AN ARTIFICIAL NEURAL NETWORK DECISION SUPPORT MODEL FOR UNIVERSITY STUDENTS PROGRESSION

ROBERT KASISI

17/04935

A thesis presented to the Faculty of Computing and Information Management for the award of the degree of Master of Science in Information Systems of KCA University

2019

DECLARATION

This thesis is my own work and it has not in part or fully been submitted or presented for award of degree or any other academic work.

Signature…………..…………………..……………. Date …………………………………..

Robert Kasisi -17/04935

This thesis has been submitted for examination with my approval as the appointed university supervisor.

Dr. Simon Mwendia

Signature: …………………………………………… Date: ………………………………

DEDICATION

This thesis is dedicated to my family for their prayers and moral support that they offered me throughout the course of my studies. To my dad and mum may God Almighty give you many more years to see the fruits you planted in your children. To my brother, Muthoka and Samuel, may you continue with this spirit and reach the highest levels of academic studies.

ACKNOWLEDGEMENT

I thank the almighty God for allowing me to undertake this study. My sincere gratitude also goes to my parents for the good upbringing. Your everlasting love has been the greatest treasure of my life. Your value for education has sincerely pushed us as a family to the limits of academic excellence as a family. To my brothers and sisters, thank you for taking your time to remind me of the goals that our parents wanted us to achieve. Special thanks also go to my Supervisor; Dr. Simon Mwendia for his invaluable guidance in writing this thesis. This document is what it is due to his dedicated professional guidance. I therefore do extend my gratitude to him for accepting to invest his knowledge in me. I am also grateful to my classmates who morally encouraged and supported me. I also thank KCAU University and the entire staff for giving me an enabling opportunity to undertake my studies.

Table of Contents

DECLARATION i

DEDICATION ii

ACKNOWLEDGEMENT iii

List of tables vii

List of figures viii

ABSTRACT ix

List of Acronyms and abbreviations x

CHAPTER ONE 1

INTRODUCTION 1

1.1 Background 1

1.2 Statement of the Problem 4

1.3 Objectives of the study 5

1.3.1 General objective 5

1.3.2 Specific objectives 5

1.4 Research questions 5

1.5 Significance of the Study 6

1.5.1 Policy 6

1.5.2 To the researchers 6

1.5.3 To the theory 6

1.6 Motivation of the study 7

1.7 Scope of the study 8

CHAPTER TWO 10

LITERATURE REVIEW 10

2.1 Introduction 10

2.2 Theoretical Literature Review ... 10

2.2.1 Related theories to the progression rate of students ... 12

a) Deferment rate of students ... 12

b) University dropout rate of students ... 13

c) Enrolment rate of students ... 14

2.2.2 Neural networks ... 15

a) The artifical Neuron network ... 21

b) Artificial Neural Network Architecture ... 21

i. The input layer ... 22

ii. The hidden layer ... 23

iii. Output Layer ... 23

2.3 Conceptual framework ... 24

2.3.1 Drop out of university students ... 24

2.3.2 Enrolment Rate of students ... 25

2.3.3 Deferment of students ... 25

2.3.4 Operationalization of Variables ... 26

2.4 Empirical study on progression rate of university students ... 27

2.5 Summary and the Research Gaps ... 28

CHAPTER THREE ... 31

RESEARCH MEHODOLOGY ... 31

3.1 Introduction ... 31

3.2 Research Design ... 31

3.2.1 Data ... 32

3.2.2 Selection of Data ... 33

3.2.3 Data Pre-processing and Transformation ... 34

3.2.4 Data mining 35

3.2.5 Model Evaluation 36

3.3 Ethical Issues 38

CHAPTER FOUR 39

RESULTS AND DISCUSSIONS 39

4.1 Introduction 39

4.1.1 Information on metadata 39

4.1.2 Descriptive analytics 40

4.2 Objective one results 41

4.4 Objective three results 44

4.5 Discussion of the Results 47

CHAPTER FIVE 50

SUMMARY, CONCLUSIONS AND RECOMENDATIONS 50

5.1 Introduction 50

5.2 Summary of findings 50

5.3 Conclusions 51

5.4 Limitation of the study 52

5.5 Recommendations 52

REFERENCES 55

APPENDIX 56

Appendix one Comparison of test and training data performance evaluation 68

List of tables

Table 2.3.1: Operationalization of variables 26

Table 4.1.1 Descriptive metadata 39

Table 4.2.1. Results for objective one 42

Table 4.4.1: Prediction accuracy of the developed model 44

Table 4.4.2. Performance measure of the developed model 45

List of figures

Figure 2.2.1 Reinforcement learning... 16

Figure 2. 2.2 Unsupervised learning (Source: (Koikkalainen, 1994)....................................... 17

Figure 2.2.3 Structure of a Perceptron (Haider & Hanif, 2009)... 21

Figure 2.2.4. Architecture of an Artificial Neural Network (Gosasang et al., 2010)............... 22

Figure 2.2.5: Conceptual Framework... 24

Figure 3.2.1: Model design approach for the study.. 32

Figure 4.1.1 scatter plot of correlation matrix... 41

Figure 4.3.1.Artificial Neural Network model for the study.. 43

ABSTRACT

This study is motivated by the recent developments in the Kenyan education sector. The government has introduced tough measures to curb cheating of K.C.S.E exams thus resulting to decreasing number of students who qualify for university placement. This means that the number of students being admitted to the university has drastically declined. The number of students achieving the minimum entry points to the universities has steadily declined. This is evident from the fact that previously the entry point to JAB programs was B plus and above but currently this has changed to C plus and above. The number of students dropping out of campus has also increased as well as the number of students deferring their studies. It is important to predict the progression rate of students in order to target potential students for early intervention. The main objective of the study was to develop an artificial neural network model for progression rate of university students. The specific objectives of the study were to determine the enrolment rate, dropout rate and deferment rate of students, to develop an appropriate artificial neural network model that uses the identified factors for predicting progression rate and to validate the developed model. Data was obtained from the Technical University of Kenya database system. The data contained information on students enrolled for the 2015 to 2018 period of study. A total of 2976 students were used for the study. The data was split into training and test set and then the artificial neural network model validated using the sigmoid activation function. The progression rate was found to be 78.5%. The study recommends that universities should have intervention programs for students who are at risk of deferment or dropping out of the university.

List of Acronyms and abbreviations

ANN	Artificial neural networks
HED	Higher education institutions
CUEA	Commission of Higher Education Authority
JAB	Joint Admissions Board
TUK	Technical University of Kenya
Pr	Progression
ENR	enrollment
Df	Deferment
Drp	Drop out
Pr	Progression rate
Enrr	enrolment rate
Dfr	Deferment rate
Drr	Dropout rate
Modelcm	model for confusion matrix
ModelTuk	A regression model used for Technical university of Kenya Data
nnet	Neural network package
caret	package used for classification and training of regression algorithms
R gui	R graphical user interface
GGally	Plotting system for the grammar of graphics in R
MSE	Mean square Error
MAE	Mean absolute error
RMSE	Root mean square error
P value	Probability value for testing significance of parameters
Neg pred value	Negative predicted value

CHAPTER ONE

INTRODUCTION

1.1 Background

Student's progression worldwide is a key concern to parents, administrators and governments. Many students enrol in universities with a hope that they will finish within the stipulated time of their courses. However this has not been the case in many instances. Many of these students end up dropping their studies along the way. It is also worth noting that students experience challenges along the journey of academics. This leads to the students deferring their studies for one or more academic years. Some of the challenges faced by the students include lack of adequate financing to carry out their studies smoothly to completion, demographic factors such as distance of the institutions making it costly for the students to be punctual for their studies, job commitments, health issues are some of the challenges that the students face among others.

Many students who enrolled for full time basis of study end up opting to weekend programs or evening classes or at worse defer their studies for some time (Hayden, 2012). In Kenya the emergence of more accessible modules of study in the early 2000 saw many students join weekend programs of study (Julie, 2018). Before the mentioned period, there used to be six major universities. These universities were located out of reach of many students. The increasing demand for study pushed the government through the ministry of education to encourage the existing universities to open up campuses in nearby populated areas. This allowed most of the students to commute to the university premise after work. However demand was still on the rise. This prompted the university management boards to develop online modules of study for those who could not still access education. During this time the number of students graduating with undergraduate degrees increased dramatically. There was also an increase in the number of people graduating with postgraduate degrees. This was deemed healthy for the country as more and more skilled work force was released into the market and the economy performance was on the rise. There was no need to worry about the progression rate of students which could affect the number of people graduating.

During the last five years there has been a turnaround in the numbers of students' in the universities (CUEA, 2018). The numbers are on a downward trend. Today many universities are not able to fill their student capacities. This has been greatly attributed to the number of students qualifying for placement in the university. The government has even lowered the entry point for government sponsored programs to c plus and above which initially used to be B plus and above. The reduced number of students enrolling in the universities has left many universities looking for alternative solutions. Many universities today are encouraging the diploma graduates to enrol and pursue degree programs to keep them afloat.The decreasing number of students has meant decreased income generation for the universities. This has led to closure of some campuses of the universities as they are not able to sustain the operation of such campuses.

Progression rate of students is perceived as a classification problem. Although the rate of enrolment of students, rate of drop out of students and the rate of deferment of students has a positive relationship with the progression rate of students, it is clearly not a linear regression issue as it involves classification which can be best done by artificial neural networks.

Artificial neural networks predictions have shown fairly good predictions in the areas that they have been previously applied such as demand forecasting, inflation forecasting as well as weather forecasting. The study proposed to use the artificial neural network method as a machine learning technique to be able to obtain accurate prediction of the progression rate of students at the universities.

Universities are diverse institutions often viewed as the highest levels of academic excellence. Every nation relies on the universities to provide the much needed intellect to be able to drive their economies. Universities are important source of new ideas, knowledge spill over and they provide human capital. On the other hand universities are often criticized as static and bureaucratic institutions, unable to cope with challenges, changes and exogenous shocks (Sav, 2016). In Sub-Saharan Africa university education plays a critical role in promoting technological advancements and improving the country's ability to capitalize economically. In western countries performance of universities cannot escape taxpayer and political scrutiny especially the public universities. Therefore studies estimating the efficiency of university education remain an important issue in times of financial challenges (Hayden, 2012).

Despite attention over the past decade on performance indicators of student completion and progression in higher education, there has been no corresponding attention paid to developing indicators of the educational process of student progression through a course (Cave et al. 1991;). According to Tinto, (1993) both the characteristics of students entering colleges or universities together with their social experiences at their institution, play a role in progression or deferment and eventual drop out. Deferment rates, drop rate enrolment rate and progress rates are just some of the student outcome performance indicators routinely used by universities and governments. Little attention has been paid however to documenting individual enrolment changes and their timing in the longitudinal process of progression through a degree course. These indicators are justified in a variety of ways. They contribute to public accountability of funding, inform prospective students, provide an information base for government policies and planning, contribute to the development of education services and so on. With current institutional interest in effectiveness and funding issues, this focus on outcomes is not unexpected. It can be argued that the pathways students take through their courses are fundamental to understanding the process of student course progression and an improvement on outcomes alone. The main challenge is that satisfactory indicators of performance need to be assessed longitudinally across a number of years.

1.2 Statement of the Problem

Student progression is a major concern in the education and policy-making communities (Demetriou & SchmitzSciborski, 2011; Tinto, 2006). Globally, about 40% of students seeking bachelor's degrees do not complete their degree within 4 years with universities losing huge revenue each year (Raisman, 2013). The same situation of poor progression rate also applies to Kenyan university students.

