Student Retention Prediction in Higher Learning Institutions: The Machakos University College Case

Joseph Mutuku Ngemu
School of Computing and Informatics,
University of Nairobi
P.O. Box 30197 – 00100, GPO Nairobi-Kenya
Email: josephngemu [AT] gmail.com

Elisha Opiyo Omulo
School of Computing and Informatics,
University of Nairobi
P.O. Box 30197 – 00100, GPO Nairobi-Kenya

William Okelo Odongo
School of Computing and Informatics,
University of Nairobi

Bernard Manderick
Artificial Intelligence Lab,
Free University of Brussels

Abstract: **Student retention has become one of the most important priorities for decision makers in higher learning institutions (HLI). Improving student retention starts with a thorough understanding of the reasons behind the attrition. In this study, using student demographic and institutional data along with several business intelligence (BI) techniques, we developed prototype to predict likelihood of student persistence or dropout. This study used classification models generated using Waikato Environment for Knowledge Analysis (WEKA). The model was built using the 10-fold cross validation, and holdout method (60% of the data was used as training and the remaining as test and validation). Random sampling techniques were used in selecting the datasets. The attribute selection analysis of the models revealed that the student age on entry, parent occupation, health of student and financial variables are among the most important predictors of the phenomenon. Results of the classifiers were compared using accuracy level, confusion matrices and speed of model building benchmarks. The study shows that identifying the relevant student background factors can be incorporated to design a prototype that can serve as valuable tool in predicting student withdrawal or persistence as well as recommend the necessary intervention strategies to adopt, leading to better education efficiency.**

Keywords-Business Intelligence, Retention, attrition, WEKA, classifiers.

I. INTRODUCTION

Business Intelligence or BI is a broad category of applications and technologies for gathering, storing, analyzing and providing access to data to help enterprise users make better business decisions. BI improves decisions by supplying timely, accurate, valuable, and actionable insights. With the rapid advancement and development of Information and Communication Technologies (ICT), organizations are now able to generate, collect and distribute huge amount data from internal and external sources, this is also happening in higher learning institutions.

As the concept of Business intelligence (BI) is steadily rising up the priority list within many institutions, it is necessary to explore the potential of BI in making better use of student data in support of student management and decision making. It is hoped that the application of BI systems will help managers and academic staff take a more proactive approach in student management and strategic planning through well informed and evidence-based decisions.

Theory and practice from many studies show that organizations' requirements to improve quality of decision-making and quality of service are largerly met by BI systems [1]. Therefore, BI has become a strategic initiative, and many business leaders now regard BI as instrumental in driving business effectiveness and innovation [2]. Moreover, BI has been used in many other sectors, for instance, in manufacturing companies, in retailing sector for user profiling, in financial services for claims analysis and fraud detection, in transportation for fleet management, in telecommunications for identifying reasons for churn and health care for outcomes analysis. However, Bi technologies have not been widely used in higher learning institution; despite that BI can also play an important role in student data analysis for decision making and strategic planning. Most of the current student information systems in higher learning institution are just a collection of student data.

The key benefit of applying business intelligent to this problem of student withdrawal is that there are multiple complex factors which influence a student's likelihood to withdraw. Business intelligent tools enables us to analyze historical data sets at an institution, identify the combination of factors which are most closely correlated with student withdrawal and build a model which allows us to predict the likelihood of individual student withdrawal in the future. This gives us a really powerful way to understand retention and a proactive way to manage retention issues.

II. PROBLEM STATEMENT

An issue of concern in higher learning institutions across the world is the retention and success of students in their studies. This is a particularly pressing issue in the context of widening participation for under-represented student groups, easing student diversity and educational quality assurance and accountability processes. As well as the personal impact and loss of life chances for students, non-completion has financial implications for students in developing countries (and their families), and for society and the economy through the loss of potential skills and knowledge.

Unfortunately, most institutions have not yet been able to translate what we know about student retention into forms of action that have led to substantial gains in student persistence and graduation. Though some have, many have not [3].

Lack of efficient educational system, lack of systems for predicting the likelihood of individual student withdrawal in the future and lack of information about the potential factors that may influence student withdrawal has been a challenge to many higher learning institutions when it comes to management of student retention issues.

Information is the new key enterprise asset as organizations across the globe not only leverage, but compete on information. But the pragmatic truth is that, while BI technologies continue to grow and mature, the promise of an efficient and effective BI environment that fits the real needs faced by higher learning institution users and decision makers day by day remains a challenge.

III. OBJECTIVES OF THE STUDY

The goal of this research is to find ways of improving the efficiency of higher learning institution systems by applying business intelligent techniques on educational databases. This can potentially reduce the incidents of student withdrawal.

Specific research objectives are: Identification of different factors which affects a student's retention rate and design a BI predictive model for higher learning institutions; Apply business intelligence concepts in the modeling process for the prediction of likelihood of dropping out or persisting; Construct a BI prototype for predicting likelihood of student withdrawal and validation of the developed model for students studying in Higher Learning Institutions.

IV. LITERATURE SURVEY

A. *Higher Education Management and Efficiency*

The efficiency of the study process can be measured by the student graduation rate, which is an important criterion in several national models of financing higher education institutions. This aspect of the efficiency of the study process ignores that the graduation rates are under the influence of external factors which are beyond the control of decision-makers at higher education institutions, which is taken into account in the study of [14].

B. *Student retention*

Student retention is one of the most important issues facing higher education today. At its core, the retention of college students is a complex issue, representing an inter play of personal, institutional, and societal factors, with likely associated detrimental costs and implications to all three audiences [5].

