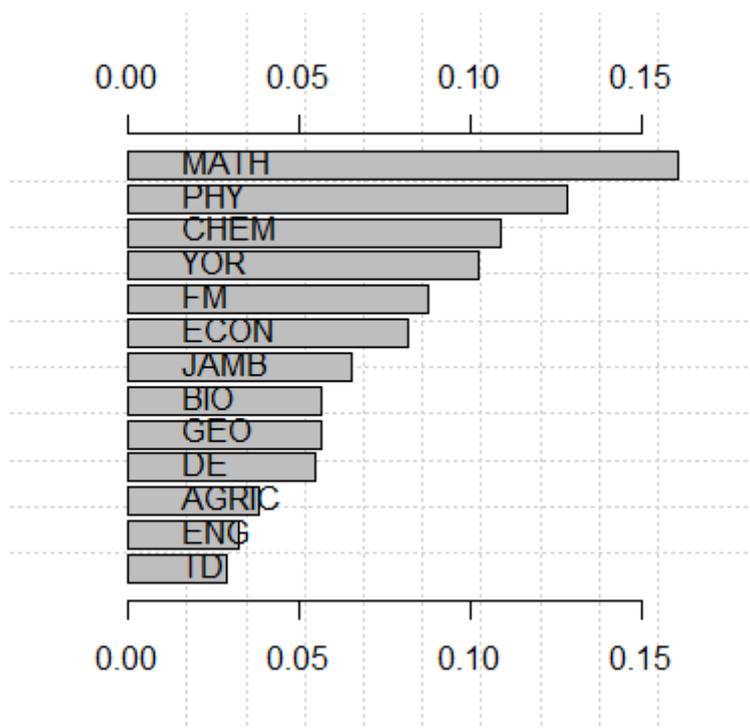

Figure 1: Importance of each input variable in the *k*-NN model

Discussion

The present study investigated the influence of prior academic performance on academic success of architecture undergraduate students. As explained in the literature review, it is evident that a number of factors influence academic success of students. However, it is worth mentioning that the current study was focused on predicting academic success of undergraduate architecture students using prior academic performance as predictors. The results show that the k-NN model is reliable. The overall accuracy of the developed k-NN model is 73.33%. The grades obtained in the following subjects: mathematics, physics, and chemistry grades at 'O' level examinations are significant predictors of academic success of architecture undergraduate students'. Surprisingly, grades obtained in local language were found to have a significant impact on academic success of architecture students. A possible explanation for this might be that lecturers explain complex architecture terms in the local language of the region where the case university is located. A note of caution is due here, since the current investigation did not entail observation of teaching process during lectures.

In this study, it was found that prior academic achievement is a significant predictor of academic success of architecture undergraduate students. This finding is in line with those obtained in previous studies (e.g. van Rooyen *et al.*, 2006; Newell and Mallik, 2011; Abisuga *et al.*, 2015). However, the coefficient of determination found in previous studies (Newell and Mallik, 2011; Abisuga *et al.*, 2015) suggest that the strength of this relationship is weak. This inconsistency may be due to the fact that these studies were based on linear regression which may not be able to capture non-linearity present in real world data. It should be noted that studies, such as Bone and Reid (2011), reported that prior academic performance in a certain combination (i.e. biology and chemistry) of subjects' influence academic performance of students in level-one biology course Molecules Genes and Cells, however, prior knowledge of biology alone had no impact on academic success in the first-year biology course. Also, the current study found that prior academic performance in mathematics had the most significant impact on academic success of architecture students. These results are consistent with those of Newell and Mallik (2011) who reported that mathematics is an important determinant for academic success in a property undergraduate programme. Overall, the findings of the present study suggest that prior academic achievement is a key determinant in academic success in architecture undergraduate degree programmes. Due to the small sample size caution must be applied, as the findings might not be generalizable to other undergraduate programmes without further testing.

In the coming years, it is likely that universities will reduce the academic requirements for gaining admission into academic programmes. This is due to low pass rates of terminal examinations at secondary school level (Kolawole and Dele, 2002; Asikhia, 2010) and the government policy targeted for attracting students from educationally disadvantage backgrounds. The need to lower admission standards to admit these cohorts of students may affect student retention rates. In addition, there is an evident need to address deficiencies in student's prior learning (e.g. mathematics) prior to entering university in order to improve student retention and academic success in architecture undergraduate programmes. This creates an additional burden for stakeholders (e.g. universities, parents, etc.) due to the cost associated with such intervention strategies. Provision of additional tutorial classes for 'weak' students is a typical example of an intervention strategy. Since the purpose of the present study is to investigate the predictability of students' academic success using admission criteria as predictors, it is evident that the developed k-NN model can provide a reliable forecast. This information would be particularly useful for stakeholders involved in making decisions regarding admission of new architecture students into undergraduate programmes at universities.