Kenyan Universities must therefore formulate strategies to attract larger student enrolments; collaborate with the private sector and development partners so as to be self-sufficient (Ogolla, Bolo, & Muchemi, 2011). Therefore is need to explore ways of reversing and addressing the above challenges through sound responses, to ensure increased progression rates of students and hence higher completion rates.

Machine learning techniques and data mining approaches have been widely used on educational data (Baker & Yacef, 2009; Siemens & Long, 2011; Siemens & Baker, 2012; Romero & Ventura, 2013; Baker & Inventado, 2014). Artificial neural network is one of the techniques that has been previously used. Boddan et al., (2013) used a neural network technology to predict students' results measured by a grade point average in the first year of study. A research by Stamos, (2008) used artificial neural network to predict students' graduation outcome. Additionally, Mohammed Awad (2018) used the same technology to predict the General High School Exam Result Level Using Multilayer Perceptron Neural Networks. The gap is that these studies did not develop an artificial neural network model for progression rate of university students. Therefore the aim of the study is to fill this knowledge gap in a Kenyan context by developing an artificial neural network model for progression rate of students. Here, we model student progression using data gathered from the registrar databases of the universities. The study will inform on status of student progression and to recommend policy interventions to reduce student attrition.

1.3 Objectives of the study

The study sought to answer the following objectives

1.3.1 General objective

The main objective of the study was to develop artificial neural network decision support model for the progression of university students.

1.3.2 Specific objectives

The specific objectives were

i) To investigate and identify factors that influence university students' progression rate.

ii) To develop an appropriate artificial neural network model that uses the identified factors for predicting progression rate.

iii) To validate the developed model.

1.4 Research questions

The study was guided by the following research questions

i) What are the factors that influence university students' progression rate?

ii) What is the appropriate artificial neural network model that uses the identified factors for predicting progression rate?

iii) What is the validated model?

1.5 Significance of the Study

The purpose of this study was to inform key players in the field of academics

1.5.1 Policy

Policy makers need to know the status of the students at the universities. This is in terms of the number of students in the university to enable the plan for financial reasons among others. The government can use information gained from this study to advise the HELB on how much financial assistance students need to avoid the problem of deferment and drop out from the university. Therefore, findings of the study will assist government and Universities management boards in formulating policy framework on the challenges facing student progression.

1.5.2 To the researchers

The study will provide important data that will act as reference point to researches that will use this study as basis for conducting research in the same area to identify gaps in this study.

1.5.3 To the theory

Theoretically, the research study was meant to help in determining and appreciating the rate of progression of students. The will enhance development of intervention programs for increasing progression rate of students. If all educational stakeholders could communicate and organize preventative programs for at-risk students, then the progression rate could increase while economic deficiencies and incarceration numbers decrease. It will also help to ascertain the need, if any, to re-orient the strategic choices of the institutions of higher learning in order to turn them into a sustainable institutions while delivering on its core mandate of offering services to the public.

1.6 Motivation of the study

Ordinarily, students who used to enroll as universities as self-sponsored, were the ones who had got a mean grade of B constant, B- and C+. But in the last two years, students with a mean score of C+ could also get university admissions under government sponsorship. Public universities admissions of students with a means grade of C+ and above has reduced the pool of students admitted in private universities and those pursing parallel degrees. Data from the Kenya National Bureau of Statistics (KNBS) shows a declined university enrollment from 564,507 to 520,893 in 2018 (KNBS, 2019). The number of KCSE candidates scoring the C+ and above grade reduced by 43,614 in 2017. The results reduced the number of students enrolling for parallel degrees (KNBS, 2017). Reports by (KNBS,2019) shows that such statistics predict a decrease of student enrollment by 8.2% and 5% of public and private universities in 2019 and 2020. This is a different case compared to the previous years where public and private university student enrollment has been increasing on expansion by institutions and the student desire for higher education. This a big drop in the enrolment rate of students and poses a serious concern to the government and other key stakeholders.

Progression rate of students is a pressing issue in Kenya. Many courses offered at the universities are lacking quorum due to students' failure to proceed to the next class. This means they become less sustainable and sometimes the whole class is forced to defer to wait for the class behind them so that they can form a quorum. This study is also motivated by the current changes happening in the university education in Kenya. The low number of students qualifying to join the universities means universities do not have enough students. If drop out and deferment rates increase then some programs might have to close. Universities need to seek for more enrolment including transitioning diploma students to degree programs.

The motivation to carry out this study at the university level rather than high school level was motivated by the increasing number of students who fail to proceed to their successive years of study. Such failure to proceed to the next year of study had cost implications to the university and to the students and the entire country as well. First several programs would lack quorum to teach and therefore the students had to choose units for other semester ahead, some of which did not have quorum. This proves to be hard for the students. The administration sometimes would have to incur the cost of teaching very few students especially if it was a graduating class. To the students failure to proceed to their next year of study implied the student would have to spent more time in campus thus attracting extra costs. Eventually the government would lack enough skilled workforces which may lead to production capacity.

1.7 Scope of the study

The study was confined to universities in Kenya and the target respondents for this study comprised of student fraternity. Data was obtained from databases at the registrar office to show the progression of the students. Data for the study was for the period between 2013 and 2018.

1.8 Organization of the Study

The study consists of five chapters. Chapter one presents the background to the study, statement of the problem, motivation of the study, objectives of the study, research questions, significance of the study,and the organization of the study. Chapter two presents a detailed literature review. The chapter contains a review of literature on the concept of progression of students in higher learning institutions, mainly the university. Other concepts discussed herein are deferment rate, dropout rate, enrolment rate, a review of artificial neural network and its architecture. The chapter also presents the empirical literature and conceptual frameworks of the study. Chapter three is the research methodology; this chapter includes the research design, data source, selection of data, data pre-processing, data mining, model evaluation and ethical consideration. Chapter four is about data analysis, presentation and interpretation,. The chapter discusses the results of the three objectives of the study based on the data collected. Chapter five presents the summary of findings, discussions, conclusions and recommendations.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

In this chapter, a review of past and related studies was presented. The aim of reviewing related literature was to form a basis for showing the status of progression rate of students as presented by other studies. Gaps that were left out by other studies were discovered and gave the study a firm foundation to show that it was necessary to undertake this study. The study reviewed literature related to progression rate of students, enrolment rate of students, dropout rate f students and deferment rate for students. The study was carried under the current trends in university education in Kenya and therefore the study believes that valuable information can be gained. In order to build the case for doing this study, several theories relating to the variable mentioned in the study were explained in detail. The study proceeded by discussing the theoretical framework, the conceptual framework and the empirical framework as the subsections of the chapter on literature review.

2.2 Theoretical Literature Review

First, a theory explains a formal and testable explanation of some events that includes explanation of the relationship of things. It is possible to build a theory through reviewing previous findings of related studies (Zikmund etal., 2010). Theories provide indicators and examples of what is incorporated in the framework. In most cases, a theoretical framework guides a research, helping to determine the variables to measure and the statistical relationships to consider in the problem being solved within a research (Trochim, 2006; Tormo, 2006). Thus, the theoretical literature assists the researcher to see the study variables. Student progression is not a widely researched area in the higher education sector, as many studies focus mainly on student dropout. In Tinto's (2006) study, he states that "there has also been a concomitant increase in the number of businesses and consulting firms that have sprung up, each of which claims unique capacity to help institutions increase the graduation rates of their students. Tinto (2006) further states that before the 1970s the reason identified for low student progression rates was the failure on the part of the student and not the institution. Students who dropped out thought to be less able, less motivated, and willing to defer the benefits of college education. Research carried out after 1970s, mostly by alexander Astin (1975, 1984), Ernest Pascarella (1980), and Patrick Terenzini (1980), focused more on the environment, in this case the institutions and the people who govern them, when determining reasons for student progression. With the focus shifting to the institutions, they began to pay more attention to the student faculty relationships mostly outside the classroom and the transition to college, by introducing extended orientation, freshman seminars, and a variety of extracurricular programs for first-year students. These programs were introduced to make them feel welcome to the new culture, community or the new environment in college. Around the 1970s, institutions held the view that students needed to break away from the society that they were in, in order to adapt to college and thereby remain in college. But later, they discovered that the gap between breaking away from society and adapting to college should be bridged through orientation programs and extracurricular programs, making them feel part of their past communities, families, churches or tribes.

Predictive modelling for student progression can be seen as early as 1975 with Tinto's model. Following Tinto's model, there have been many models introduced by researchers that consider different factors and variables to predict student progression. Some of these models focused on identifying students with high risk of dropping out from college.

In most recent years, data mining, which is recognizing patterns in large sets and then understanding those patterns, has been used to study student progression because of high accuracy and the robustness with missing data. as mentioned earlier, in the start of predictive modelling of student progression in 1975, Tinto's model considered social and academic impacts on a student's decision (voluntarily or involuntarily) to drop out from college. The model is based on Durkheim's (1961) theory of suicide, especially its notions of the cost-benefit analysis of individual decisions about investment in alternative educational activities, which comes from the field of economics of education. Tinto (1975) makes the connection with Durkheim's suicide theory by considering the case of dropping out as committing suicide and views college as a social system. In his study, he then relates all the reasons behind committing suicide to that of dropping out of a college when it is viewed as a social system. Furthermore, as colleges also consist of an academic system, he combines the academic factors to the model to be more effective and to shape it to be more suitable to the college structure. These factors are valid even for today's college structure and should be considered as inputs in the predictive models.

2.2.1 Related theories to the progression rate of students

a) Deferment rate of students

Spady was one of the first researchers to propose a widely recognised theory on student retention (Spady,1970). This theory proposes that deferment of students can be explained by looking at the student interaction with the environment within the university.

Attitudes of the students such as intellectual skills and personal interests towards academics are put forward to the expectation versus university demands. The student wishes and hopes may influence the amount of hard work that the students will put towards his/her academics.

In addition the theory surmise that the student may become discouraged and eventually lose interest towards the program they are pursuing thus this may push them to do short term courses that are not offered within the university. In this case the student may be forced to defer their studies in order to finish the sort term courses. Eventually due to lack of interest and not being able to follow up through the study process the student may end up dropping out their studies (Spady, 1970). According to Spady (1970) it is worth noting that related variables which encourage academic and social integration of students pursuing higher education are significant players in this process. The related variables include the family setting of the students, their potential when it comes to academics, their intellectual skills and pressure from peers. The immediate link to these variables is variables like environment within the university and student commitment to academics.

b) University dropout rate of students

Tinto's theory of social and academic integration is the most referred to in the area of student retention. In 1975 Tinto drew upon the work of Spady (1970) who was the first to apply Durkheim's theory of suicide to student retention. This theory is based on the assertion that the likelihood that an individual will commit suicide is predicted by the level of their integration into society (Tinto, 1975). While in Durkheim's model of suicide individuals commit suicide because they are insufficiently integrated into society, Tinto asserts that dropout occurs because students are insufficiently integrated into different aspects of the university. Tinto further contends that dropout could occur through lack of integration in either the academic or the social systems of the university (Tinto 1975). Based on further research, Tinto revised the theory in 1987 by including the three stages of moving from one community to the other. The first stage, separation, refers to the student's parting with one group to join another one. During the second stage, this is transition stage, students deal with the stresses of coping in a new, unfamiliar environment. In the last stage of incorporation students become competent in being members of the new environment (McClanahan 2004; Swail, Redd and Perna 2003). A further revision of this theory in 1993 added other variables influencing social and academic integration of university students. Such variables are adjustment, finances, isolations, learning and external commitments of the students enrolled to the university (Tinto, 1993). Tinto further revised the integration theory in 1997 by focusing on the classroom experience. From this perspective, Tinto asserts that the interaction process that takes place in the classroom determines the social and academic integration of students (Tinto 1997).

c) Enrolment rate of students

In 1980 Bean (1980) developed the psychological theory of student enrolment by asserting that background students characteristics must be considered to understand their integration into a new university environment. According to this theory, Bean (1980) contends that the intentions of students to enrolment into a university are influenced by their attitudes and behaviours. These attitudes and behaviours might affect the extent of student satisfaction within the university. The level of satisfaction might increase the level of commitment to the institution. In 1985, Bean and Metzner developed a theory on non-traditional students (these are older, part-time and commuter students). The attrition of these students is mostly affected by the external environment variables such as family responsibilities, finances and outside encouragements, rather than social integration variables such as university memberships and friends which tend to affect traditional students. In 1995, Eaton and Bean (1995) added coping behaviour as a variable into this theory, stating that students' ability to adapt to the university environment reflects their ability to cope, which is related to previous coping skills in other environments.