The most commonly referred to model in the student retention/dropout literature is Tinto's. It was first offered in a literature review [6].

[7] and other researchers [8] discuss the importance of matching students' goals and expectations to a college's mission.

Student financial issues have frequently been identified as a barrier to completion, especially by students from lower socio-economic groups [9]. They concluded that personal, emotional, and family problems, in addition to feelings of isolation and adjustment to college life, are strong barriers to retention for African American students.

C. *Student Retention Models*

Of greatest note are Tinto's Student Retention Model [6], Astin's Theory of Involvement [8], Bean's Student Attrition Model [10], and Chickering's Student Development Theory [11].

D. *Business Intelligence*

Business Intelligence (BI) systems provide a proposal that faces needs of contemporary organisations. Main tasks that are to be faced by the BI systems include intelligent exploration, integration, aggregation and a multidimensional analysis of data originating from various information resources [12].

E. *Business Intelligence in student retention.*

Higher learning institution data is massive therefore there is great need to use business intelligence to address several important and critical issues related to student retention. The patterns or trends that are discovered guide decision making such as forecasting retention and anticipating student's future fate. Business intelligence is an essential step in the process of knowledge discovery in database in which intelligent techniques are applied in order to extract patterns [2]

F. *Prediction algorithms*

1) *Decision Tree classifier (DT)*

DT is a powerful and popular classification and prediction technique [13]. [14] stress, that DT is the most common DM technique in the literature. There are several popular decision tree algorithms such as ID3, C4.5, and CART (classification and regression trees). DT is in the form of a tree structure, where each node is either a leaf node (indicating the value of the target class of examples) or a decision node (specifying a test to be carried out on a single attribute value, with one branch and sub-tree for each possible outcome of the test). DTs have many advantages such as very fast classification of unknown records, easy interpretation of small-sized trees, robust structure to the outliers' effects, and a clear indication of most important fields for prediction but DTs are very sensitive to over-fitting particularly in small data-sets [4].

2) *Multilayer perceptrons*

Perceptrons can only classify linearly separable sets of instances. If a straight line or plane can be drawn to seperate the input instances into their correct categories, input instances are linearly separable and the perceptron will find the solution. If the instances are not linearly separable learning will never reach a point where all instances are classified properly. Multilayered Perceptrons (Artificial Neural Networks) have been created to try to solve this problem [15]. [16] provided an overview of existing work in Artificial Neural Networks (ANNs).

3) Naive Bayes classifiers

Naive Bayesian networks (NB) are very simple Bayesian networks which are composed of directed acyclic graphs with only one parent (representing the unobserved node) and several children (corresponding to observed nodes) with a strong assumption of independence among child nodes in the context of their parent [17].Thus, the independence model (Naive Bayes) is based on estimating [18]:

$$R = \frac{P(i|X)}{P(j|X)} = \frac{P(i)P(X|i)}{P(j)P(X|j)} = \frac{P(i)\prod P(X,|i)}{P(j)\prod P(X,|j)}$$

Comparing these two probabilities, the larger probability indicates that the class label value that is more likely to be the actual label (if R>1: predict i else predict j). [19] first used the Naive Bayes in ML community. Since the Bayes classification algorithm uses a product operation to compute the probabilities P(X, i), it is especially prone to being unduly impacted by probabilities of 0. This can be avoided by using Laplace estimator or m-esimate, by adding one to all numerators and adding the number of added ones to the denominator [20]. The assumption of independence among child nodes is clearly almost always wrong and for this reason naïve Bayes classifiers are usually less accurate that other more sophisticated learning algorithms such as ANNs and SVMs.

4) Support Vector Machines

Support Vector Machines (SVMs) are the newest supervised machine learning technique [21]. An excellent survey of SVMs can be found in (Burges, 1998), and a more recent book is by [22]. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel tricks, implicitly mapping their inputs into high- dimensional feature spaces.

G. *Gaps to be filled*

Despite the fact that BI can play an important role in student data analysis for decision making and strategic planning and address the issues of retention, most of the current student information systems in higher learning institution are just a collection of student data. BI technologies have not been widely used in higher learning institution [2]. This study presents a BI project to generate predictive model for student retention management and construct a BI prototype for predicting the likelihood of student withdrawal .This will help decision makers to know what actions to be taken beforehand in case of drop-out issue.

H. *Conceptual Framework of the proposed system architecture*

The proposed BI Retention prediction System aims to address the challenges of student retention in higher learning institution. Thus, increasing retention has become a goal for many institutions, and a way of judging the quality of education. The proposed framework is presented in Fig.1 below.

