Graduates' Prediction System Using Artificial Intelligence

Nalina Suresh
School of Computing
University of Namibia
Windhoek, Namibia
nsuresh@unam.na

Valerianus Hashiyana
School of Computing
University of Namibia
Windhoek, Namibia
vhashiyana@unam.na

Gabriel Tuhafeni Nhinda
Informatics
Namibia University of Science and Technology
Windhoek, Namibia
gnhinda@unam.na

Ismael Stephanus
School of Computing
University of Namibia
Windhoek, Namibia
nsuresh@unam.na

Paulus Kautwima
School of Computing
University of Namibia
Windhoek, Namibia
pkautwima@unam.com

ABSTRACT

Graduation rates are an essential metric for universities to gauge the effectiveness of their programmes. Prediction models can be used to assess students' likelihood of completing their degrees, as well as to analyse the rate of completion of programmes that they offer.

Several studies have been done previously using predictive models in various universities around the world. At King Mongkut's University of Technology North Bangkok (KMUTNB), Prachuabsupakij and Wuttikamolchai developed a web application that used the Decision Tree Algorithm to predict student's graduation. Other studies made use of other machine learning algorithms such as support vector machine, neural network and classification and regression tree algorithms. These studies achieved a successful prediction rate of an average of above 70% accuracy.

The University of Namibia does not have any predictive models of any kind and neither has any study of this nature ever been done. It was for this reason that this study was conducted using the School of Computing as the case study. The study would then develop a graduates' prediction system for the school of computing.

The system is a web-based application that requires a user to log in before accessing its features, such as displaying the predicted outcomes. A sample of 500 student data was used to create the student dataset. The prediction was done using four different prediction models which were then comparatively analysed. Those prediction models are Neural Network, Decision Tree, Support Vector Machine and Random forest.

The web-based application consists of an interactive dashboard to allow the user to visualize the prediction results which can be view using different types of charts all to the user's convenience.

This study managed to attain its main object of creating a student graduation prediction system for the school of computing and the sub-objective were also met. The predictions of this study were purely based on the students' academic results in the form of credit scores and other factors were not considered. This study recommends that further research should be done with actual student data from the university's student database records that stretches for over a longer period, say about 5 to 10 years.

CCS CONCEPTS

• Prediction • Artificial Intelligence • Machine Learning • Neural Network • Decision Tree • Support Vector Machine • Random Forest • Confusion Matrix & Misclassification Error

KEYWORDS

Predictions, UNAM, School of Computing, Graduation

ACM Reference format:

Nalina Suresh, Valerianus Hashiyana, Gabriel Tuhafeni Nhinda, Ismael Stephanus and Paulus Kautwima. Graduates' Prediction System Using Artificial Intelligence. *In Proceedings of ACM Data Science, Machine Learning and Artificial Intelligence conference (DSMLAI). Namibia, 9-12 August 2021, Jackson Kaujeua Street, Windhoek, Namibia, 10 pages*. ACM Digital Library. *Association for Computing Machinery, New York, NY, United States (ISBN Number: 978-1-4503-8763-7)*

1 Introduction

A university's success is judged on the number of graduates it produces and likewise a student's academic success is judged by whether the student has graduated or not. Many universities around the world have been using predictive system to try and have a picture of how many graduate students they would produce using

ACM ISBN 978-1-4503-8763-7/20/06…$15.00 https://doi.org/10.1145/3484824.3484873

the available student data that they have gathered over long periods by applying data mining techniques. This study used the University of Namibia's school of computing as it a case study. According to the Dean, Dr Mufeti [1], the school does not have adequate statistical records on the graduation rate of students that have successfully completed any of the programs offered by the school, nor is there any statistically analysed data on its student turnover from students registering as a first year student within the school to those that graduated. This study created an Artificial intelligence system that predicts students' likelihood of graduating from within the school of computing based on records of their current academic performances using machine learning algorithms.

2 Background

Graduation rates are the key markers when discussing accountability in tertiary education. Student data from the University of Namibia is not easily and readily available on the university's website or at student records for making statistical analysis. We live in an era driven by information, this information generated by a prediction system is then used for making important business decisions that can have a big impact on an organization or institution. So, likewise, the graduates' prediction system can be used by the School of Computing for assessing student enrolment vs graduation rate turnover, determine the efficiency of its programmes by the number of graduates it transfers to the job market.

3 Statement of the problem

The school of computing, although relatively small, has been admitting more students to its programmes and this at face value would imply an increase in the number of graduating students. However, there is no way of knowing the number of students that would graduate on timely within the required minimum years for a 4-year degree program. In addition to that, the students cannot adequately gauge their academic advancement towards reaching the milestone of obtaining their qualifications.

4 Objectives

The main objective of this study was to stablish a graduates' prediction system as a novel model for the University of Namibia's School of computing using AI techniques. Further, the sub-objectives are:

To develop a model for prediction
To train the model
To evaluate the outcomes of the model

5 Literature Review

Several studies have reported and provided promising results in the prediction of students who are likely to fail in a given course. In most of these studies, data used for prediction consist of non-academic information; all of which require extra effort to collect [3]. In this section of the study, a review of the literature is presented based on previous works done relating to this study.