2.2 predictive models

This section discusses various models that were used in this project to find the best suited model to predict the progression rate at the Technical university of Kenya.

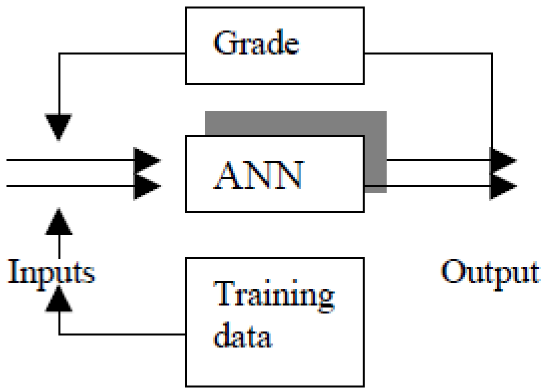
2.2.1 Classification tree

A classification tree points out regulations for splitting data into cohorts. The initial requirement divides the whole data set into some sections; yet again another requirement may be included in a sub-section, various rules to various sections, comprising a second set of data sections. Hence, therefore, a section could either be split or left intact to form a final cohort. The tree shows the initial split into sections as branches resulting from a base and following sections as branches resulting from parts on older branches. The leaves of the tree are the ultimate cohorts, the nodes that are not split. Due to degenerate explanation, trees are usually constructed upside down, emulating an organizational chart. For this model to be significant, the figures in a leaf should coincide regarding another target measure, such that the model represents the dissociation of a mixture of data into specified cohorts.

2.2.2 Neural networks

An artificial neural network consists of many single processors, which interact through a dense web of interconnections. This processor has many names, such as a processing element, a node, a unit, a cell, an artificial neuron, or just a neuron. A neuron has two tasks. It computes one output y, which is sent to the other neurons or outside the network. The neuron determines its output value by applying a transfer function (Freeman and Skapura 1991). Then it updates a local memory that is, weights and other types of data called data variables (Hecht-Nielsen 1990). The neurons are organized into layers. The first layer is called the input layer and the last layer is the output layer. The inner layers, one or more, are known as hidden layers. The input neurons receive input values from outside the ANN's environment, whereas the output neurons send their output values there. A hidden or an output neuron receives input signals from the incoming connections and values from its local memory.

An important and attractive feature of an ANN is its learning capability, which allows the network to adapt to its environment. Learning or training means that an ANN tries to find an appropriate set of weights, which allows the network to carry out the desired task (Rumelhart et al. 1994). An ANN learns from training examples that are provided from the environment. The weights of the network change after every training example. The learning consists of different learning paradigms and algorithms/rules. A priori knowledge of the task and the data influences the selection of learning paradigms when modeling an ANN. A learning paradigm refers to a model of the environment in which an ANN operates (Haykin 1994). The most common learning paradigms are supervised learning, reinforced learning, and unsupervised learning. In supervised learning a teacher has some knowledge of the environment that is unknown to an ANN (Figure 1). The teacher expresses this knowledge with training examples, which consist of input variables together with desired target values (Hecht- Nielsen 1990). The network processes its output values from the input variables and compares them with the target output values. If an error that is a difference between outputs and targets exists, the network adjusts the weights by a small amount in some direction in a step-by-step manner until the error is at an acceptable level. Supervised learning is an instructive feedback system. After the network has been trained, it will be able to deal with the environment alone. One disadvantage of this paradigm is that it cannot learn new strategies without a teacher and new training examples (Haykin 1994).



Source (Koikkalainen, 1994)

Figure 2.2.1 Reinforcement learning

In reinforcement or graded learning the training examples are given to a network without any desired outputs (Figure 2.2.1). In addition to the training data inputs, the network occasionally receives a grade, a performance score, from its environment. This grade tells how well the network has done overall since it was last graded (Hecht- Nielsen 1990). The reinforcement learning is on-line learning without a teacher. This paradigm is an evaluative feedback system, since it evaluates the system's behaviour. However, it does not indicate if an improvement is possible or the way that the system should change its behaviour (Haykin 1994). The reinforcement learning is a special case of supervised learning (Hertz et al. 1991).

In unsupervised learning neither a teacher nor a grade oversees the learning process (Figure 2.2.2). Therefore, the network is given only the training data inputs from which the network organises itself into some useful configuration (Hecht-Nielsen 1990). The input vectors are classified according to their degree of similarity. The similar input vectors activate the same output cluster. The user is responsible for giving an interpretation to the clusters.

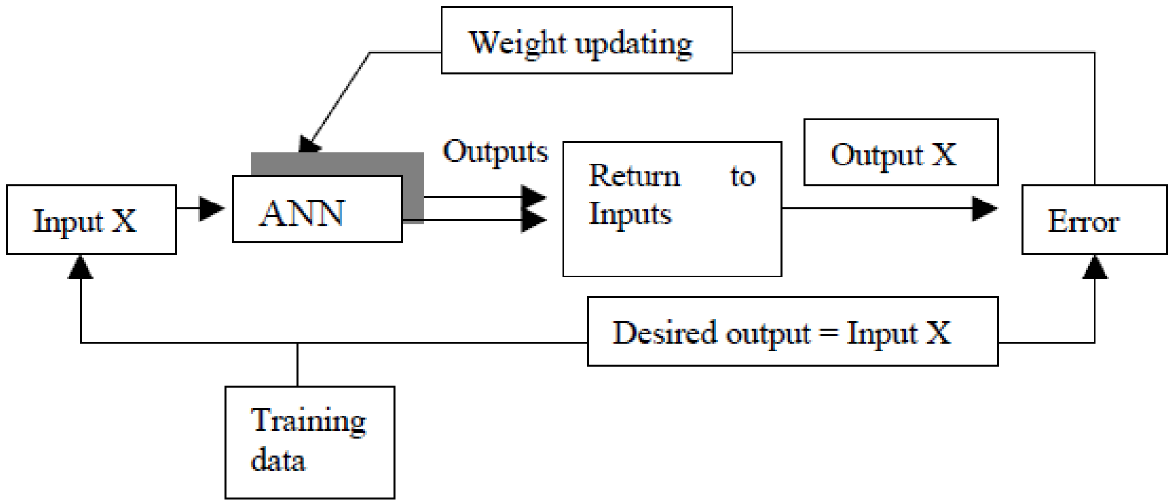


Figure 2. 2.2 Unsupervised learning (Source: (Koikkalainen, 1994)

The hybrid use of learning paradigms may provide a better solution than one paradigm alone. For example, when similar input vectors produce similar outputs, it may be rational to categorise the inputs first with unsupervised learning and use that information for the supervised learning (Hertz et al. 1991). A learning algorithm is a set of well-defined rules for the solution of a learning problem. Several alternative learning algorithms exist and they all have their own advantages. The differences between them are based on various weight adjustments (Haykin, 1994).

Three different learning algorithms, which suit the three different paradigms, are described. Back propagation algorithm has become the most popular one for prediction and classification problems (Sohl and Venkatachalam 1995). This algorithm is used in the supervised learning paradigm (Haykin 1994) and it operates on a multi-layered perceptron network. For a given input vector, it generates the output vector by a forward pass. Then, the difference between the output vector and the desired target vector, the root mean square error (RMSE) is back propagated through the ANN to modify the weights for the entire neural network. Bolzman learning may be described with a Bolzman machine. The neurons in it constitute a recurrent or feedback structure with symmetric weights. It learns its weights in order to determine an appropriate value for them at the stable state (Holmström and Kohonen 1993). This algorithm is suitable for reinforcement learning.

In competitive learning, all the output neurons compete against each other. One of them will be a winner in accordance with a chosen metric and only it will be activated. The winner's weight vector is updated to correspond more closely to the input vectors. This algorithm can discover features that may be used to classify a set of input vectors. This algorithm is suitable for unsupervised learning (Haykin 1994).

Neural networks were developed to mimic human reasoning, hence the term neural. In practice and especially in the area of machine learning neural networks form a brain chain of the analysis in the study. A neural network as presented in this study is composed of a sequence of layers that pass information for processing and updating from one layer to another. The information if validated through a learning algorithm. These layers of the neural network include the input layer, the hidden layer and output layer. In a nutshell the input layer takes the predictor variables, the hidden layer performs complex computations and learns the algorithm for updating the network to give accurate results and the output layer gives the computed result. A neural network can be divided into two major types, the Biological neural network and the artificial neural network. The study uses the artificial neural network to perform analysis necessary for the study. Allen (1997) notes that several other studies have been conducted in attempt to predict enrolment rate of students. Some of these studies have featured in the world conferences on statistical approaches. However these studies did not use the neural network approach to model the progression rates. Some of the notable studies that this study came across used multiple linear regression approach Scheckley (2001), the method of path regression (Tinto, 2000), and the method of discriminant analysis (Roweton, 1994). These methodologies produced varying results in terms of producing previously undetected patterns of progression. It is notable that the accuracy of these approaches is less than that produced when using artificial neural networks. Byers (2002) noted that ANNs produced more accurate results in a study that they sued to predict students' attrition in a high school perspective.

Artificial neural networks produced more accurate results in a study conducted by Barker et al,.(2004). Their study which used 59 variables as inputs included demographic characteristics, academic desire and information regarding attitude of the student achieved a prediction accuracy of 63.4% when using test data and a prediction accuracy of 67.5% when using training set data. Other areas that artificial neural networks have been applied with success are in predicting customer churn behaviour (Tsai, et al(2015). Huang et al., used machine learning to forecast customer churn in telecommunication and finance area of study. Their study showed that ANNs achieved fairly good results. Artificial Neural networks (ANNs) have been widely acclaimed to solve many classification and decision making challenges by easily modelling parametric and non-parametric processes. In addition, these ANNs are able to capture non-linear and noisy data by transforming the input

(Bahrammirzaee, 2010). ANNs are easy to integrate with information systems, can learn automatically how to perform prediction and decision making without human intervention. ANN therefore are practical since they are precise and are able to capture any data movements with a high degree of accuracy making them suitable for use in predicting progression rate of university students (Gupta & Kashyap, 2015;Somaratna, Arunatilaka, & Premarathna, 2010)

ANNs are also attractive for use because of their robustness and ease of adapting their performance to any changing characteristic of the modelled system, fast processing speed and ease of maintenance (Karlik & Olgac, 2010).Bahrammirzaee (2010) concludes that ANN models produce better and more accurate outputs.

