Prediction of Acceptance of offers for Academic places using Data Mining

By

Raj Man Shrestha

A THESIS SUBMITTED TO MACQUARIE UNIVERSITY

for the degree of

Masters of Research

Department of Computing

November 2014

Declaration

I certify that the work in this thesis entitled PREDICTION OF ACCEPTANCE OF OFFERS FOR ACADEMIC PLACES USING DATA MINING has not previously been submitted for a degree nor has it been submitted as part of the requirements for a degree to any other university or institution other than Macquarie University. I also certify that the thesis is an original piece of research and it has been written by me. Any help and assistance that I have received in my research work and the preparation of the thesis itself have been appropriately acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

_____

Raj Man Shrestha

Acknowledgements

I would like to thank Ms. Enid Zhang for the information and her expertise regarding the process of international student recruitment.

I would like to thank my supervisors: Prof. Mehmet Orgun and Dr. Peter Busch, for their guidance and encouragement during the preparation of this thesis.

Abstract

The thesis examines the validation of prediction models of acceptance of offers by students in the context and settings of international students in a large Australian University using data mining techniques. Earlier works in the enrolment management have examined various classification problems such as inquiry to enrolment, persistence, graduation using the data and the settings of a particular university. The data and settings from different institutions are often different which implies that, in order to find out which models and techniques are applicable at a given university, the dataset from that university needs to be used in the validation efforts. A dataset comprising the offers to students from a large Australian university where around 3,500 new international students commence their studies every year was analyzed. The important predictors for the acceptance of offers were the chosen course and the faculty, whether the student was awarded any forms of scholarship and also the visa assessment level of the country by immigration department. The prediction models were developed using logistic regression, decision trees and neural networks and their performances were compared. The prediction model by neural networks produced the best result.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Introduction and Context of this Study

Educational data mining (EDM) is an interdisciplinary research field which explores the data from educational settings to better understand the students, the settings at which they learn, to gain insights into, explain educational phenomena as well as enhancing the decision processes in higher learning institutions and streamlining efficiency in the decision making process.

Enrolment management which is a part of EDM helps academic institutions to improve ties to prospective client groups, attract students into and through the institution, predict the number of students in the upcoming study periods and also investigate the retention and academic success of the students.

Data mining in education is a process of uncovering hidden trends and patterns using a combination of an explicit knowledge base, sophisticated analytical skills and academic domain knowledge, producing new observations from existing observations. In contrast to traditional analytical studies which are often in hindsight and aggregate in nature, data mining is forward looking and provides an ability to gain a deeper understanding of the patterns previously unseen using current available reporting capacities (Luan, 2002).

Data driven analytical tools can be used to predict whether a student who has been offered a place in an academic institution will accept an offer for a place or not. This helps to segment students into groups who are very likely to enroll, who will not enroll and those who are at the margin and require more information to assist with their decision-making. These groups can be offered specific incentives and/or additional recruitment efforts. By directing the recruiting efforts to a group of students who are more likely to accept an offer, improves the enrolment yield can be improved and the recruitment process more efficient and targeted.

This project aims at studying predictive models for identifying those international applicants who are more likely to accept an offer of a place, and those applicants who

are less likely to accept the offer at a large Australian university, using data mining techniques. Classification techniques are used to identify the international student profiles of interest. From the international student profiles, a list of features important for the purpose is identified. An analysis of the different approaches to predictive modeling for the purpose is carried out.

1.2 Objective of the study

The thesis examines the validation of prediction of acceptance of offers for academic places in the context and settings of international students in an Australian university. The students considered are those applicants who apply for the undergraduate and postgraduate coursework places.

Developing predictive models to assist in the marketing, recruitment, and admission of prospective students holds the potential in making the best strategic decisions not only for the organization, but for the prospective students as well. The scarcity of fiscal resources, in concert with the increasing number of educational alternatives, will necessitate the use of business intelligence to compete in the higher education marketplace.

The prospective students apply for the courses they are interested in using paper-based forms or online application in the admissions portal of the university. During the submission of the application, they mainly submit their academic transcripts, certificates of the courses they have completed and English language proficiency test results. Some students apply on their own and are referred to as direct applications while others students apply through agents who may help students during applications processing as well as visa lodgment. The applications by the students are assessed on the basis of the documents they have submitted by the respective department or faculty. The qualified candidates are offered a place at the university. There are three different types of offers that can be made to a student; a full unconditional offer (Q, Qualified); an offer packaged with an English course (QPAC, Qualified with Package); or a full offer with conditions such as waiting for last semester's results or transcript (FWC , Full offer With condition). Some students accept the offer by paying the course commencement fees while others do not accept the offer. There is an email follow up process for those students who do not accept the offer.

The prediction of students who will accept the offer will help to segment students into groups and act on those students who are more likely to accept the offer. Being able to target direct mail, or to eliminate some students from mailings, has the potential of saving institutional resources.

1.3 Overview of the study

The following chapters provide details of the dataset, preprocessing, modeling and results.

Chapter 2 details the background of the educational data mining and enrolment management. The classification techniques together with the three models that will be used logistic regression, decision trees and neural networks are discussed. Also model validation techniques commonly used in classification are discussed. A short description of the RapidMiner software which has been used in the project is also included.

Chapter 3 discusses the related works that have been carried out in the student enrolment domain using different statistical and data mining models.

Chapter 4 discusses the data set and analysis of the attributes in relation to the target label attribute of acceptance of offers. The data preprocessing, data preparation for modeling and feature selection using weight correlation and principal component analysis is discussed.

Chapter 5 discusses the modeling and the results of the prediction models as well as a comparison of the models using validation techniques based on receiver operating characteristic (ROC) curves and lift charts.

Chapter 6 discusses the results of the experiments, usefulness of the model in a particular institutional setting.

Chapter 7 summarizes the findings of the study and examines future extensions of the work for incorporating retention and academic success.

Chapter 2: Background

2.1 Introduction

This chapter will provide an overview of educational data mining and enrolment management, classification models and model evaluation techniques. RapidMiner 5.3[1] is an open source software which has been used in the project for data preprocessing, model development and evaluation. A short overview of RapidMiner software will be included in the chapter.

2.2 Education Data Mining and Enrolment Management

Educational data mining (EDM) is concerned with developing, researching, and applying computerized methods by exploring the data from educational settings using a combination of explicit knowledge base, sophisticated analytical skills and academic domain knowledge to gain insights into, explain educational phenomena as well as enhance the decision processes in higher learning institutions and to streamline efficiency in the decision making process.

Educational data miners can use classification and prediction on areas like alumni, institutional effectiveness, marketing and enrolment management. Institutional effectiveness questions like how do students learn best, the courses that are often taken together, the learning experiences that most contributive to overall learning outcomes can be answered by data mining. Marketing, powered by data mining, may boost enrolments by figuring out the students the college has not reached and the students likely to be interested in receiving more information in a particular program area. Enrolment management powered by data mining can quickly identify the prospective student, persistence of the student and can also predict the academic success of the student.

2.3 Classification and prediction

The benefits of data mining are its ability to gain deeper understanding of the patterns previously unseen using currently available reporting capacities. Classification involves

[1] https://rapidminer.com/documentation/

predicting a certain outcome based on a given input using an algorithm which processes a training set containing a set of attributes and the respective outcome or label attribute. The algorithm tries to discover relationships between the attributes which would make it possible to predict the outcome for other unseen data sets.

Prediction from data mining allows the academic institutions an opportunity to act before a student drops out, or to plan for resource allocation from knowing how many students will enroll. The prediction of students who will accept the offer will help to segment students into groups and act on those students who are more likely to accept the offer. Being able to target direct mail or to eliminate some students from mailings has the potential of saving institutional resources.

2.4 Prediction Models

The widely used models for enrolment prediction are the statistical model; logistic regression and data mining models; decision tree and neural networks.

2.4.1 Logistic regression

Logistic regression is an appropriate technique because of the dichotomy of the dependent variable. The model is estimated using maximum likelihood estimation. Maximum likelihood is a way of finding the smallest possible deviance between the observed and predicted mean through the use of calculus. With maximum likelihood, the computer uses different iterations, in which it tries different solutions until it gets the smallest possible deviance or best fit.

DesJardins (2002) described the logistic regression model used in the study as

$$\log \frac{P}{1-P_i} = \alpha + \beta_i X_i + \delta_i Y_i + \gamma_i Z_i + \varepsilon_i$$

where P_i is the probability that student i will choose to enroll; X_i is a vector of personal and demographic characteristics; Y_i is a vector of prior educational characteristics, college intentions and preferences; Z_i is an institutional level variable; α, β_i, δ_i, and γ_i are estimated coefficients; and εi represents a random error term that is logistically distributed. The dependent variable is the logarithm of the odds that a particular student will enroll in the study institution.

2.4.2 Decision Tree

Han, Kamber and Pei (2006) described a decision tree as a flowchart-like tree structure, where each internal node (nonleaf node) denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node.

The construction of decision tree classifiers does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery. The tree generated is intuitive and easily interpretable. The other advantages of decision trees are the ability to handle high dimensional data, fast learning and also have good accuracy.

The popular decision tree algorithms are ID3 (Iterative Dichotomiser), C4.5 (a successor of ID3), and Classification and Regression Trees (CART). These algorithms adopt a greedy or non backtracking approach in which decision trees are constructed in a top-down recursive divide-and-conquer manner which starts with a training set of tuples and their associated class labels. The training set is recursively partitioned into smaller subsets as the tree is being built.

An attribute selection measure is a heuristic for selecting the splitting criterion that best separates a given data partition of class-labeled training tuples into individual classes. The three attribute selection measures for tree construction are information gain or *Entropy*, *Gain Ratio* and *Gini Index*. The *Gini Index* considers a binary split for each attribute. Although biased towards multivalued attributes, it gives reasonably good results in practice (Han, Kamber and Pei 2006).

2.4.3 Neural Networks

Neural networks are suitable for problems whose inputs are both categorical and numeric, and where the relationships between inputs and outputs are not linear or the input data are not normally distributed. Neural networks have been shown to be very promising systems in many forecasting applications and business classification applications due to their ability to “learn” from the data, their non-parametric nature (i.e., no rigid assumptions), and their ability to generalize.

A neural network is composed of processing elements or neurons grouped in layers to form the network's structure. The three layers of a neural network are: input, intermediate (called the hidden layer), and output. A hidden layer is a layer of neurons that takes input from the previous layer and converts those inputs into outputs for further processing. Several hidden layers can be placed between the input and output layers.

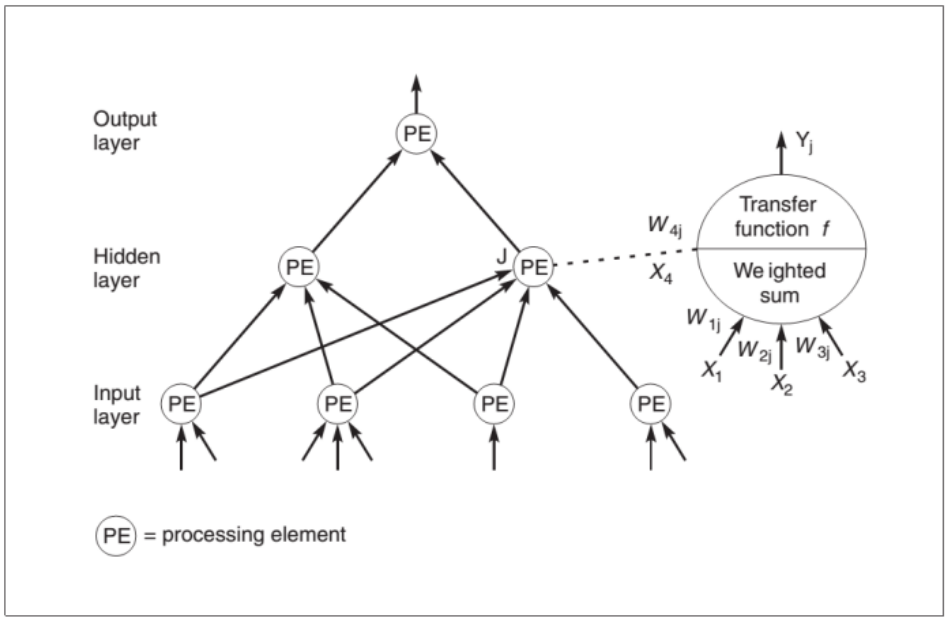


Fig 2.1: Neural network with One Hidden Layer

Each input corresponds to a single attribute. Connection weights express the relative importance of each input to a processing element and a network learns through repeated adjustments of weights. The summation function computes the weighted sums of all the input elements entering each processing element. A summation function multiplies each input value by its weight and totals the values for a weighted sum Y. The formula for n inputs in one processing element is:

$$Y = \sum_{i=1}^{n} X_i W_i$$

Where, W_i is the connection weight for the input X_i.

