



2017

A Model for the classification of student neediness using artificial neural networks

Eunice Engefu Manyasi
Faculty of Information Technology (FIT)
Strathmore University

Follow this and additional works at <https://su-plus.strathmore.edu/handle/11071/5631>

Recommended Citation

Manyasi, E. E. (2017). *A Model for the classification of student neediness using artificial neural networks* (Thesis). Strathmore University. Retrieved from <http://su-plus.strathmore.edu/handle/11071/5631>

A Model for the Classification of Student Neediness using Artificial Neural Networks

By Eunice Engefu Manyasi

**Submitted in partial fulfilment of the requirements of the Degree of Master of Science
in Information Technology (MSc. IT)**

Faculty of Information Technology

Strathmore University

Nairobi, Kenya

June, 2017

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement

Declaration

I declare that this work has not been previously submitted and approved for the award of a Degree by this or any other University. To the best of my knowledge and belief, the research Thesis contains no material previously published or written by another person except where due reference is made in the research proposal itself.

© No part of this thesis may be reproduced without the permission of the author and Strathmore University.

Eunice Manyasi Engefu

.....
.....

Approval

The thesis of Eunice Engefu Manyasi was reviewed and approved by the following:

Professor Ismail Ateya Lukandu
Senior Lecturer, Faculty of Information Technology
Strathmore University

Dr. Joseph Orero (PhD)
Dean, Faculty of Information Technology
Strathmore University

Professor Ruth Kiraka
Dean, School of Graduate Studies
Strathmore University

Abstract

Financial aid has been used worldwide to assist students at higher learning institutions finance their education. The aid has majorly been offered by the government, private companies and non-governmental institutions in form of loans, grants, scholarships and work study programs. It has made great progress in increasing the enrolment rate of students to higher learning institutions. The aid is usually given to applicants who have been selected after applying for the aid, and a committee ensuring that they have meet the set criteria to be awarded. Currently the number of applicants applying for financial aid has increased leading to challenges of errors and bias in the selection and award process due to overwhelming data which becomes too complex for the committees to analyse. This has led to some more deserving students not receiving the financial aid due to inaccuracies. Artificial intelligence has been applied in various fields for the analysis and classification of huge amounts of data. It has been applied in finance to predict the credit rating of customers which uses a similar concept in classification of applicants. The research sought to apply machine learning to in the selection and award process of needy students. Historical financial aid data which was labelled as awarded and not awarded, was used to train the feed forward neural network learning model. The inputs used included parents occupation and income, family income and family spending. The research employed experimental research to determine the variables that best identified the needy students and qualitative research to get the ideas and opinions of participants with regards to the study. The model accurately classified 2955 instances as true positives and 18 instances as true negative out of 3043 instances, giving it a 97.6% accuracy.

Key terms: Artificial neural networks, Financial aid allocation, Supervised learning.

Dedication

I dedicate this work to Almighty God for giving me good health and strength to be able to do the masters, my friend Nancy Gacheri for always encouraging and my parents for their full support, my sister Emily Manyasi and brother Alex Dodi Manyasi for encouraging and believing in me.

Acknowledgment

I am grateful for the support received from my supervisor Prof Ismail Ateya through his guidance on research. I would like to acknowledge my colleagues Miss Victoria Wasonga, Mr Bernard Alaka and Mr Stephen Obonyo for encouragement and guidance through the research. I also acknowledge my family members for making it easy for me to work on the project by supporting in their own special ways. I thank the financial aid office at Strathmore University for providing me with the information I needed for my research.

List of Abbreviations/Acronyms

ANN	-	Artificial neural networks
DFD	-	Data flow Diagram
EFC	-	Expected Family Contribution
HELB	-	Higher Education Loans Board
SQL	-	Structured Query Language
UI	-	User Interface
UML	-	Unified Modelling Language

Table of contents

Declaration.....	ii
Approval.....	ii
Abstract.....	iii
Dedication.....	iv
Acknowledgment.....	v
List of Abbreviations/Acronyms.....	vi
Table of contents.....	vii
List of figures.....	xi
List of tables.....	xii
List of Equations.....	xiii
Chapter 1: Introduction.....	1
1.1 Background.....	1
1.2 Problem Statement.....	2
1.3 Research Objectives.....	2
1.4 Research Questions.....	3
1.5 Justification.....	3
1.6 Scope.....	3
1.7 Limitation.....	4
Chapter 2 Literature Review.....	5
2.1 Introduction.....	5
2.2 Financial Aid in Higher Education.....	5
2.3 Allocation of Financial Aid to Applicants.....	6
2.4 Machine Learning Methods used to Solve Classification Problems.....	7
2.5 Classification using Artificial Neural Networks (ANNs).....	8
2.6 Models used in Identifying neediness.....	9

2.6.1	Integrating RFM Model and Cluster for Students Loan Subsidy Valuation	9
2.6.2	College Student Scholarships and Subsidies Granting: A Multi-Modal Multi-Label Approach	10
2.6.3	Multinomial Logistic Regression Model for Classifying HELB Applicants.....	11
2.6.4	Means Testing based on Least Squares Regression Predictor	13
2.7	Limitations of the methods used	14
2.7.1	Using linear regression predictor	15
2.7.2	Using RFM model.....	15
2.8	Conceptual Framework	15
Chapter 3 Research Methodology		17
3.1	Introduction	17
3.2	Research Design.....	17
3.3	Location of the study.....	17
3.4	Target Population	17
3.5	Research Instruments	18
3.5.1	Questionnaires.....	18
3.5.2	Personal Interviews	18
3.6	Data Collection Procedure	18
3.6.1	Interview process	18
3.6.2	Questionnaire process	19
3.7	Data Analysis Procedure	19
3.8	System Design.....	19
3.8.1	Data Flow Diagrams	19
3.8.2	Use Case Diagrams	19
3.8.3	System Sequence Diagram and Collaboration Diagrams	20
3.8.4	Class Diagrams	20
3.8.5	Database Schema which includes ERD	20

3.9	System Implementation.....	20
3.10	System Testing	20
3.11	System Evaluation/Validation.....	20
Chapter 4 Design and Architecture		21
4.1	Introduction	21
4.2	Data Analysis	21
4.3	Requirement Analysis	21
4.3.1	Functional Requirements	21
4.3.2	Non-Functional Requirements	22
4.4	Design Phase	22
4.4.1	System Architecture.....	22
4.4.2	Use Case Diagram.....	23
4.4.3	Data Flow Diagram.....	26
4.4.4	Sequence Diagram	27
4.4.5	Class Diagram.....	28
Chapter 5 Implementation and Testing		30
5.1	Introduction	30
5.2	Model Environment.....	30
5.2.1	Web Application	30
5.2.2	Neural Network Components	31
5.3	Model Implementation	31
5.3.1	Data Input.....	31
5.3.2	Data Normalization.....	34
5.3.3	Network training	34
5.4	Model Architecture	35
5.5	Validation and Testing	35
5.5.1	Model Testing	35

5.5.2	System Testing.....	36
5.5.3	Acceptance Testing.....	37
Chapter 6	Discussions and Findings	38
6.1	Introduction	38
6.2	Findings of the Research.....	38
6.3	Discussions.....	39
6.4	Contributions of the model to research	39
6.5	Shortfalls of the Research.	39
Chapter 7	Conclusions and Recommendations.....	40
7.1	Conclusions.....	40
7.2	Recommendations	40
7.3	Suggestions for Future Studies.....	41
References.....		42
Appendix A	: Turnitin Report.....	45
Appendix B	: Financial Aid Interview Guide	46
Appendix C	: Financial Aid Interview Feedback	47
Appendix D:	Verify Applicants Information Use Case Description	49
Appendix E	: Allocate Financial Aid Award Use Case Description	50

List of figures

Figure 2-1 Classification Procedure using Machine Learning Algorithms	8
Figure 2-2 Framework to Evaluate Loan Subsidy Valuation	10
Figure 2-3 CS ³ G Approach	11
Figure 2-4 Conceptual Framework	16
Figure 4-1 System Architecture	23
Figure 4-2 Use case diagram.....	24
Figure 4-3 Data Flow Diagram	27
Figure 4-4 Sequence Diagram	28
Figure 4-5 Class Diagram	29
Figure 5-1 Login page.....	32
Figure 5-2 Application Form	33
Figure 5-3 View Applicants.....	33
Figure 5-4 Model Architecture	35
Figure 5-5: Confusion Matrix	36

List of tables

Table 2-1 Financial Need Calculation	7
Table 4-1 Apply for financial aid use case	25
Table 5-1 System Testing	36
Table 5-2 Acceptance Testing	37
Table 6-1 Confusion Matrix.....	38

List of Equations

Equation 2-1 Multinomial Regression	12
Equation 2-2 Multinomial Regression	12
Equation 2-3 Least Square Regression	13
Equation 5-1 Normalization.....	34

Chapter 1: Introduction

1.1 Background

Student financial aid programs were introduced to assist students pay for their higher education and related expenses, Johnstone (2003). The programs started after the introduction of cost sharing of higher education, where the costs of education which were fully paid by the government were now shared with parents and government (Johnstone, 2003). They offer financial assistance in forms of scholarships, grants, loans and workstudy programs. According to Gross, Hossler, & Ziskin, (2007) recent studies have been done by various scholars on the impact of financial aid in enrolment and participation in higher education, to low and middle income students, where it was found that financial aid has led to increase in their participation.

In order for students to be given financial assistance, Ye, Zhan, Li, Huang, & Jiang, (2016) state that the allocation of financial aid goes through two main stages: application and allocation. The allocations mainly depend on student grades and their backgrounds. According to Ngolovoi (2008) financing institutions such as the higher education loans board in Kenya use means testing method to allocate financial aid based on student financial ability. Means testing involves the analysis of income information and categorical indicators such as parent's occupation, and number of supported siblings and parental income, to determine the financial ability of the applicant (Ngolovoi, 2008). The process involves getting applicants personal information in addition to supporting documents to show their financial status.

The allocation of financial aid has similar structure to the credit ratings where borrowers are classified according to their risks. In credit ratings, learning models such as artificial neural networks (ANN) and support vector machines (SVM) have been applied in prediction and have improved the level of accuracy (Saitoh, 2016) . Learning models have been used in other fields such as speech , vision, and pattern recognition to perform complex functions. Artificial neuron networks are trained using past data where they learn the patterns among given inputs and the expected outputs. After training, the networks can be used to predict the values of new inputs (Jha, 2017) . Wabwoba & Mwakondo, (2011) have applied neural networks to determine students qualified to join public universities under government

sponsorship. Allocation of financial aid to applicants can also be done using artificial neural networks to improve accuracy and save on human resource capital.

1.2 Problem Statement

Financial aid award is an important decision made by the financial aid offices, mainly based on financial need and students' grade. The programs have limited resources hence maximize on ensuring that they reach the low income students, to expand their access to higher education (Hopper, 2001). This requires accuracy in identifying the needy students, and allocating them financial assistance in addition to ensuring that the financially able applicants are not misclassified as needy.

Currently, financial aid is allocated based on committees who determine the applicant to be awarded financial aid based on a criteria outlined by the financial aid institution. The committees go through the financial information given by the applicant, together with the categorical indicators such as secondary school attended by applicant, parent's age and number of dependent siblings (Johnstone , 2003). After analysing various applications the committees decide on which applicants should be allocated the aid. With increase in the number of applicants applying for financial aid, the amount of data to be analysed by the committees has also increased, leading to higher chances of inaccuracy and bias, hence students who need the aid are left out. This has led to the need of using a learning model in the classification system to increase accuracy in classifying applicants hence enable them award loans to the right applicants.