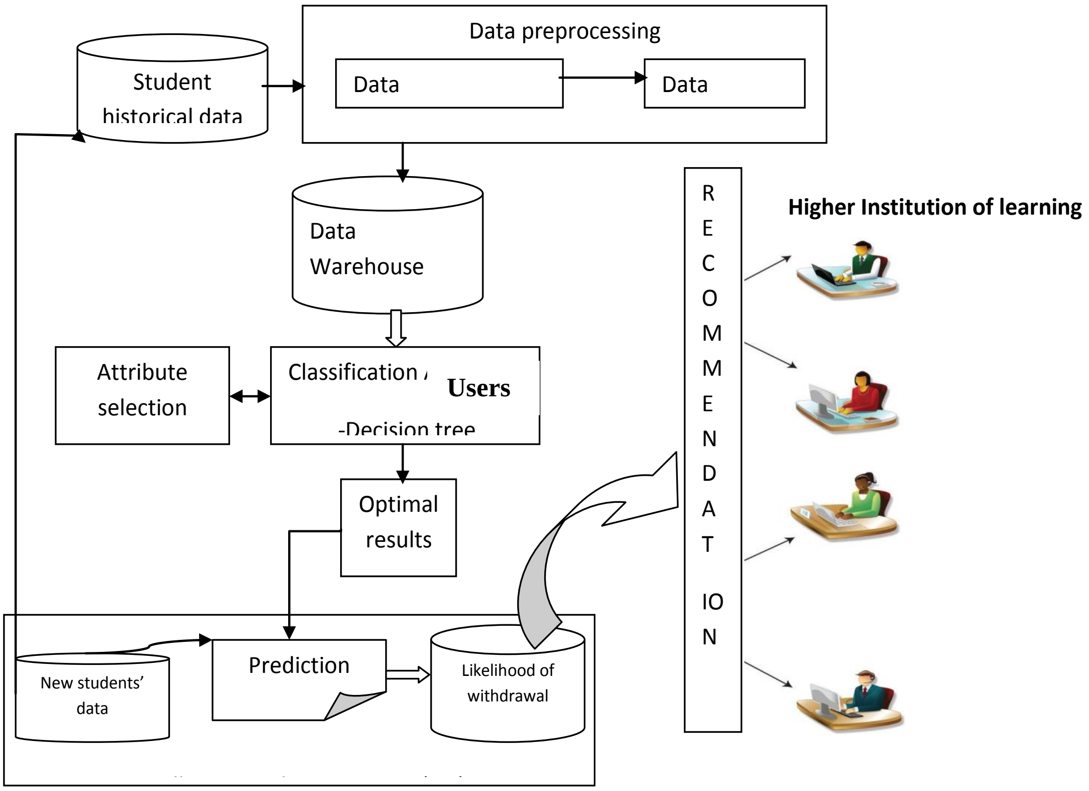


Figure 1: Block diagram of the proposed student retention predictive System

The proposed retention predictive system aggregates three components:

1) Data Acquisition and Storage component

The Data Acquisition and Storage component responsible for storing the students' data, gathered from different data sources in a data warehouse.

Data warehousing (DW) is playing a major role in the integration process in BI. [13] Suggest that data mining support BI including classification and prediction. The rapidly expanding volume of historical and real-time data contributes to the demand for and provision of data mining tools and it has become a critical role for advanced analytics in BI [23].

2) Model building component

The Model building component, responsible for obtaining knowledge about the students, through appropriates classification algorithms such as decision trees. Classification (also known as classification trees or decision trees) is a data mining algorithm that creates a step-by-step guide for how to determine the output of a new data instance. The tree it creates is exactly that: a tree whereby each node in the tree represents a spot where a decision must be made based on the input, and you move to the next node and the next until you reach a leaf that tells you the predicted output.

3) Intelligent Predictive System component

The Intelligent Predictive System component responsible for mapping the pattern in the rules generated with the new student data to predict likelihood of withdrawal or persistence.

V. METHODOLOGY

Spiral model methodology was used in the system specification, system design and implementation.

A. Overview of spiral model

The spiral model methodology is a systems development lifecycle model which combines the features of the Prototyping Model and the Waterfall Model and has detailed process for specifying, designing, and implementing prototypes [24]. The spiral model is favored for large, expensive and complicated projects.

B. Overview of WEKA

WEKA is an acronym for Waikato Environment for Knowledge Analysis, which is a free and open source software used to mine data. WEKA implements different algorithms which include Decision Trees, Artificial Neural Networks, and Logic Regression. WEKA allows the GUI user to select the four different ways to work with. These four ways include Explorer, Experimenter, Knowledge Flow or a simple CLI. WEKA only accepts data in ARFF (Attribute-Relation File Format) formats which is an ASCII text file that describes a list of instances sharing a set of attributes.

This data is processed by the different algorithms exhibited in the WEKA GUI chooser and from these different outcomes one is able to know which algorithm is best for the predictive model.

C. Methodological framework

The business intelligence model considered in our study was based on supervised learning (classification) techniques given that labeled training data was available. Classification is the process of finding a model that describes and distinguishes data, classes or concepts for the class of objects whose class labels is known. Our methodology consists of data collection, data-preprocessing, building classification model using training data and evaluation of the generated models using test data. Trained and tested model was then used to score incoming data. In this study we used student data from Machakos university college database having 270 attributes and 14 instances. It consist of attributes like DFP,AOE, PO, HTH etc, these attributes predict the likelihood of a student withdrawal. Also different classifiers were applied in the classification such as decision tree, naïve bayes, support vector machine and multilayer perceptrons. Below Fig.2 is a knowledge flow environment of the models design of the prototype.