5.1 Related Studies

In 2015, Ismail and Abdulla [4] did a study on the Design and Implementation of an Intelligent System to predict student Graduation using the students Accumulated Grad-Point Average (AGPA). They stated that "Since Accumulated Grad-Point Average (AGPA) is crucial in the professional life of students, it is an interesting and challenging problem to create profiles for those students who are likely to graduate with low AGPA. Identifying these kinds of students accurately will enable the university staff to help them improve their ability by providing them with special academic guidance and tutoring. According to Ismail and Abdulla, they used a large and feature rich dataset of marks of high secondary school subjects to develop a data-mining model to classify the newly enrolled students into two groups; "weak students" (i.e., students who are likely to graduate with low AGPA) and "normal students" (i.e., students who are likely to graduate with high AGPA). They investigated the suitability of evolving fuzzy clustering methods to predict the ability of students graduating in five disciplines at Sultan Qaboos University in the Sultanate of Oman. A solid test has been conducted to determine the model quality and validity. The experimental results showed a high level of accuracy, ranging from 71%-84%. This accuracy revealed the suitability of evolving fuzzy clustering methods for predicting the students' AGPA. The objective of this system was to predict student grades, which describe the learning behaviour of students that is useful for helping the management and academic advisors to plan for allocating resources that improve the weak student's academic levels. Ismail and Abdulla however noted that their system had a weakness, in that they used a small size dataset, and the accuracy may have been affected by it.

In their paper, Prachuabsupakij and Wuttikamolchai [5], created an intelligent system to predict student's graduation using data mining techniques from data collected from King Mongkut's University of Technology North Bangkok – KMUTNB. Data contains the student's registration system in the academic year of 2005-2009 (544 instances). Then, a decision tree algorithm is applied to choose the best prediction and analysis to construct the predictive model. Finally, the predictive model is deployed to the web application, which is implemented in the final stages using the Yii framework, PHP, MySQL, and Apache Web server. The experimental results showed that the predictive model correctly predicted with 98.03% accuracy, which confirms that the intelligence system helps students to plan a program of study that will provide the opportunity to graduate in four years. Their results indicated that their system achieved 98.03% accuracy in predicting student's graduation, by making use of the Decision Tree Algorithm.

According to Tahyudin, Utami and Amborowat [6], the percentage of students who graduate on time is one of the elements of accreditation of a study program. The purpose of this research is to compare the several data mining classification algorithms, especially the Decision Tree (DT), Naive Bayes (NB), Artificial

Neural Network (ANN), Support Vector Machine (SVM) and Logistic Regression (LR) algorithms with cross-validation evaluation and T-Test to predict the graduation student on time. The method used is the comparison method. Based on the comparison of performance score and t-test, the SVM algorithm is the appropriate algorithm that was used to predict the student's graduation on time. The level of accuracy to predict the SVM algorithm is high (almost 100% with excellent classification category). But the algorithm which has a higher t-test value than the others is the SVM algorithm. Thus, the SVM algorithm is the best algorithm that can be used to predict the graduation student on time. Tahyudin et al [6] suggested that for further studies, other classification algorithms such as K-Nearest Neighbourhood, ID3, CHAID, etc need to be tried.

Karamouzi and Vrettos [7], created an Artificial Neural Network (ANN) for predicting student graduation outcomes. They used data from students enrolled at Waubonsee College, a community college located in the State of Illinois, for 5 years (Fall of 1997 to Spring of 2002). The network was developed as a three-layered perceptron and was trained using the backpropagation principles. For training and testing various experiments were executed. In these experiments, a sample of 1,407 profiles of students was used. The sample represented students at Waubonsee College, and it was divided into two sets. The first set of 1,100 profiles was used for training and the remaining 307 profiles were used for testing. Two sets of results were obtained after the test was done on the network. The first test was done using a training data set of 1,100 profiles. The mean square error achieved was 0.18 and the network was able to correctly classify 148 out of 172 successful graduates (86.04%) and 633 out of 928 unsuccessful graduates (68.21%). The second test was done using a testing data set of 307 profiles which produced a mean square error of 0.22 on both successful and unsuccessful graduates. The network successfully classified 26 out of 37 successful graduates (70.27%) and 179 out of 280 unsuccessful graduates (66.29%). The high percentage rate of the training data can be because of the small dataset which is sometimes an indicator that the network has been over trained (i.e., the network simply memorizes the training set and is unable to generalize the problem). Karamouzi and Vrettos [7] suggested that follow-up work should include more profiles in the training set to refine the network's performance. This is useful for exploring cross-validation as a method for determining when to stop the network's training to avoid overtraining and improved predictive power.

Azeez, Awe and Omasebi [8] did a study in which the goal was to have early detection of graduating students' cumulative grade point average (CGPA) before they eventually graduate. Classification and regression tree (CART) and linear regression were the algorithms used to carry out the prediction model. Also, a novel algorithm: Difference Level (DL) was designed and incorporated into the system. The system works by taking the differences of each level grade point average and adding the resultant values together; then subtracted from their final year first-semester result to give a predicted graduating cumulative grade point average. Data analysis was performed on datasets of a specific graduated class. The dataset was obtained from Joseph Ayo Babalola University (JABU) Exams and Records unit. This study found that students with a risk of graduating with a low CGPA can be predicted at the end of the final year's first semester. The study aimed to help students in improving their ability in getting a better graduating CGPA.