Conclusion

The present study investigated the impact of prior academic achievement on academic success in an undergraduate architecture degree programme. The study is based on data collected from 101 students who completed a taught undergraduate architecture degree programme at the Olabisi Onabanjo University between 2011 and 2014. In order to evaluate the generalizability of the models, the training dataset was fitted to two models (i.e. Linear discriminant analysis and k-NN) and the developed models were used to forecast the test dataset (i.e. previously unseen data). This guaranteed that valid inferences can be drawn for theory testing.

Based on the results of the current study, it is evident that prior academic achievement is a significant predictor of academic success in this undergraduate architecture programme in Nigeria. In addition, prior academic performance in mathematics, physics, chemistry, and local language are significant determinants of academic success in the undergraduate architecture programme. This is in line with previous studies (Abisuga *et al.*, 2015) which have acknowledged the role of prior academic performance in academic success of undergraduate students.

Fellows and Liu (2015) assert that studies based on prediction, such as the one reported here, are a valid means for testing theories.The findings of this study suggest that there is a need to develop and implement strategies aimed at improving academic performance of students in terminal examinations at secondary school level. This can be achieved by, for example, additional investments in teacher training, motivation of students, and innovative teaching methods. This will ensure that there is significant improvement in the knowledge gained and the academic performance in secondary school subjects, which could substantially improve academic success of students, and impact positively on student retention rates at university level.

There are obvious limitations that affect the generalization of the results of this kind of study. The current study is focused entirely on using prior academic performance as predictors of academic success for architecture undergraduate students. The influence of other factors (such as teaching method, learning style, etc.) on the academic success of architecture undergraduate students is not discussed here, but may have the potential to add to this debate if further research were conducted in this area. Despite this limitation, the application of modeling techniques ensures that valid inferences can be drawn. The present study provides a benchmark against which similar future studies can be evaluated. Therefore, it can be suggested that the developed k-NN model could be a useful decision support tool that can be used by stakeholders in the process of selecting new candidates and identifying those students that need additional support in undergraduate architecture programmes at Nigerian universities.

References

Abisuga, A.O., Olanrewaju D.O. and Oyekanmi O.O. (2015) Pre-Qualification Academic Requirement as a Predictor of Academic Performance in a Building Technology Programme: A Case of Lagos State Polytechnic. *Covenant Journal of Research in the Built Environment,* 3(1), p.44-53.

Allen, M. and Carter, C. (2007) Academic success determinants for undergraduate real estate students. *Journal of Real Estate Practice and Education*, 10(1), p.149-160.

Asikhia, O. A. (2010) Students' and teachers' perception of the causes of poor academic performance in Ogun State secondary schools: Implication for counselling for national development. European Journal of Social Sciences. 13(2): 229—242

Bone, E. K. and Reid, R. J. (2011) Prior learning in biology at high school does not predict performance in the first year at university. *Higher Education Research and Development*, 30(6), p.709-724. doi: https://doi.org/10.1080/07294360.2010.539599

Coleman, S. (2007) Testing theories with qualitative and quantitative predictions. *European Political Science*, 6(2), p.124–133. doi: https://doi.org/10.1057/palgrave.eps.2210122

Cortez, P. (2010) Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool. In: P. Perner, ed., *Advances in Data Mining Applications and Theoretical Aspects*, 10th Industrial Conference on Data Mining, Berlin, Germany, 12-14 July, 2010. Germany: Springer. LNAI 6171. p.572-583. doi: https://doi.org/10.1007/978-3-642-14400-4_44

Cortez, P., Cerdeira, A., Almeida, F., Matos, T., and Reis, J. (2009) Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4), p.547-553. doi: https://doi.org/10.1016/j.dss.2009.05.016

Cortez, P. (2015) *A tutorial on the rminer R package for data mining tasks*, Teaching Report, Department of Information Systems, ALGORITMI Research Centre, Engineering School, University of Minho, Guimaraes, Portugal, July 2015. Available at: <https://repositorium.sdum.uminho.pt/bitstream/1822/36210/1/rminer-tutorial.pdf> [Accessed 9 June 2016].