The transformation (transfer) function combines the inputs coming into a neuron from other neurons and then produces an output based on the choice of the transfer function. One of the popular transfer functions is a sigmoid function, which is an S-shaped transfer function in the range of 0 to 1.

$$Y_T = {}^{1}/_{(1 + e^{-Y})}$$

where Y_T is the transformed or normalized value of Y.

The transformation function modifies the output levels to reasonable values and is performed before the output reaches the next level. The outputs of a network contain the solution to a problem (Sharda, Aronson and King, 2008).

2.5 Model Evaluation

After the training of the model, the model fit is determined on test or holdout or validation sample to score students on the basis of their probability to enroll. In the holdout method, the given data are randomly partitioned into two independent sets, a training set and a test set. Typically, two-thirds of the data are allocated to the training set, and the remaining one-third is allocated to the test set. The training set is used to derive the model, whose accuracy is estimated with the test set.

The cross validation technique has been used in many earlier works. In k-fold cross-validation, the initial data are randomly partitioned into k mutually exclusive subsets or folds, each of approximately equal size. Training and testing is performed k times. In iteration i, partition Di is reserved as the test set, and the remaining partitions are collectively used to train the model. For classification, the accuracy estimate is the overall number of correct classifications from the k iterations, divided by the total number of tuples in the initial data.

The use of 2X2 table of the hits and misses of a prediction rule also called a classification table helps to determine the predictive ability of the model. There will be two types of mis-classifications, classification of enrolled as non-enrolled and classification of non-enrolled as enrolled. The results obtained using a 2×2 classification table is sensitive to the choice of the cutoff score. By default the threshold value is 0.5, which may not be appropriate when the number of enrolled and non-enrolled students in the training sample is not balanced. The choice of an appropriate cutoff value depends on the costs of incorrectly classifying enrollees or non-enrollees. Using a low cutoff score will result in relatively more students who actually did not enroll being classified as enrollees. Failures to predict enrollments represent a direct economic loss to the university, since students that would have enrolled - are

categorized as students that will not enroll. The subsequent lack of attention by the admissions staff could alter the student enrollment decision and result in a probable loss of a student enrollment. Correspondingly, failures to predict non-enrollment result in a greater quantity of students that must be handled by the current admissions office resources (DesJardins, 2002).

Another important tool that has been used for model validation in research is Receiver Operating Characteristic (ROC) curves. ROC curves are a useful visual tool for comparing two classification models. An ROC curve shows the trade-off between the true positive rate or sensitivity (proportion of positive tuples that are correctly identified) and the false-positive rate (proportion of negative tuples that are incorrectly identified as positive) for a given model. Any increase in the true positive rate occurs at the cost of an increase in the false-positive rate. The area under the ROC curve is a measure of the accuracy of the model.

The lift chart measures the effectiveness of models by calculating the ratio between the results obtained with a model, and the result obtained without a model. The result obtained without a model is based on randomly selected records. The chart consists of a lift curve (response after predictive model) and a baseline (response before predictive model). The greater the area between the lift curve and the baseline, the better is the model which represents the gains from using the predictive model.

2.6 RapidMiner Software

RapidMiner is the open source system for data mining, machine learning and predictive analytics. With access to over 1,500 operators, the Java-based visual environment of RapidMiner allows for rapid data mining process development. RapidMiner uses a client/server model with the server offered as Software as a Service or on cloud infrastructures. RapidMiner functionality can be extended with additional plugins. The Rapid Miner extensions marketplace provides a platform for developers to create data analysis algorithms and publish them to the community.

RapidMiner 5.3 has been used for data preprocessing, developing and testing the model which is available under an OSI-certified open source license. RapidMiner provides

operators for importing, exploring and transforming data; such as models for logistic regression, decision tree, neural networks; model evaluation steps such as ROC curve and lift chart. RapidMiner also provides structures that express the control flow of the process. The workflow involves loading a dataset, choosing an operator for operations and set parameters, connecting with the next operator and producing results and visualizations. Some terminologies specific to RapidMiner as well as screenshots are included in Appendix C.

Chapter 3: Literature Review

3.1 Introduction

There are many previous studies on the student recruitment domain which have been carried out using different statistical and data mining models. The studies range from predicting the enrolments for a study period, enrolment from admitted students and also predicting enrolments from inquiries. Student Enrolment has been also treated as a resource allocation problem trying to have a better enrolment yield and efficiency from the admission team.

For the prediction of student enrolment macro level studies such as aggregate high school characteristics as well as micro level studies at the individual student level have been carried out. The important features that may be relevant are high school size, whether the university was the first choice or not, distance a student needs to travel, the student major etc. Comparison of data mining models such as logistic regression and neural networks have been carried out for the admission domain.

3.2 Previous works

Walczak and Sincich (1999) performed a micro-level study that examines the probability of an individual student enrolling at a specific academic institution so that admissions counselors can more effectively dedicate their time and resources to converting applicants into actual enrollments. A feedforward backpropagation neural network model was developed to predict the enrollment decision of individual students and compared the relative performance with logistic regression models. The main predictor variables identified were location (in-state, out-of-state, international), distance cumulative grade point average (GPA), financial aid request, source of inquiry, application submission delay, campus visit, and desired major.

The average prediction accuracy of the neural networks for enrolment and non enrolment was 89.39% and 72.83% respectively. Neural networks effectively reduced the total number of students that should be allocated admission staff resources by 50% to 60%. Neural networks showed performance improvement over logistic regression

and can also handle noisy data as well as being easy to retrain the model with new classification data sets, on change of admission policies.

Walzack (1998) treated enrolment issues as a resource allocation problem. The size of the application pool taxes the resources of the admissions staff. A neural network system that categorizes students with regard to enrollment yield potential would enable counselors to allocate their time more productively and maximize enrollment yield.

A comparative study of six different supervised learning neural network architectures has been performed to analyze quantitative differences in the categorization capabilities of the different neural networks for the enrollment application domain. Backpropagation networks produced the best overall performance and achieved 56% reduction in counselor case load (Walczack & Sincich, 1999).

Thomas, Dawes, and Reznik (2001) used regression analysis to estimate students' probability of enrollment and designed an experiment to identify those predicted enrolment groups on whom recruitment efforts had the greatest impact. Naturally students with relatively low enrollment probabilities are more susceptible to increased recruitment efforts. Additional attention appears to have had no effect on students with the highest enrollment probabilities but are more effective. The model included four kinds of variables: demographic, academic, geographic, and behavioral. The highly significant variables positively related to enrollment are high-yield high school average, high-yield SAT score, high-yield math SAT, high-yield verbal SAT, high-yield zip code, and open house attendance.

The model's predictive power has been assessed by measuring goodness-of-fit through classification tables that compare results predicted by the model with students' actual enrollment decisions. After the recruitment intervention, the students having 30% chance and 80% chance of enrolling increased their probability by 10% and 5% respectively. The analysis of admitted students by enrollment probability also confirmed that targeting recruitment efforts to the students most likely to enroll, does *not* focus attention on those with the strongest academic credentials, as enrollment probability is inversely related to average SAT score.

DesJardins (2002) built a model using logistic regression to predict enrollment out-of-sample, and used the results to segment admitted students into groups so that different

recruitment and marketing interventions can be applied to different groups. The model correctly predicted 64.9% of enrolments and 66.4% of non-enrolments.

The data of students who were admitted for enrollment in the fall of 1999 and the students who were admitted for the fall 2001 class of a large, public research institution which enrolls approximately 3,800 new students each autumn were used. The important features considered were high school size, whether the university was the first choice or not, residents of the home state or not, have previous enrolment from the family or not. Logistic regression has been used because of the dichotomous nature of the outcome (enrolled or not). Cross validation has been used to assess the efficacy of the model and a decile grouping has been used to test the goodness of the fit of the model.

It was observed that the students who have a high admissions index score are less likely to enroll than the students with lower admission index score. Hence rather than focusing on students who are very likely to enroll, students who are at the margin or the fence sitters should be the focus for enrolment.

Goenner and Pauls (2006) built a predictive model for inquiry to enrolment and test the predictions out of sample. In comparison, inquiry data, such as by returning a request for information form, has limited information than applications made up by integrating with geographic and demographic data.

A logistic regression model was used to estimate the probability that individuals enroll, while controlling for a number of factors theoretically relevant to the decision. A Bayesian Modelling average has been used for variable selection methods. The important features were types of interaction (ACT submission, college visit, internet), distance, program interest. The model with a cutoff of 0.5 correctly predicts 89% of student enquiries.

Chang (2006) worked on the applying data mining of enrolment using C&R Tree, Neural Networks and Logistic Regression using Clementine. Demographic fields such as gender, ethnicity, age, region, origin, as well as academic factors such as high school GPA, high school rank, ACT and SAT scores, attempted programs, degree or non-degree seeking status and communication activities are the main attributes used. In all the prediction models regions, communication types, and certain sources of contact were the important features.

Model fitting involved many reiterations between training and testing data sets as well as several model evaluation processes. Through the confusion matrix, the number of hits or misses by each predictive model was evaluated and compared. The probability of correct prediction from C&RT, neural network, and logistic regression models was about 74 percent, 75 percent, and 64 percent respectively, supported by results from the corresponding testing groups: 67 percent, 71 percent, and 58 percent. The study also evaluated the predictions from the three modeling nodes by examining their level of agreements i.e. the number of cases predicted by the nodes that were the same by all three. Individual models varied in their performances. Agreements between the three models were 66 percent; of those, 82 percent agreed with the actual enrollment. The result showed that data-mining predictive models, C&RT or neural networks provide better solutions than the traditional logistic regression in predicting admission yields.

Johnson (2008) studied the aggregate-level high school effects on three outcome variables: the odds of admitted students to enroll, one year persistence and timely completion. The student personal profile information such as gender, ethnicity, family income, first generation to attend college, SAT score, high school GPA, as well as the aggregate school characteristics such as percent of SAT takers, school poverty (free lunch) and ethnic composition (within a 60 miles radius) were the important features used.

The research was based on a doctoral/research university which enrolled about 12,000 students annually. Two level Bernoulli model using Hierarchical Linear and Nonlinear Modeling (HLM) software was used for the analysis. Separate models were built to estimate the effects of individual-level and aggregate-level school characteristics for the three models. It was found that the high percentage of SAT takers in the high school, as well as students located within 60 miles of the institution, have greater odds of enrolment.

Bogard (2013) combined enrolment with retention and worked on model development to score university students based on their probability of enrolment and retention, so that staff and administrators could work to recruit students that not only have an average or better chance of enrolling but also succeeding once they enroll. Decision trees had been used for the enrolment model using SAS® Enterprise Miner with optimization based on average square error, limiting leaf size to 50 and the number of

rules to 3 to improve interpretation. The results of Neural Networks using multilayer perceptron, Decision Trees and Gradient Boosting have been compared for a retention model of which Gradient Boosting provided the best performance.