There is need for a financial aid classification system that uses artificial neural networks to allocate financial aid to applicants, since it has been proven to be accurate in performing complex classification problems (Jha, 2017). This will enable the process of selection and award of financial aid to the more efficient and accurate.

1.3 Research Objectives

- i. To investigate the data requirements for financial aid application.
- ii. To investigate challenges faced by the financial aid office using the current systems of selection and award of financial aid.

- iii. To review the methods, systems, frameworks and models used in classification of applicants
- iv. To develop a prototype that allocates financial aid to applicants using artificial neural networks
- v. To test the prototype developed using data of financial aid applicants.

1.4 Research Questions

- i) What are the data requirements for financial aid application?
- ii) What are the challenges associated with the current systems of selection and award of financial aid.
- iii) What are the methods, systems, frameworks and models that have been used to classify applicants?
- iv) How will the prototype be designed?
- v) How will the prototype be tested?

1.5 Justification

According to Johnstone (2003), the introduction of cost sharing to higher education led to an increase in tuition fees making it expensive to access higher education, hence more people depend on financial aid for financial support. Financial aid allocation is an important resource allocation decision, which affects the number of talented students enrolled in higher education from poor backgrounds. Thus, a classification system which accurately identifies the needy students will ensure that more students from poor background have access to higher education. This research can also be applied to other social programs that have an aim of distributing limited resources by targeting the needy in the society.

1.6 Scope

The research focused on the process of selecting and allocating of financial aid to applicants based on need and student performance. The allocation was dependent on past criteria of allocations done by the financial aid office.

1.7 Limitation

The researcher was limited to carrying out the research at the Strathmore University financial aid office as it was easily accessible. The determination of the amount of financial aid to allocate individual applicants, in addition to determining whether or not to award would have been more appropriate for the research, however, due to limited dataset, it was not included in the research.

Chapter 2 Literature Review

2.1 Introduction

The growth of demand for higher education, and increased costs of tuition fees has led to an increase in the number of ways to assist students financially, majorly through financial aid institutions. This chapter looks at the selection financial aid applicants, to determine which student needs financial assistance more, and classification which determines the amount of aid they will get. It also highlights the challenges faced in identifying the applicants who need the aid, and discusses the process followed in awarding the aid. Empirical models and frameworks that have been proposed by other researchers, to classify applicants and allocate aid are also discussed.

2.2 Financial Aid in Higher Education

In universities, financial aid has been used as a means to finance the education of students, mainly in terms of grants, loans and scholarship. In Kenya the higher education loans board is the largest students financing institution. According to Gichuhi (2015) HELB has financed over 300,000 students both privately and publicly sponsored, since its inception in 1995 by majorly giving student loans. In addition to HELB, the government has created the constituent development fund (CDF) and local authority transfer funds to finance students education through bursaries. The bursaries and loans are given at a competitive basis where the most needy and bright students are given higher consideration.

The provision of financial assistance to higher education has assisted students in various ways. According to Johnstone & Marcucci (2010) financial aid has enabled applicants from low income families participate in higher education, based on a combination of means, merit and financial need. Student loans have also provided performance incentives to students at the universities. Otieno (2002) states that student loans have been used to promote equality of opportunity to qualified students irrespective of their backgrounds, reduced dropout rates; as a result, contributed to national development.

Financial aid data, has also been used by learning institutions to make enrolment decisions. According to Thanh & Haddawy (2007) financial aid has been used to achieve a number of enrollment objectives including having a diverse student population, maximizing

tuition revenue and attracting strong students, hence the accurate allocation of the aid is crucial

Gichuhi (2015) states that despite the efforts put by the government and other financial institutions to finance higher education, there is a critical concern on the number of people from rural population and economically disadvantaged population. The funds for financial aid are usually limited, and come from government, private companies and international agencies, and hence, the accurate allocation of the funds is important to ensure its effectiveness (Salmi, 2001).

2.3 Allocation of Financial Aid to Applicants

Financial aid is allocated to applicants on various basis specific to the financing institution. The process of awarding financial aid goes through four stages which include: application, verification, classification and allocation. During application, applicants are required to have meet minimum requirements as set out by the financing institution and submit required documents as set out in the application form. The most common financial aid is need based. The financing institutions usually request for documents and indicators that will be used to determine the financial status of the applicant (Ngolovoi, 2008). To ensure that the documents given are valid, the financial aid institutions request the institutions that issue the documents to validate them. According to Johnstone & Marcucci (2010) the proper indicators used to measure a persons ability to pay can be chosen by considering: a) The cost and effort involved in verifying the indicator, b) its manipulability c) Efficiency in targeting the correct group. The most common indicators used are income and assets, in addition, categorical indicators are used to supplement the means test.

After applicants submit their applications and the required documents, the information given is usually verified to ensure that it is truthful and correct, to enable only the eligible students receive financial aid. The verified information is then used to calculate the financial need. According to UniversityBrown (2017) financial need is calculated by getting the total cost of attendance, less the expected family contribution (EFC). The financial need is used to

classify the applicants into categories. The table below shows how financial need is calculated.

Cost of Attendance	-(Minus)	EFC	=(Equals)	Financial Need
<ul style="list-style-type: none"> • Tuition & Fees • Room & Board • Books & Supplies • Misc. Expenses • Travel 		<ul style="list-style-type: none"> • Parent Contribution • Student Contribution from Assets • Student Contribution from Summer Earnings 		<ul style="list-style-type: none"> • Student Loans • Work-Study • Federal or State Grants • University Scholarship

Table 2-1 Financial Need Calculation (UniversityBrown, 2017)

The classification of students based on financial need, has been faced with challenges, which include having errors of inclusion and exclusion. In inclusion errors, the non-poor are included as needy applicants leading to financial aid being given to wrong applicants while in the exclusion errors, the poor applicants are not identified hence excluded from getting the financial aid, (Johnstone et al 2010). These errors increase with the use of inaccurate methods of classifying students hence the need for more accurate methods. To overcome these problems, a number of machine learning models and frameworks have been proposed to identify needy students and allocate them aid.

2.4 Machine Learning Methods used to Solve Classification Problems

Machine learning is a field in artificial intelligence which has largely been used in data mining, to analyse and identify relationships between patterns in large datasets, hence solve difficult problems. Machine learning algorithms have been applied in systems to solve the problems, improving their efficiency. They include: Decision trees, neural networks, K Nearest neighbour and Support Vector machines.

The algorithms have been used to solve classification problems such as credit rating of customer. According to Kotsiantis (2007), the process of applying the algorithms to solve the classification problems starts with the collection of a dataset which is then pre-processed to remove noise. The data is then divided into a training set and a test set. An algorithm is then selected to be used for training and the values evaluated using the test set. The algorithm is selected based on the type of the problem and the accuracy of the algorithm in solving the

problem. The figure below illustrates the process of applying machine learning algorithm to solve a real-world problem.

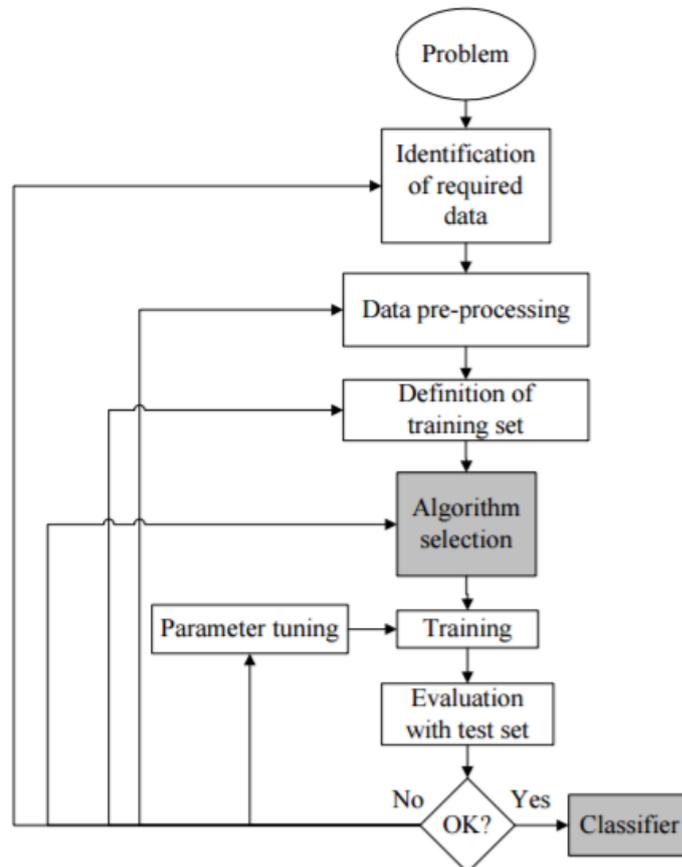


Figure 2-1 Classification Procedure using Machine Learning Algorithms (Kotsiantis, 2007)

2.5 Classification using Artificial Neural Networks (ANNs)

Artificial neural networks (ANNs) are composed of connected neurons, which receive input signals, processes the signals and produce an output. The neurons learn from example just like a human brain, and capture functional relationships between data, (Hamzic et al ,2016). Neurons in artificial neural networks are arranged in three layers which include the input layer, hidden layer and output layer. The learning in neurons is accomplished by adjusting of the weights until the difference between the desired output and the network output is negligible (Pattanayak et al, 2013).

Artificial neural networks involves two steps namely training and learning. The training of feed forward networks uses labelled training sets that is ones that have and input with their corresponding output. The inputs are selected carefully as they highly influence the success of the training. As the neural network learns, it constructs a mapping of the input and output, as it adjusts the weights and biases for each iteration based on minimizing the error between the actual output and the expected output, until an acceptable value is reached (Iranmanesh & Zarezadeh, 2008)

2.6 Models used in Identifying neediness

This section introduces some of the models that have been proposed to identify needy students. The models include: Integrating RFM model and Cluster for Students Loan Subsidy Valuation, College Student Scholarships and Subsidies Granting: A Multi-Modal Multi-Label Approach and Multinomial Logistic Regression Model for Classifying HELB applicants.

2.6.1 Integrating RFM Model and Cluster for Students Loan Subsidy Valuation

Recency frequency and Monetary (RFM) model and clustering algorithm has been used by Dan, (2008) to build a framework for identifying needy students and determine their loan valuation. RFM uses information about customers past behavior in terms of Recency, how long ago the customer made the last purchase; Frequency, the number of purchases made by the customer; and Monetary, the amount spent by the customer; to predict the trends of a particular customer. The research is based on a university in China where students get food supplies from a canteen which provides them with a smartcard, and the canteen database stores records of their transactions. The RFM model used information from the canteen database to identify needy students whereby the three attributes were calculated as:

R (Recency) as ratio=average cost non-grain foods/average cost staple,

Frequency= the number of monthly consumptions,

Monetary=Monthly consumption

The attributes were given weights using analytic hierarchy process (AHP), after which they were analysed through RFM binning, using descriptive statistics such that poverty

stricken students had a high RFM score. The result was used to calculate RFM score as follows:

$$\text{RFM Score} = (\text{Recency score} * \text{Recency Weight}) + (\text{Frequency score} * \text{Frequency Weight}) + (\text{Monetary Score} * \text{Monetary Weight})$$

The scores were clustered and segmented using K-Means clustering and T-test. Poverty stricken students were clustered into groups according to their needs. The figure below represents the framework used in identifying needy students for loan subsidy valuation.