Curtis, D. A., Lind, S. L., Plesh, O., and Finzen, F. C. (2007) Correlation of admissions criteria with academic performance in dental students. *Journal of Dental Education*, 71(10), p.1314-1321.

Fellows, R. F. and Liu, A. M. (2015) *Research methods for construction.* 4th Edition, John Wiley and Sons, United Kingdom.

Ferreira, I.S., Simoes, M.R., and Maroco, J. (2012). The Addenbrooke's Cognitive Examination Revised as a potential screening test for elderly drivers. Accident Analysis and Prevention, 49:278-286. doi: https://doi.org/10.1016/j.aap.2012.03.036

Guillermo, M. A., Grau, D., Acevedo, S. C. and Rivera, T. E. (2014) How the Design and Assessment of Instruction/Learning Strategies Influence the Academic Performance of Civil Engineering Students at Universidad Industrial De Santander, Colombia. *Journal of Construction Engineering and Management*, 140(4), 10.1061(ASCE) CO.1943-7862.0000709, B4014004. doi: https://doi.org/10.1061/(ASCE)CO.1943-7862.0000709

Herminio, R.P. (2005) Factors Influencing Students' Academic Performance in the First Accounting Course: A Comparative Study between Public and Private Universities in Puerto Rico, Florida. PhD thesis. Argosy University.

Holt, D. T., Bleckmann, C. A., and Zitzmann, C. C. (2006) The graduate record examination and success in an engineering management program: A case study. *Engineering Management Journal*, 18(1), p.10-16. doi: https://doi.org/10.1080/10429247.2006.11431679

Iqbal, S.M., Gopal, A., Sankaranarayanan, P.E., and Nair, A.B. (2016) Classification of Selected Citrus Fruits Based on Color Using Machine Vision System. *International journal of food properties*, 19(2), p.272-288. doi: https://doi.org/10.1080/10942912.2015.1020439

Izenman, A. J. (2008) *Modern Multivariate Statistical Techniques.* London: Springer. doi: https://doi.org/10.1007/978-0-387-78189-1

Kirby, N. F. and Dempster, E. R. (2014) Using decision tree analysis to understand foundation science student performance: Insight gained at one South African university. *International Journal of Science Education*, 36(17), p.2825-2847. doi: https://doi.org/10.1080/09500693.2014.936921

Kolawole, C.O.O., and Dele, A. (2002) An examination of the national policy of language education in Nigeria and its implications for the teaching and learning of the English language. Ibadan Journal of Education Studies, 2(1), p.12-20

Ling, Y. Y., Khai Ng, P., and Leung, M. Y. (2010) Predicting the academic performance of construction engineering students by teaching and learning approaches: Case study. *Journal of Professional Issues in Engineering Education and Practice*, 137(4), p.277-284. doi: https://doi.org/10.1061/(ASCE)EI.1943-5541.0000060

Long, G., Mawdesley, M. J., and Scott, D. (2009) Teaching construction management through games alone: a detailed investigation. *On the Horizon*, 17(4), p.330-344. doi: https://doi.org/10.1108/10748120910998443

McKenzie, K. and Schweitzer, R. (2001) Who succeeds at university? Factors predicting academic performance in first-year Australian university students. *Higher Education Research and Development*, 20(1), p.21–33. doi: https://doi.org/10.1080/07924360120043621

Neame, R. L. B., Powis, D. A., and Bristow, T. (1992) Should medical students be selected only from recent school-leavers who have studied science? *Medical Education*, 26(6), p.433-440. doi: https://doi.org/10.1111/j.1365-2923.1992.tb00202.x

Newell, G. and Mallik, G. (2011) The importance of mathematics background and student performance in a property degree. *Pacific Rim Property Research Journal*, 17(2), p.313-328. doi: https://doi.org/10.1080/14445921.2011.11104330

Ogunlana, S. O., Li, H., and Sukhera, F. A. (2003) System dynamics approach to exploring performance enhancement in a construction organization. Journal of construction engineering and management, 129(5), p.528-536. doi: https://doi.org/10.1061/(ASCE)0733-9364(2003)129:5(528)

Parker, J. D. A., Summerfeldt, L. J., Hogan, M. J., and Majeski, S. A. (2004) Emotional intelligence and academic success: Examining the transition from high school to university. *Personality and Individual Differences*, *36*(1), p. 163-172. doi: https://doi.org/10.1016/S0191-8869(03)00076-X

Parsian, M. (2015) *Data Algorithms: Recipes for Scaling Up with Hadoop and Spark.* O'Reilly Media, Inc.