Beaulac and Rosenthal [9], used and analyzed a large dataset containing every course taken by every undergraduate student in a major University in Canada over 10-years. The study constructed two classifiers using random forests. To begin, the first two semesters of courses completed by a student are used to predict if they will obtain an undergraduate degree. Secondly, for the students that completed a program, their major is predicted using once again the first few courses they have registered to. A classification tree is an intuitive and powerful classifier and building a random forest of trees improves this classifier. Random forests also allow for reliable variable importance measurements. These measures explain what variables are useful to the classifiers and can be used to better understand what is statistically related to the students' situation. The results were two accurate classifiers and a variable importance analysis that provides useful information to university administrations.

5.2 Discussion

A thorough review and analysis of the literature interestingly reveals that even with the different methods and technologies, there is a combined successful prediction rate of 82.74%. from the literature, it has been determined that care should be taken during training of the algorithms with any given dataset, and given the fact that one of the factors that influence the accuracy in predictions by machine learning algorithms accuracy is the size of the dataset, this study has tried to not over train the algorithms because the dataset size was relatively small.

Comparative to previous studies mentioned above that looked at other factors such as demography, financial circumstances etc as attributes in the composition of their datasets, this study however only considered the academic credits that students obtain to make predictions and analysis. Like some of the previous studies, this study also implemented different predictive models such as random forest, support vector machine, decision tree and neural network.

6 Methodology

In this section, the research methods and approached applied in this study are discussed in section 6.1 and 6.2.

6.1 Research Methodology

The research design adopted by this study is applied research. Applied research is the process of quantifying how well we applied the knowledge we have learned from basic science to solving some problem, it is a process to rigorously understand and quantify how effective an engineered system is at solving the problem for which it was designed. Applied research includes designing, implementing, and testing systems [10] and one might say that the goal of the applied scientist is to improve the human condition, rather than to acquire knowledge for knowledge's sake [11].

According to Building trees interactively has proved popular in applied research, and data exploration is based on experts' knowledge about the domain or area under investigation and relies on interactive choices (for how to grow the tree) by such experts to arrive at "good" (valid) models for prediction or predictive classification [12]. Figure 1 shows an overview of the applied research design processes [13].

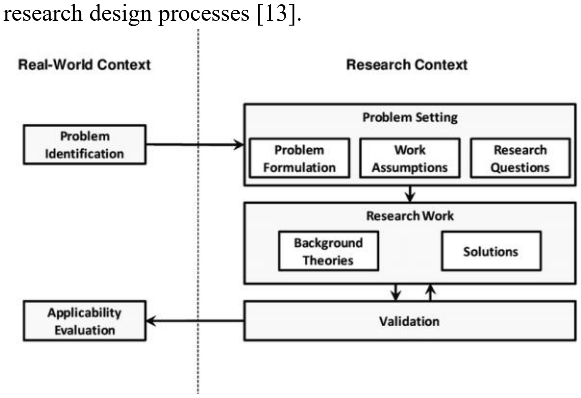
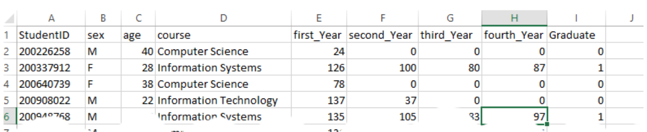


Figure 1: Overview of the applied research process

6.1.1 Data

The data used in the study was primarily in the form of a student dataset. The student dataset was generated using randomly generated attributes in a CSV file. The dataset was made of the following data attributes; 1. StudentID, 2. Sex, 3. Age, 4. Course, 5. The academic level (1st year to 4th year), and 6. The determining factor if the student graduates or not.

Figure **2** shows a snippet of the student dataset that was used.



StudentID	sex	age	course	first_Year	second_Year	third_Year	fourth_Year	Graduate
200226258	M	40	Computer Science	24	0	0	0	0
200337912	F	28	Information Systems	126	100	80	87	1
200640739	F	38	Computer Science	78	0	0	0	0
200908022	M	22	Information Technology	137	37	0	0	0
[illegible]	M	[illegible]	Information Systems	135	105	33	97	1

Figure 2: A snippet of the Student Dataset

The attributes of the student dataset are described in Table 1 below.

Table 1: Dataset attribute description

Attribute	Description	Type	Value
StudentID	Unique student identifier	Numeric	from 9999 - 2019999
Gender	The gender of the student	Binary	Male or Female
Age	The age of the student	Numeric	from 18 -50
Course	The course the student enrolled for	Nominal	CS, IS or IT
first_Year	First year academic credits obtained	Numeric	from 0 to160
second_Year	Second year academic credits obtained	Numeric	from 0 to160
third_Year	Third year academic credits obtained	Numeric	from 0 to160
fourth_year	fourth year academic credits obtained	Numeric	from 0 to160
Graduate	Graduation outcome status	Binary	yes = 1 or no = 0

After the dataset was created and randomly populated, using the inbuilt MS Excel functions (RANDBETWEEN and INDEX) see Figure 3, the data had to be cleaned to make it represent the real world. For example, a student is deemed to have passed his/her first year if they have obtained a minimum of 112 credits, 80 credits to move from the second year to the third year and so forth [14]. So, the dataset had to be cleaned for it to give a fair representation.

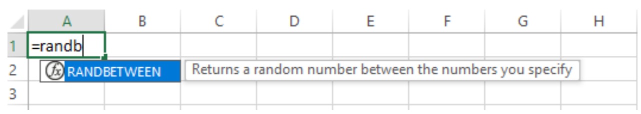


Figure 3: The RANDBETWEEN function in MS Excel

A similar approach was taken to create a database which tables consist of student academic performance for each module offered by the school of computing as shown in

Figure **4**. This data was generated to obtain the best performing student for each module.