In order to model student progression rates the ANN was adopted approach for this study. ANN consists of input, hidden and output layers. The input layer acts as the independent variable. In this study the enrolment rate, deferment rate and rate of school dropout are the inputs. Information is processed within a hidden layer. For this case, the progression will be the result of the output layer. Neural networks were developed from research efforts to create systems with cognitive abilities inspired by the information processing structure of the human nervous system. The biological neuron learns by adjusting to the synaptic connections between the millions of neurons (Hyndman & Athanasopoulos, 2012). The neurons within the output and the input layers depend on the input and the outputs of the problem. The quantity of units in the hidden layer on the other hand depends on the problem complexity or the concept to be represented and solved (Gosasang et al.,2010). The weights received from the input layer and the adjusted weights are critical in determining the solutions final output value. The ANN input layer receives input in the form of a data vector. This data vector is then weighted and passed through the connections to the hidden layer.

For each hidden neuron the incoming products of the weights and the input values are summed up and passed on to an activation function that generates an output value. This output becomes input for the next layer and so on depending on the quantity of hidden layers within the ANN (Stahl&Jordanov, 2012).

There are two types of neural networks; the biological neural and artificial network. This study concentrates on neural network.

a) The artifical Neuron network

The artificial neuron known as a perceptron is modelled after the biological neuron with inputs or an input layer which are the the dendrites represrented as an approximated matematical summation. The output axons that connect to other neurons are represented as the output of one neuron (figure 3). The inputs are the study variables, the deferment rate, the enrolment rate, and the rate of university drop out of the students.The artificial neural network assigns weights to the iputs upon receiving them(Kumar Ram, & Hanmanthu, 2014. They are then summed upat the transfier function. This gives thenet input which is then subjected to the activation fucntion for final processing at the hidden layer.

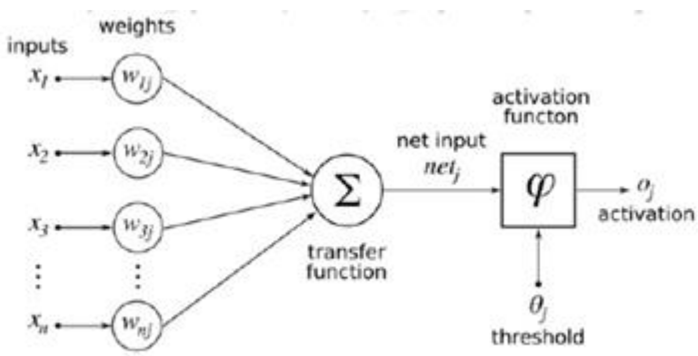


Figure 2.2.3 Structure of a Perceptron (Haider & Hanif, 2009)

b) Artificial Neural Network Architecture

The architecture of a neural network is composed of various layers as depicted in figure 2.2.4.

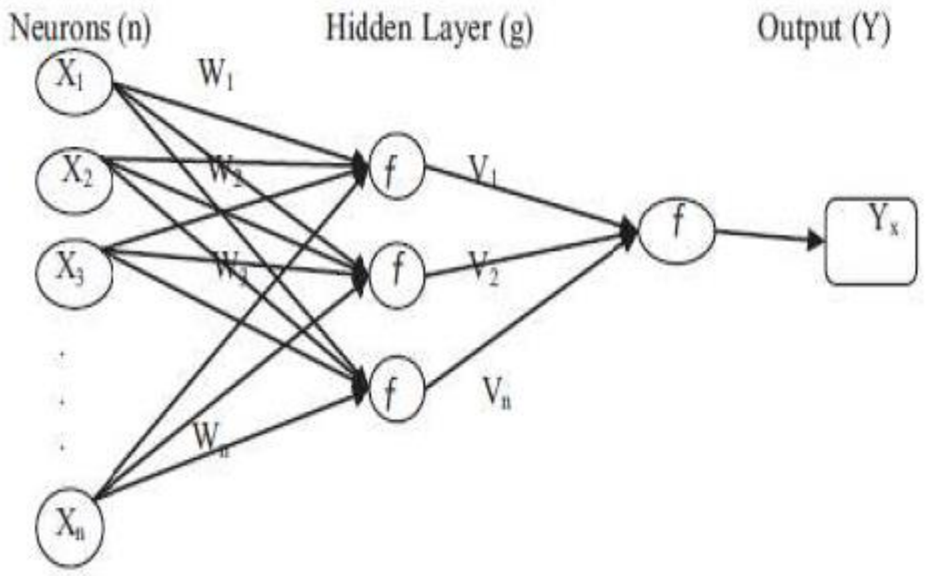


Figure 2.2.4. Architecture of an Artificial Neural Network (Gosasang et al., 2010)

The architecture points to the composition of perceptrons and the kind of links permitted (Choudhary & Haider, 2011).

Three layers define the structure of artificial neural networks. Theses layers are the input layer, the hidden layer and the output layer. The first layer of interest is the input layer which takes in the input variables also referred to as the independent variable.

i. The input layer

The input layer is then linked to the hidden layer which performs complex computations on the entered input variables. The neural network assigns some weights to the inputs entered into the network. The weights are usually for modifying the network during the learning process. These weights range from -1 to 1. A constant adjustment is made on the weights as the network undergoes learning process.

ii. The hidden layer

Little is known on what happens in the hidden layer. For this reason it is sometimes known as the black box. Summations of the weights is performed and parameterized and then used within an activation function.

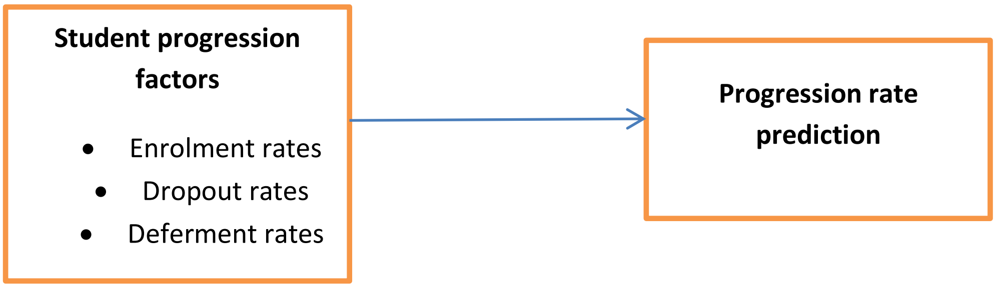
iii. Output Layer

The processed information is displayed in the output layer. In other words the output layer composes the dependent variable.

The information in the layers is processed and moved from inputs through the black box and the assigned a single or multiple Boolean values 1 or 0. The final results from the summation and activation layer give the value of the output (Haider & Hanif, 2009).

2.3 Conceptual framework

Firstly, conceptual framework explains a brief explanation of the variables being studied. The diagram illustrates the relationship amongst the study variables of the study as shown in Figure 2.2.5. The outline shows how the ANN neural network was used.



Independent Variables **Dependent variable**

Figure 2.2.5: Conceptual Framework

2.3.1 Drop out of university students

Student drop out is a major issue (Lau, 2003; Talbert, 2012). A research by Schmitz, (1993) indicates that dropout rates are less at vocational schools than at HED institutions. Students taking short-lived courses to have a reduced risk of school drop-out (Murtaugh, 1999). Despite the fact that many universities concentrate on the increased retention rates minimal evaluation of university retention initiatives exist and lacks evidence to confirm the effectiveness of the initiatives (Hossler, 2008). The results are partly contributed by the slow adoption of advanced data management systems in the universities. Nevertheless, the new and low cost analytics solutions have led to increased interest in the use of analytics to gain timely insights regaining the drivers of student drop out (Pirani, 2005; Talbert, 2012).

2.3.2 Enrolment Rate of students

Enrolment of students is based on academic merits mainly student's score at KCSE. Academic factors have proved to be beneficial for university success and have been useful during the admission selection (Schnell, 2003). Academic merits during the university enrolment instil motivation and self-confidence for achievement (Lotkowski, 2004).

The utilization of descriptive analytics focuses majorly on the analysis of admission processes, pre-university factors besides their implications on graduation rates. (Campbell, 2007) notes that not all the students who obtain university placement actually enrol to the same program. Many students opt to change the course they were admitted to do. Others change to their university of choice. This shows that the number of students reporting to university may be less as compared to the number that received university placement for a particular course.

2.3.3 Deferment of students

Students' deferment in universities is of a particular concern and it affects the performance after the student enrols the campus (Hossler, 2008). University students fail to go to class, take courses from the recommended sequence and change their curriculum (Schnell, 2003).This puts pressure on the students as the exams approach. When the students are unable to cope up with pressure of not having adequate preparation they opt to defer their studies. This eventually affects their progression to the next academic year.

2.3.4 Operationalization of Variables

The variables of the study were operationalized as per table 2.3.1

Table 2.3.1: Operationalization of variables

Variables	**Indicator**	**Values (Data)**
University progression factors	• Enrolment rates • Dropout rates • Deferment rates	• Number of students

2.4 Empirical study on progression rate of university students

This section presents some empirical studies that have been done using neural networks in the university setting. These studies are introduced here in order to cement the need for the current study. A major issue within the higher learning education is the poor admission results. Students leave universities for different reasons such as poor background understanding of the field of study, very low grades and the incapacity of passing an examination, lack of financial resources. Predicting students' results is an important problem for the management of the universities who want to avoid the phenomenon of early school leaving.

Boddan et al., (2013) used a neural network to predict the students' results measured by the grade point average in the first year of study. The input data were the students profile at the enrolment time at the university including information about the student age, the GPA at high school graduation, the gap between secondary school graduation and university enrolment. After training the network the authors obtained an MSE of about 1.7%.

A study by Stamos(2008) used artificial neural network to predict student graduation results. The ANN network was developed as a three-layered perceptron and trained with the back propagation principles. During the training and testing, the experiments used a sample of 1,407 of students' profiles. The selected sample represented learners at Waubonsee College and it was split to two sets. One set had 1,100 profiles was used as the training set and the remaining 307 used as the testing set. The average predictability rate for the two sets was 77% and 68%, respectively. Mohammed Awad (2018) used artificial neural network technique in Prediction of General High School Exam Result Level Using Multilayer Perceptron Neural Networks. The findings of the study showed that that the ANN method was more accurate than logistic regression technique.

None of these studies have focused on developing an artificial neural network for the progression rate of students.

2.5 Summary and the Research Gaps

In this chapter a detailed analysis of the variables of study and their relationships within a selected theoretical framework has been presented. It has also been demonstrated how drop out, school drop- out and deferments can affect progression rate of students. Several theories have been discussed that explain the three factors considered in the study. A conceptual framework has also been presented to show the relationship between variables in the study. The gaps in the literature point to the need to develop a model for progression rate of the university students.

Summary of Literature Review

The problem concerning progression of students is a complex and multifaceted problem that is not easily resolved. In order to get a clearer understanding of the issue, Wells et al (1989) classified the factors into four categories. These included student indicators, school indicators, family indicators and community indicators. Studies showed that student factors that lead to student progression rate are varied and include deferment, dropout and enrolment. All these factors influence the decision of student's ability to progress with their studies. An analysis of research showed that student engagement and motivation play huge roles in academic achievement. Engagement can be broken up into four different categories. These categories are academic engagement, behavioural engagement, psychological engagement, and social engagement. The research showed that engagement in school was what kept students motivated to learn. As soon as disengagement began, then the downward spiral is what causes students to fall behind and eventually leave school. Understanding what causes the different types of disengagement and how to prevent them from occurring are the stepping stones to solving the epidemic of lack of student progression. Motivational theories also support the need to create environments that encourage successful students. Motivational theories can show how some educational practices are detrimental to students' futures. These theorists also shed light on how to use motivational techniques to reach all students no matter if they are on track to graduate or at-risk for quitting.

Preventative intervention programs are what research reveals is the solution to the progression rate crisis. Targeting students early and getting them active in a preventative program increases the likeliness of them to graduate. Programs, no matter if they are alternative school programs or those that focus on transitional times, using all resources and getting everyone on board in the support program is what makes them more successful. Whether the program is embedded within the school or an outside agency, the support program must be implemented and carried out in order to save students at-risk.