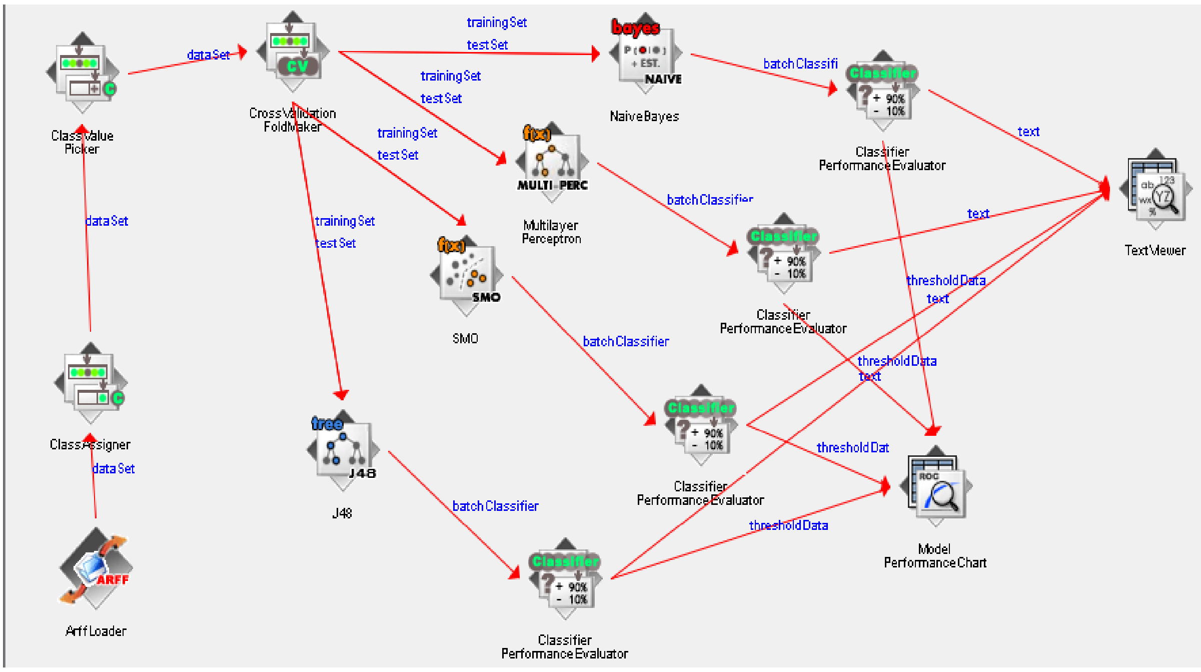


Figure 2: Models knowledge flow environment design of the prototype.

D. Sources of Data and Target Population

This study used survey-based secondary data provided by Machakos University College. Machakos University College is reliable source since is the public higher learning institution that keeps data about students so as to analyze and deduce information from the data which later enables them to make sound decisions. The outcome of the patterns are to help policy makers, educational administrators and the affected, to be able to make timely and rational decisions.

E. Sampling

Sampling was however applied at the data preprocessing stage and to reduce any biases in the sampling process, resample technique was used. The sampling methodology thus involved the generation of a random subsample of the dataset using sampling without replacement (to ensure an equal chance for every individual attribute to be selected).

F. Description of the Basic Dataset

The Basic Dataset in this study refers to the raw dataset that was sourced from the University database systems and files for this study as shown in Table 1.

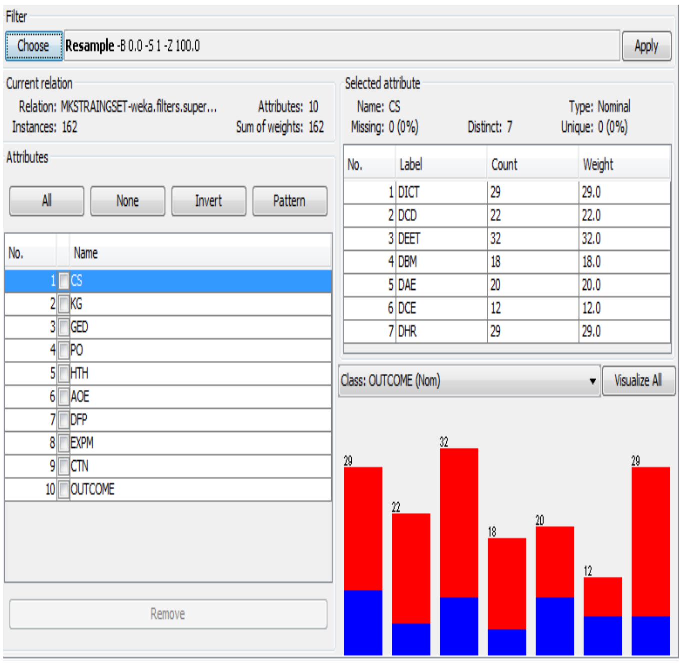
Table 1: Description of the variables of the dataset

VARIABLES	DESCRIPTION	POSSIBLE VALUES
CS	Course taken	DICT,DCE,DAE,DCD,DBM,DEET, DHR
KG	KCSE grade	C-,C+,B-,B+
GED	Gender	F,M
FEQ	Fathers Education qualification	DEG, SEC.CERT, DIP, PRI.CERT,NONE, MSC, DR
MEQ	Mothers Education qualification	DEG, SEC.CERT, DIP, PRI.CERT,NONE, MSC, DR
DFP	Difficulties in fees payment	NO, YES
PO	Parents' occupation	GOK,UNEMPL,SEMPLOY,NGO
MSP	Marital status of parents	MARRIED, SEPARETED,SINGLE
SP	Sponsor/guardian	PARENT,SELF,,SCHOLARSHIP,ORG
AOE	Age on Entry	BELOW 20, ABOVE 20
EXM	Whether course expectations are met	YES, NO
HTH	Health	GOOD, FAIR, POOR
CTN	Course match	APPROPRIATE, NOT-APPROPRIATE
OUTCOME	Actual outcome	PERSIST, DROPOUT

G. Data Preprocessing: Transformation and Selection of attributes

The attributes from the original dataset are not necessarily of the most analytical relevance in the indication and revealing of pattern. Transformations are attribute filters that are done to realize new attributes that could be of increased predictive power. Other filters implemented in this study was remove, a preprocessing technique that omits a range of attributes from the dataset one at time that have lower ranks to improve the accuracy of the classification algorithm.

The input data was randomly divided into three datasets: a training data set, test data set and validation set. The training data set was used to build the model. Model was then tested using test data to compute a realistic estimate of the performance of the model on unobserved data. We used a ratio of 60% of the data used for training, and 30% for testing, and 10% for validation following standard data mining practice as in Fig.3.