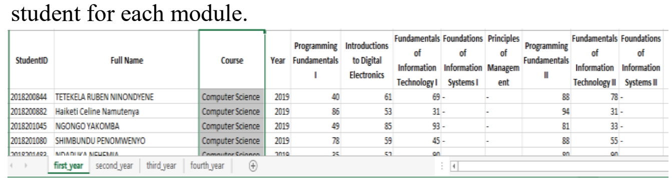


Figure 4: An abstract of Database Tables

6.2 Software Development Methodology

The system for this study was developed using Rapid Application Development (RAD). RAD refers to a development lifecycle designed to give much faster development and higher quality results than the traditional lifecycle. It is designed to take maximum advantage of powerful development software [15]. Rapid application development (RAD) is an agile project management strategy popular in software development [16] and it is good for project such a web-based or mobile applications [17]. Figure 5 illustrates the lifecycle of the Rapid Application Development lifecycle.

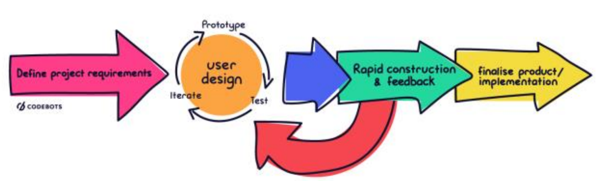


Figure 5: RAD Lifecycle

6.2.1 Design Phases

Phase 1: Requirements gathering

This first phased involved gathering all the requirements needed to develop the prediction system. Data gathering also took place during this stage, where determinations were made as to which data will be useful and which data was to be discarded. A sample of 500 students was used, which comprised of 188 Computer science students, 159 Information Technology students and 153 Information Systems students.

Phase 2: Planning and Prototyping

At this phase, determination as to how the data obtained in the first phase was going to be used. Dataset creation and data cleaning were done during the phase. The tools that would be required to develop the system were also identified. Prototyping and mapping out the system were also done here.

Phase 3: Development

This phase involved the development in small incremental stages. The different prediction algorithms were used to make prediction outcomes and compare the results with each other.

7 Results

In this section, the results obtained from the different prediction models are presented. As mentioned earlier the prediction models used in this study are Neural Network, Decision Tree, Random Forest, and Support Vector Machine. Predictions were done in RStudio. In all the cases the dataset was partitioned in the ratio of 70:30. With 70% being the training dataset and 30% making up the testing dataset.

Neural Network

Artificial Neural Networks are mathematical models that try to simulate the basic actions of the human brain. Information to be processed is passed among neurons based on the structure and synapse weights, hence producing a network behaviour [21]. Three different neural networks were created and compared with each other.

Neural network No. 1

This neural network only consisted of one node in the hidden layer. The generated neural network is shown in

Figure **6** below.

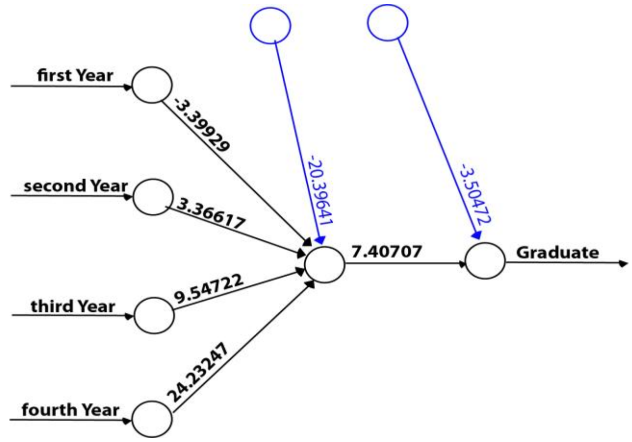


Figure 6: Neural network with one node in the hidden layer

Predictions

After the prediction algorithm was run, the first 6 prediction results were obtained, which predicted the probability of a student to graduate as illustrated in Figure 7 below.

```
> head(output$net.result)
        [,1]
2 0.95232386
3 0.02917817
4 0.02917817
7 0.02917820
8 0.97993752
9 0.97958896
```

Figure 7: First 6 predictions

The predicted values were then compared with the training data (see appendix) and the result of the first result is shown in

Figure **8**. Comparing the two results, it is clear that the predicted outcome that there is approximately 0.95 chance that the outcome is that the student will graduate matches that obtained from the training data that states that the student will graduate.

```
head(training[1,])
StudentID sex age              course first_Year second_Year third_Year fourth_Year Graduate
200337912   F  28 Information Systems  0.7716049   0.7575758      0.625   0.6796875        1
```

Figure 8: Result from training data

• Machine Learning Training Data

The confusion matrix and misclassification error for this neural network using the training data is illustrated in Figure 9. The result is that 138 students did not graduate which the model also correctly predicted; 199 students graduated which the model also correctly predicted. However, there was a misclassification error of 5 students that graduated but the model predicted them not to have graduated. Similarly, 7 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately 0.034.

```
pred1   0   1
    0 138   5
    1   7 199
> 1-sum(diag(tab1))/sum(tab1)
[1] 0.03438395
>
```

Figure 9: Confusion Matrix & Misclassification error – training data Testing Data

The confusion matrix and misclassification error for this neural network using the testing data is illustrated in Figure . The result is that 50 students did not graduate which the model also correctly predicted; 93 students graduated which the model also correctly predicted. However, there was a misclassification error of 3 students that graduated but the model predicted them not to have graduated, similarly, 5 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately 0.052.

```
pred2  0  1
    0 50  3
    1  5 93
> 1-sum(diag(tab2))/sum(tab2)
[1] 0.05298013
>
```

Figure 10: Confusion Matrix & Misclassification error – Testing Data

Neural network No. 2

This neural network only consisted of three nodes in the hidden layer. The generated neural network is shown in Figure 101 below.