While this chapter shed light on the research of the progression rate crisis, factors that cause students to quit, and intervention programs used to keep students in school, supplementary research could further explain the progression rate crisis. In the next chapter information gathered from this review of literature was used and the researcher built upon the investigation with explanations of the study.

CHAPTER THREE

RESEARCH MEHODOLOGY

3.1 Introduction

This chapter presents details of the methodology adopted for the study. This chapter presents the research design which is the blueprint for conducting a study detailing the design approach. The research design also covers the research paradigm selected for the study. This chapter also presents the data source which explains the population of the study. It also covers the data selection and sampling procedures. Data pre-processing which focuses on the assembling of the data and ensuring data is in the form required for analysis is also discussed. Data mining is also presented which describes the process of obtaining the data and using the R statistical software to perform descriptive and predictive analytics. Data transformation is also discussed which describes ensuring that the data is scaled in the required format for use in the analysis. Then the Knowledge gap is also presented for presenting the output results of the study. The last item in this chapter is the ethical consideration which describes how the study obtained the relevant permission to use the data for analysis.

3.2 Research Design

Research design can be referred to as a plan which gives guidance on any given research and aids in data collection, analysis as well as interpretation of observations. It can also be used by researchers as a blueprint which helps one to make decision on methods and tools to use in collection of information as well as its evaluation, so as to be able to answer the questions guiding the research (Cooper & Schindle, 2014). This study employed Fayaz model for data mining shown in figure 3.2.1

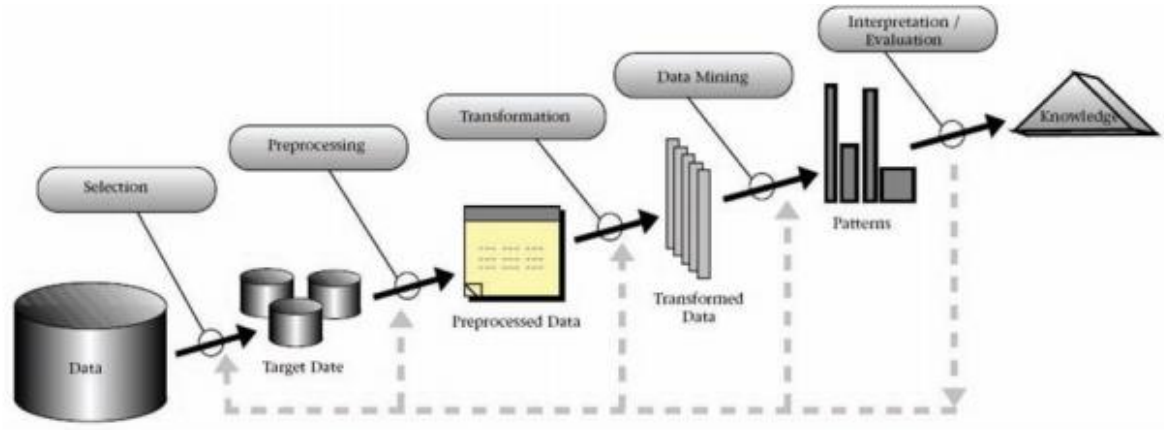


Figure 3.2.1: Model design approach for the study

Source: Fayyad et al. (1996)

The various stages in the design approach were described in the next sections.

3.2.1 Data

The data used for modelling was obtained from Technical university of Kenya student admissions database. The study conducted a census study and the population of the study was 2976 students drawn from the Faculty of Applied Sciences and Technology. According to (Kadam, 2010) the larger the sample size the more accurate is the data. Therefore this study was deemed to yield accurate results since the entire population was used. The database contained information on enrolment, deferment and number of students who had dropped out. The data was stored in an excel data sheet. The excel data sheet had column data for number of enrolling students from 2015 to 2018, the number of students dropping from their studies, the number of students deferring during the course of their studies, the rate of enrolment of the students, the rate of deferment of the students and the dropout rate of the students. The independent variables for the study were the enrolment rate, deferment rate and dropout rate of students. The dependent variable was progression rate of students. The study defined drop out as the student who opts not to continue with their study. Deferment was defined as the students who failed to do his/ her exams but had made official communications expressing that he/she would be deferring studies. The study used three semesters in one academic year, that is, January –April, May –August and September –December semesters for 2015 through to 2018 academic years.

3.2.2 Selection of Data

The study conducted a census study since all the student population was used in the study. After obtaining data from the university data base and storing in an excel database, the study sought to work with a representative dataset by performing data reduction. One method of data reduction which this study adopted is numerosity reduction technique, where the data was replaced or estimated by alternative, smaller data representation methods such as means, clustering, sampling, and the use of histograms. The study then undertook data cleaning activity. Data cleaning is the process of detecting and removing corrupt or inaccurate records from a database. This process was done by use of binning technique. Binning methods sorts a data value by consulting its "neighbour- hood," that is, the values around. The sorted values were distributed into a number of "buckets," or bins. The data was portioned into equal frequency bins and then smoothing by means was performed.

Technical university of Kenya is composed of Three Faculties namely: faculty of engineering sciences and technology, faculty of social sciences and technology and faculty of applied sciences and technology. The study used convenience sampling technique to use data from the faculty of applied sciences and technology. This represented a third of the faculties and therefore thirty three per cent of the population. According to Mugenda and Mugenda, (2013) ten per cent of the population is representative of the population. Information contained in the database showed that there cases where the students had received university placement and failed to enrol for their studies. There were also cases where students enrolled in second year meaning that some of the students had first done diploma courses and were therefore eligible to enrol for university degree programs at second year. The study therefore performed data cleaning to take care of the students by filling the missing first year records with averages of the first year enrolment data. The database also contained a mix of registration numbers where for instance student information appeared more than once in the database due to repeated registration numbers. This was also handled using dimension reduction technique. Data reduction was performed to help increase efficiency in handling of the data for analysis. The data was then split into training and test data. This method of splitting the sample has been effectively used by Pesaran, Pick, and Timmermann, (2011) and Andic and Ogunc (2015).

3.2.3 Data Pre-processing and Transformation

This involves identification of the data to be used in training the model, test data to test model, training and the validation data to measure the output error. It is critical to identify the type of data to be put for each data set as identified in the research sampling procedure. The sample taken was representative of the identified population.

The data was converted into comma delimited (csv) format files to have three different csvs containing the data on enrolment of students, deferment and drop out of the students. all three of these files were linked through the student ids. After data input files were prepared, data cleaning was done in the following steps:

i. Data type conversion (character to numeric, character to factor)

ii. Aggregated data to student id level

iii. Identified outliers and removed them appropriately

iv. Cleaned/ removed missing values

v. Merged dataset into one dataset

Admission information and term results information datasets were compiled together for pre-processing while the third dataset courses and grades from data-mart table were treated separately. Then all datasets were merged at the end of the process.

Having read the csv files, as the first step of the data preparation process, all the character type data were converted into numeric data.

```
Enrolment<- read.csv(file.choose(), header= TRUE)

Deferment<-read.csv(file.choose(), header=TRUE)

Dropout<-read.csv(file.choose(), header=TRUE)

Mydata<-as.data.frame(Enrolment,Deferment,Dropout)

Attach(mydata)

Mydata

Enrolment;Deferment;Dropout
```

3.2.4 Data mining

This process involves feeding inputs to the model for processing in order to train the model on the type of input data and the expected output of the training session. As mentioned earlier data was stored in an excel data sheet. The excel sheet was then saved as a CSV file (comma delimited) to make sure that it was possible to read it into R statistical software. After reading the data into R software the next step was to scale the data in a [0,1] interval to ensure efficiency in working with the data. The data was then split in to a ratio of 80: 20 where eighty per cent of data was training data while 20 per cent of the data was test data. According to Dobbin (2011), twenty per cent of the data should be treated as test data while 80% of the data should be treated as training data.

The training data was fed into the ANN model via the identified model neurons. The training process was done over a number of iterations until the artificial neural network algorithm converged. This was done for three trials while reducing the error rate and adjusting the input weights.

3.2.5 Model Evaluation

This process involves the use of a test data set to check whether the system will be properly trained by observing the actual model output versus the expected output. By using the validation data set, any disparities in the output capture by error performance measures were used to adjust the weights of the neurons for purposes of fine tuning the model. The model output data was tested through repeated trials on the training phase. This was done to ensure that the error rate was reduced to the global minimum for the model in order to provide the most optimal input weights for each variable input. The model was evaluated by calculating measures of performance as identified in the study chapter 2. This was done using the root mean squared error and the mean absolute error. These measures were calculated against the actual recorded progression rate values in comparison with the progression rate values forecasted by the model over the same period of time. The model was run for three trials and the average accuracy of the model determined. A confusion matrix was obtained to show the various performance measures and the accuracy of the model. Output of the results was displayed at the Knowledge a gap. Several performance measures discussed in the next sections below were used in this study

Accuracy

Accuracy is how probable it is on average for the prediction of the model to be correct. This is calculated as the proportion of the number of correctly classified cases to the total number of cases. Correctly classified cases are the total of true positives and true negatives. The following equation represents the calculation of accuracy, n being the total number of cases.

$$\text{Accuracy} =$$

$$\frac{\text{True Positives} + \text{True Negatives}}{\text{n}}$$

Sensitivity

Sensitivity is how probable it is to classify a case as true when it is actually true. This is calculated as the proportion of correctly classified true cases within actual true cases.

Specificity

Specificity is how probable it is to classify a case as false when it is actually false. This is calculated as the proportion of correctly classified false cases within actual false cases.

Positive predictive value

Positive predictive value is how probable it is for a case to be actually true when the model predicts it to be true. In the context of this study, the probability would be how likely it is that a student will progress, when the model identifies the student as a student who will progress. This is calculated as the proportion of correctly classified true cases within predicted true cases.

Negative predictive value

Negative predictive value is how probable it is for a case to be actually false when the model predicted it to be false. In the context of this study, it is how likely it would be for a student to dropout or defer his or her studies, when the model identifies the student as a student who will dropout or defer. This is calculated as the proportion of correctly classified false cases within predicted false cases.

3.3 Ethical Issues

Research ethics is critical since it guides the interactions with people, organizations and institutions (Christensen, Johnson, & Turner, 2014). The study sought authorization for data collection from the Technical University of Kenya registrar admissions by explaining the purpose and importance of study. Privacy and confidentiality was employed to ensure that the data collected from respondents was kept safe, free from interference and protected from unwanted use.

CHAPTER FOUR

RESULTS AND DISCUSSIONS

4.1 Introduction

This chapter deals with the analysis and results of the data collected for the study. The detailed analysis focused on findings is presented based on the three objectives and research questions of the study clearly demonstrated. First, it evaluates the descriptive analysis of the variables under study, the means and standard deviations and the development of the neural network model, model validation and finally prediction of progression rate of the students. Finally, the chapter reviews the discussion of the findings of the study.

4.1.1 Information on metadata

Descriptive metadata comprising mean and standard deviations of variables enrolment rate, deferment rate and dropout rates were analysed and presented in table 4.1.1.