Filter

Resample -B 0.0 -S 1 -Z 100.0

Current relation

Relation: MKSTRAINGSET-weka.filters.super... Attributes: 10
Instances: 162 Sum of weights: 162

Selected attribute

Name: CS Type: Nominal
Missing: 0 (0%) Distinct: 7 Unique: 0 (0%)

No.	Label	Count	Weight
1	DICT	29	29.0
2	DCD	22	22.0
3	DEET	32	32.0
4	DBM	18	18.0
5	DAE	20	20.0
6	DCE	12	12.0
7	DHR	29	29.0

Attributes

No.		Name
1	□	CS
2	□	KG
3	□	GED
4	□	PO
5	□	HTH
6	□	AOE
7	□	DFP
8	□	EXPM
9	□	CTN
10	□	OUTCOME

Class: OUTCOME (Nom)

Figure3: Data preprocessing

H. Attribute selection

Attribute selection searches through all possible combinations of attributes in the data and finds which subset of attributes works best for prediction. The attributes relating to students' family background factors and previous academic achievement were considered. The attributes used in this study was ranked in order of importance using information gain and gain ratio measures. Information gain evaluates the worth of an attribute by measuring the information gain with respect to the class whereas gain ratio evaluates the worth of an attribute by measuring the gain ratio with respect to the class.

I. Model Building

The model building supported in this study is a classification in the search for the perfect model. The population for which a model is built is further divided into three sets: training, testing and validation. The ratio of the sample population is set at approximately 60%: 30%: 10% with the motivation to avoid occurrence of over-fitting and thus increase model accuracy and applicability in the performance dataset.

J. Model Validation

Model validation, in most cases relies on stakeholder and data based techniques. In this study, we investigate the usage and automation of the model validation process.

K. Modeling Techniques and Tools Used

The BI model considered in our study was based on supervised learning (classification) technique. The software tool used was WEKA an open-source and free software used for knowledge analysis and downloadable from the internet and used under the GNU license. WEKA implements different machine learning algorithms. The presentation of results and the development of the prototype were done using JAVA while the data will be stored in JavaDB.

VI. DATA ANALYSIS AND RESULTS

A. Predictive model/ Basic Classification Results using WEKA

In the classification we used J48, Naïve bayes, Multilayer perceptron and SVM. These classification algorithms were selected because they are considered as “white box” classification model, that is, they provide explanation for the classification and can be used directly for decision making. Each classifier belongs to a different family of classifiers implemented in WEKA. J48 relate to Decision trees, the multilayer perceptron belong to neural networks, Naïve bayes belongs to Bayesian network and SMO belong to support vector machine. Since they are from different classifiers family, they yielded different models that classify differently on some inputs. Attribute importance analysis was carried out to rank the attributes by significance using Information gain and gain ratio attribute evaluators. Ranker’s Search method was used to achieve this. The outcome is presented in Table3 and Figure13. The ranking of both attribute evaluators was done using ranker search method. Among the attributes used in this study, it was discovered that DFP, AOE, PO and HTH are the best four attributes. The outcome of both evaluators is similar as shown in Table 2.

Table 2: Attributes ranking using information gain and gain ratio

GAIN RATIO				INFORMATION GAIN			
s/n	Attribute	Value	Rank	s/n	Attribute	Value	Rank
7	DFP	0.42436	1	7	DFN	0.35036	1
6	AOE	0.15285	2	6	AOE	0.13401	2
4	PO	0.06074	3	4	PO	0.11784	3
5	HTH	0.03477	4	5	HTH	0.05483	4
9	CTN	0.02686	5	2	KG	0.04203	5
2	KG	0.01728	6	9	CTN	0.0232	6
8	EXPM	0.01301	7	8	EXPM	0.01122	7
3	GED	0.00399	8	1	CS	0.00792	8
1	CS	0.00293	9	3	GED	0.00394	9

Attribute ranking (with respect to the class attribute) according to information gain and gain ratio criteria show that DFP, AOE, PO and HTH are the best attributes. These attributes outperform other attributes in their contribution to the outcome of students’ withdrawal or persistence in HLI as shown in Fig.4 below.

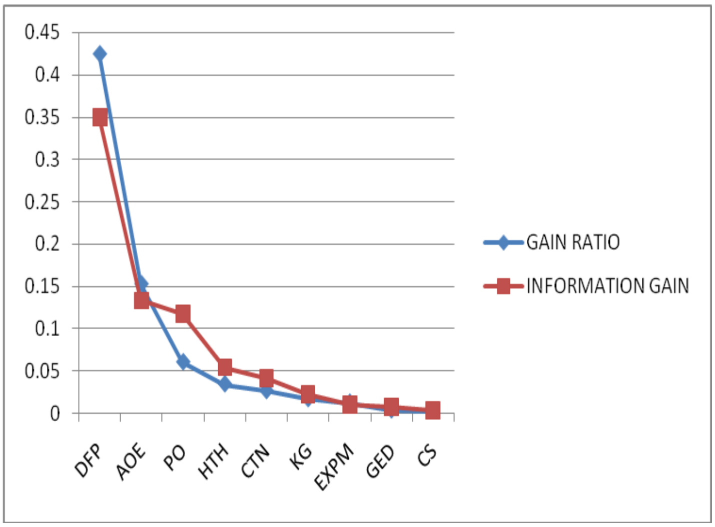


Figure 4: Information gain and gain ratio of the attributes for attribute selection

B. Comparison of learning algorithms

No single learning algorithm can uniformly outperform other algorithms over all datasets. Features of learning techniques are compared in Table 3 below from the models built.