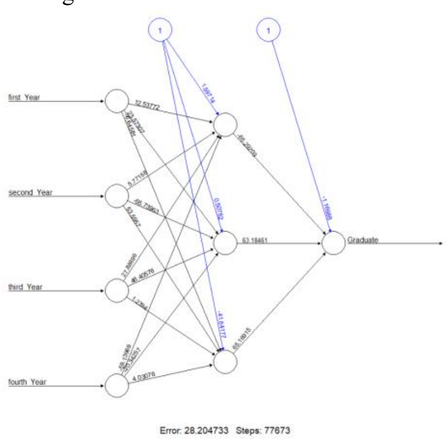


Figure 101: Neural network with three nodes in the hidden layer

Predictions

After the prediction algorithm was run, the first 6 prediction results were obtained, which predicted the probability of a student to graduate as illustrated in Figure 12 below.

```
head(output$net.result)
         [,1]
9.561591e-01
3.755163e-02
2.059491e-02
5.644931e-14
1.000000e+00
1.000000e+00
```

Figure 112: First 6 predictions

The predicted values were then compared with the training data (see appendix) and the result of the first result is shown in Figure . Comparing the two results, it is clear that the predicted outcome that there is approximately a 9.56 chance that the outcome is that the student will graduate matches that obtained from the training data that states that the student will graduate.

```
head(training[1,])
 StudentID sex age             course first_Year second_Year third_Year fourth_Year Graduate
 200337912   F  28 Information Systems  0.7716049   0.7575758      0.625   0.6796875        1
```

Figure 13: Result from Training Data

Confusion Matrix & Misclassification Error

Training Data

The confusion matrix and misclassification error for this neural network using the training data is illustrated in Figure 14. The result is that 142 students did not graduate which the model also correctly predicted; 199 students graduated which the model also correctly predicted. However, there was a misclassification error of 5 students that graduated but the model predicted them not to have graduated, similarly, 3 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately 0.022.

```
pred1   0   1
    0 142   5
    1   3 199
> 1-sum(diag(tab1))/sum(tab1)
[1] 0.02292264
>
```

Figure 14: Confusion Matrix & Misclassification error – training data

Testing Data

The confusion matrix and misclassification error for this neural network using the testing data is illustrated in Figure 15. The result is that 50 students did not graduate which the model also correctly predicted; 93 students graduated which the model also correctly predicted. However, there was a misclassification error of 3 students that graduated but the model predicted them not to have graduated, similarly, 5 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately 0.072.

```
pred2  0  1
    0 51  7
    1  4 89
> 1-sum(diag(tab2))/sum(tab2)
[1] 0.07284768
>
```

Figure 15: Confusion Matrix & Misclassification error – testing data

Neural network No. 3

This neural network only consisted of two hidden layers, the first hidden layer has three nodes, and the second hidden layer has two nodes. The generated neural network is shown in Figure 16 below.

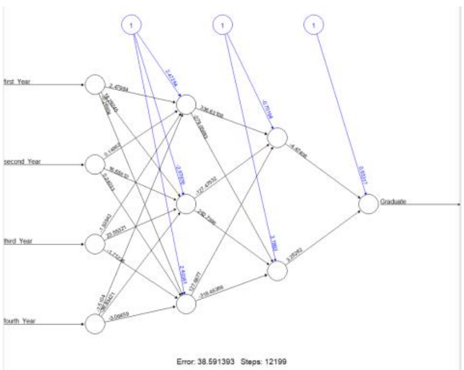


Figure 16: Neural network with two hidden layers

Predictions

After the prediction algorithm was run, the first 6 prediction results were obtained, which predicted the probability of a student to graduate as illustrated in Figure 17 below.

```
head(output$net.result)
      [,1]
0.98501764
0.02817079
0.02817079
0.02817079
0.98501764
0.98501764
```

Figure 17: First 6 predictions

The predicted values were then compared with the training data (see appendix) and the result of the first result is shown in Figure . Comparing the two results, it is clear that the predicted outcome that there is approximately 0.985 chance that the outcome is that the student will graduate matches that obtained from the training data that states that the student will graduate.

```
head(training[1,])
StudentID sex age              course first_Year second_Year third_Year fourth_Year Graduate
200337912   F  28 Information Systems  0.7716049   0.7575758      0.625   0.6796875        1
```

Figure 18: Result from Training Data Confusion Matrix & Misclassification Error

Training Data

The confusion matrix and misclassification error for this neural network using the training data is illustrated in Figure 19. The result is that 142 students did not graduate which the model also correctly predicted; 197 students graduated which the model also correctly predicted. However, there was a misclassification error of 7 students that graduated but the model predicted them not to have graduated, similarly, 3 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately 0.028.

```
pred1   0   1
    0 142   7
    1   3 197
> 1-sum(diag(tab1))/sum(tab1)
[1] 0.0286533
>
```

Figure 19: Confusion Matrix & Misclassification error – training data

Testing Data

The confusion matrix and misclassification error for this neural network using the testing data is illustrated in Figure 20. The result is that 50 students did not graduate which the model also correctly predicted; 92 students graduated which the model also correctly predicted. However, there was a misclassification error of 4 students that graduated but the model predicted them not to have graduated. Similarly, 5 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately 0.059.

```
pred2  0  1
    0 50  4
    1  5 92
> 1-sum(diag(tab2))/sum(tab2)
[1] 0.05960265
>
```

Figure 20: Confusion Matrix & Misclassification error – testing data

Decision Tree

The decision trees were plotted using two different libraries within RStudio. The plot was done using the library known as "party", while the second plot done using the library known as "rpart".