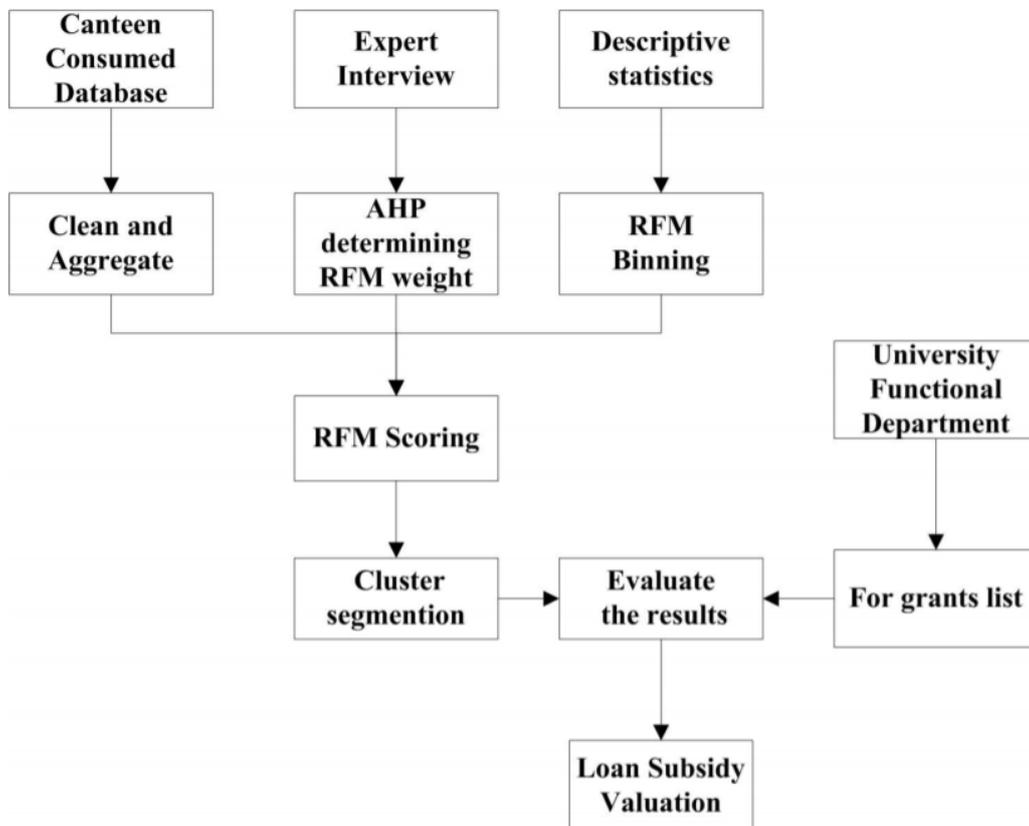


Figure 2-2 Framework to Evaluate Loan Subsidy Valuation (Dan,(2008))

2.6.2 College Student Scholarships and Subsidies Granting: A Multi-Modal Multi-Label Approach

A Multi-modal Multi-label approach has been proposed by Han-Jia Ye et al ,(2016) to determine types and amount of granting, given to students in universities. This approach

uses students information collected through multi-modal channels; which include their behavior of internet usage, campus consumption, daily trajectory together with their enrollment information, which collects students personal and family information. The approach organises and transforms these features/ modalities in a label-specific manner (LIFT) which are then used to create a multi-modal multi-label approach to make predictions for students scholarships and subsidies allocations using a semi-supervised method. On each modality, classifiers are learned, with regularizers to weigh the different modalities and discover the correlations between the features. Each modality produces a prediction result which is regularized using ‘rank’ operator, to ensure modal consistency. The diagram on figure 2-3 illustrates the process used by the multi-modal , multi label approach to allocate subsidies and scholarships to applicants, known as the College Student Scholarships and Subsides Granting (CS³ G) approach.

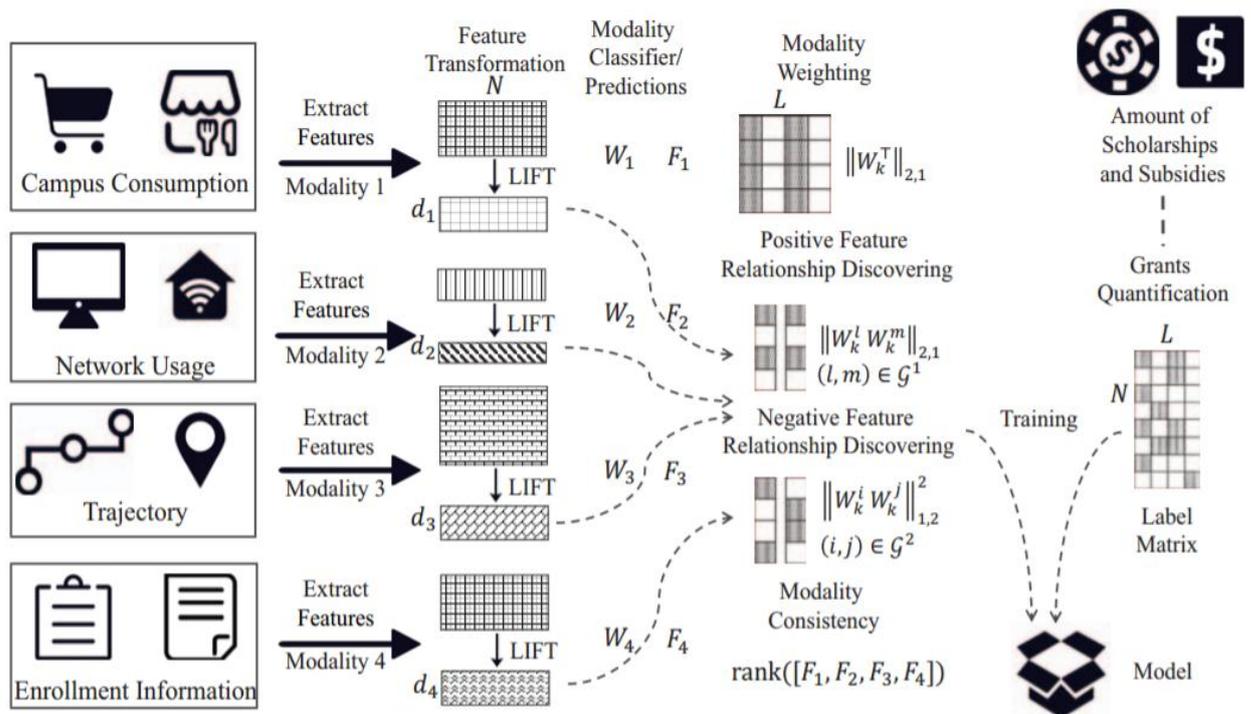


Figure 2-3 CS³ G Approach (Han-Jia Ye et al, (2016))

2.6.3 Multinomial Logistic Regression Model for Classifying HELB Applicants

A study by Muriithi et al, used multinomial logistic regression model to classify HELB applicants into categories for loans and predict their appropriate allocations. The data was collected using questionnaires from HELB applicants from Kenyatta University, Jomo Kenyatta University Of Agriculture and technology (JKUAT) and Kimathi University College of Technology (KUCT.) The data was used to assess financial need by passing it through a regression model. After assessing financial need, applicants are divided into categories which are used to allocate financial aid.

HELB allocates loans using family income and secondary school attended by an applicant. An applicant who has a single parent with income equal to or less than 250,000 is eligible for a loan of Ksh 55,000 and bursary of Ksh 7,000, if female and males Ksh 6,000, however, the secondary school attended is used to modify these amounts Johnstone & Marcucci,(2010). This research study used only three key indicators which included house worth, wealth and cost of fees.

The models were generated that included all predictor variables that were useful in predicting the response variable (Muriithi, Njoroge, et al 2013) .The multinomial regression equation used is given below:

$$P(Y_i = r) = \frac{\exp(\beta_{r0} + z_i' \beta_r)}{1 + \sum_{s=1}^m \exp(\beta_{s0} + z_i' \beta_s)}$$

Equation 2-1 Multinomial Regression (Muriithi & Okongo, (2013))

which can be equivalently written as:

$$\text{Log} \frac{P(Y_i = r)}{P(Y_i = k)} = \beta_{r0} + z_i' \beta_r.$$

Equation 2-2 Multinomial Regression (Muriithi & Okongo, (2013))

In the analysis, they used the following predictor variables:

x_1 House worth- Computed from the total materials of the house they live in

x_2 Wealth - Sum of property owned by the parents of an applicant

x_3 Amount of fees - Cost of education of sibling of an applicant each year.

The response variable was:

$Y_i = 1$ if awarded Ksh 35,000, 2 if awarded Ksh 40,000, 3 if awarded Ksh 45,000, 4 if awarded Ksh 50,000, 5 if awarded Ksh 55,000, 6 if awarded Ksh 60,000.

The model correctly classified 79.8% of cases under study

2.6.4 Means Testing based on Least Squares Regression Predictor

This is a method of calculating welfare, using the least squares regression analysis of income proposed by proposed by Grosh and Baker (Castano, 2002). The method is based on the assumption that income is an appropriate measure of testing the welfare on individuals, and that variables related to housing and individual's conditions are highly correlated with the income measurements. The means test predictor uses a regression in which the income or consumption is taken as the dependent variable while the explanatory variable comprises of a set of welfare related, individual or household measures. Equation to obtain the predictor is:

$$y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_p X_{pi} + E_i$$

Equation 2-3 Least Square Regression (Castano, (2002))

Where:

y_i is the log monthly/annual household consumption per person (log per capita consumption expenditure is found to work well in regressions).

X_i is a set of variables describing the household, demographic and durable/asset characteristics, and the location variable, characteristics of the dwelling and its residents

X_j is a set of variables describing the economic environment of the household (these include employment status, other sources of income, mode of transportation, etc.).

β_1 and β_2 are parameters to be estimated.

E_i is the random error term assumed to be normally distributed with mean $(u) = 0$ and constant variance (σ^2) .

This regression equation is estimated using Ordinary Least Squares (OLS) method. The predictor is obtained using the equation is an approximate measurement of household welfare, and using the poverty line, potential benefits of a social program example education and health can be chosen. Using regression analysis, consumption expenditure (as a measure of household welfare) is regressed unto a set of explanatory variables that are correlated with household/individual welfare. The explanatory variables used were both qualitative and quantitative, while the regression method calculated only quantitative hence the regression method was limited. (Castano, 2002)

2.6.4.1 Procedure used in the Linear Regression

The original set of variables in the broad categories are introduced into a weighted OLS regression of (log) per capita monthly/annual consumption expenditure. Checks are made on the set of variables for possibility of multiple correlation, where some variables are included and some removed, to reduce problems of multiple correlation. A step wise regression is performed on the variables, to eliminate variables that are not statistically significant don't contribute to the model's overall explanatory power (R^2). Based on the set of variables entered in the regression, different models evolve, and each is analysed for targeting errors, which is the degree of under coverage, before the most efficient model is selected.

Under coverage is calculated by dividing the number of persons likely to be excluded with by the total number of persons who should get benefits. Leakage is calculated by dividing the number of persons incorrectly identified as being eligible by the number of persons served by the program. The estimates of the explanatory variables in the model of choice become the weights of the scoring formula which is then used to determine the eligibility of potential beneficiaries for the programme at hand.