Poole, A., Catano, V. M., and Cunningham, D.P. (2007) Predicting performance in Canadian dental schools: the new CDA structured interview, a new personality assessment, and the DAT. *Journal of Dental Education*, 71(5), p.664-676.

R Core Team (2015) *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Roberts, A. S. (2007) Predictors of Future Performance in Architectural Design Education. *Educational Psychology,* 27(4), p.447-463. doi: https://doi.org/10.1080/01443410601104361

Runeson, G. (2011) The methodology of building economics research. In: G.D. Valence, ed. 2011. *Modern Construction Economics*, Abingdon: Spon Press. p.191-212.

Sánchez, A. S., Iglesias-Rodríguez, F. J., Fernández, P. R. and de Cos Juez, F. J. (2016) Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders. *International Journal of Industrial Ergonomics*, 52, p.92-99. doi: https://doi.org/10.1016/j.ergon.2015.09.012

Shahiri, A. M. and Husain, W. (2015) A Review on Predicting Student's Performance Using Data Mining Techniques. *Procedia Computer Science*, 72, p.414-422. doi: https://doi.org/10.1016/j.procs.2015.12.157

Shmueli, G. (2010) To explain or to predict? *Statistical Science*, 25(3), p. 289-310. doi: https://doi.org/10.1214/10-STS330

Sohn, D. (1977) Affect-generating powers of effort and ability self-attributions of academic success and failure. *Journal of Educational Psychology*, 69(5), p.500-505. doi: https://doi.org/10.1037/0022-0663.69.5.500

Ting, S.M. R. (2001) Predicting academic success of first-year engineering students from standardized test scores and psychosocial variables. *International Journal of Engineering Education*, 17(1), p.75-80. Available at: http://www.ijee.ie/articles/Vol17-1/Ijee1168.pdf

Tinoco, J., Correia, A. G., and Cortez, P. (2011) Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time. *Construction and Building Materials*, 25(3), p.1257-1262. doi: https://doi.org/10.1016/j.conbuildmat.2010.09.027

van Rooyen, P., Dixon, A., Dixon, G., and Wells, C. (2006) Entry criteria as predictor of performance in an undergraduate nursing degree programme. *Nurse Education Today*, 26(7), p.593-600. doi: https://doi.org/10.1016/j.nedt.2006.02.002

Verma, J.P. (2013) *Data Analysis in Management with SPSS Software.* New Delhi, India: Springer. doi: https://doi.org/10.1007/978-81-322-0786-3

Whyte, D. G., Madigan, V. and Drinkwater, E. J. (2011) Predictors of academic performance of nursing and paramedic students in first-year bioscience. *Nurse Education Today*, 31(8), p.849-854. doi: https://doi.org/10.1016/j.nedt.2010.12.021

Win, R. and Miller, P.W. (2005) The effects of individual and school factors on university students' academic performance. *Australian Economic Review*, 38(1), p.1–18. doi: https://doi.org/10.1111/j.1467-8462.2005.00349.x

Yoon, B. and Chang, H. (2014) Potentialities of data-driven nonparametric regression in urban signalized traffic flow forecasting. *Journal of Transportation Engineering*, 140(7), 04014027. doi: https://doi.org/10.1061/(ASCE)TE.1943-5436.0000662

York, T. T., Gibson, C. and Rankin, S. (2015) Defining and Measuring Academic Success. *Practical Assessment, Research and Evaluation*, 20(5), p.1-20. Available at: http://pareonline.net/getvn.asp?v=20andn=5

Young, A.S. (1989) Pre-enrollment factors and academic performance of first-year science students at a Nigerian university: a multivariate analysis. *Higher Education*, 18(3), p.321-339. doi: https://doi.org/10.1007/BF00138187

Zhu, M., Chen, W., Hirdes, J. P. and Stolee, P. (2007) The K-nearest neighbor algorithm predicted rehabilitation potential better than current Clinical Assessment Protocol. *Journal of clinical epidemiology*, 60(10), p.1015-1021. doi: https://doi.org/10.1016/j.jclinepi.2007.06.001