Plotting and Predicting with "party" Library are shown in Figure 21 to Figure 24.

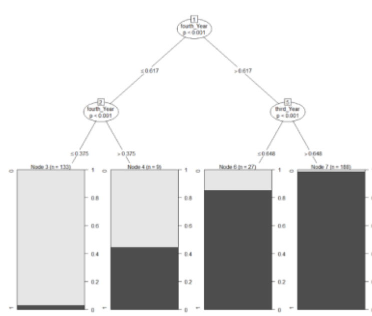


Figure 21: Plot with party library

```
> plot(tree)
> #Predict
> predict(tree, test, type= '
[[1]]
[1] 0.1481481 0.8518519

[[2]]
[1] 0.96992481 0.03007519

[[3]]
[1] 0.1481481 0.8518519

[[4]]
[1] 0.01595745 0.98404255

[[5]]
[1] 0.96992481 0.03007519

[[6]]
[1] 0.01595745 0.98404255

[[7]]
[1] 0.01595745 0.98404255

[[8]]
[1] 0.01595745 0.98404255

[[9]]
[1] 0.01595745 0.98404255
```

Figure 22: Predictions with party library

Plotting and Predicting with "rpart" Library

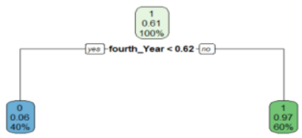


Figure 23: plot with rpart library

```
> predict(tree1, test)
            0          1
2  0.03255814 0.96744186
4  0.94366197 0.05633803
5  0.03255814 0.96744186
8  0.03255814 0.96744186
11 0.94366197 0.05633803
16 0.03255814 0.96744186
20 0.03255814 0.96744186
21 0.03255814 0.96744186
24 0.03255814 0.96744186
26 0.94366197 0.05633803
31 0.03255814 0.96744186
32 0.94366197 0.05633803
34 0.03255814 0.96744186
37 0.94366197 0.05633803
50 0.03255814 0.96744186
53 0.94366197 0.05633803
58 0.03255814 0.96744186
59 0.94366197 0.05633803
65 0.03255814 0.96744186
67 0.03255814 0.96744186
68 0.94366197 0.05633803
69 0.03255814 0.96744186
71 0.03255814 0.96744186
73 0.94366197 0.05633803
```

Figure 24: Prediction with rpart library

Confusion Matrix and Misclassification error

Train data

The confusion matrix and misclassification error for this neural network using the training data is illustrated in Figure 25. The result is that 134 students did not graduate which the model also correctly predicted; 208 students graduated which the model also correctly predicted. However, there was a misclassification error of 8 students that graduated but the model predicted them not to have graduated. Similarly, 7 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately 0.042.

```
    0   1
0 134   8
1   7 208
> 1-sum(diag(tab))/sum(tab)
[1] 0.04201681
>
```

Figure 25: Confusion Matrix & Misclassification error – Train Data

Test data

The confusion matrix and misclassification error for this neural network using the test data is illustrated in Figure **26**26. The result is that 56 students did not graduate which the model also correctly predicted; 82 students graduated which the model also correctly predicted. However, there was a misclassification error of 2 students that graduated but the model predicted them not to have graduated. Similarly, 3 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately 0.0349.

```
testPred  0  1
       0 56  2
       1  3 82
> 1-sum(diag(tab))/sum(tab)
[1] 0.03496503
>
```

Figure 26: Confusion Matrix & Misclassification error – test data

Random Forest

The random forest model was created consisting of 300 trees (ntree) as shown in Figure 27. The default ntree in RStudio is 500. This model has an accuracy rate of approximately 95%.

```
Call:
 randomForest(formula = Graduate ~ ., data = train, ntree = 300)
               Type of random forest: classification
                     Number of trees: 300
No. of variables tried at each split: 2

        OOB estimate of  error rate: 5.32%
Confusion matrix:
    0   1 class.error
0 129  12  0.08510638
1   7 209  0.03240741
>
```

Figure 27: Random forest information

Figure 28 shows the error rate of the Random Forest. When the trees are few, the error rate is high and as the trees increase, the error rate decreases.

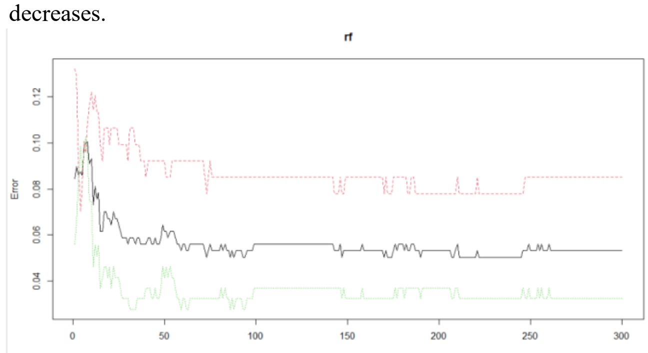


Figure 28: Error rate of Random Forest

The following Figure show a plotted graph ranking variable in order of importance.