Table 4.1.1 Descriptive metadata

	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
enrolment	2937.75	610.033	-.230	.637	-.851	1.232
Drop out	69.42	37.621	.415	.637	-1.590	1.232
deferment	191.42	79.598	.322	.637	-1.248	1.232
progression	2676.92	645.243	-.233	.637	-.764	1.232
Enrolment rate	1.0083	.38624	1.684	.637	1.730	1.232
Dropout rate	.0767	.04185	1.299	.637	1.085	1.232
Deferment rate	1.1183	.34300	-.607	.637	-.085	1.232
Progression rate	.2558	.08775	.159	.637	-.565	1.232

The results in table 4.1.1 indicate that on average 2937 students enrolled annually from 2015 to 2018. This number was however fluctuating at a standard deviation of 610 during the study period. The skewness and kurtosis values were -0.230 and -0.851 respectively. This indicated that the data was normally distributed as the recommended value is between -3 and 3 for normally distributed data. The results indicated that on average 69 students dropped out of their course programs annually during the study period of 2015 to 2018. This number was however fluctuating at a standard deviation of 37.6 during the study period. The skewness and kurtosis values were -0.415 and -1.590 respectively. This indicated that the data was normally distributed as the recommended value is between -3 and 3 for normally distributed data. The results indicated that on average 191 students' deferred their course programs annually during the study period of 2015 to 2018. This number was however fluctuating at a standard deviation of 79.598 during the study period. The skewness and kurtosis values were -0.322 and -1.248 respectively. This indicated that the data was normally distributed as the recommended value is between -3 and 3 for normally distributed data. The results indicated that on average 2676 students progressed through their course programs annually during the study period of 2015 to 2018. This number was however fluctuating at a standard deviation of 645 during the study period. The skewness and kurtosis values were -0.233 and -0.764 respectively. This indicated that the data was normally distributed as the recommended value is between -3 and 3 for normally distributed data.

4.1.2 Descriptive analytics

Based on the results of the study data, a correlation of the three variables namely, enrolment rate, deferment rate, dropout rate and progression rate was evident. This was demonstrated in figure 4.1.1.The figure shows the amount of correlation of the independent variables where deferment was highly correlated with progression. This was followed enrolment rate and then by dropout rate

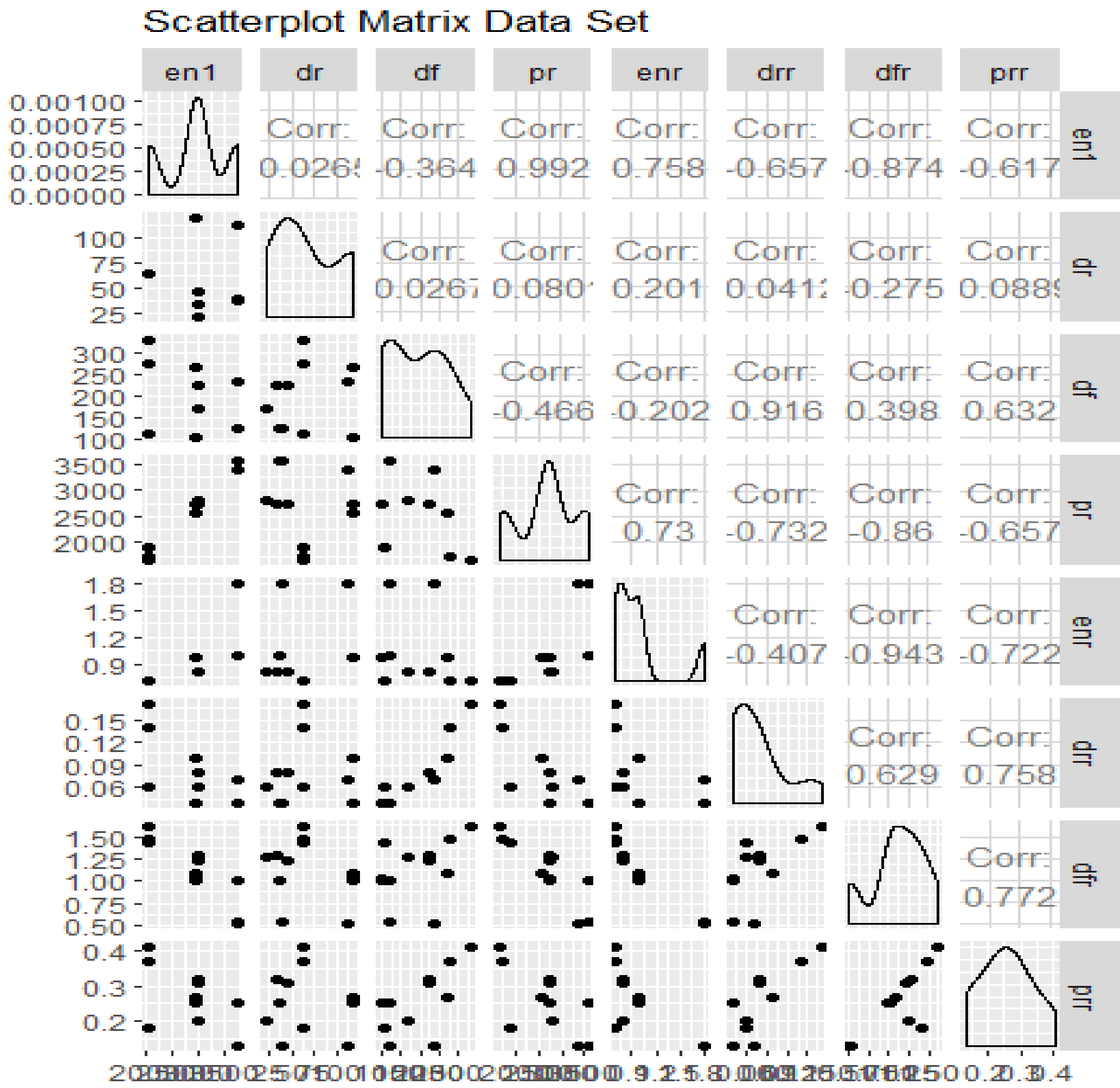


Figure 4.1.1 scatter plot of correlation matrix

4.2 Objective one results

The objective was to investigate and identify factors that influence university students' progression rate. It was found that three factors were significant in predicting the progression rate of students. These were the enrolment rate of students, dropout rate of students and the deferment rate of students. Results for this objective were presented in table 4.2.1.

Table 4.2.1. Results for objective one

Variable	Beta coefficient	significance	Remark
Intercept	0.8167		
Enrolment rate	-1.00099	P=0.000***	Significant
Dropout rate	-0.2774	P=0.020**	Significant
Deferment rate	-0.98575	P=0.020**	Significant

The study used the data of interest to obtain regression coefficients pertaining to enrolment rate of students, dropout rate of students and progression rate of students in Technical University of Kenya, faculty of science and technology (table 4.2.1). The artificial neural network was configured to obtain a combination of enrolment rate of students, dropout rate of students and deferment rate of the students. The regression analysis showed that enrolment rate achieved alpha of 1.00099 per cent which was significant at the 5 per cent level while the dropout rate and deferment was less than 1 per cent. For every 1 per cent decrease in dropout rate the progression rate increased by 27.7 per cent. For every 1 decrease in deferment rate the progression rate increased by 98.6 per cent. For every 1 per cent decrease in enrolment rate the progression rate increased by 100.099 per cent. This implies that the enrolment rate was more significant in predicting the progression rate of students.

4.3 Objective two results

The objective was to develop an appropriate artificial neural network model that uses the identified factors for predicting progression rate. The results of this objective were presented in figure 4.3.1

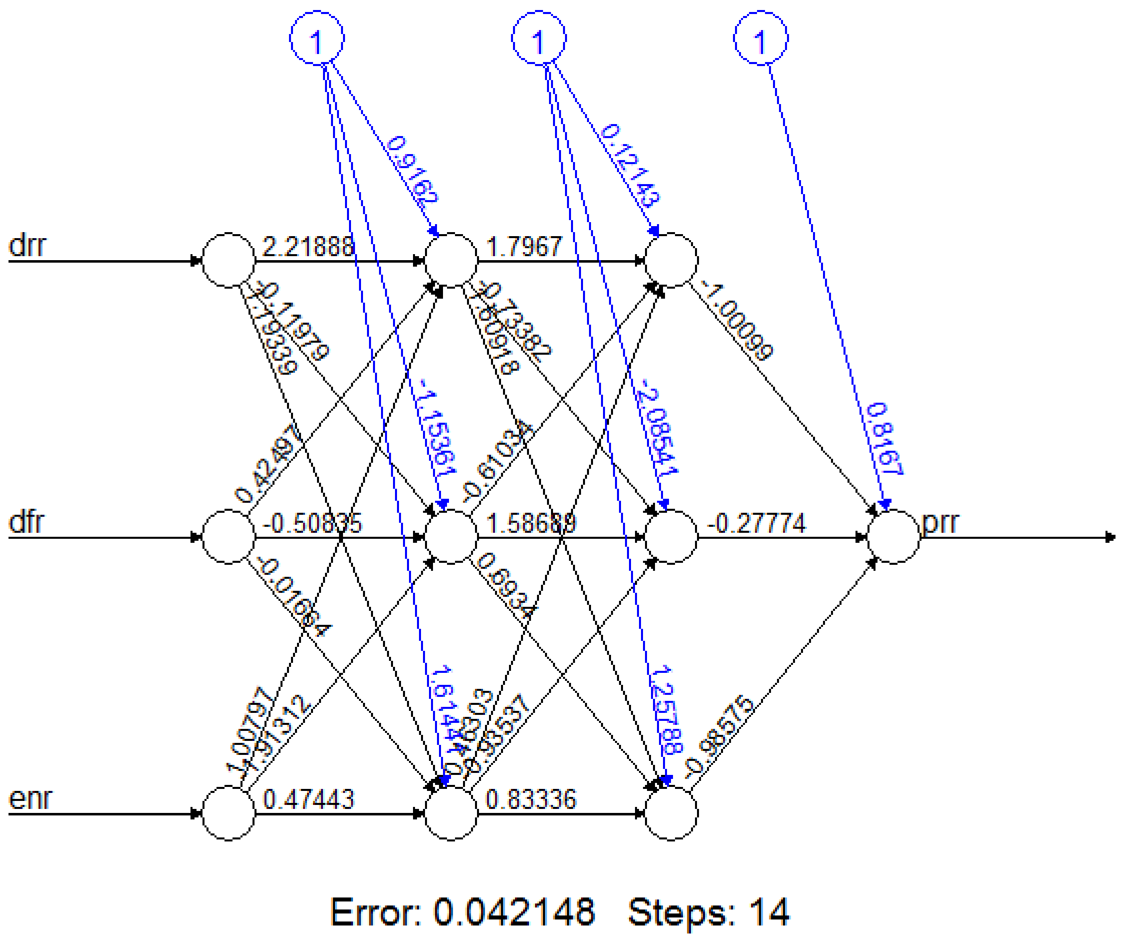


Figure 4.3.1.Artificial Neural Network model for the study.

The results in figure 4.3.1 showed that enrolment rate achieved alpha of 1.00099 per cent which was significant at the 5 per cent level while the dropout rate and deferment was less than 1 per cent. For every 1 per cent decrease in dropout rate the progression rate increased by 27.7 per cent. For every 1 decrease in deferment rate the progression rate increased by 98.6 per cent. For every 1 per cent decrease in enrolment rate the progression rate increased by 100.099 per cent. This implies that the enrolment rate was more significant in predicting the progression rate of students.

The development of the artificial neural network model for the study involved enhancing the network to learn the inputs through back propagation algorithm. This learning process was controlled and monitored through the analysis stage. The process underwent several steps. First the input variables were entered into the artificial neural network thorough the input layer of the neural network. The effect of introducing the inputs was propagated forward into the succeeding layers that is, the hidden layer and the output layers. The network outputs were then calculated. The learning process involved adjusting the weights of the neural network to reduce error in the network computation. The error generated was signalled back to the preceding layers and the weights were adjusted appropriately. The minimum error attained for test data was 0.042 while the mean absolute error obtained using the training set was 0.041. This suggested that the model had been adequately validated.