2.7 Limitations of the methods used

There are various methods that have been used to identify the needy students. This subsection looks at a few of the limitations faced by the models.

2.7.1 Using linear regression predictor

Linear regression has been used to identify needy students by Castano (2002). The method has a few disadvantages according to Flom, (2017) linear regression has a few disadvantages: firstly it is limited to linear relationships where it assumes a straight line between the dependent and independent variable. It is also sensitive to outliers, where the outlier can have a huge effect on the results, example if a student has characteristics which are very different from other students could affect the results of the linear regression. Secondly data used in the linear regression must be independent, where they should not have any correlation. Students coming from the same geographical area and the same background tend to have similar characteristics which would not be identified by the linear regression.`

2.7.2 Using RFM model

RFM analysis is a technique used to group or segment existing customers based on historic behaviour. This model has been used by Dan, (2008) to identify needy students by checking their spending rate in the canteen. Using this method to identify needy students can have some errors caused by student's behaviours example some students may not buy food from the canteen because they are on a diet or some may prefer to get their food from home.

2.8 Conceptual Framework

The figure 2.4 illustrates the conceptual framework used to develop the research prototype. As the figure shows, the applicants' information will be collected and stored in a database after which it will be cleaned and normalized. The inputs will be selected from the data which include the parents and guardians' occupation, students' grades, and amount of aid requested by student, the family income and the family spending. The data will be divided into training set and test set where 60% of the data will be training set while 40% will be test set. The training set will be used together with the learning algorithm to train the model. The trained model will later be applied to the test set data to determine new values.

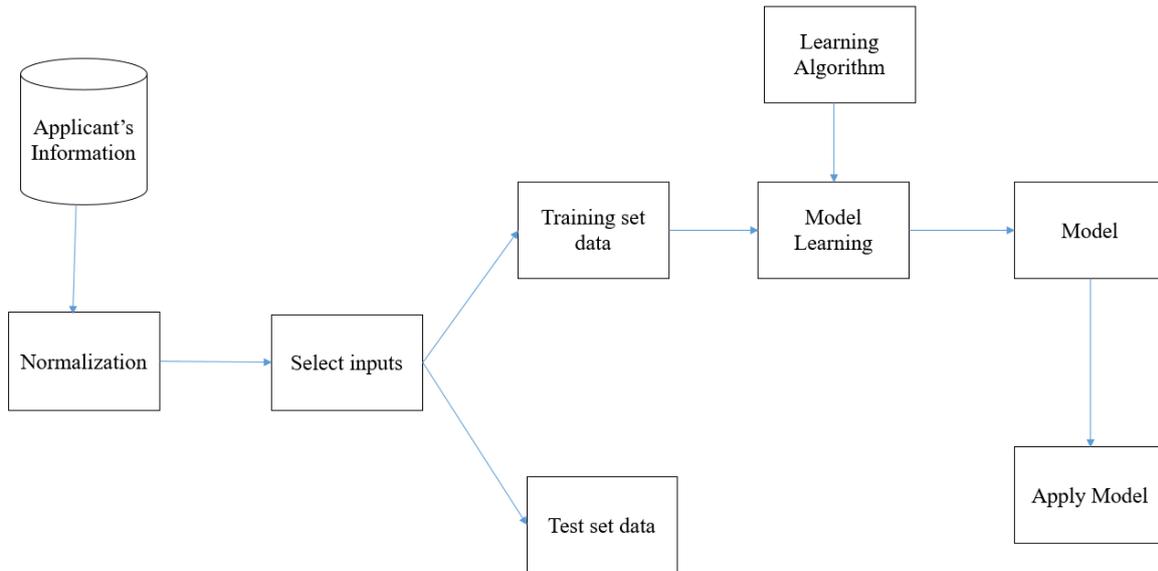


Figure 2-4 Conceptual Framework

Chapter 3 Research Methodology

3.1 Introduction

The study is aimed at finding out challenges experienced by student financial aid institutions in the process of selection and award of financial aid to applicants. Primary and secondary data were used to facilitate the research. Interviews guides were prepared, and were guided by the research objectives and a random population chosen to be interviewed.

3.2 Research Design

Parahoo (1997), describes a research design as “a plan that describes how, when and where data are to be collected and analysed”. This study focuses on identifying bright needy students to be awarded financial aid and uses an experimental research approach to find out which variables best identify needy students and qualitative research approach to understand the process followed to identify needy students, and explore the feelings and experiences of people of applying for financial aid. In the experimental design approach inputs such as students grade, parents occupation, family income, family spending are considered to find out how best they can be used to determine the neediness of a student. The experimental samples were taken from historical data, where there were two groups with significant differences, one with most needy students and another with well of students. A part of the historical data is used to test the experiment and new data is also used to see how well the model can determine neediness of the new data.

3.3 Location of the study

The study was carried out in financial aid offices of a higher learning institution, focusing on financial aid given to undergraduate students applying to the institution. The main focus was Strathmore University as it had the data that the researcher needed and had cooperative staff members who created an enabling environment for doing the research.

3.4 Target Population

The target population will involve two categories of the population, the first will be applicants who applied for the financial aid and were awarded aid, and applicants who are currently applying for the financial aid. The applicants who already have the financial aid will be used in the research to help the researcher identify patterns and characteristics of applicants who receive the financial aid, and the new applicants will be used by the researcher to test if the system will accurately determine whether or not the applicant

deserves to get financial aid. This would assist in comparing the effectiveness having the system and its accuracy.

3.5 Research Instruments

In this study, the researcher was the primary data collection instrument, which allowed the participants to give a more in-depth information regarding the financial aid process. Data was collected through interviews and questionnaires. These methods of data collection assisted the researcher to observe non-verbal communication and clarify unclear questions. The research comprised of two sections: collecting historical data and collecting participants' opinion on the working of the financial aid offices.

3.5.1 Questionnaires

Online surveys will be used to get feedback from the population to enable the researcher to come up with a prototype that addresses all or most of the challenges. The questionnaires will be given to applicants of the financial aid, the financial aid staff and the sponsors. Through the questionnaires, the researcher would be able to know: the sponsors view on the operations of financial aid and challenges faced, the financial aid staff view on the procedure followed in the financial aid process and the challenges faced, the challenges faced by the applicant while applying for financial aid.

3.5.2 Personal Interviews

The researcher will conduct personal interviews with the financial aid staff to fully understand the process followed by financial aid, and can easily get clarifications. Interviews can help the researcher know if the respondent is truthful and get first-hand information.

3.6 Data Collection Procedure

The researcher was the main research instrument in this research. The role of the researcher was to collect information through interviews and questionnaires. The process of data collection is discussed in section 3.6.1 and 3.6.2.

3.6.1 Interview process

The researcher introduced herself to the participants to establish a rapport, and informed them the purpose of the research study. She used an interview guide to ensure that

all areas were covered. The researcher asked questions during the interview for clarification taking into account both the verbal and non-verbal probing techniques, and also took notes of the ideas and opinions given. At the end of the interview, the researcher summarised it by restating in her own words the ideas and opinions of the interviewee to ensure understanding.

3.6.2 Questionnaire process

The questionnaires were created using the Survey Monkey tool and were sent to 10 individuals via their personal emails. The responders were students who had tried applying for financial aid and those who would like to apply for financial aid to give their opinions on the difficulties faced.

3.7 Data Analysis Procedure

Data analysis involves organisation of data. After data collection the analysis of the data commenced. The data collected via questionnaires were analysed using the survey monkey tool as all the responses will be collected using the tool. The tool enabled the researcher to summarise the data and make sense out of it. The data was categorised into groups and reports created from the information collected from the responses. These were then integrated with the feedback received from the interviews to come up with a conclusion of the core challenge faced in the financial aid award procedure hence proposed an experimental solution.

3.8 System Design

3.8.1 Data Flow Diagrams

The system will use data flow diagrams to show the flow of data. The system intends to get input data from the user which will be stored and retrieved for calculating means of the applicant's household'.

3.8.2 Use Case Diagrams

The use case diagrams will be used to show the interaction of the users with the system. The users will interact with the system while entering the applicant's information and awarding of the financial aids.

3.8.3 System Sequence Diagram and Collaboration Diagrams

Sequence diagrams will be used to show the flow of events in the system from the time the user logs in and applies for financial aid to the classification of applicants as eligible or non-eligible and awards of financial aid.

3.8.4 Class Diagrams

The research has used class diagrams to describe the structure of the system by showing classes, relationships and attributes and methods in the system structure.

3.8.5 Database Schema which includes ERD

The ERD has been used to show the database design according to the requirement analysis. It would show the entities that exist in the database and the relationships among the entities. The database schema shows the entities in the database, their attributes and data types.

3.9 System Implementation

The prototype system has been implemented using artificial neural networks to enable accurate awards of financial aid. The system will be piloted at the financial aid offices of Strathmore University to enable the researcher to get the required feedback from the financial aid experts to improve the system further and later implement it at national levels.

3.10 System Testing

The system will be tested by experts in the financial aid office to verify if the classification is accurate. The supporting documents given by the applicants will also be analysed to verify the accuracy of the system. The system will also be compared to the previously used systems to analyse its performance.

3.11 System Evaluation/Validation

The system was to be tested to ensure that the algorithm used is accurate, and validated to ensure that user has minimum chance of making mistakes while inputting data.

Chapter 4 Design and Architecture

4.1 Introduction

The researcher collected data through interviews held at the financial aid office. The data was analysed and user requirements drawn from the information received. The prototype was design using diagrams which included use case diagram, sequence diagram and data flow diagrams.

4.2 Data Analysis

The data was collected by doing interviews with the financial aid staff at the financial aid office. They assisted the researcher to better understand the challenges faced, and to come up with the user requirements that would help in building the prototype. The target population consisted of people who have worked in the financial aid office and have awarded financial aid to various applicants. From the interviews done 50% of the respondents thought given the current way of awarding financial aid, having an intelligent system would greatly improve the efficiency of the process. Most respondents highlighted the selection of applicants as a challenge given more than 100 applicants applying at the same time. In addition to that accounting of the donors' funds and tracking of the sponsored students' academic progress was identified as a challenge. In conclusion, the interviews were a major instrument in designing the prototype. The researcher choose to design a prototype that would solve the selection and award process of the financial aid after which more modules would be considered.

4.3 Requirement Analysis

The system requirements were both functional and non-functional which were concluded after conducting interviews. The requirements formed the basis of the expected functionalities of the proposed system.

4.3.1 Functional Requirements

The functional requirements of the system include the functionalities that are necessary for the system they are: System authentication: where the system should be able to ensure that only the authorised users are able to login in, and should only view data according to their user level. This would promote confidentiality and integrity of the data; applying for financial aid: After the applicant has been admitted to the university, he will be eligible to apply for financial aid, using the financial aid system. To ensure that the applicant is admitted

he will be required to login using his admission number. This function makes it easier for the financial aid office to review the applicant information and any supporting documents submitted; Verify applicants information: The financial aid staff should be able to record in the system whether he has verified the applicants information, after doing interviews and background checks; Receive notifications: The financial aid applicant should be able to receive notifications of the progress of the application, while the financial aid staff should be notified when applicants send their applications; Classify and award applicants: The financial aid staff should be able to log in to the system and award applicants using this functionality. This will enable eligible students to be awarded with financial aid using the system; Check application status: After applying for aid the applicant will be able to check his application status and know whether he is eligible for the award or not.