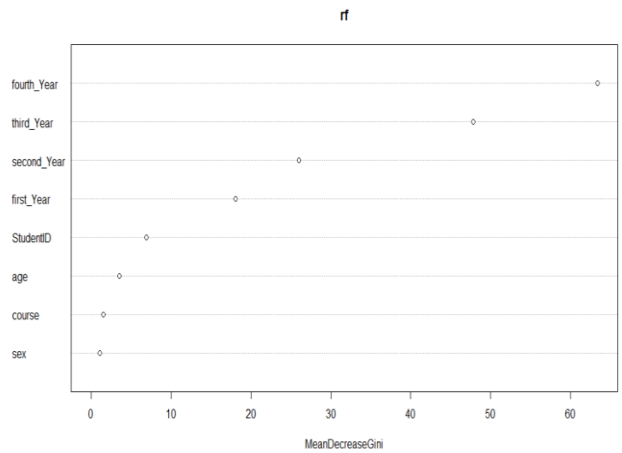


Figure 29: Variable Importance

Confusion Matrix

Training Data

The confusion matrix and misclassification error for this random forest using the training data is illustrated in Figure 30. The result is that 141 students did not graduate which the model also correctly predicted; 215 students graduated which the model also correctly predicted. However, there was a misclassification error of 1 student that graduated but the model predicted them not to have graduated. Similarly, 0 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately less than 1%.

```
          Reference
Prediction   0   1
         0 141   1
         1   0 215

               Accuracy : 0.9972
                 95% CI : (0.9845, 0.9999)
    No Information Rate : 0.605
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9941

 Mcnemar's Test P-Value : 1

            Sensitivity : 1.0000
            Specificity : 0.9954
         Pos Pred Value : 0.9930
         Neg Pred Value : 1.0000
             Prevalence : 0.3950
         Detection Rate : 0.3950
   Detection Prevalence : 0.3978
      Balanced Accuracy : 0.9977

       'Positive' Class : 0

>
```

Figure 30: Confusion Matrix – Training Data

Testing Data

The confusion matrix and misclassification error for this random forest using the testing data is illustrated in Figure **31**31. The result is that 56 students did not graduate which the model also correctly predicted; 82 students graduated which the model also correctly predicted. However, there was a misclassification error of 1 student that graduated but the model predicted them not to have graduated. Similarly, 3 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately 4%.

```
          Reference
Prediction  0  1
         0 56  2
         1  3 82

               Accuracy : 0.965
                 95% CI : (0.9203, 0.9886)
    No Information Rate : 0.5874
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.9277

 Mcnemar's Test P-Value : 1

            Sensitivity : 0.9492
            Specificity : 0.9762
         Pos Pred Value : 0.9655
         Neg Pred Value : 0.9647
             Prevalence : 0.4126
         Detection Rate : 0.3916
   Detection Prevalence : 0.4056
      Balanced Accuracy : 0.9627

       'Positive' Class : 0

>
```

Figure 31: Confusion Matrix – Testing Data

Support Vector Machine (SVM)

The summary for the support vector machine model created is shown in Figure **32**32.

```
> summary(mymodel)

Call:
svm(formula = Graduate ~ ., data = data)


Parameters:
   SVM-Type:  C-classification
 SVM-Kernel:  radial
       cost:  1

Number of Support Vectors:  99

 ( 56 43 )


Number of Classes:  2

Levels:
 0 1
```

Figure 32: summary of SVM

Confusion Matrix

The confusion matrix and misclassification error for this SVN for the prediction is illustrated in Figure 33. The result is that 184 students did not graduate which the model also correctly predicted; 294 students graduated which the model also correctly predicted. However, there was a misclassification error of 6 students that graduated but the model predicted them not to have graduated. Similarly, 16 students did not graduate but the model wrongly predicted them to have graduated. The misclassification error was approximately less than 1%.

```
         Actual
predicted   0   1
        0 184   6
        1  16 294
> 1-sum(diag(tab))/sum(tab)
[1] 0.044
>
```

Figure 33: Prediction Confusion Matrix

8 Discussions

Looking at the results of the three neural networks, the network with the best performing network with the training data was neural network 2 with three nodes in the hidden layer. This layer produced a misclassification error approximately 0.022 of Figure 14, students graduated but the model predicted them not to have graduated and 3 students did not graduate but the model wrongly predicted them to have graduated. On the testing data, neural network 1 with only one node in the hidden layer had the best score during prediction and had misclassified 3 students that graduated as not to have graduated and 5 students that did not graduate but the having them graduated. Overall, the results indicate that on average the best neural network is the one with one node in the hidden layer with a misclassification error of 0.034 for training and 0.052 for testing.

The second model was the decision tree model. This model scored a misclassification error rate of 0.042. The decision tree model's confusion matrix indicates that it wrongly predicted 8 students to have graduated while they did not graduate, and it further wrongly predicted that 7 students did not graduate while they did graduate.
The random forest model had an accuracy of almost 95% which is a very good score. However, its confusion matrix indicates that the model misclassified during prediction that 12 students have graduated while they did not graduate which translated into a misclassification error rate of 0.085 and the model predicted that 7 students did not graduate while they graduated which translated into a misclassification error rate of 0.032.
The Support Vector Machine had a misclassification error rate of 0.044. During prediction, the model's confusion matrix indicates that there was a misclassification of 6 students wrongly classified as having graduated while they did not graduate, and it further misclassified 16 students as not having graduated while they did graduate.
From the results obtained from the four different prediction models, it can be said the best model for this study was the random forest prediction model, this is because since it has a prediction accuracy rate of approximately 95%.