4.4 Objective three results

The objective of this study was to validate the developed model. Results of this objective were presented in table 3 and table 4.4.1

Table 4.4.1: Prediction accuracy of the developed model

Actual	prediction
0.25	0.2524
0.2	0.2522
0.25	0.2524
0.28	0.2521
0.23	0.2523
0.32	0.2521
0.27	0.2522
0.37	0.2519
0.23	0.2523
0.31	0.2522
0.25	0.2524
0.251	0.2517

Table 4 shows the actual values of the model and the predicted values after training the neural network. After testing the model the prediction accuracy was 73% while the prediction accuracy was 78.5% after training the model.The performance of the model is presented in table 4.4.2.

Table 4.4.2. Performance measure of the developed model

MSE	MAE	Accuracy	Sensitivity	Specificity	Pos.Pred.Value	Neg.Pred.Value
0.0421	0.214	0.7816	0.92	0.42	0.797	0.71

The study carried out further analysis to improve and fine tune the results earlier obtained. This was carried out to improve the predictive accuracy of the artificial neural network. Improving accuracy of the neural network involved relooking at the parameter estimated and eliminating the ones that were not significant. The study was of the idea that if the less significant predictor variables could be dropped then the error term generated would interestingly reduce and therefore improve the predictive ability of the neural network. So this procedure of dropping the less influential variable was carried out and the remaining contributing variables were retained in the model. A sensitivity analysis of the model was carried out and the model with the least mean square error was believed to be the optimal model. The analysis was done throughout by validating the training set and this enhanced measuring the effects of small changes in each of the variable parameters as they related to the output. Input parameters that had least significance were treated as unrealistic and therefore were eliminated from the neural network. A Confusion matrix of the trained model was obtained showing the various performance measures as shown in Table 4.3

The study used the following R code to obtain a confusion matrix for progression rate of students using TUK data

```
modeltuk <- train(pr ~ ., data=mydata, method="nnet")

modeltuk
```

```
modelcm <- confusionMatrix(predict(modeltuk, newdata=trainset,
type="raw"),trainset$df_Enrolled)

modelcm
```

The Neural Network model built with original set of variables had an accuracy of 78.5 per cent for the training set and 73.0 per cent for the test data.

After optimizing the network's structure and training the network within 4 epochs, the study tested the network's predictive power on the training data set. The mean square error achieved was 0.041 and the network was able to correctly classify 2321 out of 2976 students progressing through their academic studies (78.5%).

The neural networks solves classification type problems that is why it was important to define the number of neurons at each layer, that is, the input, the hidden and the output layers. However, it is worth noting that it is not easy to define appropriately the number of neurons in the hidden layer. Therefor the study used trial and error method to finally adopt a 3, 1 layer setting. This set up had the least amount of absolute error of 0.041063generated. With this amount of error the study achieved an optimal performance of 78%

The equation of the model was therefore

$$\mathrm{prr} = 0.8167 - 1.00099\mathrm{drr} - 0.2774\mathrm{dfr} - 0.98575\mathrm{enr}$$

The precision in the prediction of the progression rate was higher based on enrolment class than the dropout class and deferment class, which was mainly because the data attributions were not sufficiently comprehensive. Related research suggests that there are many factors affecting students dropping out as well as deferment rate and that if more students enrol into the universities then it is possible to achieve higher progression rates, hence an increase in the dropout rate and deferment rate of students has an impact in the prediction accuracy of progression rates of students.

The average predictability rate for the training and test sets was 78%. After comparing the study results with results achieved by other researcher the study was highly encouraging. As it is shown in Figure 2, in this run the network's error reaches to its least amount after four epochs. Table 3 shows the correct classification rate for training and network test data. Result of classification using multiple layer perceptron proves satisfactory performance of the network in identifying progression rate of students.

```
roundedresults<-sapply(results,round,digits=4)
> roundedresults
```

4.5 Discussion of the Results

This study looked at the progression rate of students pursuing higher levels of education at the universities. The study was significant because there has been a great concern on the future of the university learning in Kenya. There has been a growing rate of students failing to proceed and graduate in their courses. This concern has been public spoken about by the key stakeholders in the institutions of higher learning. The government through the ministry of higher education has gone to the extent of introducing more curriculum based trainings with a few to making the students not drop their studies. Similarly universities have witnessed parents coming to the universities to inquire about the progress of their children. Many of the parents have been shocked to find that their students were actually discontinued in their early years in the courses. Other parents have come to the universities complaining that their children have not been attending lessons only to find that these students had applied to defer their studies without knowledge of the parent. Therefore it was important to carry out this study to understand the status of the education at the university as a country so that appropriate interventions can be put in place.

This study established that enrolment rate of the students at the university had actually dropped significantly, thus affecting the number of the students who were progressing with their studies. This was in conformity to a research presented in the University news (2018) report who stated that the rate of enrolled was gradually decreasing ever since drastic education reforms were instituted by the government. The rate of deferment of the students was also on the rise as cited by Julie (2018) in their report that reported on factors affecting academic performance of the students. In the study by Julie (2018) it was reported that deferment and dropout had largely increased and thus leading to negative performance of the students.

(Cave et al. 1991) also reported that deferment and dropout had largely increased and thus leading to negative performance of the students.

A similar study by (Turri & Agasisti, 2016) reported that deferment of students was a key contributor to student under performance and led to increased drop out from college.

Before proceeding with the analysis the study obtained metadata about the study. It was found that skewness and kurtosis ranged from -3 to 3 an indication that the data was normal. Related studies have also reported almost similar figure. This is also the figure recommended by Shapiro Wilk (2007) who asserts that real data should be normally distributed.

The predictive analytics carried out by the study showed a certain degree of correlation among the variables in the study. This implied that there was a significant association between a deferment and drop out. It can therefore be stated that a person differing his or her studies is more likely to eventually dropout of his or her studies. This is an essential rule used in association mining to show that variables in a study usually have a close connection

This study was also successful in developing an artificial neural network model. The model was presented in figure five and it was found that this model was optimal since it had the least absolute error. According to Hasith (2016) an optimal model is one that gives the minimum absolute error as a measure of performance.

Objective three of this study was to validate the developed model. This was also a success in this study since the study obtained an accuracy of 78.5 per cent which was deemed to agree with previous studies. Hasith (2016) conducted a study to determine the factors that lead to drop out of students in higher learning institutions. The study accuracy of the neural network model used by Hasith (2016) was about 80 per cent, which was a difference of about one point five per cent from the current study. Overall this study was a success since the objectives of the study were achieved.

Although students spend more time in the education system than anyone else, their voices are often overlooked and limited when it comes to literature and research on education. These students' experiences during their studies, along with their individual perspectives, resulted in the emergent themes of this study.

CHAPTER FIVE

SUMMARY, CONCLUSIONS AND RECOMENDATIONS

5.1 Introduction

The study sought to determine the progression rate of students using artificial neural networks Summaries of the study were drawn from the analysis in chapter four and conclusions were drawn from the findings and recommendations made based on the objectives of the study. The recommendations were made to area of policy interest and for further studies.

5.2 Summary of findings

The study focused on determining the progression rate of students at the university level. The control variables for the study were the enrolment rate, the dropout rate and the deferment rate of the study. These variables were analysed using the artificial neural network model. The accuracy of the neural network showed a high predictive ability. In terms of accuracy the network showed a high predictive ability for the training set data as opposed to the test data. Progression rate of students was highly determined by the number of deferment of students. This was supported by a high average number of students deferring as opposed to those who were completely dropping out of their course programs. The enrolment rate was at an all-time low for the universities. This was deemed to pose a threat to continued proper functioning of the universities. This study proposes policy intervention as rate of enrolment was significantly leading to reduced number of students progressing through their academics. In the long run, reduced progression rates of students may lead to alarmingly low number of students being absorbed into the job market. As a result this may lead to reduced workforce hence low economic production.

5.3 Conclusions

This study made important contributions to the knowledge base pertaining to students' progression in the universities which are currently facing crisis of student numbers. The study found that there is a large number of students who are deferring their studies and therefore not able to finish their four year courses within the stipulated time. This means that a lot needs to be done by university management to intervene and help reduce the number of such cases. Also it was realized that the numbers of the student enrolment was on a downward trend and urgent measures were needed to be put in place to mitigate this trend.

The study used university as target population since there had been public outcry on decreasing number of students enrolling in the university. Findings are therefore of great help to stake holders to know the position of enrolment, the dropout rate and deferment rate of students. Because if the enrolment of students keeps going down and no close monitoring of the number of students finishing successfully, it is possible that in the near future there will also be public outcry on the reduction of skilled labour force.

Valuable information can be deduced from the findings of this study since many universities are struggling with small student numbers. Some university departments are facing shut down due to the lack of students enrolling or even drop out and deferments. It is highly likely that if necessary interventions and measures are not taken in the coming years, then universities may lay off some of their staff for lack of students to teach.

The study was successful in developing an artificial neural network model for the progression of students. The model had a good prediction accuracy that was in agreement with past studies as described in the discussion of results. The model was able to show the significance of the study variables.The model was validated and a minimum absolute error obtained which was comparable with related studies as presented in the discussion of results.

5.4 Limitation of the study

This study did not obtain data regarding the gender of the students and their background characteristics which could also be factors that can contribute to student progression. As a result the study could not differentiate progression based on gender of students. The study used data from one university that is Technical university of Kenya and the data was collected from one faculty out of a possible three faculties. This was due to time constraint and availability of resources to collect data.

5.5 Recommendations

This study brought interesting findings that can be believed to bring a positive change if implemented. The study used data from one university that is Technical university of Kenya and the data was collected from one faculty out of a possible three faculties. This was due to time constraint and availability of resources to collect data. The study recommends that future studies can be done using data from other studies to ascertain the progression rate of students.

The study recommends that aspects of drop out of students should be followed carefully and there should be close monitoring of the students and their behaviours in order to identify those are at risk of dropping out. The on the issue of deferment proper mechanisms should be put in place in order to identify students wishing to defer their studies and their reasons should be taken seriously even if it is involving the parents and the university management. Issues of deferment of students should not be left to be sole decisions of the students.

An important follow-up of this work is the identification of the student predictor variables that are significant in predictability accuracy. This is of value for three reasons; First of all, it is of value to the government through the ministry of education, to the university management board and to the faculty committee since if they know the reasons that students do not progress to the succeeding academic years and within the appropriate time period, then they can directly intervene. Future studies may be conducted from the perspectives of enhancing the attributions and improving the learning behaviour. Learning behaviour data can be obtained from the academic management boards to facilitate the input attributions for the predictive model, thereby achieving higher progression.

The government should provide more resources to help increase the number of students enrolling to the university and reduce the rate of deferment of students due to lack of financial aid. University management boards should pay a keen attention on the matters to help discover leading causes of deferment and drop out of students. The findings of this study showed that the deferment of students is on the rise. A separate study can be done to help identify the main problem that leads to the increasing deferment of students.

Students' voices are vital components needed to enlighten educators about the complex problem of dropping out. This study shows clearly that early monitoring of progress, academic support, and a safe and inviting learning environment are what is needed for changes to occur. Educators will be more capable to prevent students from leaving if they will listen to the voices of dropouts. A students' voice is a great source for educators that seek answers to the dropout crisis because their responses could create stronger supports for those at-risk for dropping out. Results of these findings call for the need to monitor progress often and early communication between all stakeholders, increase academic support, and create safe and inviting learning environments. The insights of these findings point to a successful future for other at-risk students.