For enrollment predictions, students were classified into four categories: Least Likely, Unlikely, Average, and Most Likely. For retention, students were classified (from most likely to drop out to least likely) into four different categories: Double Red, Red, Yellow and Green. Based on actual historical enrolment and retention data, these classifications performed a good job discriminating between the groups of students. Cross tabulation of Propensity to Enrol and Attrition risk has been produced such that a unique strategy can be tailored for each type of student classified into each cell based on the model results.

The following table lists the models used in the papers discussed above.

Table 3.1: Models used across different previous works in enrolment prediction

Title of Papers	Model(s) Used
An analytic Strategy to assist institutional recruitment and marketing efforts	Logistic regression
Neural Network models for a resource allocation problem	Neural networks Comparison of multilayer perceptron backpropagation, radial basis function, counterpropagation general regression, fuzzy ARTMAP, linear vector quantization
Enrollment, persistence and graduation of in-state students at a public research university: does high school matter?	Two level Bernoulli model Using Hierarchical Linear and Nonlinear Modeling (HLM) software
A predictive model of inquiry to enrollment	Logistic regression Bayesian Model Averaging for the linear combination of the variables
Applying Data Mining to Predict College Admissions Yield: A Case Study	C&RT (Decision Tree) Neural networks (multilayer perceptron) Logistic Regression
Using predictive modeling to target student recruitment: Theory and practice	Logistic Regression
A comparative analysis of regression and neural networks for university admissions	Logistic Regression Neural Networks
A Data Driven Analytic Strategy for Increasing Yield and Retention at Western Kentucky University Using SAS Enterprise BI and SAS Enterprise Miner	Decision Trees Neural Networks using multilayer perceptron Gradient Boosting

3.3 Conclusion

Earlier works suggest that prediction of enrolment is a well-defined classification problem and macro level studies such as effects of aggregate high school factors, as well as micro level studies such as scoring individual students for probability of enrolment, have been carried out.

The important predictor variables identified are high school size, whether the university was the first choice or not, distance a student needs to travel, the major of the student, financial aid and so forth. The statistical models; logistic regression and data mining model and neural networks are the main regression tools used in enrollment prediction domain. Neural networks are more suitable for the purpose because of their capability to learn from the data, ability to generalize, capacity for handling noisy data and ease to include changes in the rule.

For model validation and providing prediction accuracy, classification table is the main tool used. The choice of the cutoff point is important as lower cut-off points result in lower overall accuracy while higher cut-off points substantially underestimate the number of students who enroll. It depends on the misclassification costs. Enrollment managers are more concerned with incorrectly classifying actual enrollees.

The above research lays the conceptual and theoretical foundations for this project, prediction of offer acceptance for academic places using data mining. There are some institution wide differences in the admission process as well as in the definition of certain terms. Since most of the studies have been carried out in US universities, their terminological definitions of *Admitted* and *Enrollment* are similar to *Offers* and *Acceptances* in the Australian universities.

Chapter 4: Data – Descriptive Analysis

4.1 Introduction

The dataset consists of international students' admissions data from a large Australian university. The admissions database stores international students' applications information, tracks the applications through the stages such as assessment, offers, status such as accepted, withdrawn. A de-identified dataset has been created from the admissions database and provided which consists of the offers made to international students for undergraduate and postgraduate coursework studies. The dataset did not contain any identifiable information of the student such as studentid, name, email address or date of birth.

The chapter consists of three main sections. The first section discusses the structure of the data file, attributes and descriptive analysis of the dataset. The second section discusses the data preprocessing and transformation carried out in the dataset. The third section discusses the relevance of the attributes that determines the acceptance or rejection of the offers.

4.2 Descriptive Analysis

4.2.1 The Nature of the data set

The data set contains 24,283 rows of the offers made to the international students and whether the offer was accepted or not from the year 2008 to 2013 for both undergraduate and postgraduate programs. The data set comprises 29 columns and consists of a mix of continuous, nominal and categorical values. The column names and properties are listed in Table A.1 of Appendix A. The dataset is in Microsoft Excel format.

The applications by the students are assessed on the basis of the documents they have submitted. The qualified candidates are offered a place at the university. There are three different types of offers that can be made to a student; a full unconditional offer (Q, Qualified);an offer packaged with an English course (QPAC, Qualified with Package);

or a full offer with conditions such as waiting for last semester's results or transcript (FWC , Full offer With condition). On average 21.1% of the offers made are packaged with an English language course, 75.5% offers made were full offers whereas only 3.4% offers were full offers with condition(s). The acceptance rate of packaged offers at 58%, was slightly better than the acceptance rate for full offers at 57%.

The average acceptance rate for the years 2008, 2009 and 2010 was around 60% whereas in 2012 and 2013 the rate dropped to 53.2% and 50% respectively. Out of the two semesters in a year, the number of offers sent in the second semester is about 24% more in average than the number of offers sent in the first semester. The acceptance rate is slightly better by 1.5% in the first semester in comparison to the second semester.

4.2.2 Attribute Analysis

The dataset consists of three types of information; geo-demographic, academic and application related.

i. Geo-demographic information

The geo-demographic information includes gender, age, country of citizenship, economy level of the country, visa assessment level for the country etc.

a. Gender

The distribution of male and female students was 42% and 58% respectively. 56.1% of the male students accepted the offer, whereas 58.3% female students accepted the offer.

b. Age

The age of the students ranged from 18 to 69. The average age of the applicants for undergraduate places is 25.6 and the average age of the applicants for postgraduate places is 29.3. About 49.6% of offers sent to the students were in the 26-30 age group, whereas 25.8% of offers were sent to the students in the 18-25 age group and only 17.8% of offers were sent to students in the 31-35 age group. Less than 8% of offers were made to the students above 35 years of age. The acceptance rate of 47.5% was

the highest among the 18 - 25 year old age group, while the acceptance rate of 39.8% was the lowest in the 31-35 year old age group.

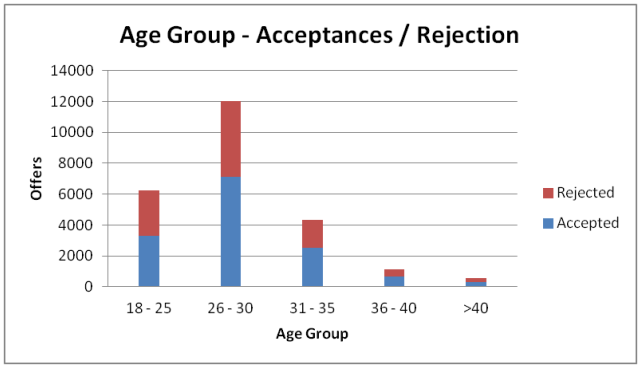


Figure 4.1: Age Group – Acceptances / Rejection

c. *Country of Citizenship*

Offers were sent to students from 145 different countries. The number of offers sent to China was 10,253 which is 42.2% of all offers sent. The second and third country on the basis of number of offers sent were India and South Korea with 5.55% and 4.59% of offers respectively. China and South Korea had 66% and 68% acceptance rates respectively, whereas India had a 44% acceptance rate.

d. *Economy Level of the country*

According to the World Bank classification, economies of the world are divided into four income groupings on the basis of Gross National Income (GNI) per capita in U.S. dollars converted from local currency: low income (GNI $1,025 or less), lower middle income (GNI $1,026 to $4,035), upper middle income (GNI $4,036 to $12,475) and high income (GNI $12,476 or more). The GNI reflects the average income of a country's citizens. It also tends to be linked with other indicators that measure the social, economic, and environmental well-being of the country and its people. The acceptance rate of 62.8%, is highest in the upper income level group and lowest (42%) in the lower middle income group, whereas in low income group and high income groups, the acceptance rate is 52.7% and 55.9% respectively.

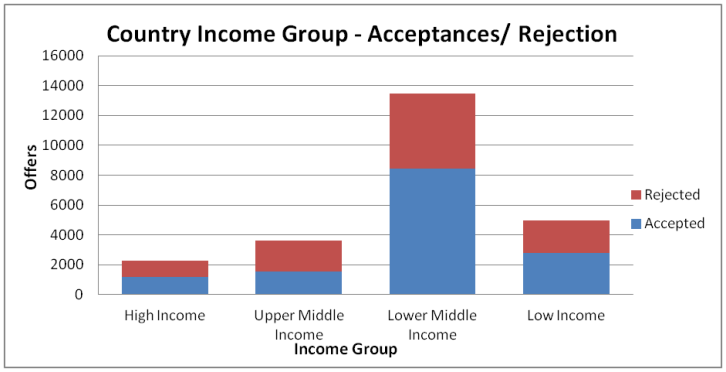


Figure 4.2: Income Group – Acceptances / Rejection

e. Visa Assessment Level

According to the Department of Immigration - Australia[2] , although each student visa application is considered on its individual merits, assessment levels facilitate this process, allowing fast decision-making and efficient service to student-candidates, maintaining the integrity of Australia's immigration program. There are three assessment levels in the student visa program. They align student visa requirements to the immigration risk posed by applicants from a particular country studying in a particular educational sector. Assessment level 1 represents the lowest immigration risk and assessment level 3 the highest risk. The higher the assessment level, the greater the evidence an applicant is required to demonstrate, to support their claims for the granting of a student visa.

The rate of acceptances for assessment Levels 1, 2 and 3 are 61.5%, 49.3% and 41.8% respectively.

ii. Academic

The academic information includes the year, study period for which the application was made, course and faculty related information.

a. Course Study Level

In 2013 and 2012, the percentage of offers for postgraduate places was 78.1% and 76.4% respectively, of which 53.3% and 55.4% were accepted.

[2] http://www.immi.gov.au/Study/Pages/student-visa-assessment-levels.aspx

In the same years, the percentage of offers for undergraduate places was 21.9% and 23.6% respectively, of which 38.4% and 46.1% were accepted

Table 4.1: Study Level – Acceptances / Rejection

	PostGraduate					UnderGraduate				
Year	Offers	Accepted	Accepted %	Rejected	Rejected %	Offers	Accepted	Accepted %	Rejected	Rejected %
2008	2609	1736	66.5%	873	33.5%	1222	670	54.8%	552	45.2%
2009	3027	1975	65.2%	1052	34.8%	1097	605	55.2%	492	44.8%
2010	3208	2026	63.2%	1182	36.8%	909	464	51.0%	445	49.0%
2011	2841	1637	57.6%	1204	42.4%	824	415	50.4%	409	49.6%
2012	3117	1727	55.4%	1390	44.6%	964	444	46.1%	520	53.9%
2013	3470	1850	53.3%	1620	46.7%	972	373	38.4%	599	61.6%

b. *Faculty and Courses*

Among the four faculties (name changed to A, B, C and D for anonymity) , Faculty A comprise of 67.3% of all the offers whereas Faculty B has 12.1% offers, Faculty D and Faculty C has 12.1% and 11.0% offer respectively.

In 2013, at the postgraduate level, out of the top 10 most popular courses on the basis of the number of acceptances, 8 courses were from the Faculty A, 1 course was from the Faculty B and 1 course was from the Faculty D.

In 2013, the acceptance rate for Faculty A was 53.5%. The Faculty B, C and Faculty D acceptance rates were 32.4%, 46.9% and 55.7% respectively.

Table 4.2: Faculty – Acceptances / Rejection

Faculty	2013 Offers	Accepted	Accepted %	Rejected	Rejected %
Faculty A	3058	1636	53.5%	1423	46.5%
Faculty B	639	207	32.4%	432	67.6%
Faculty C	424	199	46.9%	225	53.1%
Faculty D	334	186	55.7%	148	44.3%

iii. Application Related

The application related information include how many days or months ahead of the course start date the application was made, the GPA of the student when the student applied for the course, and whether the student received exemptions of units, and whether the student received any form of scholarship or not.

a. GPA

There were many records which had no value for GPA specified. Only 15,468 records had a GPA recorded, whereas 8,815 records did not have GPA recorded. The GPA ranged from 0 to 4. Most of the GPAs recorded were in the range of 2 to 4. The percentage of students who had their GPAs in the range of 2-3 and 3-4 were 46.4% and 51.2% respectively. The acceptance rates for the two groups were 61.0% and 58.4% respectively.

b. *CPS (Consideration of Prior Studies)*

Some students are exempted from some units based on their previous studies. Students apply for exemption of certain units, which is further assessed by the faculty / department and the student gets unit exemptions. Only 4,109 records or 16.9% records had exemptions from one or more units. Of the students who had an exemption, 58% accepted the course and 42% rejected the course.

c. *Scholarship*

The dataset records whether any form of scholarship has been awarded to the student and does not differentiate the types of scholarship.