4.3.2 Non-Functional Requirements

The non-functional requirements specifies the properties of the system which will enable it efficiently perform the functional requirements. They include: Reliability: where the system should be able to be relied upon by the users and keep a copy of the data to ensure that in case of system failure there is data which can be used for recovery; Availability: where the system should be able to be accessed by the user when he wants to access it and have a minimal failure rate to ensure high availability; Extensibility: The system should be able to handle future growth and improvement and the source code should be well written to allow the system to be extended; Integrity: The financial aid data and results of the financial aid awards should not be able to be altered in any way, without recording the user who made the change; Security: The system should be able to maintain confidentiality and integrity of information by ensuring that only authorised users are able to view and edit data.

4.4 Design Phase

The section describes the design diagrams that were used to come up with the prototype. The diagrams include: Use case diagram: This is used to show how the various users will interact with the system and the functions they will perform; Data flow diagram: shows the flow of data in the system; Sequence diagram: illustrates the sequence of events in the system; Class diagram: shows the interaction of classes and their relations.

4.4.1 System Architecture

The figure 4-1 shows the general architecture of the system, which comprises of three components to facilitate the working of the system.

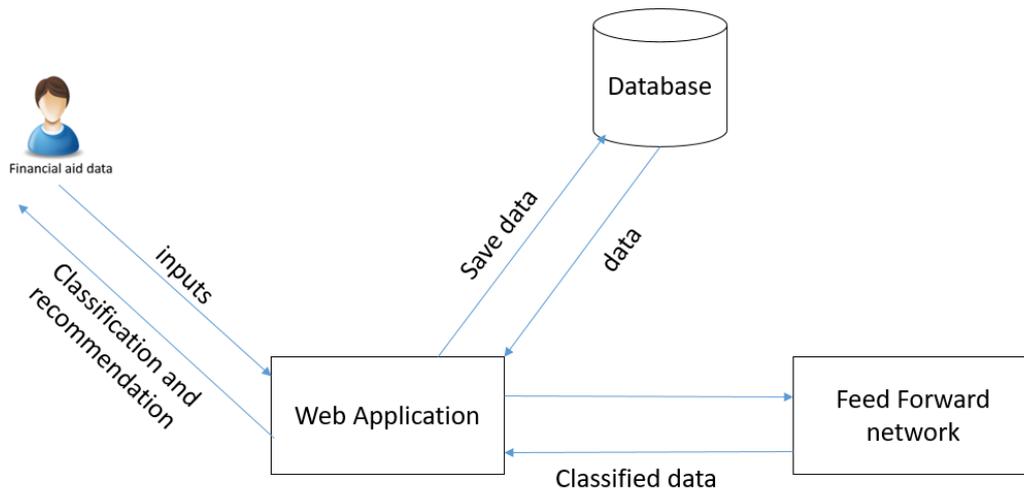


Figure 4-1 System Architecture

The architecture has three main basic components which include the web application module provides an interface where the financial aid applicants make their applications. During the applications, the applicant gives in information which include: personal information: date of birth, county, name, gender; academic information: school attended, grades; Parents Information: age, type of job, income; and the amount of aid they would like to apply for. The information is then saved in a database, from which they are retrieved and normalized using the min max scalar, to be in a format that can be used by the machine learning algorithm. The normalized data is then passed through the feed forward network to classify and award applicants. The results are then displayed on the interface of the web application.³

4.4.2 Use Case Diagram

The use case diagram on figure 4-2 illustrates the users of the system and the functionalities that they are able to perform in the system. The system will have actors which include:

- i) Financial aid staff: Responsible for managing the financial aid applications, and the system.

- ii) Financial aid applicant: This user interacts with the system to apply for financial aid. The application involves uploading of information which are critical for the classification of applicants.
- iii) System: The system will be an actor as it would perform operations critical to the award of the financial aid applicants.

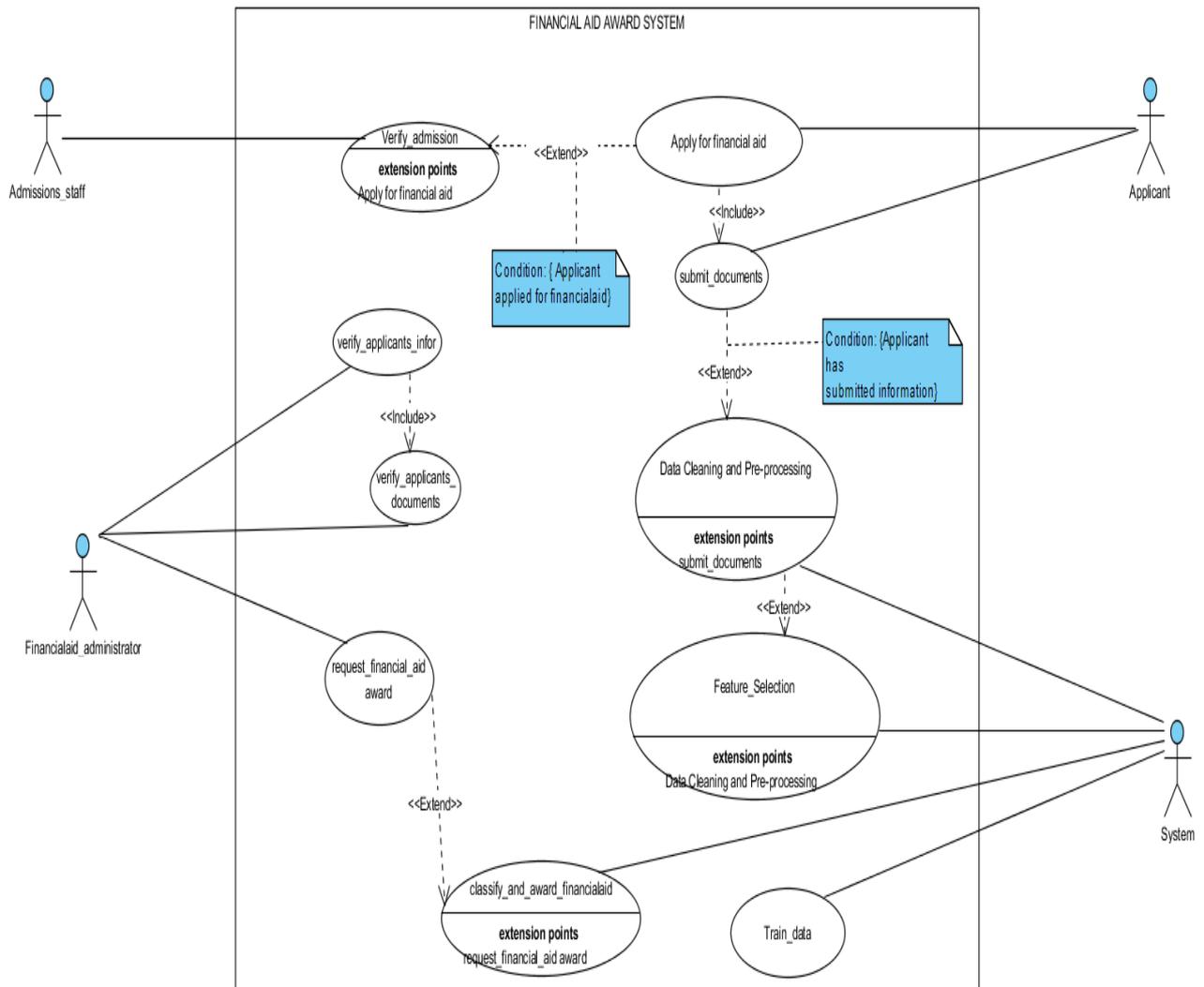


Figure 4-2 Use case diagram

4.4.2.1 Use Case Description

The use case descriptions describes the functionalities of the actor/system user and the pre-conditions and post-conditions requirements. The table 4-1 describes application for

financial aid use case. The use case description of allocation of financial aid award is on Appendix D of the document.

a) Applying for financial aid

Table 4-1 Apply for financial aid use case

Use case name: Apply for financial aid
Description : An applicant accesses the system and inputs the relevant information indicated on the system, and applies for financial aid
Primary Actor : Student (Financial aid applicant)
Pre-Conditions : Student has logged on to the system
Post-Conditions : Student has applied for financial aid
Success Scenario : <ul style="list-style-type: none"> i) Student selects apply for financial aid from the menu ii) System displays a list of text fields indicating which information is needed from the student iii) Student fills in the information and clicks submit iv) System saves the information and displays a confirmation message
Extensions: Student does not fill all required information: <ul style="list-style-type: none"> a) The system informs the user to ensure all fields are filled, highlighting the fields that have not been filled

Frequency of use : Every admission period
Status : Pending review by the financial aid office

The table 4-1 describes the use case scenario of applying for financial aid. The primary actor is the financial aid applicant. After logging in to the system the applicant applies for financial aid by filling in the required information. The pre-condition of applying for aid is to have been admitted for a course in the university. After applying for the aid the applicants is informed that his applications have been received.

4.4.3 Data Flow Diagram

The diagram on figure 4-3 shows the flow of data in the financial aid system. The financial aid applicant submits his information which is required for the application of financial aid. The information is saved in the database, from where it is normalized and saved in its normalised form. Feature selection is performed on the normalized data where features to be used in the selection process are selected. The data is partly used in the training of the model.

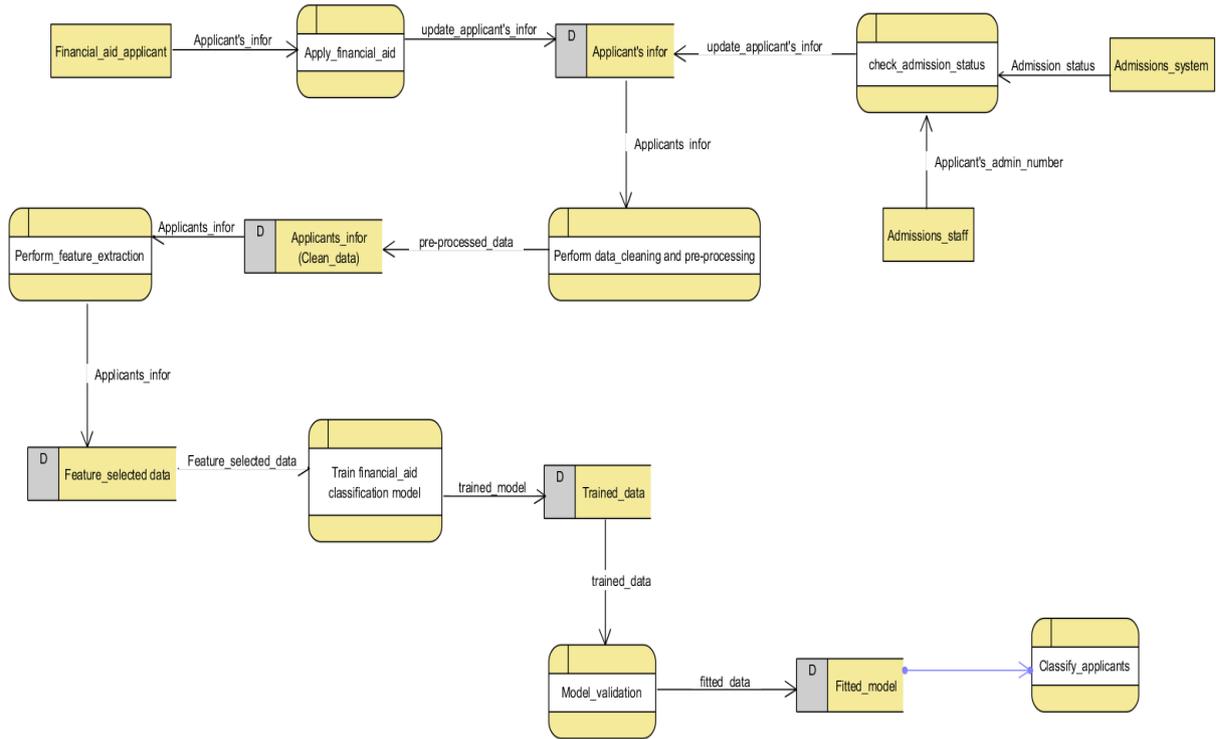


Figure 4-3 Data Flow Diagram

4.4.4 Sequence Diagram

The diagram on figure 4-4 below shows the flow of information in the system, from the time the user logs in to the classification of an applicant. The financial aid applicant applies for financial aid, after being admitted to the university, where the financial aid applications are stored together. . During application, the financial aid applicant is required to provide information, which will be used by the system to classify and award aid, after it has been verified by the financial aid staff. The admission staff verifies that the applicants are admitted to the university. The financial aid staff verifies the applicants' information and initializes the normalization of the data. The normalized data is used to train the model. The model is used for classifying and awarding new applicants.