The study used artificial neural network model for progression rate of students, this study was successful in applying the model. Future studies can also be conducted to compare the results with those of the artificial model. Models using classification trees and using other data mining softwares such as Weka can be used to compare the accuracy of prediction with that of the artificial neural network model developed herein.For objective three other performance measures aside those mentioned in this study such as the Akaike Information Criteria and the Shwarz Bayesian Criteria can be used to validate the models of progression rate of the study.

The study recommends that future studies can be done on influence of student characteristics such as gender and family background on student progression since this study did not explore this aspect. It would be interesting to compare progression rate of students for female and male students.

REFERENCES

Malhotra, N. K. & Birks, D. F. (2006).Marketing research: an applied approach. 3rd ed. Prentice Hall/ finance times

Matambo, O. K. (2013). 2013 Budget speech Gaborone: Government Printers.

Miller. D. & Seller, M. (2007). School Facilities and Pupils' Academic Achievements.

Transkei University Research Study, South Africa: Unpublished.

MOEST, (2010). Sectional Paper No 1 2010 Policy Framework for Education,

Training and Research. Meeting the Challenges of Education Training and

Research in Kenya in the 21st Century, Nairobi.

Momoh, A. J. (2010). Effects of instructional resources on students' performance in

West Africa School Certificate Examinations (WASCE).International Journal

of Inclusive Education. 6(2).113-113-126.

Mosha, M. A. (2014).Factors influenceing students' performance in English Language in Zanzibar Rural and Urban Secondary School. In J. Edu and Practice.5 (35), 200- 20

Mugenda, O. M., & Mugenda, A. G. (2012). Research methods dictionary. Act press Nairobi.

Mugenda, O., and Mugenda, A., (2003).Research Methods, Quantities and qualitative approaches. Act press Nairobi.

APPENDIX

```
library(neuralnet)

library(tidyverse)

library(neuralnet)

library(GGally)

mydata<-read.csv(file.choose(),header=TRUE)

mydata

attach(mydata)

##########

scaleddata<-scale(mydata)

# TRAINING AND TEST DATA

trainset <- mydata[2:5, ]

testset <- mydata[6:11, ]

# The predictor vars must be scaled data for the ANN fitting

ism.scaled <- as.data.frame(scale(mydata))

min.medv <- min(mydata$x1)

max.medv <- max(mydata$x1)

# response var must be scaled to [0 < resp < 1]

ism.scaled$medv <- scale(mydata$x1, center = min.medv, scale = max.medv - min.medv)
```

```
#Neural Network

nn <- neuralnet(y ~ x1 +x2, data=mydata, hidden=c(2,1), linear.output=FALSE,
threshold=0.01)

nn$result.matrix

plot(nn)

nn$result.matrix

#Test the resulting output

temp_test <- subset(mydata, select = c("x1","x2"))

head(temp_test)

nn.results <- compute(nn, temp_test)

nn.results

results <- data.frame(actual = mydata$y, prediction = nn.results$net.result)

results

roundedresults<-sapply(results,round,digits=4)

roundedresults

roundedresultsdf=data.frame(roundedresults)

attach(roundedresultsdf)

table(actual,prediction)

#Validation
```

```
results <- data.frame(actual = mydata$x1, prediction = nn.results$net.result)

results

#Progression rate

predicted=results$prediction * abs(diff(range(x1))) + min(x1)

actual=results$actual * abs(diff(range(x1))) + min(x1)

comparison=data.frame(predicted,actual)

deviation=((actual-predicted)/actual)

comparison=data.frame(predicted,actual,deviation)

accuracy=1-abs(mean(deviation))

accuracy

#5 layer perceptron(OPTIONAL)

nn <- neuralnet(y ~ x1+x2,data=mydata, hidden=c(5,2), linear.output=TRUE, threshold=0.01)

nn$result.matrix

n <- names(mydata)

f <- as.formula(paste("y ~", paste(n[!n %in% "x3"], collapse = " + ")))

nn <- neuralnet(f,data=mydata,hidden=c(5,3),linear.output=T)

plot(nn)

pr.nn <- compute(nn,mydata[,1:3])

pr.nn_ <- pr.nn$net.result*(max(mydata$x1)-min(mydata$x1))+min(mydata$x1)
```

```
test.r <- (mydata$x1)*(max(mydata$x1)-min(mydata$x1))+min(mydata$x1)

MSE.nn <- sum((test.r - pr.nn_)^2)/nrow(mydata)

print(paste(MSE.nn))

cleandata <- mydata

cleandata <- na.omit(cleandata)

#scaling

apply(cleandata,MARGIN = 2, FUN = function(x)sum(is.na(x)))

maxs = apply(cleandata, MARGIN = 2, max)

mins = apply(cleandata, MARGIN = 2, min)

scaledData = as.data.frame(scale(cleandata, center = mins, scale = maxs - mins))

summary(scaledData)

#Splitting data in 80:20 ratio

train = sample(1:nrow(scaledData), nrow(scaledData)*0.8)

test = -train

training_Data = scaledData[train,]

testing_Data = scaledData[test,]

dim(training_Data)

dim(testing_Data)

#neural net
```

```
library(neuralnet)

n   <- names(training_Data)

f   <- as.formula(paste("x1   ~", paste(n[!n  %in%    "x1"], collapse    =   "   +   ")))

neuralnet_Model <- neuralnet(f,data = training_Data, hidden = c(2,1))

plot(neuralnet_Model)

neuralnet_Model$result.matrix

pred_neuralnet<-compute(neuralnet_Model,testing_Data)

pred_neuralnet.scaled   <- pred_neuralnet$net.result *(max(scaledData$x1)-
min(scaledData$x1))+min(scaledData$x1)

real.values <- (testing_Data$x1)*(max(cleandata$x1)-min(cleandata$x1))+min(cleandata$x1)

MSE.neuralnetModel  <- sum((real.values - pred_neuralnet.scaled)^2)/nrow(testing_Data)

MSE.neuralnetModel

plot(real.values, pred_neuralnet.scaled, col='red',main='Real   vs  predicted',pch=18,cex=0.7)

abline(0,1,lwd=2)


library(neuralnet)

library(tidyverse)

library(neuralnet)

library(GGally)
```

```
mydata<-read.csv(file.choose(),header=TRUE)

mydata

attach(mydata)

names(mydata)

##########

# Scale the Data

scale01 <- function(x){

  (x - min(x)) / (max(x) - min(x))

}


mydata<- mydata%>%mutate_all(scale01)

# Split into test and train sets

# Split into test and train sets

set.seed(12345)

Yacht_Data_Train <- sample_frac(tbl = mydata, replace = FALSE, size = 0.80)

Yacht_Data_Test <- anti_join(mydata, Yacht_Data_Train)

Yacht_Data_Test

set.seed(12321)

Yacht_NN1 <- neuralnet(prr ~enr, data = Yacht_Data_Train)
```

```
plot(Yacht_NN1, rep = 'best')

NN1_Train_SSE <- sum((Yacht_NN1$net.result - Yacht_Data_Train[, 7])^2)/2

paste("SSE: ", round(NN1_Train_SSE, 4))

Test_NN1_Output <- neuralnet::compute(Yacht_NN1, Yacht_Data_Test[, 1:6])$net.result

NN1_Test_SSE <- sum((Test_NN1_Output - Yacht_Data_Test[, 7])^2)/2

NN1_Test_SSE



mydata.train=Yacht_Data_Train[,1:8]

Yacht_NN2$net.result - mydata.train[5]



set.seed(12321)

Yacht_NN2 <- neuralnet(prr ~dfr, data = Yacht_Data_Train,act.fct = "logistic")

## Training Error

Test_NN2_Output <- neuralnet::compute(Yacht_NN2, Yacht_Data_Test[,7:8])$net.result

NN2_Test_SSE <- sum((Test_NN2_Output - Yacht_Data_Test[, 7])^2)/2



set.seed(12321)

Yacht_NN3 <- neuralnet(prr ~drr, data = Yacht_Data_Train,act.fct = "logistic")

## Training Error
```

```
Test_NN3_Output <- neuralnet::compute(Yacht_NN3, Yacht_Data_Test[,6:7])$net.result

NN2_Test_SSE <- sum((Test_NN3_Output - Yacht_Data_Test[, 7])^2)/2

# Bar plot of results

Regression_NN_Errors <- tibble(Network = rep(c("NN1", "NN2","NN3"), each = 2), DataSet
= rep(c("Test", "Train"), time = 3),  SSE = c(NN1_Train_SSE, NN1_Test_SSE,
NN2_Train_SSE, NN2_Test_SSE,NN3_Train_SSE, NN3_Test_SSE))

Regression_NN_Errors %>% ggplot(aes(Network, SSE, fill = DataSet)) +  geom_col(position
= "dodge") +   ggtitle("Regression ANN's SSE")



library(neuralnet)

library(tidyverse)

library(neuralnet)

library(GGally)

mydata<-read.csv(file.choose(),header=TRUE)

mydata

attach(mydata)



library(caret)

my.grid <- expand.grid(.decay = c(0.5, 0.1), .size = c(5, 6, 7))

prestige.fit <- train(pr ~ enr, data = trainset,
```

```
    method = "nnet", maxit = 1000, tuneGrid = my.grid, trace = F, linout = 1)

prestige.predict <- predict(prestige.fit, newdata = trainset)

prestige.predict

prestige.rmse <- sqrt(mean((prestige.predict - trainset$enr)^2))

prestige.rmse

prestige.fit <- train(pr ~ enr, data = trainset,method = "nnet", maxit = 1000, tuneGrid =
my.grid, trace = F, linout = 0)

prestige.fit

library(caret)

final <- factor(c("true", "true", "true", "true", "Normal"))

levels(final)

test <- factor(c("true", "true", "true", "true"))

levels(test)

levels(test) <- c("true", "Normal")

confusionMatrix(final, test)



# example of a confusion matrix in R

library(caret)
```

```
expected <- factor(c(1, 1, 0, 1, 0, 0, 1, 0, 0, 0))

predicted <- factor(c(1, 0, 0, 1, 0, 0, 1, 1, 1, 0))

results <- confusionMatrix(data=predicted, reference=expected)

print(results)

library(nnet)

install.packages("e1071")

library(caret)

library(e1071)

data(rock)

enr <- as.factor(enr)

nnclas_model <- nnet(enr ~ df, data = mydata, size = 4, decay = 0.0001, maxit = 500)

x <- mydata[, 1:3]

y <- mydata[, 4]

yhat <- predict(nnclas_model, x, type = 'class')

confusionMatrix(as.factor(yhat), y)

library(neuralnet)

data<-read.csv(file.choose(),header=TRUE)

data

attach(data)
```

```
x1

x2

x3

##########

scaleddata<-scale(mydata)

#Data Normalization

#Again, we normalize our data and split into training and test data:

# MAX-MIN NORMALIZATION

normalize <- function(x1) {

 return ((x1 - min(x1)) / (max(x1) - min(x1)))

}

maxmindf <- as.data.frame(lapply(data, normalize))

> #Progression rate

> predicted=results$prediction * abs(diff(range(drr))) + min(drr)

> actual=results$actual * abs(diff(range(drr))) + min(drr)

> comparison=data.frame(predicted,actual)

> deviation=((actual-predicted)/actual)
```

```
> comparison=data.frame(predicted,actual,deviation)

> accuracy=1-abs(mean(deviation))

> accuracy

[1] 0.785834

> predicted=results$prediction * abs(diff(range(drr))) + min(drr)
```

Appendix one Comparison of test and training data performance evaluation

Performance	**Test data**	**Training data**
Mean sq error	0.552148	0.042148
Mean abs error	2.6745	2.14
MIN ABS ERROR	0.0458226	0.000228226
MAX ABS ERROR	3.008	1.15
PERCENT CORRECT	73.0	78.5