Scholarship is one of the main attributes which determines whether the student will accept the offer or not. Only 807 or 3.3% records had a scholarship recorded. Of the students who had a scholarship, 78.1% students accepted the offer, whereas 21.9% rejected the offer.

d. *Duration of the application process ahead of course commencement*

To test the effects of the time period in which the students apply for a course, the number of months before the applied course study period was calculated. Some 46% of applications were made within 3 months of session start. The acceptance rate for the applications made within the first, second and third months were 77.8%, 63.1% and 52.9% respectively. Offers which were made for a course 1 year or more before actual commencement were accepted by 24.4% of students.

e. Application Channel

Some students apply on their own and are referred to as direct applications. Some students apply through agents. In short, 25.5% of the total number of offers sent belonged to direct applicants and 74.5% of the offers belonged to the students who applied through agents, clearly highlighting the important role of agents in the international student recruitment process. The acceptance rate of direct and agent applicants are 60.4% and 56.3% respectively.

Table 4.3: Channel – Acceptances / Rejection

Channel	Offers	Accepted	Accepted %	Rejected	Rejected %
Direct Applicant	6188	3740	60.4%	2448	39.6%
Agent Assisted	18095	10191	56.3%	7904	43.7%

4.3 Data Pre-Processing

The original data was pre-processed to compute null and missing values, handle outliers, transform and select important predictors.

The attributes *Gender*, *Age*, *Study Level*, *Assessment Level* and *GPA* had missing values. The attribute *Gender*, *Age* and *Study Level* had 2, 1 and 23 records respectively and will be excluded to minimize the effect in the prediction model as it is assumed that those values are mandatory values and missing of those values may be the result of some error. The *Assessment Level* attribute had 174 missing values and will be replaced by the most common assessment value 2. The attribute *GPA* had 8815 missing values and will be replaced by the mean GPA.

The original data provided did not contain country income group and assessment level. The country income group has been included for the dataset by combining data from World Bank classification, 2012 based on Gross National Income (GNI) per capita of 2011. The four income groups are Low income (GNI $1,025 or less), Lower middle income (GNI $1,026 to $4,035), Upper middle income (GNI $4,036 to $12,475) and High income (GNI $12,476 or more). The country of citizenship of each offer was mapped with country income group according to the World Bank classification.

Student visa assessment level for Subclass 573 Higher Education was imported from the government site of immigration. The country of citizenship of each offer was mapped with the assessment level specified by the department of immigration, Australia. Where data was presented in numeric code and description, the numeric code was used in the logistic regression and neural network model as they require data in numeric form. For decision making the description fields were used as it helps in easier interpreting of the decision tree.

4.4 Feature Selection

Correlation of the attributes with respect to the label attribute Acceptance Status was carried out to calculate the relevance of the input dataset with respect to the label attributes. Principal component analysis was run for the full data set to find the effect of each attribute in the acceptance or rejection of an offer.

4.4.1 Correlation of attributes:

The weight by correlation step provided by RapidMiner was used to calculate the relevance of the attributes by computing the value of correlation for each attribute of the input dataset with respect to the label attribute. This weighting scheme is based upon correlation and it returns the absolute or squared value of correlation as the attribute weight. The higher the weight of an attribute, the more relevant it is considered.

A correlation is a number between -1 and +1 that measures the degree of association between two attributes (called X and Y). A positive value for the correlation implies a positive association. In this case, large values of X tend to be associated with large values of Y and small values of X tend to be associated with small values of Y. A negative value for the correlation implies a negative or inverse association. In this case, large values of X tend to be associated with small values of Y and vice versa.

The following table shows the correlation of the attributes with the label attribute Acceptance status.

Table 4.4: Weight by Correlation of Attributes

Attribute	Weight
AssessmentLevel	0.68263916
CountryId	0.433360424
CourseStudyLevel	0.370111962
FacultyId	0.328343834
HasScholarship	0.315527423
CountryIncomeGroupId	0.272970216
GPA	0.219137952
Age	0.174026919
channelid	0.133169865
GenderId	0.069509805
courseId	0.064568902
semester	0.044613391
OfferStatus	0.010834942

The important attributes, having positive correlation with the label attribute Acceptance status and weight greater than 0.1 from the above correlation table are *Assessment Level*, *CountryId*, *CourseStudyLevel*, *FacultyId*, *HasScholarship*, *CountryIncomeGroupId*, *GPA*, *Age* and *ChannelId*. Also the *Channelid*, *Genderid*, *Courseid* , *Semester* and *OfferStatus* have shown positive correlations with the label attribute Acceptance status.

4.4.2 Principal Component Analysis

Principal Component Analysis (PCA) is a technique which allows reducing the dimension of a dataset by identifying a few of the most influential parameters, if they exist. This sort of variable screening or feature selection will make it easy to apply other predictive modeling techniques and also make the job of interpreting the results easier. PCA captures the parameters which explain the greatest amount of variation in the dataset. It does this by transforming the existing variables into a set of "principal components" or new variables which have the following properties:

i. They are uncorrelated with each other;

ii. They cumulatively contain/explain a large amount of variance within the data; and

iii. They can be related back to the original variables via weightage factors.

Original variables with very low weightage factors in their principal components can be removed from the dataset.

Using Eigenvalues, information can be obtained about the contribution to the data variance coming from each principal component individually and cumulatively. When Principal Component Analysis was applied on the dataset, the first two principal components, PC1 and PC2 allowed for 99% variance.

Table 4.5: Eigen Values of Principal Components

Component	Standard Deviation	Proportion of Variance	Cumulative Variance
PC 1	19743.68	0.993	0.993
PC 2	1285.454	0.004	0.997
PC 3	1074.715	0.003	1
PC 4	218.032	0	1
PC 5	118.944	0	1
PC 6	4.138	0	1
PC 7	1.008	0	1
PC 8	0.719	0	1
PC 9	0.567	0	1
PC 10	0.519	0	1

Table 4.6: EigenVectors of Attributes in Principal Components

Attribute	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10
applicationid	-0.999	0.032	0.003	0	0.001	0	0	0	0	0
GenderId	0	0	0	0	0	0.008	0.063	0.072	-0.131	0.686
Age	0	0	-0.001	-0.003	0	-0.999	0.008	-0.001	0.001	0.004
CountryId	0	0.004	-0.04	-0.998	-0.045	0.003	0	0.001	0	0
CountryIncomeGroupId	0	0	0	0	0	0.009	0.748	-0.194	0.254	-0.402
AssessmentLevel	0	0	0	0	0	-0.001	-0.646	-0.077	0.17	-0.463
year	0	0	0	0	0.002	0.007	0.053	-0.052	-0.662	-0.209
semester	0	0	0	0	-0.001	-0.003	-0.019	0.202	0.617	0.159
AppliedDaysBefore	0.001	0	-0.009	-0.045	0.999	0	-0.001	-0.003	0.002	0.001
courseId	-0.033	-0.992	-0.121	0.001	-0.001	0	0	0	0	0
FacultyId	0.001	0.121	-0.992	0.041	-0.007	0.001	0	0	0	0
CourseStudyLevel	0	0	0	0	0	-0.042	0.007	0.069	-0.114	-0.019
channelid	0	0	0	0	0.001	0.016	0.002	0.116	0.06	0.074
GPA	0	0	0	0	0	0.006	-0.021	0.02	0.017	0.157
CPS	0	0	0	0	0	0.015	0.018	-0.039	-0.022	0.042
HasScholarship	0	0	0	0	0	-0.002	0	0.004	-0.02	-0.005
OfferStatus	0	0	0	0	-0.003	-0.003	-0.095	-0.935	0.066	0.223
AcceptanceStatus	0	0	0	0	-0.001	-0.006	0.075	0.118	-0.219	-0.05

Only those real parameters which have significant weightage contribution to the each of the first 2 PCs were considered. These will ultimately form the subset of reduced parameters for further predictive modeling. The important attributes from PC1 and PC2 are *Countryid, AppliedDaysBefore, Courseid*, and *Facultyid*.

4.5 Conclusion

The acceptance of offers made by the students is a complex process and interdependent on multiple attributes. The average acceptance rate has dropped from 2008 to 2013, and in 2013 the average acceptance rate was 50%. The acceptance rate of packaged offers is 58% which is slightly better than the acceptance rate for full offers. Out of the two semesters in a year, the number of offers sent in the second semester is about 24% more in average, than the number of offers sent in the first semester but the acceptance rate is similar.

Although the offers were sent to 145 different countries, there are certain countries that dominate from which the university draws students, such as China, India and South Korea. The acceptance rate of 62.8% is highest in the upper income level group, and least (42%) in the lower middle income group. Since assessment levels of a country represents the immigration risk level and the higher the assessment level, the greater the evidence an applicant is required to demonstrate, to support their claims for the granting of a student visa, the acceptance rate for countries in assessment level 1, which is 61.5% is the highest, compared to 49.3% and 41.8% for Assessment Levels 2 and 3.

Among international students, more than 75% students are postgraduate students and the acceptance rate of postgraduate students (55.4%) is also better than the acceptance rate of undergraduate students which is 46.1%. Most offers are sent from the Faculty A (more than 60%). The acceptance rate is 53.5%. The total number of offers sent belonged to direct applicants, and 74.5% of the offers belonged to students who applied through agents - clearly highlighting the role of agents in the international student recruitment process. The acceptance rate of direct and agent applicants are 60.4% and 56.3% respectively. The students who have been awarded a scholarship have a better chance of accepting the offer. Some 78.1% students who received a scholarship also accepted the offer. Also, of the students who had an exemption, 58% accepted the course and 42% rejected the course.

The dataset provided was mostly clean. The original data was pre-processed to compute null and missing values, handle outliers, transform and select important predictors. The records with missing values in *Gender, Age and StudyLevel* were excluded. The missing

values in *Assessment Level* attribute was replaced by the most common assessment value and missing values in *GPA* was replaced by the mean *GPA.*

The original data provided did not contain country income group and assessment level. The *Country Income Group* has been included for the dataset by combining data from World Bank classification (2012) based on Gross National Income (GNI) per capita of 2011. The student visa assessment level for the countries has been integrated from the government website of immigration.

Correlation of the attributes with respect to the label attribute *Acceptance Status* was carried out to calculate the relevance of the input dataset with respect to the label attribute. The important attributes, having positive correlation with the label attribute *Acceptance Status* and weight greater than 0.1 from the above correlation table are: *Assessment Level, CountryId, CourseStudyLevel, FacultyId, HasScholarship, CountryIncomeGroupId, GPA, Age* and *ChannelId*.

Principal Component Analysis (PCA) was run for the full data set to find the effect of each attribute in the acceptance or rejection of the offer. PCA captures the parameters which explain the greatest variation in the dataset. When PCA was applied on the dataset, the first two principal components, PC1 and PC2 allowed for 99% variance. The important attributes from PC1 and PC2 *are Countryid, AppliedDaysBefore, Courseid,* and *Facultyid*.

Chapter 5: Prediction Models

5.1 Introduction

The classification models used in the analysis include logistic regression, decision trees and neural networks, which are among the common models used in the enrolment prediction domain. Different models have specific data requirements. While logistic regression and neural networks require only numeric data, decision trees could also use text or nominal data. With decision trees, the focus is on readability and interpretability of the prediction model. Logistic regression and neural network models were focused to improve prediction accuracy.

RapidMiner 5.3[3] has been used for developing and testing the prediction models which is available under a OSI-certified open source license. RapidMiner is a powerful advanced analytics platform for data mining, machine learning and predictive analytics. Parameter settings for these models are discussed for each model. A series of parameter settings were tried and only the best set is shown in the report. Some terminologies specific to the RapidMiner are included in Appendix C.