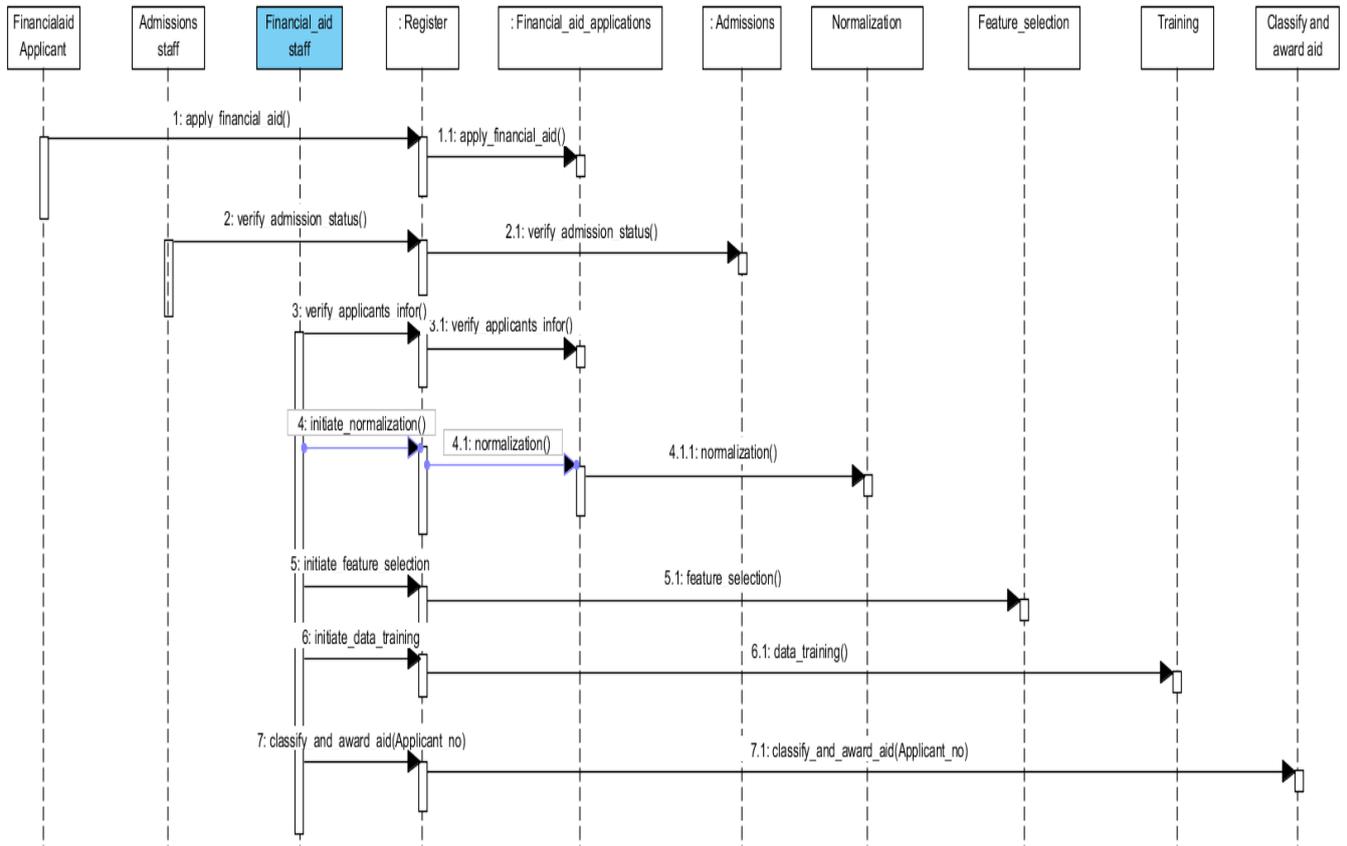


Figure 4-4 Sequence Diagram

4.4.5 Class Diagram

The diagram on figure 4-5 describes the class diagrams used in creation of the financial aid system prototype. The system has two users: Financial aid staff and the financial aid applicant. One financial aid staff is selected to manage the users. A donor donates money to the financial aid office, which is used for financial aid awards. A financial aid applicant is awarded financial aid one type of financial aid.

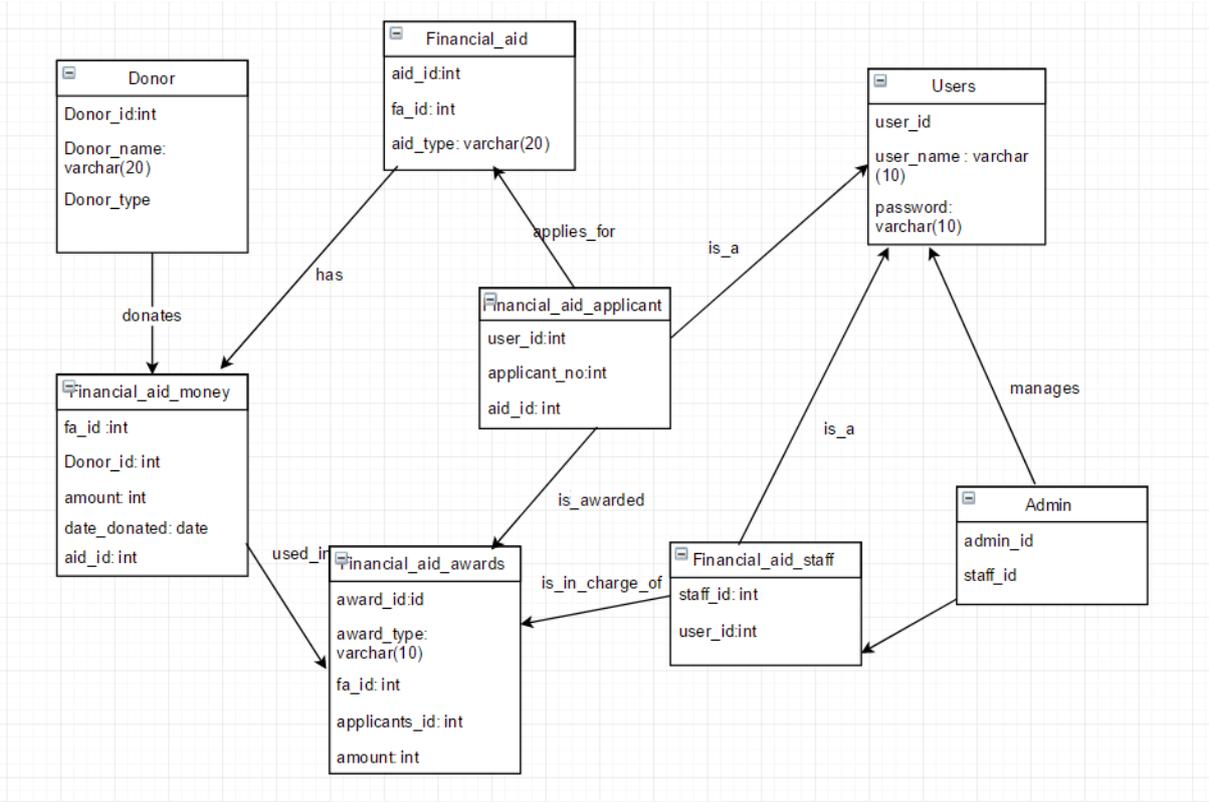


Figure 4-5 Class Diagram

Chapter 5 Implementation and Testing

5.1 Introduction

The prototype model was implemented through two main activities. The first step was to capture the data from the application by using a web application. The information was saved in a database from which it was obtained for allocation. The information was passed through a learning model which was created using artificial neural networks algorithm to allocate the awards. After allocation the financial aid staff receives a notification of the award.

5.2 Model Environment

The section discusses the components used in the implementation of the prototype. The prototype consisted of two parts the interface, which was created as a web application and the neural network model which was created using python Scikit learn library.

5.2.1 Web Application

The interface of the prototype was a web based application built using PHP as the server side language. The application was divided into various modules which include:

5.2.1.1 Users Module

The purpose of this module is to register users of different levels to the system. The financial aid applicants must have applied and admitted for a course in the university to be able to login to the system using their admission numbers. The financial aid staff must be registered to the system to be able to have rights to view the applicants' information.

5.2.1.2 Financial Aid Application Module

This module will be used by financial aid applicants to apply for financial aid. The applicants submit their personal information through this module. The information submitted will be viewed by the financial aid staff. The information will then be saved in the MYSQL database management system and will retrieved be used in the classification process.

5.2.1.3 Classification Module

This module will be used for classification of financial aid applicants. It will be available in the financial aid staff user view. The staff will be able to classify applicants through this module. The results of the module will be saved in database and displayed on the web application. The module uses neural network algorithm for classification of applicants to determine whether to award or not.

5.2.2 Neural Network Components

The prototype was created used the artificial neural network model for classification of applicants. The model consists of three layers which included the input layer, hidden layer and the output layer as described below.

5.2.2.1 Input Layer

This layer consists of features selected to be used in the model. It is the first layer of the neural network where the input values are collected, to be used in the neural network. The input layer also includes a bias neuron which is used to improve the accuracy of the output by shifting the activation function/decision boundary to the left or right. In the model, the information supplied by the applicant through the web application as selected as inputs to this layer.

5.2.2.2 Hidden Layer

This is the second layer of the neural network after the input layer. The inputs collected at the input layer are multiplied by weights which were in the range of -1 and +1 and summed together to produce a single output. The output was then passed through an activation function to ensure that the final output is within a threshold value. The model used the rectified linear unit activation function to transfer the input to an output of either 0 or 1.

5.2.2.3 Output Layer

The output layer produces the results of the neural network. The output categories depends on the specifications given. The output layer consisted of the targets which were the decisions to be made with regards to the financial aid. The model had two target outputs: awarded or not awarded, which were to be used as recommendation to the financial aid staff, for decision making.