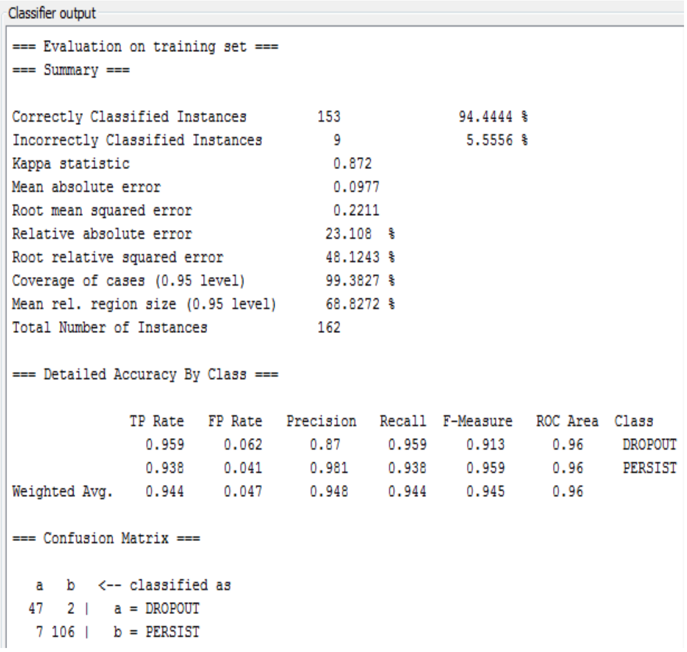
Table 3: Comparison of learning algorithms

S no.	Algorithm	Execution time on 10 –fold cross validation	Accuracy on 10 –fold cross validation	Recall On 10 –fold cross validation
1	J48	0.02sec.	94.8	94.4
2	Naïve Bayes	0.03sec	90.1	90.1
3	Multilayer perceptron	2.23sec	93.4	93.2
4	SVM	0.08 sec	90.3	90.1

Based on all the benchmarks used to measure the algorithms employed in this study, it is discovered that J48 performance is better than all other algorithms. We focus on designing our predictive system on the most suitable algorithm which is J48 in this domain.

C. Training data set

To produce the model a training data was used, we used a data set with known output values and use this data set to build our model as in Fig.5. Then, whenever we have a new data point, with an unknown output value, we put it through the model and produce our expected output. However, this type of model takes an entire training set and divide it into two parts, i.e about 60-70% of the data is taken and put into our training set, which we use to create the model; then the remaining data set is put into a test data set, which we use immediately after creating the model to test the accuracy of our model.



```
=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances         153               94.4444 %
Incorrectly Classified Instances         9                5.5556 %
Kappa statistic                          0.872
Mean absolute error                      0.0977
Root mean squared error                  0.2211
Relative absolute error                 23.108  %
Root relative squared error             48.1243 %
Coverage of cases (0.95 level)          99.3827 %
Mean rel. region size (0.95 level)      68.8272 %
Total Number of Instances              162

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                 0.959    0.062     0.87     0.959    0.913      0.96    DROPOUT
                 0.938    0.041     0.981    0.938    0.959      0.96    PERSIST
Weighted Avg.    0.944    0.047     0.948    0.944    0.945      0.96

=== Confusion Matrix ===

   a   b   <-- classified as
  47   2 |   a = DROPOUT
   7 106 |   b = PERSIST
```

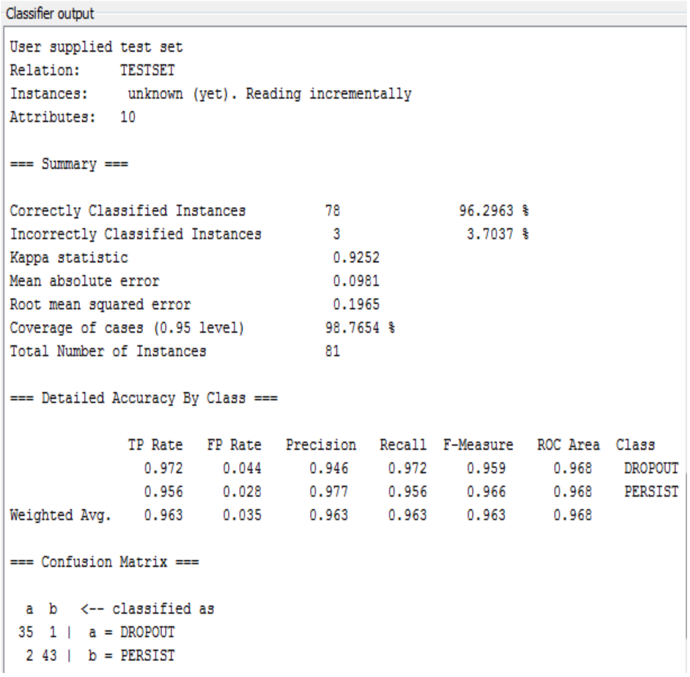
Figure 5: Evaluation on training set

1) *Interpretation of results of the training data set*

The model classifies 153 instances correctly with an accurate rate of 94.4%, this indicates that the results obtained from training data are optimistic and can be relied on for future or new predictions.