The dataset has been divided into training and testing datasets. The ten-fold cross-validation technique has been used to assess the efficacy of the model and ensure that the predicted outcome is not the result of overfitting on a single dataset, but tested on a dataset randomly partitioned into 10 mutually exclusive subsets or folds, each of which is approximately equal size.

Receiver operating characteristic (ROC) curve and lift charts have been used to compare the models and decide which model performs the best. The ROC curve defines the area between true positives and false positives.

5.2 Logistic Regression

The logistic regression model in RapidMiner uses the Java implementation of the myKLR which is a tool for large scale kernel logistic regression based on the algorithm of Keerthi et al. (2003) and the code of mySVM. This learning method can be used for both regression and classification and provides a fast algorithm and good results for

[3] https://rapidminer.com/documentation/

many learning tasks. The logical regression operator cannot handle nominal attributes; it can be applied on data sets with numeric attributes.

The different types of supported kernel types are dot, radial, polynomial, neural, ANOVA, epachnenikov, gaussian combination and multiquadric. When the different kernel types were tested, ANOVA kernel type produced the best result. The ANOVA kernel is defined by raised to power d of the summation of exp(-g (x-y)) .

$$k(x,y) = \sum_{k=1}^{n} \exp(-\sigma(x^k - y^k)^2)^d$$

where g is gamma and d is degree. Gamma and degree are adjusted by the kernel gamma and kernel degree parameters respectively.

The kernel cache parameter (table 5.1) specifies the size of the cache for kernel evaluations in megabytes. C is the SVM complexity constant which sets the tolerance for misclassification, where higher C values allow for 'softer' boundaries and lower values create 'harder' boundaries. A complexity constant that is too large can lead to overfitting, while values that are too small may result in over-generalization. We use the value 200 for kernel cache and 1 for c in the model.

Table 5.1 Logistic Regression Model Parameters

Logistic Regression (2) (Logistic Regression)	
kernel type	anova
kernel gamma	1.0
kernel degree	2.0
kernel cache	200
C	1.0
convergence epsilon	0.001
max iterations	100000

Using the *ANOVA* kernel type, the prediction accuracy for acceptance of offers is 67.33% and 67.86% for rejection of offers (table 5.2).

Table 5.2 Logistic Regression Predictive Performance

Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	4373	9011	67.33%
Predicted to Reject	7381	3495	67.86%
Class Recall	62.80%	72.05%	

The weight table generated (table 5.3) shows that the HasScholarship variable has the highest weight. The other important attributes are Faculty, Age, Country Income Group, GPA and CourseStudyLevel.

Table 5.3 Logistic Regression Weight Table

Attribute	Weight
HasScholarship	0.588907
FacultyId	0.1870228
Age	0.1823
CountryIncomeGroupId	0.1120958
GPA	0.1084293
CourseStudyLevel	0.101902
courseId	0.0951721
CPS	0.0885565
GenderId	0.0347624
channelid	0.0231332
OfferStatus	-0.0624488
year	-0.0928392
semester	-0.0948121
AssessmentLevel	-0.1973965
CountryId	-0.3180795
AppliedDaysBefore	-1.2895003

In order to see how different the prediction model performed on different subgroups of the applicants, the models were separately run with Semester 1 and Semester 2 records and undergraduate and postgraduate students separately (tables 5.4 and 5.5 - a and b).

Table 5.4 a and b Logistic Regression Predictive Performance (Semesterwise)

Semester 1			
Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	1890	4869	72.03%
Predicted to Reject	4512	1968	69.63%
Class Recall	70.47%	71.22%	

Semester 2			
Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	2217	4518	67.08%
Predicted to Reject	2735	1552	63.80%
Class Recall	55.22%	74.44%	

Table 5.5 a and b Logistic Regression Predictive Performance (Study Levelwise)

Undergraduate			
Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	1076	1736	61.73%
Predicted to Reject	2015	904	69.04%
Class Recall	65.19%	65.77%	

Postgraduate			
Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	3174	7633	70.63%
Predicted to Reject	5089	2633	65.90%
Class Recall	61.58%	74.35%	

The prediction accuracy for acceptance for Semester 1, 72.03% was better than for Semester 2, 67.08%. Similarly the prediction accuracy for acceptance for postgraduate level, 70.63% was better than for undergraduate level, 61.73%.

5.3 Decision Trees

A decision tree generates a tree structure for classification and supports both nominal and numerical data. A decision tree has its root at the top and grows downwards. This representation of the data has the advantage when compared with other approaches of being meaningful and easy to interpret. The goal is to create a classification model that predicts the value of a target attribute or class label based on several input attributes of the dataset.

In RapidMiner, an attribute with a label role is predicted by the Decision Tree operator. Each interior node of the tree corresponds to one of the input attributes. The number of the edges of a nominal interior node is equal to the number of possible values of the corresponding input attribute. The outgoing edges of numerical attributes are labeled with disjoint ranges. Each leaf node represents a value of the label attribute given the values of the input attributes represented by the path from the root to the leaf. Decision trees are generated by recursive partitioning which means repeatedly splitting on the

values of attributes. The attribute used to split is selected depending upon a selection criterion which can be selected by the criterion parameter.

It can have one of the following values:

i. information_gain: The entropy of all the attributes is calculated. The attribute with the minimum entropy is selected for split. This method has a bias towards selecting attributes with a large number of values.
ii. gain_ratio: It is a variant of information gain. It adjusts the information gain for each attribute to allow the breadth and uniformity of the attribute values.
iii. gini_index: This is a measure of impurity of an ExampleSet. Splitting on a chosen attribute gives a reduction in the average gini index of the resulting subsets.
iv. accuracy: An attribute is selected for split that maximizes the accuracy of the whole tree.

When the different splitting criteria were tested, gini_index produced the best result.

The maximal depth parameter is used to restrict the size of the Decision Tree. The tree generation process is not continued when the tree depth is equal to the maximal depth. The maximal depth 20 was used in the model. The confidence parameter specifies the confidence level used for the pessimistic error calculation of pruning and was set at 0.25.

The prediction accuracy for the acceptance of offers (table 5.6) was 63.79% and 60.01% for the rejection of offers. The decision trees produced have been included in the Appendix B.1

Table 5.6 Decision Tree Predictive Performance

Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	4858	8559	63.79%
Predicted to Reject	6507	4336	60.01%
Class Recall	57.25%	66.38%	

5.4 Neural Networks

The neural network operator in RapidMiner learns a model by means of a feed-forward neural network - trained using a back propagation algorithm (multi-layer perceptron). This operator cannot handle nominal attributes.

A neural network consists of an interconnected group of artificial neurons, and it processes information using a connectionist approach to computation. A feed-forward neural network is an artificial neural network where the information moves in only one direction - forward, from the input nodes, through the hidden nodes (if any) to the output nodes. A back propagation algorithm is a supervised learning method which can be divided into two phases: propagation and weight update. The two phases are repeated until the performance of the network is good enough. In back propagation algorithms, the output values are compared with the correct answer to compute the value of some predefined error-function. By various techniques, the error is then fed back through the network. Using this information, the algorithm adjusts the weights of each connection in order to reduce the value of the error function by some small amount. After repeating this process for a sufficiently large number of training cycles, the network will usually converge to some state where the error of the calculations is small. In this case, one would say that the network has learned a certain target function.

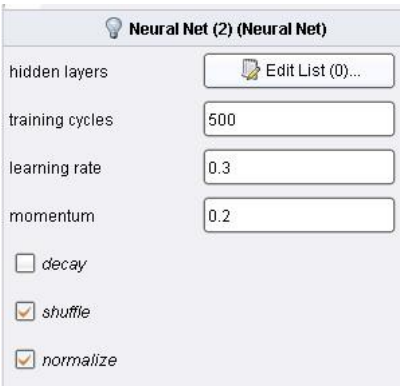
A multilayer perceptron (MLP) is a feed-forward artificial neural network model that maps sets of input data onto a set of appropriate output. An MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes back propagation for training the network. This class of networks consists of multiple layers of computational units, usually interconnected in a feed-forward way. In many applications the units of these networks apply a sigmoid function as an activation function.

In this operator, the usual sigmoid function is used as the activation function. Therefore, the values ranges of the attributes should be scaled to -1 and +1. This can be done through the normalize parameter. The type of the output node is sigmoid if the learning data describes a classification task, and linear if the learning data describes a numerical regression task.

The main parameter - hidden layers, describes the name and the size of all hidden layers. The user can define the structure of the neural network with this parameter. Each list entry describes a new hidden layer. Each entry requires the name and size of the hidden layer. If the hidden layer size value is set to -1, the layer size would be calculated from the number of attributes of the input example set. In this case the layer size will be set to (number of attributes + number of classes) / 2 + 1. If the user does not specify any hidden layers, a default hidden layer with sigmoid type and size equal to (number of attributes + number of classes) / 2 + 1 will be created and added to the net.

Normalize is an expert parameter. The Neural Net operator uses a usual sigmoid function as the activation function. Therefore, the value range of the attributes should be scaled to -1 and +1: This can be done through the normalize parameter (table 5.7). Normalization is performed before learning. Although it increases runtime, it is necessary in most cases.

Table 5.7 Neural Network Model Parameters



The prediction accuracy for the acceptance of offers (table 5.8) is 68.09% and 65.84% for rejection of offers.

Table 5.8 Neural Network Predictive Performance

Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	5300	11308	68.09%
Predicted to Reject	5038	2614	65.84%
Class Recall	48.73%	81.22%	

The neural network used by the model with 1 hidden layer is shown below.

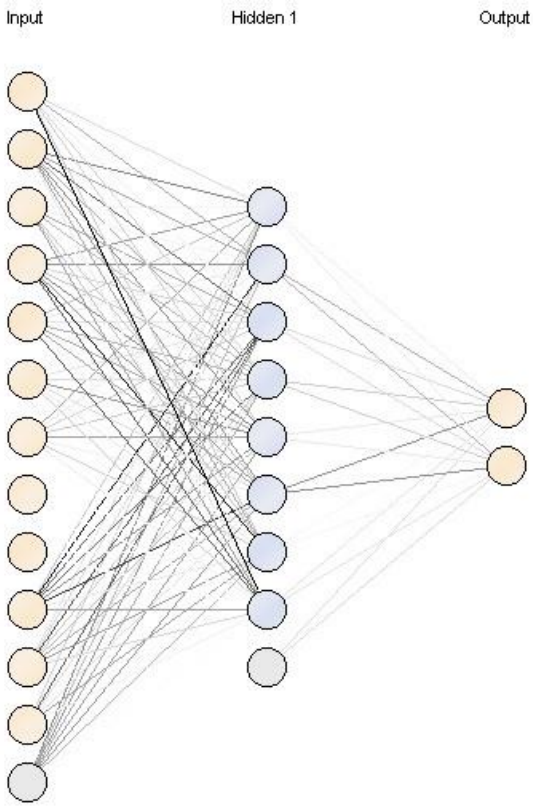


Fig 5.1 Neural Network used by the model

As with the logistic regression mode, the neural network model was run separately with Semester 1 and Semester 2 records, and Undergraduate and postgraduate student records.

The prediction accuracy for Semester 2 was better (69.15%), where for Semester 1 the result was 67.89%. Similarly, the prediction accuracy for undergraduate and postgraduate when run separately, was 65.39% and 68.40% respectively (table 5.8 a and b).