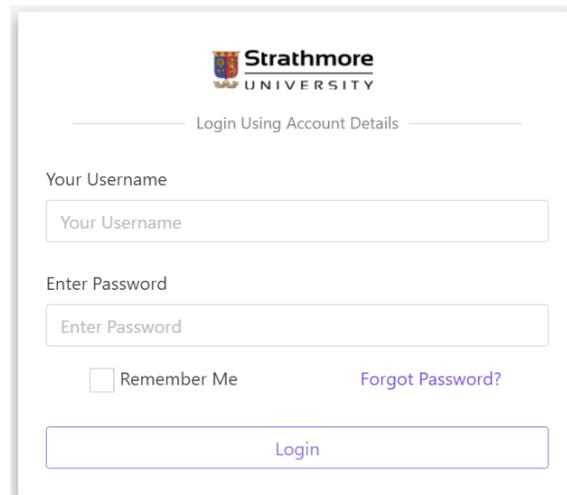
5.3 Model Implementation

This section describes how the prototype was built using the proposed model. The user interface was developed as a web based application, which was used for the collection and display of data while the model was created using python's scikit learn library, which was used for creation of the classification model. The application and the model were integrated to form the proposed prototype.

5.3.1 Data Input

A web application system was created to be used by both the financial aid applicant and the financial aid staff. The users had to login by providing their credentials for authorisation hence ensuring that the data was secure. The two users had different views as

both of them were interacting with the system to perform different functions. The applicant was to use the system to apply for financial aid and get receive any communication and feedback from the financial aid office, while the financial aid staff was to view the applications, verify them and initiate classification and award. The data inputs were validated using the interface on insertion to ensure that there were no missing or incorrect data. The figure 5-1 shows the login page of the web application.



Strathmore
UNIVERSITY

Login Using Account Details

Your Username

Your Username

Enter Password

Enter Password

Remember Me [Forgot Password?](#)

Login

Figure 5-1 Login page

Once the user logs in depending on the user level, they get different views. The financial aid applicant has an option of applying for financial aid where he fills in the form below. The details entered are saved in the MYSQL database. The form on figure 5-2 illustrates the application form used by the applicant.

The form is divided into four steps: 1. Personal, 2. Academic, 3. Parents Information, and 4. Financial Aid. The 'Personal' step is currently active and contains the following fields:

- Student Number : * (text input)
- First Name : * (text input)
- Last Name : * (text input)
- Date of Birth : (text input with format mm/dd/yyyy)
- Nationality : * (dropdown menu with 'Select Nationality')
- County : * (dropdown menu with 'Select County')
- Gender: (dropdown menu with 'Select Gender')
- Marital Status: (dropdown menu with 'Select Status')

Figure 5-2 Application Form

The financial aid staff was able to view all applications. After viewing the staff was able to do a background check by inviting the applicant for an interview. After the checks, the applicants' information is categorised as verified. The verified applications are then eligible for classification. The figure 5-3 illustrates the staff's view of the application.

The screenshot shows a 'List Of Applicants' interface. At the top, there are export options: Copy, Excel, CSV, PDF, and a search bar. Below is a table with the following data:

Name	Nationality	Secondary School	Grade	Current Status	Action
Eunice Kemunto Kerubo	kenyan	Ngara girls	A-	admitted	
Jonathan Maweu Mutiso	kenyan	Lenana school	B+	admitted	
Martin Mwenda Mururi	kenyan	Nairobi school	A	admitted	
miki winnie munene	kenyan	kiambu high	B	active	
Vamua omundo Kilwaki	kenyan	kisumu girls	A-	admitted	

Showing 1 to 5 of 5 entries

Navigation: Previous | 1 | Next

Figure 5-3 View Applicants

5.3.2 Data Normalization

The data goes through the normalization process to ensure that the values are scaled within the range of 0 to -1. The data in the ordinal values in the database were saved in numerical form using ids as primary keys. The data in the database was then fetched and passed through a normalizer. All values are then converted to range between 0 and -1. The dataset was normalised using the min max normalizer which scales the values between 0 and -1. The min max scalar normalizer uses the equation outlined on Equation 5-1:

$$Y = \frac{z - \min(z)}{\max(z) - \min(z)}$$

Equation 5-1 Normalization

5.3.3 Network training

The model implemented supervised learning where it was given a labelled dataset where the network was given the inputs together with their outputs. In order for the network to produce an output, it needed to be trained using the training set. The dataset was obtained from historical data in financial aid office where past applicants were awarded or not awarded the financial aid.

This was required by the network to predict the values of a new dataset. A financial aid dataset of 5000 records was used in the model. It was divided into 60% training set, 20% validation set and 20% test set. The training set was used to train the network by updating the weights and bias. Training epoch was set to 2000 epoch. The validation set was used to measure the generalization of the net, while the test set was used to measure the performance of the neural network. The network used 8 inputs which included: father occupation, mother occupation, secondary school grade, county, guardian occupation and aid percentage, family income and family spending. There were two outputs which were categorised as either award or not award. The network had one hidden layer with three hidden nodes.

5.4 Model Architecture

The diagram on figure 5-4 illustrates the architecture of the proposed model. The web application was used to collect the data which was stored in the database. Data collected from the database was converted to binary form between the values of -1 and +1. The normalized data was then passed to the neural network classifier, and the output values obtained.

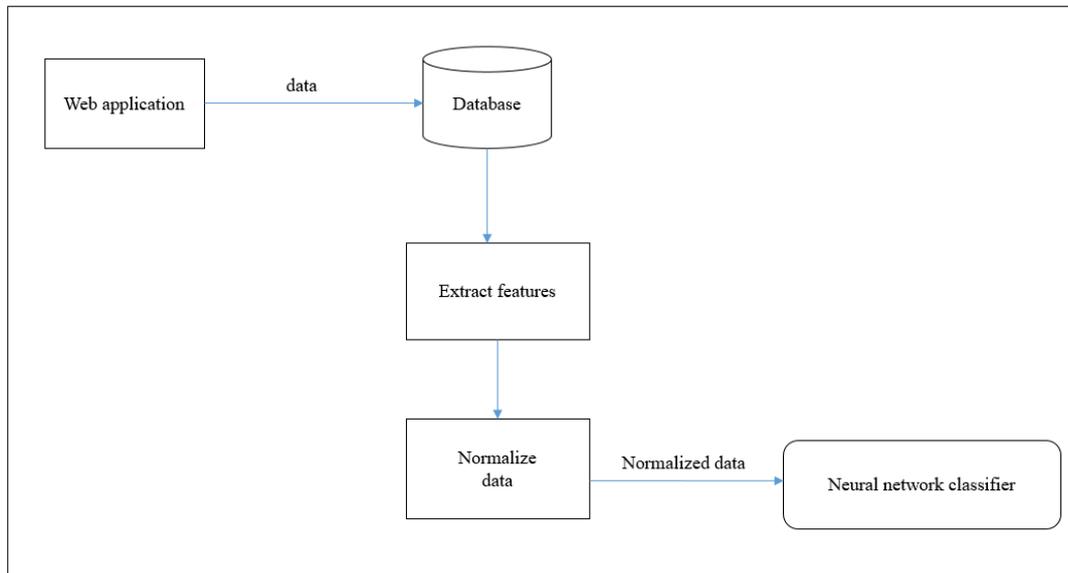


Figure 5-4 Model Architecture

5.5 Validation and Testing

This section highlights the various tests that were done to ensure that the system requirements were implemented with minimum errors. The testing was divided into three stages: model testing, system testing and acceptance testing.

5.5.1 Model Testing

Testing the model was important to ensure that it works as expected with high level of accuracy. During testing, a total of 3043 instances of the historical data was used. The data was passed through the model that had been learned after training, and the predicted results were compared to the actual results. This enabled the researcher to find the accuracy of the model by comparing the number of true positives and true negatives to the total test cases. The researcher found that the number of true positives was 2955 and true negatives was 18 giving an accuracy of 97.6%. The model was then fed with new data to classify new applicants which was obtained from volunteer students.

N=3043	Predicted: Awarded financial aid	Predicted: Not awarded financial aid
Actual: Awarded financial aid	2955	40
Actual: Not awarded financial aid	30	18

Figure 5-5: Confusion Matrix

5.5.2 System Testing

Testing the system is done to ensure that its requirements are met and the modules are working as expected. System testing can be divided into unit testing and integration testing. The unit tests are done by the programmer during development to ensure that the modules created have no errors, while integration testing is done to ensure integration between the modules. The table 5-2 shows some of the system tests.

Test id	Test Case	Priority	Results
1	Does the system save inputs in the database?	High	The inputs were correctly saved in the database
2	Can the user view his records?	High	The users were able to view their records
3	Is the user able to search through records?	medium	The users were able to search for a record using their id
4	Does the system give feedback to the user?	High	The users received feedback from the system for every action performed
5	Can the user update his records?	Medium	The users were able to make changes to their records.

Table 5-1 System Testing

5.5.3 Acceptance Testing

This testing is carried out by the user to reveal any errors with respect to the user requirements, before the system is implemented. This was done to ensure that the system was easy to use and user friendly. The table below shows some of the user acceptance tests.

Test id	Test Case	Priority	Results
1	Does the application meet the user requirements?	High	The application was able to meet the expected functional and non-functional requirements
2	Is the application easy for the user to understand?	High	The application was easy for the user to navigate
3	Is the user interface user friendly?	Medium	The user interface was acceptable.

Table 5-2 Acceptance Testing

Chapter 6 Discussions and Findings

6.1 Introduction

The financial aid model was implemented using features relating to the student which the applicant submitted through the web application portal. Previous awards done by the committees were used in the training of the neural network to classify new inputs. The correctness of the model was tested using precision, accuracy and the error rate that was obtained. The allocation of financial aid solely using committees was prone to bias and high inaccuracies in the event of a large dataset of applications. The process was also time consuming which led to delays in decision making, and subjective hence the intention of having an effective and fair system was lost.

6.2 Findings of the Research

A confusion matrix on table 6-2 was obtained from the classification of financial aid applicants, was used to analyse the results of the model. There were a total of 3043 instances used in the training and testing of the neural network. A total of 2955 were correctly classified, 40 instances which were actually awarded were wrongly predicted as not awarded, 30 instances that were actually not awarded were wrongly predicted as awarded while 18 instances were wrongly predicted as not awarded. The correctly classified instances accounted for 97.6% of the population while the incorrectly classified accounted for 2.3%.

N=3043	Predicted: Awarded financial aid	Predicted: Not awarded financial aid
Actual: Awarded financial aid	2955	40
Actual: Not awarded financial aid	30	18

Table 6-1 Confusion Matrix

6.3 Discussions

Muriithi et al (2013) used multinomial logistic regression model to classify HELB applicants into categories for loans, where he used house worth and wealth as predictor variables. The model correctly classified 79.8% of the cases under study. This research used artificial neural networks to predict on whether to award or not to award the applicants. It used data from Strathmore University Financial aid, to train the networks. The variables used include parents' occupation, Student grades, Secondary School attended, gender, county, percentage of aid applied for, marital status and the mode of study, whether full time or part-time. The model had an accuracy of 97.6% in predicting the awards based on the past data.

6.4 Contributions of the model to research

The model contributed to the research of automating the financial aid systems in Kenya to ensure the scarce resources are distributed fairly. The model concentrated on the award of financial aid to undergraduate applicants. The automation of the selection and award process in the financial aid sector would ensure that there is more transparency and efficiency, and accuracy. The model would also assist the financial aid offices to be more accountable to donors as they would have recorded the use of the contributed money and would also be able to report on the number and characteristics of the applicants who were awarded the aid.