D. *Test data set*

The test data was created to control over fitting, after the model is created it is tested to ensure that the accuracy of the model built does not decrease with the test set as in Fig.6. This ensures that our model will accurately predict future unknown values.



```
User supplied test set
Relation:     TESTSET
Instances:     unknown (yet). Reading incrementally
Attributes:   10

=== Summary ===

Correctly Classified Instances          78               96.2963 %
Incorrectly Classified Instances         3                3.7037 %
Kappa statistic                          0.9252
Mean absolute error                      0.0981
Root mean squared error                  0.1965
Coverage of cases (0.95 level)          98.7654 %
Total Number of Instances               81

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                 0.972    0.044     0.946    0.972    0.959     0.968    DROPOUT
                 0.956    0.028     0.977    0.956    0.966     0.968    PERSIST
Weighted Avg.    0.963    0.035     0.963    0.963    0.963     0.968

=== Confusion Matrix ===

  a  b   <-- classified as
 35  1 |  a = DROPOUT
  2 43 |  b = PERSIST
```

Figure 6: Evaluation on test set

2) *Interpretation of results of the test data set*

The model classifies 78 instances correctly with an accurate rate of 96.3%, this indicates that our model will accurately predict future unknown values.

E. *Models perfomance.*

Fig.7 is a Receiver Operating Characteristic curve (or ROC curve.) It is a plot of the true positive rate against the false positive rate for the different possible cutpoints of a diagnostic test. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the model. Based on the threshold curves used to measure the algorithms employed in this study, it is discovered that J48 performance is better than all other algorithms.

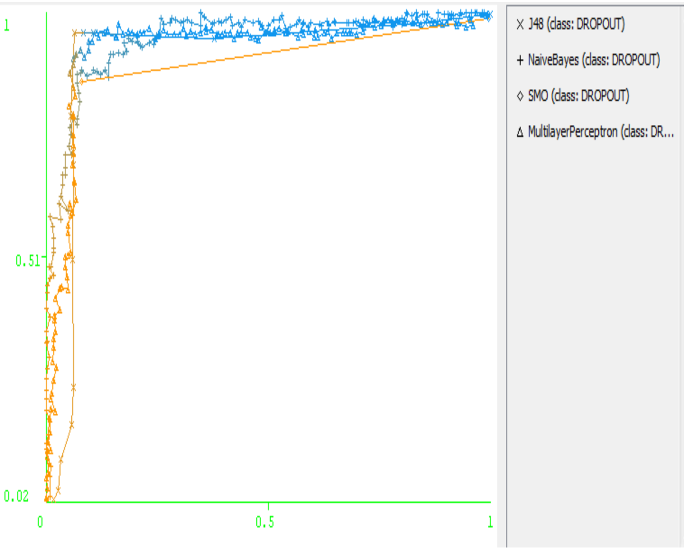


Figure7: ROC curve for Classifiers performance comparison

F. *Tree visualization*

Fig.8 is the graphical representation of the classification tree.

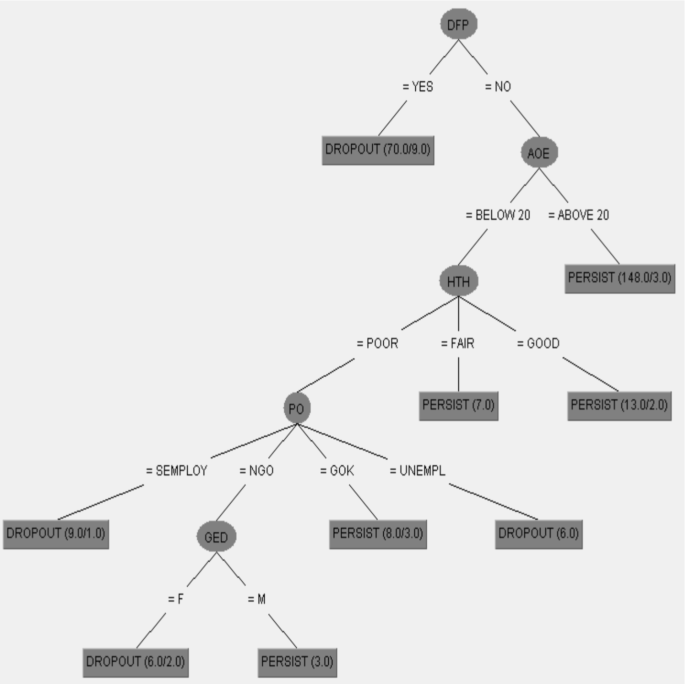
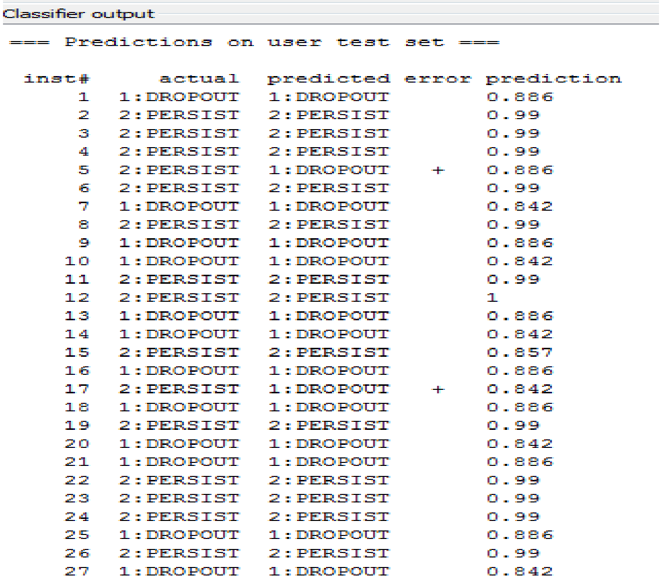