Table 5.8 a and b Neural Network Predictive Performance (Semesterwise)

Semester 1			
Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	2765.5162	5847.585	67.89%
Predicted to Reject	3036.9837	1743.915	63.52%
Class Recall	52.34%	77.03%	

Semester 2			
Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	2276.7829	5103.134	69.15%
Predicted to Reject	2353.3299	1382.753	62.99%
Class Recall	50.83%	78.68%	

Table 5.9 a and b Neural Network Predictive Performance (Study Levelwise)

Undergraduate			
Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	1096	2071	65.39%
Predicted to Reject	1921	900	68.10%
Class Recall	63.67%	69.71%	

Postgraduate			
Prediction	Actual Outcome		Class Precision
	Accepted	Rejected	
Predicted to Accept	4104	8885	68.40%
Predicted to Reject	3217	2066	60.89%
Class Recall	43.94%	81.13%	

5.5 Models Performance Comparison

ROC curves are useful visual tools for comparing two or more classification models. An ROC curve shows the trade-off between the true positive rate or sensitivity (the proportion of the positive tuples that are correctly identified) and the false-positive rate (the proportion of the negative tuples that are incorrectly identified as positive) for a given model. Any increase in the true positive rate occurs at the cost of an increase in the false-positive rate. The area under the ROC curve is a measure of the accuracy of the model. The generated ROC curve for logistic regression and neural network is shown in figure 5.2 below.

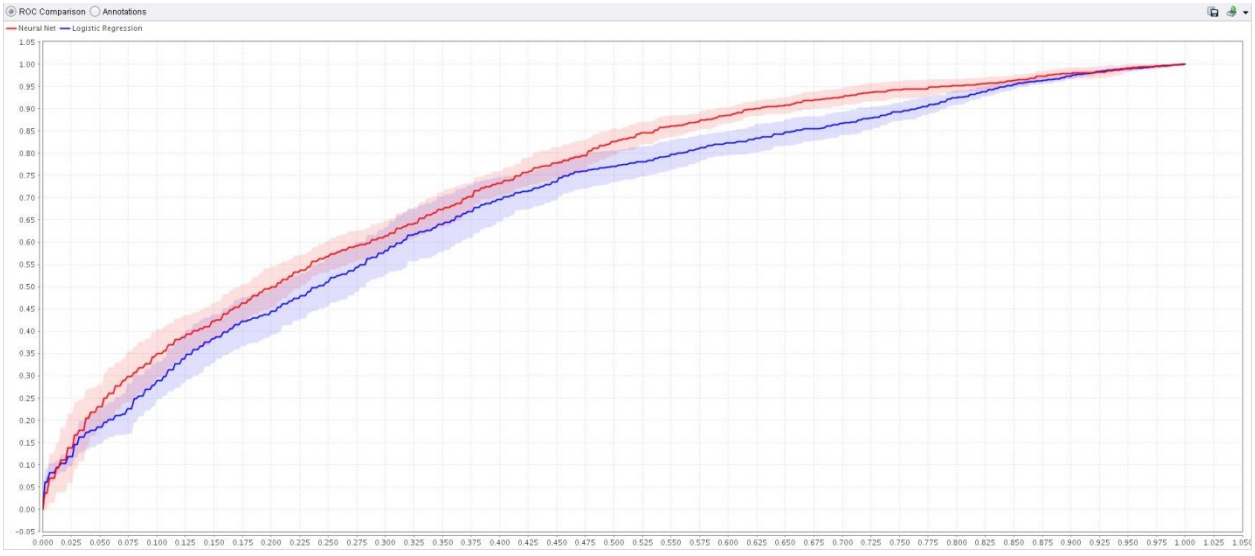


Fig 5.2 ROC Curve for Logistic Regression and Neural Network

If 50% of the population is tested; in the case of logistic regression, around 72% of the students would be correctly predicted to accept the offer. In the case of a neural network 77% of the students would be correctly predicted to accept the offer. The red line

representing the neural network is slightly steeper than the blue line for logistic regression.

Also, the lift charts generated from RapidMiner showed that the neural network is better than logistic regression. Lift chart expresses the cases in batches sorted by confidence for true and the percentage that are predicted correctly. The lift chart generated from Neural Network model, the first batch of the students who accepted the offer was predicted with 92% accuracy where as logistic regression predicted with 83% accuracy. The lift charts have been included in Appendix 5.2

5.6 Conclusion

The classification models have been developed using Logistic Regression, Decision Trees and Neural networks. While logistic regression and the neural network models were focused to improve prediction accuracy, the decision tree model was focused on readability and interpretability of the prediction model.

The weight table generated by logistic regression showed that the HasScholarship variable has the highest weight. The other important attributes are Faculty, Age, Country Income Group, GPA and CourseStudyLevel. Using the ANOVA kernel type, the prediction accuracy for acceptance is 67.33%, and 67.86% for rejection of offers. The models were separately run with semester 1 and semester 2 records and undergraduate and postgraduate student records. The prediction accuracy for acceptance for Semester 1 at 72.03%, was better than for Semester at 67.08%. Similarly the prediction accuracy for acceptance for postgraduate level was 70.63%, which was better than for undergraduate level at 61.73%.

The decision tree model used gini_index as the splitting criteria. The prediction accuracy for the acceptance of offers was 63.79%, and 60.01% for the rejection of offers.

A feed-forward neural network trained by a back propagation algorithm (multi-layer perceptron) was used to train the model. The prediction accuracy for the acceptance of offers was 68.09% and 65.84% for rejection of offers. The prediction accuracy for semester 2 was better at 69.15% while for semester 1 it was less at 67.89%. Similarly

the prediction accuracy for undergraduate and postgraduate when run separately, was 65.39% and 68.40% respectively.

The performance of logistic regression and neural network were quite similar, but neural network models performed slightly better than logistic regression with 68.09% of prediction accuracy for acceptance, and 65.84% for rejection of offers. Logistic regression predicted 67.33% correctly for acceptance, and 67.86% for the rejection of offers. The decision tree predicted 63.79% correctly for acceptance, and 60.01% for the rejection of offers. The decision tree helped in the interpretation and readability of the model.

ROC Curves and lift charts were used for model performance comparisons. In the ROC curve, the red line representing the neural network is slightly steeper than the blue line for logistic regression. ROC curves showed that the neural network model was slightly better than logistic regression. Also, by comparing the confidence level and prediction accuracy of the batches in the lift charts generated from logistic regression and neural networks, neural network is better than logistic regression.

Chapter 6: Conclusions

The thesis examined the validation of prediction models of acceptance of offers by students in the context and settings of international students in a large Australian University using data mining techniques. Prediction models for acceptance of offers have been successfully developed using the data and settings of an Australian university with a focus on coursework-based international students. The acceptance of offers made by the students is a complex process and interdependent on multiple attributes. An analysis with a dataset comprising offers to students from a large metropolitan Australian university revealed that the average acceptance rate has dropped from 2008 to 2013 and the current average acceptance rate is around 50%. Although the offers were sent to prospective students from 145 different countries, there are certain countries from which the university draws large numbers of students, such as China, India and South Korea. The acceptance rate of 62.8%, is highest in the upper income level group and the lowest (42%) in the lower middle income group. The acceptance rate for countries in assessment level 1 (61.5%) is the highest, compared to 49.3% and 41.8% for Assessment Levels 2 and 3.

The number of offers sent to applicants for postgraduate studies is three times more than the number of offers sent to applicants for undergraduate studies and the acceptance rate for postgraduate places at 55.4% is also better than acceptance rate for undergraduate places which is 46.1%. One of the faculties which offers the business program is the coursework strength of the university, representing more than 60% of all offers. Of the total number of offers sent, 74.5% of the offers belonged to prospective students who applied through agents - clearly highlighting the role of agents in the international student recruitment process. The acceptance rate of direct and agent applicants are 60.4% and 56.3% respectively. The students who have been awarded scholarships have a better chance of accepting the offer. Some 78.1% students who received a scholarship accepted the offer. Also, of the students who had an exemption, 58% accepted the offer whereas 42% rejected the offer.

The dataset used in evaluation was mostly clean. The original data was pre-processed to compute null and missing values, handle outliers, transform and select important

predictors. The original data provided did not contain country income group and assessment level, which were integrated from external sources; namely the World Bank classification as well as the government website of immigration respectively. Correlation of the attributes with respect to the label attribute Acceptance Status as well as Principal Component Analysis was carried out for feature selection. The evaluation identified several important attributes which impact acceptance of offers, namely, *Assessment Level*, *CountryId*, *CourseStudyLevel*, *FacultyId*, *HasScholarship*, *CountryIncomeGroupId* and *ChannelId.*

The prediction models have been developed using Logistic Regression, Decision Trees and Neural networks. While logistic regression and neural network models were focused to improve prediction accuracy, the decision tree model focused on readability and interpretability of the prediction model.

Using the ANOVA kernel type of logistic regression, the prediction accuracy for acceptance is 67.33%, and 67.86% for rejection of offers. With the decision tree model using gini_index as the splitting criterion, the prediction accuracy for the acceptance of offers was 63.79% and 60.01% for the rejection of offers. A feed-forward neural network trained by a back propagation algorithm produced a prediction accuracy of 68.09% for acceptance of offers and 65.84% for rejection of offers.

The models when run separately with Semester 1 and Semester 2 records and Undergraduate and postgraduate students, proved slightly better suggesting that multiple models can be created semester-wise or study level wise to produce better results.

The performance of models created by Logistic Regression and Neural Networks were quite similar but the neural network model with 68.09% prediction accuracy for acceptance and 65.84% for rejection of offers, performed slightly better than Logistic Regression. Logistic Regression predicted 67.33% correctly for acceptances and 67.86% for rejection of offers. The decision tree predicted 63.79% correctly for acceptances and 60.01% for rejection of offers. The decision tree helped in interpretability and readability of the model. Receiver Operating Characteristic (ROC) Curves and lift charts were used for model performance comparison.

In conclusion, the thesis examined the validation of prediction models regarding acceptances of offers by international students in the context of an Australian university. The prediction accuracy rate for the acceptance of offers of the developed prediction model at 68%, is better than the average acceptance rate of 50%. The important predictors for the acceptance of offers were the chosen course and the faculty, whether the student was awarded any form of scholarship and also the visa assessment level of the country by immigration department. The comparison of the prediction models developed using logistic regression, decision trees and neural networks, showed that the prediction model by neural networks produced the best result.

With regard to future research directions, it remains to be seen whether other models could also be applied to the problem of prediction of acceptance of offers and whether they provide better (or worse) performance compared to the three models compared here. The research can also be extended to incorporate the prediction of the quality of students in terms of academic success and likely timely completion, along with the prediction of acceptance of offers to target the students who are more likely to be successful and will complete their course on time. Such outcomes help the university attain quality objectives, whilst addressing student retention issues.