6.5 Shortfalls of the Research.

The research study had the following limitations:

- i. The features used in creation of the model were limited to the available data.
- ii. The model was limited to a section of the financial aid process which was the allocation of the aid.
- iii. The model was also tested with data from one financial aid office, and the variations of ways of allocating financial aid in other institutions was not put into consideration.
- iv. The model was trained using past awards which were done through committees, which may be subject to bias.

Chapter 7 Conclusions and Recommendations

7.1 Conclusions

This study focused on improving the functioning of the financial aid office by automating the process of selection and awarding of financial aid. In an interview with a financial aid staff at Strathmore University financial aid office, it was highlighted that the office faces a number of challenges which could be solved by using machine learning algorithms to automate the systems. In addition to select and award process the other major challenge was tracking the students' performance and to measure the value of the aid to the student after receiving the aid. The financial aid staff has to personally follow up on each student, which is not easy to track without a system.

The researcher choose to focus more on the select and award process and used artificial neural network to make suggestions to the financial aid office on which applicants deserved to be awarded financial aid. The research focused on the characteristics of the applicants and the occupations of the parents and guardian and past data on the award decisions that had been made by the committee to applicants with similar characteristics, to make a decision on awards.

By analysing the results obtained using the machine learning algorithm, the researcher concludes that the financial aid office should include automated systems in their work to improve performance of the office and ensure that money donated is well utilized. Another conclusion from the research is that the accuracy of the results depends upon the number of key attributes used. The working of the financial aid offices affects the society as a whole, as the improvement in its operations will be able to raise the living standards of the people hence reduce poverty through the provision of education to the deserving students.

7.2 Recommendations

After the research the researcher had the following recommendations:

- i. The features used from the dataset could be added to improve the accuracy of the output.
- ii. The research should be able to accurately suggest the amount of money to be awarded.
- iii. The research should be able to recommend awards on various categories given the students characteristics.

7.3 Suggestions for Future Studies

In future the research can be expanded to include:

- i. The research should include the donors who contribute the money used in financial aid to enable them track the usage of the money contributed.
- ii. The research could also make performance predictions for the financial aid based on the characteristics of the student, to make a decision on whether to award the aid.
- iii. The application should be expanded to inform the financial aid office on the students' performance
- iv. The research could be expanded to use financial aid information to make enrolment predictions in universities.

References

- Castano, E. (2002). Proxy Means Test Index for targeting social programs: Two Methodologies and Empirical Evidence. *Lecturas de economia*, no 56.
- Dan, Z. (2008). Integrating RFM model and cluster for students loan subsidy Valuation. *2008 International Seminar on Business and Information Management*, (pp. 461-464).
- Gichuhi, L. (2015). Alternative Methods of Financing Higher Education in Kenya . *International Journal of Scientific Research and Innovative Technology* , 57-67.
- Gross, J. P., Hossler, D., & Ziskin, M. (2007). Institutional Aid and Student Persistence: An Analysis of the Effects of Institutional Financial Aid at Public Four-Year Institutions. *Nasfaa Journal of Student Financial Aid*, 28-39.
- Hamzic, A., & Avdagic, Z. (2016). Multilevel Prediction of Missing Time Series Dam Displacements Data Based on Artificial Neural Networks Voting Evaluation. *IEEE International Conference on Systems, Man and Cybernetics* (pp. 002391-002396). Hungary: IEEE.
- Han-Jia Ye, Zhan, D.-C., Li, X., Huang, Z.-C., & Jiang, Y. (2016). College Student Scholarships and Subsidies Granting: A Multi-Modal Multi-Label Approach. *IEEE 16th International Conference on Data Mining* (pp. 559-568). Nanjing: IEEE.
- Hopper, R. (2001). Students loans in Comparative and International Perspective. *International higher education*, 10-12.
- Iranmanesh, S. H., & Zarezadeh, M. (2008). Application of Artificial Neural Network to Forecast Actual Cost of a Project to Improve Earned Value Management System. *International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering* , 658-661.
- Jha, G. K. (2017, April 1). *Artificial Neural networks and its application*. Retrieved from Indian Agricultural Statistics Research Institute: http://www.iasri.res.in/ebook/EBADAT/5-Modeling%20and%20Forecasting%20Techniques%20in%20Agriculture/5-ANN_GKJHA_2007.pdf

- Johnstone , D. B. (2003). Higher Education Finance and Accessibility:Tuition Fees and Student Loans in Sub Saharan Africa. *Improving Tertiary Education in Sub-Saharan Africa: Things That Work!* Accra.
- Johnstone, B., & Marcucci, P. (2010). *Targeting Financial Assistance to Students in Higher Education: Means testing with special emphasis on low and middle income countries.* World bank Contract No 0007728373.
- Kotsiantis, S. (2007). Supervised Machine Learning: A review of Classification Techniques. *Informatica*, 249-267.
- Muriithi, D. K., Njoroge, G. G., Okongo, M. O., & Njoroge, E. W. (2013). Classification of Higher Education Loans using Multinomial Logistic Regression Model. *Journal of Mathematical Sciences: Advances and Applications*, Volume 22, 1-17.
- Muriithi, D., & Okongo, G. (2013). Classification of higher education loans using multinomial regression model. *Journal of mathematical sciences*, Volume 22, 1-17.
- Ngolovoi, M. (2008). *Means Testing Students Loans in Kenya.* (Doctoral Dissertation, State University of New York at Buffalo). Retrieved from ProQuest Dissertations Publishing .(UMI No 3307681)
- Otieno, W. (2009). Higher Education System In Kenya. Kenya.
- Pattanayak, S., & Nandi, R. (2013). Identification of Spectrum Holes using ANN Model for Cognitive Radio Applications. *IEEE* (pp. 133-137). India: IEEE.
- Saitoh, F. (2016). Predictive Modeling of Corporate Credit Ratings Using a Semi-Supervised Random Forest Regression. *IEEE*. Kanagawa, Japan.
- Salmi, J. (2001). Student loans:The World Bank Experience. *International Higher Education*, 10-12.
- Thanh, L. V., & Haddawy, P. (2007). Deriving Financial Aid Optimization Models from Admissions Data. *37th ASEE/IEEE Frontiers in Education Conference*, 7-12.
- UniversityBrown. (2017, March). *Financial Aid*. Retrieved from Brown University: <https://www.brown.edu/about/administration/financial-aid/needs-analysis-calculating-expected-family-contribution-efc>

- Wabwoba, F., & Mwakondo, F. (2011). Student selection for university course admission at the joint admission board (Kenya) using trained neural networks. *Journal of Information Technology Education*, 334-347.
- Ye*, H.-J., Zhan*, D.-C., Li, X., Huang*, Z.-C., & Jiang*, Y. (2016). College Student Scholarships and Subsidies Granting: A Multi-Modal Multi-Label Approach. *IEEE 16th International Conference on Data Mining*, 559-568, DOI 10.1109/ICDM.2016.97.

Appendix A : Turnitin Report



Digital Receipt

This receipt acknowledges that Turnitin received your paper. Below you will find the receipt information regarding your submission.

The first page of your submissions is displayed below.

Submission author: Eunice Manyasi
Assignment title: Masters Theses - 2017 Examinatio..
Submission title: Thesis Draft
File name: A_model_for_the_classification_o...
File size: 1.25M
Page count: 64
Word count: 11,980
Character count: 74,224
Submission date: 06-Apr-2017 01:00PM
Submission ID: 794166795



Appendix B : Financial Aid Interview Guide

1. Which types of financial aid do you provide?
2. Whom do you provide financial aid? (Undergraduates, Graduates, 1st degrees)
3. What is the basis of allocating financial aid? (Need based, Merit Based)
4. Who is eligible for financial aid?
5. What is the application procedure of financial aid?
6. What information do financial aid programmes need about an applicant?
7. How is the information of applicants collected and stored?
8. How do you verify the information presented?
9. What is the procedure of allocating the financial aid? (Committees, Approval ,...)
10. What is the basis of allocating the financial aid (County, School, Religion)
11. Which information is heavily relied on when determining eligibility and awarding of financial aid?
12. What happens to applicants with missing information?
13. Key features that determine the amount of aid to be given to an applicant.

Appendix C : Financial Aid Interview Feedback

1. Which types of financial aid do you provide?

- Scholarships, Loans, Work Study programs

2. Whom do you provide financial aid? (Undergraduates, Graduates, 1st degrees)

- Financial assistance is provided to students pursuing higher education

3. What is the basis of allocating financial aid? (Need based, Merit Based)

- Financial aid is mainly allocated based on financial need and merits.

4. Who is eligible for financial aid?

- A student admitted for a course at Strathmore University and meets the outlined criteria for the financial aid applied for.

5. What is the application procedure of financial aid?

- Applicant gets an admission letter
- After admission the applicant applies for financial aid where he provides his personal information together with supporting documents to prove his financial need.
- The applicants are then invited for interviews which is further used to assess the applicants.
- After the interviews a financial aid committee goes through the applications and determine the applicants who are most suitable to receive awards.

6. What information do financial aid programmes need about an applicant?

- Data collected includes personal information such as age, secondary school grade, parents/guardian occupation, and supporting documents such as parents' payslips, and bank account statements, letter from area chief or religious leader.

7. How is the information of applicants collected and stored?

- The supporting documents are stored as hard copies while personal information is stored in spreadsheet files.

8. How do you verify the information presented?

- The supporting documents and interviews are used as verification of information. The government, religious leaders and secondary school principals also assist in verification of the provided information.

9. What is the procedure of allocating the financial aid? (Committees, Approval)

- Allocation of financial aid is done through committees who agree based on a set criteria.

10. What is the basis of allocating the financial aid (County, School, Religion)

- The basis of allocating aid varies from one financial aid institution to another depending on the type of financial aid applied for.

11. Which information is heavily relied on when determining eligibility and awarding of financial aid?

- All information is important and required to make decisions

12. What happens to applicants with missing information?

- The applicant is mostly disqualified.

13. Key features that determine the amount of aid to be given to an applicant.

- The main determinants depend on the type of financial aid. If the aid is need based it depends on the financial need of the applicant, merit based financial aid are given based on the performance of the students and some are given based on extra-curriculum activities.

Appendix D: Verify Applicants Information Use Case Description

Use case name: Verify applicant's information
Description : A financial aid staff verifies the information submitted is true and records in the system as verified
Primary Actor : Financial aid staff
Pre-Conditions : Applicant has been admitted and has applied for financial aid
Post-Conditions : Applicants information has been verified
Success Scenario : <ul style="list-style-type: none">i) Financial aid staff views the information submitted by the applicantii) A background check is done by the financial aid office to ensure that the information is trueiii) Financial aid staff records in the system that information is verifiediv) Applicants status changes verified and system displays a confirmation messagev) Applicant is notified that his information is verified
Frequency of use : Every admission period
Status : Verified by the financial aid office

Appendix E : Allocate Financial Aid Award Use Case Description

Use case name: Allocate financial aid award
Description : A financial aid staff requests the system to classify and award financial aid to an applicant
Primary Actor : Financial aid staff
Pre-Conditions : The applicant's information must be verified by the financial aid staff
Post-Conditions : Applicants has been classified
Success Scenario : <ul style="list-style-type: none">i) Financial aid staff views a list of applicants whose information has been verified.ii) Financial aid staff clicks on the classify buttoniii) The system responds stating that classification is in progressiv) System classifies the applicant and a notification is sent to the applicant
Frequency of use : Every classification period
Status : Applicants have been classified