8 Conclusions and Recommendations

In conclusion, this study was motivated by having a system in place where the school of computing can predict and track the graduation rate of students enrolled with the courses offered by the school, which can provide critical information that can help the school ensure retention and completion of the student's studies. The first-year result of the students is the started point for the prediction and this prediction builds on as the student progresses through their academic life, the fourth-year results are ultimately the most important attributes that will determine if a student will be able to graduate or not and this has been proven by the variable importance generated by the random forest model.
This study managed to attain its main object of creating a student graduation prediction system for the school of computing and the sub-objective were also met. The predictions of this study were purely based on the students' academic results in the form of credit scores and other factors were not considered.
This study recommends that further research should be done with actual student data from the university's student database records that stretches for over a longer period, say about 5 to 10 years.

9 Limitations of the study

This study was limited to four prediction models, and all four of them are classification models. There are other more complex models that this study did not look at. Another limitation of the study is that it was done by a complete novice in the field of Artificial Intelligence (AI) and Machine Learning (ML); thus, a great deal of time was spent on learning the concepts and tools.

ACKNOWLEDGMENTS

Thanks, are also reserved to the institutions, School of Computing University of Namibia (UNAM) for making resources available for conducting this study.

REFERENCES

[1] K. Mufeti, "School Overview," [Online]. Available: https://www.unam.edu.na/faculty-of-science/school-of-computing.

[2] E. C. Ploutz, "Machine Learning Applications in Graduation Prediction at the University of Nevada," Las Vegas, 2018.

[3] R. Umar, T. Susnjak, A. Mathrani and S. Suriadi, "On predicting academic performance with process mining in learning analytics," *Journal of Research in Innovative Teaching & Learning,* vol. 10, no. 2, pp. 160-176, 2017.

[4] S. Ismail and S. Abdulla, "Design and implementation of an intelligent system to predict the student graduation AGPA," Australian Educational Computing, 2015.

[5] W. Prachuabsupakij and O. Wuttikamolchai, "An Intelligent System to Predict Student's Graduation," in *The 8th International Conference on Science, Technology and Innovation for Sustainable Well-Being (STISWB VIII)*, 2016.

[6] I. Tahyudin, E. Utami and A. Amborowati, "Comparing classification algorithm of data mining to predict the graduation students on time," in *Information Systems International Conference (ISICO)*, 2013.

[7] S. T. Karamouzis and A. Vrettos, "An artificial neural network for predicting student graduation outcomes," in *Proceedings of the World Congress on Engineering and Computer Science*, 2008.

[8] T. O. Azeez, A. C. Awe and P. A. Omosebi, "Predicting Students' Graduating Cumulative Grade Point Average Using Difference Level, Classification and Regression Tree and Linear Regression Algorithm," 2018.

[9] C. Beaulac and J. S. Rosenthal, "Predicting university Students' academic success and major using random forests," *Research in Higher Education,* vol. 60, no. 7, 2019.

[10] T. W. Edgar and D. O. Manz, Research Methods for Cyber Security, Elsevier Science, 2017.

[11] Lawrence Berkeley National Laboratory, "Basic vs. AppliedResearch," [Online]. Available: https://www.sjsu.edu/people/fred.prochaska/courses/ScWk170/s0/Basic-vs.-Applied-Research.pdf. [Accessed 20 10 2020].

[12] R. Nisbet, G. Miner and K. Yale, Handbook of Statistical Analysis and Data Mining Applications (Second Edition), Academic Press, 2018.

[13] S. Sentilles, Towards Efficient Component-Based Software Development of Distributed Embedded Systems, M¨alardalen University Press Licentiate Theses, 2009.

[14] University of Namibia, Science Faculty Prospectus, 2020.

[15] J. Martin, Rapid application development, New York. Macmillan Pub. Co, 1991.

[16] Lucidchart Content Team, "4 Phases of Rapid Application Development Methodology," Lucidchart, 2020. [Online]. Available: https://www.lucidchart.com/blog/rapid-application-development-methodology. [Accessed 20 11 2020].

[17] Brodie O'Carroll, "App development. What is Rapid Application Development (RAD)?" code bots, 04 February 2020. [Online]. Available: https://codebots.com/app-development/what-is-rapid-application-development-rad. [Accessed 20 November 2020].

[18] SkillsYouNeed, "Ethical Issues in Research," SkillsYouNeed, 2011. [Online]. Available: https://www.skillsyouneed.com/learn/research-ethics.html. [Accessed 06 11 2020].

[19] D. B. RESNĠK, "What is Ethics in Research and Why is it Important," 2008.

[20] K. Beck, Test Driven Development: By Example 1st Edition, 2002.

[21] J. S. Bassi, E. G. Dada, A. A. Hamidu and M. . D. Elijah, "Students Graduation on Time Prediction Model Using Artificial Neural Network," *IOSR Journal of Computer Engineering (IOSR-JCE)* , vol. 21, no. 3, pp. 28-35, 2019.

[22] C. R. Kothari, Research methodology: Methods and Techniques, New Age International, 2004.