Appendix A

Original Data – Prior to any transformations

1. Summary of data

Table A.1 – Column Level Summaries

Field	Description	Type	Column statistics	Range	Missings
Applicationid	Unique Identifer	integer	avg = 189513.336 +/- 19733.219	[159424.000 ; 233900.000]	0
Gender	Gender of the applicant	binominal	mode = F (14081), least = M (10200)	M (10200), F (14081)	2
GenderId	Genderid 1 for Male, 2 for female	integer	avg = 1.580 +/- 0.494	[1.000 ; 2.000]	2
Age	Age of the applicant	integer	avg = 28.396 +/- 4.579	[5.000 ; 69.000]	1
Country	Country of Citizenship	polynominal	mode = China (10253), least = Uruguay (1)	China (10253), India (1348), South Korea (1114)..	0
CountryId	CountryId of Citizenship	integer	avg = 324.402 +/- 222.136	[4.000 ; 895.000]	0
CountryIncomeGroup	Income Group of the country (High, Upper Middle, Lower Middle, Lower)	polynominal	mode = Upper middle income (13470), least = Low income (2253)	Upper middle income (13470), High income (4939), Lower middle income (3621), Low income (2253)	0
CountryIncomeGroupId	Income Group Id of the country	integer	avg = 2.869 +/- 0.840	[1.000 ; 4.000]	0
AssessmentLevel	Assessment Level of the country (1, 2, 3)	integer	avg = 1.382 +/- 0.769	[1.000 ; 4.000]	174
year	Course Start Year	integer	avg = 2010.553 +/- 1.724	[2008.000 ; 2013.000]	0
semester	Study Period (Semester 1 , Semester 2)	integer	avg = 1.425 +/- 0.494	[1.000 ; 2.000]	0

ApplicationDate	Date of application	Date		[12/3/2005 ; 15/10/2013	0
AppliedDaysBefore	Number of days applied before the course starts	integer	avg = 131.044 +/- 120.844	[0.000 ; 1937.000]	0
courseId	Course Id of the applied course	integer	avg = 2108.804 +/- 1434.367	[3.000 ; 4911.000]	0
CourseCode	Course Code of the applied course	polynominal	mode = MACCG(CPA) (1981), least = PC-SOCH (1)	MACCG(CPA) (1981), MCOMM;FIN (1021), BCOMM (925)….	0
CourseTitle	Course title of the applied course	polynominal	mode = Master of Accounting (CPA Extension) (2720), least = Postgraduate Certificate in Social Health (1)	Master of Accounting (CPA Extension) (2720), Master of Accounting (Professional) (1202)…	0
Faculty	Faculty owning the applied course	polynominal	mode = Faculty of Business and Economics (16043), least = Macquarie Graduate School of Management (301)	Faculty of Business and Economics (16043), Faculty of Science (2937), Faculty of Arts (2329)….	0
FacultyId	FacultyId owning the applied course	integer	avg = 1692.750 +/- 1077.345	[1011.000 ; 4011.000]	0
StudyLevel	StudyLevel (Undergraduate, Postgraduate)	binominal	mode = Postgraduate (18272), least = Undergraduate (5988)	Postgraduate (18272), Undergraduate (5988)	23
CourseStudyLevel	StudyLevel (Undergraduate: 1, Postgraduate: 2)	integer	avg = 1.753 +/- 0.431	[1.000 ; 2.000]	23
Channel	Application Channel (Direct, Agent)	binominal	mode = Agent (18095), least = Direct (6188)	Direct (6188), Agent (18095)	0
channelid	Application Channel (Direct:1 , Agent:2)	integer	avg = 1.745 +/- 0.436	[1.000 ; 2.000]	0
GPA	Course GPA of the previously completed degree	numeric	avg = 3.033 +/- 0.484	[0.000 ; 4.000]	8815
CPS	CPS (Consideration of Previous Studies: Unit Exempt (0, 1))	integer	avg = 0.169 +/- 0.375	[0.000 ; 1.000]	0
HasScholarship	Has been offered Scholarship or not	integer	avg = 0.033 +/- 0.179	[0.000 ; 1.000]	0

ApplicationStatus	Offer Type Q (Qualified), FWC(Full offer With condition) , QPAC(Qualified with Package)	binominal	mode = Q (18423), least = QPAC (5086)	Q (18423), QPAC (5086)	774
OfferStatus	Offer Type Q:8, FWC: 7, QPAC: 6	integer	avg = 7.549 +/- 0.816	[6.000 ; 8.000]	0
AdmissionStatus	AdmissionStatus A (Accepted), DP (Deposit Paid) , W (Withdrawn)	polynominal	mode = A (13760), least = DP (6)	A (13760), W (1621), NULL (8731), DEF (165), DP (6)	0
AcceptanceStatus	AcceptanceStatus Accepted:1, Rejected:0	integer	avg = 0.574 +/- 0.495	[0.000 ; 1.000]	0

Appendix B

1. Decision Tree generated by the model

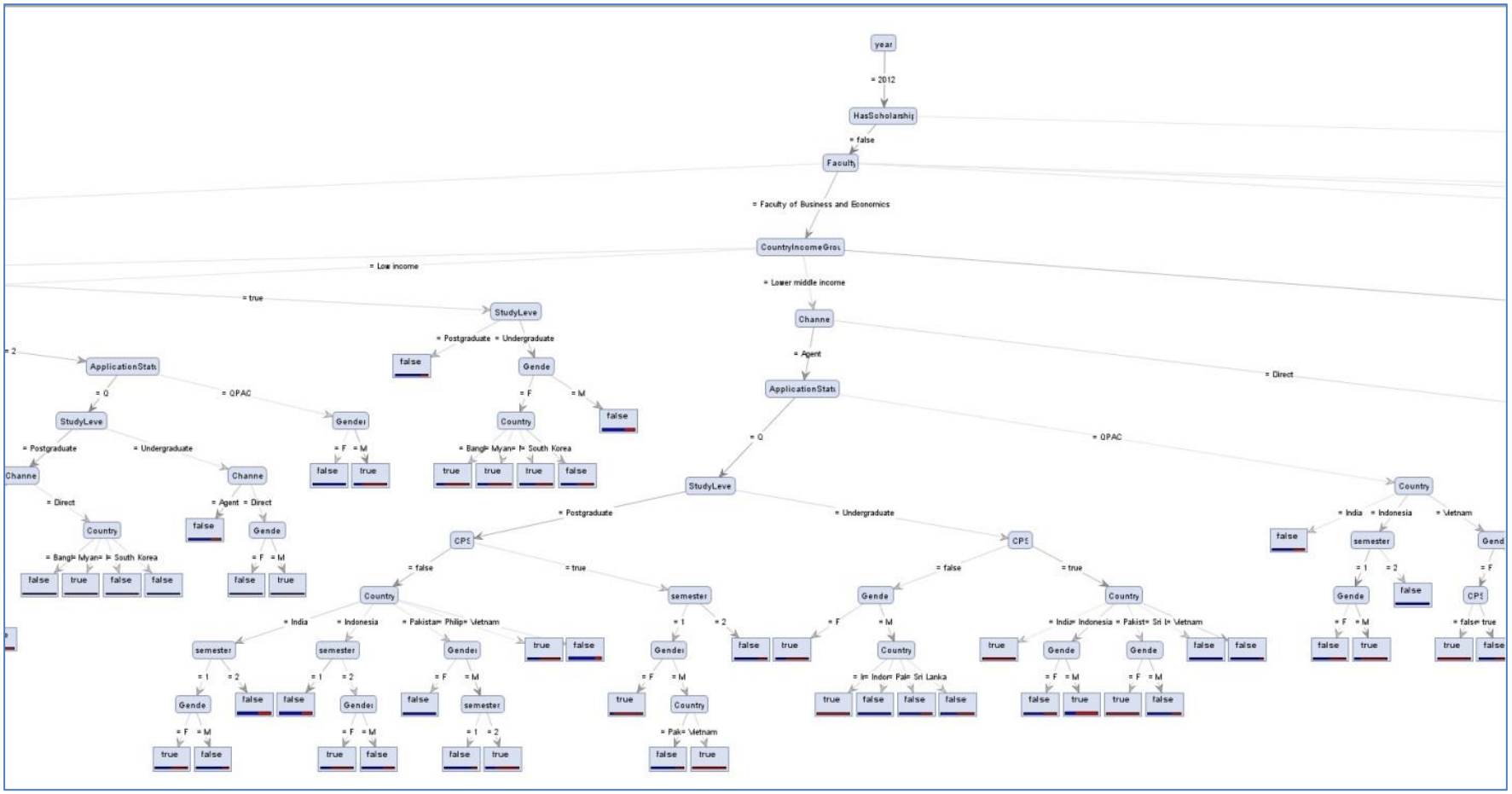


Fig B.1 Decision Tree

2. Decision Tree branching

```
year = 2012

|   HasScholarship = false
|   |   Faculty = Faculty of Arts
|   |   |   CountryIncomeGroup = High income
|   |   |   |   Channel = Agent
|   |   |   |   |   ApplicationStatus = Q
|   |   |   |   |   |   StudyLevel = Postgraduate: false {false=21, true=15}
|   |   |   |   |   |   StudyLevel = Undergraduate
|   |   |   |   |   |   |   semester = 1: false {false=16, true=7}
|   |   |   |   |   |   |   semester = 2
|   |   |   |   |   |   |   |   CPS = false: false {false=2, true=0}
|   |   |   |   |   |   |   |   CPS = true: true {false=1, true=3}
|   |   |   |   |   ApplicationStatus = QPAC: false {false=4, true=1}
|   |   |   |   Channel = Direct
|   |   |   |   |   StudyLevel = Postgraduate
|   |   |   |   |   |   ApplicationStatus = Q
|   |   |   |   |   |   |   CPS = false
|   |   |   |   |   |   |   |   semester = 1: true {false=5, true=11}
|   |   |   |   |   |   |   |   semester = 2
|   |   |   |   |   |   |   |   |   Gender = F: true {false=2, true=3}
|   |   |   |   |   |   |   |   |   Gender = M: false {false=5, true=1}
|   |   |   |   |   StudyLevel = Undergraduate
|   |   |   |   |   |   semester = 1
|   |   |   |   |   |   |   ApplicationStatus = Q
|   |   |   |   |   |   |   |   CPS = false
|   |   |   |   |   |   |   |   |   Gender = F: true {false=1, true=2}
|   |   |   |   |   |   |   |   |   Gender = M: false {false=3, true=1}
|   |   |   |   |   |   |   |   CPS = true: false {false=4, true=1}
|   |   |   |   |   |   semester = 2
|   |   |   |   |   |   |   Gender = F: false {false=4, true=3}
|   |   |   |   |   |   |   Gender = M: true {false=1, true=4}
|   |   |   CountryIncomeGroup = Low income
|   |   |   |   semester = 1
|   |   |   |   |   Channel = Agent: false {false=10, true=0}
|   |   |   |   |   Channel = Direct
|   |   |   |   |   |   ApplicationStatus = Q
|   |   |   |   |   |   |   CPS = false
|   |   |   |   |   |   |   |   StudyLevel = Postgraduate
|   |   |   |   |   |   |   |   |   Country = Kenya: false {false=2, true=0}
|   |   |   |   |   |   |   |   |   Country = South Korea: true {false=1, true=1}
|   |   |   |   |   |   |   StudyLevel = Undergraduate: false {false=2, true=2}
|   |   |   |   semester = 2
|   |   |   |   |   StudyLevel = Postgraduate
|   |   |   |   |   |   CPS = false
|   |   |   |   |   |   |   Gender = F: false {false=6, true=6}
|   |   |   |   |   |   |   Gender = M
|   |   |   |   |   |   |   |   ApplicationStatus = Q: false {false=6, true=0}
|   |   |   |   |   |   |   |   ApplicationStatus = QPAC: true {false=0, true=2}
|   |   |   |   |   StudyLevel = Undergraduate: true {false=0, true=6}
|   |   |   CountryIncomeGroup = Lower middle income
|   |   |   |   Gender = F
|   |   |   |   |   Channel = Agent
|   |   |   |   |   |   semester = 1: false {false=9, true=2}
|   |   |   |   |   |   semester = 2: true {false=3, true=3}
|   |   |   |   |   Channel = Direct: false {false=12, true=1}
|   |   |   |   Gender = M
|   |   |   |   |   Channel = Agent
|   |   |   |   |   |   ApplicationStatus = Q
|   |   |   |   |   |   |   StudyLevel = Postgraduate: false {false=3, true=0}
|   |   |   |   |   |   |   StudyLevel = Undergraduate
|   |   |   |   |   |   |   |   semester = 1: true {false=1, true=2}
|   |   |   |   |   |   |   |   semester = 2: false {false=2, true=1}
|   |   |   |   |   Channel = Direct
|   |   |   |   |   |   CPS = false: false {false=7, true=6}
|   |   |   |   |   |   CPS = true: true {false=0, true=3}
```

Fig B. 2 Decision Tree

3. Lift Charts

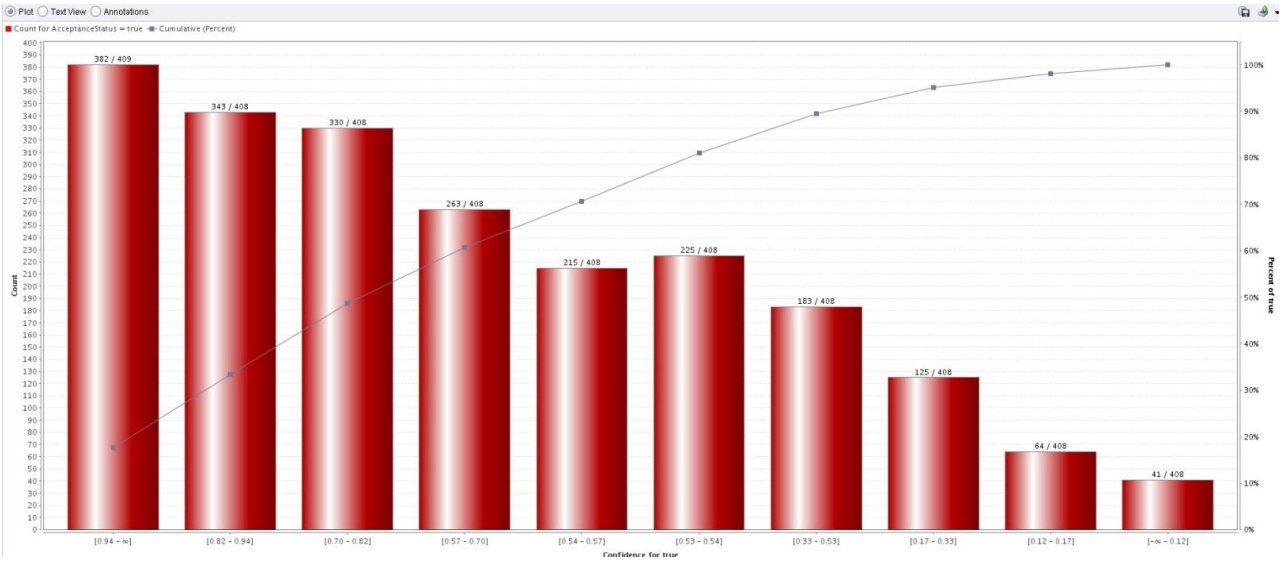


Fig B.3 Lift Chart for Neural Network

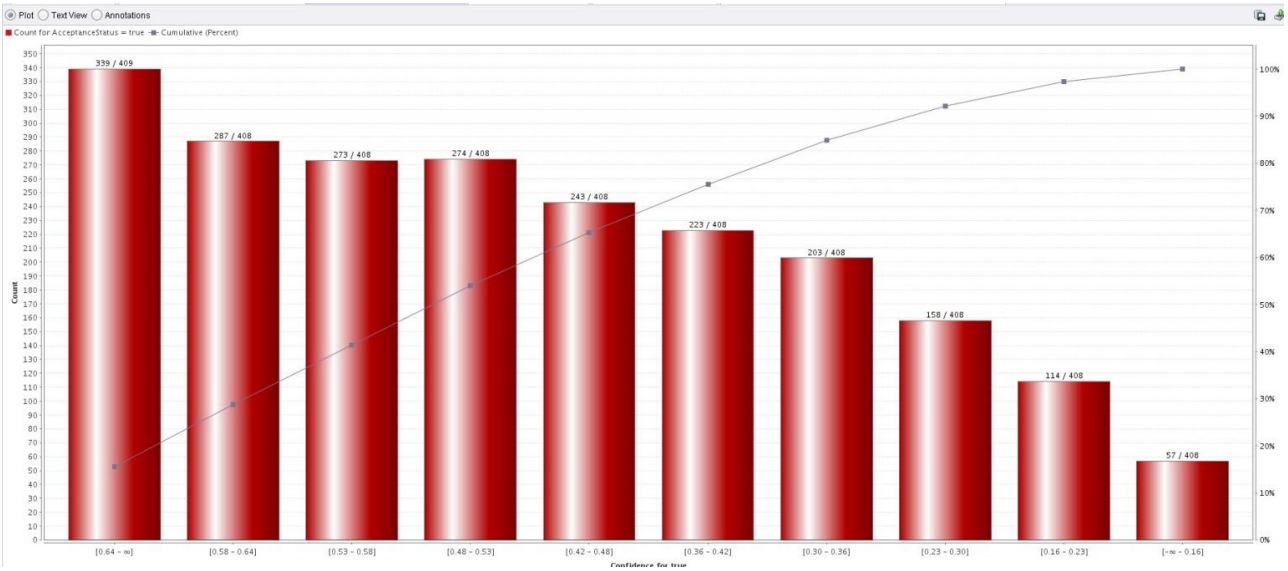


Fig B.4 Lift Chart for Logistic Regression

Appendix C

1. RapidMiner Terminologies

Table C.1 – RapidMiner Key Terms

Term	Explanation
Example	An example is a single row of data.
Example set	An example set is a set of one or more examples.
Attribute	An attribute is a column of data.
Type	This is the type of an attribute. It can be real, integer, date_time, nominal(both polynominal and binominal), or text.
Repository	A repository is a location where processes, data, models,and files can be stored and read either from the RapidMiner GUI or from a process.
Operator	An operator is a single block of functionality available from the RapidMiner Studio GUI that can be arranged in a process and connected to other processes. Each operator has parameters that can be configured as per the specific requirements of the process
Process	A process is an executable unit containing the functionality to be executed. The user creates the process using operators and joins them together in whatever way is required.
Role	An attribute's role dictates how operators will use the attribute. The default role is regular.
ID	This is a special role that indicates an identifier for an example. Some operators use the ID as part of their operation.
Label	A label is the target attribute to be predicted in a data mining classification context. This is one of the special role types for an attribute.

2. RapidMiner Screenshots

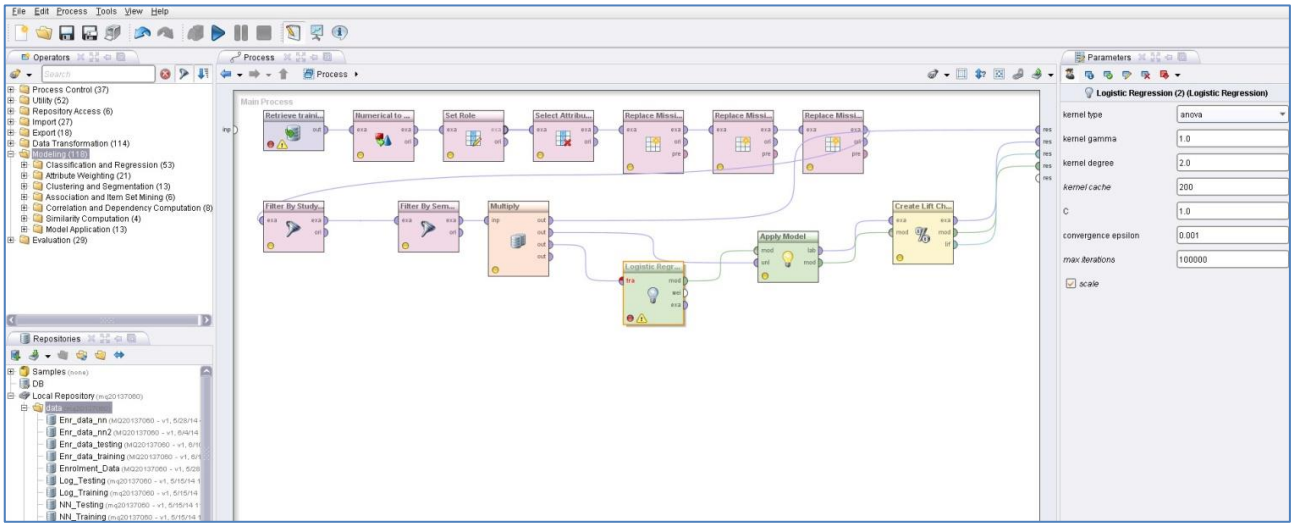


Fig C.1 RapidMiner Model Development

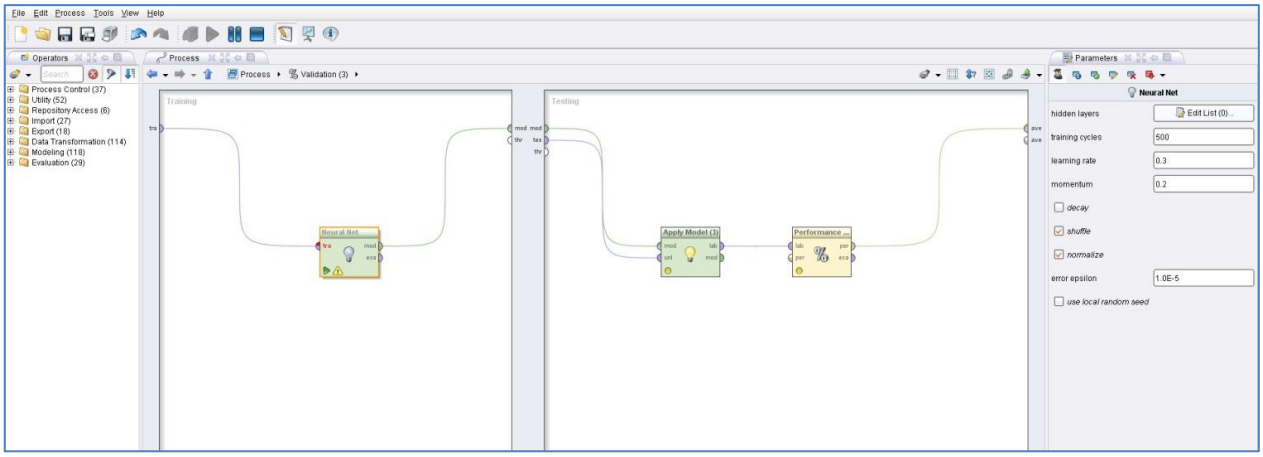


Fig C.2 RapidMiner Cross Validation

References:

Bogard, M., (2013). A Data Driven Analytic Strategy for Increasing Yield and Retention at Western Kentucky University Using SAS Enterprise BI and SAS Enterprise Miner. SAS Global Forum 2013 Proceedings.

Chang, L., (2006). Applying Data Mining to Predict College Admissions Yield: A Case Study. *New Directions for Institutional Research* 131:53-68.

DesJardins, S., (2002). An analytic strategy to assist institutional recruitment and marketing efforts. *Research in Higher Education* 43(5):531-553.

Goenner, C., & Kenton P., (2006). A predictive model of inquiry to enrollment. *Research in Higher Education* 47(8) :935-956.

Han, J., Micheline K. & Jian, P., (2006). Data mining: concepts and techniques. Morgan Kaufmann

Johnson, I., (2008). Enrollment, persistence and graduation of in-state students at a public research university: does high school matter?. *Research in Higher Education* 49(8) :776-793.

Sharda, R., Aronson, J. & King, D., (2008). Business intelligence: A Managerial Approach. Pearson Prentice Hall Upper Saddle River U.S.A.

Thomas, E., Dawes, W., & Reznik, G., (2001). Using predictive modeling to target student recruitment: Theory and practice. *AIR Professional* File 78.11.

Walczak, S., & Sincich, T., (1999). A comparative analysis of regression and neural networks for university admissions. *Information Sciences* 119(1) :1-20.

Walczak, S., (1998). Neural network models for a resource allocation problem. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 28(2) :276-284.

Romero, C., &Sebastian, V., (2013). Data mining in education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 3. 12-27.

Luan, J., (2002). Data mining and its applications in higher education. New directions for institutional research 2002.113 : 17-36.

Romero, C., & Sebastián, V., (2010). Educational data mining: a review of the state of the art. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 40.6: 601-618.

Baker, R., & Kalina,Y. (2009). The state of educational data mining in 2009: A review and future visions. JEDM-Journal of Educational Data Mining 1.1: 3-17.

Crows, T. (1999). Introduction to data mining and knowledge discovery. Two Crows Corporation 36

Amburgey, W. O., & Yi, J. (2011). Using Business Intelligence in College Admissions: A Strategic Approach. International Journal of Business Intelligence Research (IJBIR) 2.1: 1-15.

North, M., (2012). Data Mining for the Masses. A Global Text Project Book

Akthar, F., & Hahne, C. (2012). RapidMiner 5: Operator Reference. Rapid-I GmbH.

Chisholm A., (2013). Exploring Data with RapidMiner. Packt Publishing