Figure 8: Model classification tree

G. *Using the classification Algorithm in our Dataset*

Classification is used to find a model that segregates data into predefined classes. Classification is based on the features present in the data. The result is a description of the present data and a better understanding of each class in the database. Thus classification provides a model for describing future data. Prediction helps users make a decision. Predictive modeling for knowledge discovery in databases predicts unknown or future values of some attributes of interest based on the values of other attributes in a database as in Fig.9.



```
=== Predictions on user test set ===

 inst#     actual   predicted error prediction
     1   1:DROPOUT  1:DROPOUT       0.886
     2   2:PERSIST  2:PERSIST       0.99
     3   2:PERSIST  2:PERSIST       0.99
     4   2:PERSIST  2:PERSIST       0.99
     5   2:PERSIST  1:DROPOUT   +   0.886
     6   2:PERSIST  2:PERSIST       0.99
     7   1:DROPOUT  1:DROPOUT       0.842
     8   2:PERSIST  2:PERSIST       0.99
     9   1:DROPOUT  1:DROPOUT       0.886
    10   1:DROPOUT  1:DROPOUT       0.842
    11   2:PERSIST  2:PERSIST       0.99
    12   2:PERSIST  2:PERSIST       1
    13   1:DROPOUT  1:DROPOUT       0.886
    14   1:DROPOUT  1:DROPOUT       0.842
    15   2:PERSIST  2:PERSIST       0.857
    16   1:DROPOUT  1:DROPOUT       0.886
    17   2:PERSIST  1:DROPOUT   +   0.842
    18   1:DROPOUT  1:DROPOUT       0.886
    19   2:PERSIST  2:PERSIST       0.99
    20   1:DROPOUT  1:DROPOUT       0.842
    21   1:DROPOUT  1:DROPOUT       0.886
    22   2:PERSIST  2:PERSIST       0.99
    23   2:PERSIST  2:PERSIST       0.99
    24   2:PERSIST  2:PERSIST       0.99
    25   1:DROPOUT  1:DROPOUT       0.886
    26   2:PERSIST  2:PERSIST       0.99
    27   1:DROPOUT  1:DROPOUT       0.842
```

Figure 9: predictions on user test set

VII. CONCLUSIONS AND RECOMMENDATIONS

A. *Conclusions*

In our research study we were able to build a prototype that has the ability to load a model and fetch data for prediction from the database of the Higher learning institution. A number of classification models were considered as specified in the literature review and compared in analysis stage out of which we chose to use the decision tree (J48) classifier model because of its performance in adapting it to the data collected. We developed a J48 classifier that integrates an information gain and gain ratio in Waikato Environment for Knowledge analysis (WEKA) Tool Kit and trained it on a preprocessed dataset from a HLI. The results obtained from experiments with the classifier (see Chapter 4 above) show that the classifier is capable of performing classification with an accuracy of 94.4% for dataset obtained from the HLI. Finally, we integrate the techniques and methods developed into a Java based application for use in predicting the likelihood of a student withdrawing in future.

Further this research has shown that it is possible to predict the dropout for different students. The study has revealed some advantages of J48 model over naïve bayes, multilayer perceptron and support vector machine over other models. One of the advantages realized is that J48 could predict with more accuracy on small volumes of data with noise.

ANN and DT are the methods widely adopted mostly due to their prevalence in the field of BI and proven ability to form models across wide range of application area. More so with advancement in BI the two have proven to be the most versatile and accurate. Also compared with other techniques, they are well established for adoption in performance prediction.

By gaining a deep understanding of student retention patterns and tendencies, we are enabled to predict which students are most likely to dropout, or those who are most likely to persist. By identifying these students and future prediction of their further outcome, the faculty and managerial decision maker can utilize necessary action and directly or indirectly intervene by providing extra academic counseling, and financial aid. Therefore the Higher Learning institution management system is enabled to improve their policy making, setting new strategies, and having more advanced decision making procedures. The final result of such model is improving the quality of higher educational system.

Few patterns which we came across during the course of the study are listed below:

- If difficult in fees payment = YES, then outcome = DROPOUT
- If difficult in fees payment = NO, student health = GOOD, then outcome = PERSIST
- If difficult in fees payment = YES, age of entry of student < 20 years, and parent occupation= self employed OR unemployed, then outcome, = DROPOUT
- If difficult in fees payment = NO, student health=poor, and parent occupation =GOK, then outcome= PERSIST
- If student health = poor, age of student< 20 years, parent occupation =NGO, and gender= female then outcome= DROPOUT

B. *Contributions*

This research contributes to the body of knowledge. Further a number of Business Intelligence models have been evaluated on their performance in retention prediction for HLI. The research has revealed that BI technologies can be used efficiently in HLI to enhance education efficiency. On the other hand, the researcher has proposed a system that can be adopted by HLI to perform student retention prediction for better education efficiency. The finding of this research have important implication for HLI specifically registrar. Any HLI that needs to establish its policy upon future dropout prediction may use this finding. Big volumes of past student data are available to many HLI. This data can be a rich source of knowledge, if only properly used. This can be very beneficial for the HLI using BI to extract knowledge and useful information from this available source of data. Thus, one of the managerial implication of this research is to inform managers about the advantages and importance of BI in their strategic planning

C. *Limitations faced*

We were not able to collect more information associated to the student social and cultural factors.

D. *Future work*

The future scope of the system may provide facilities of generation of more reports to evaluate the retention issue. It can be implemented on a wide basis for all the Higher Learning institutions in Kenya, by associating students' personal information with test score and social-cultural factors in determining retention. Functionalities for accommodating other classifiers other than the J48 classifier can be developed into the application. These classifiers include Naïve bayes, Support Vector Machines and Multilayer perceptron. Results from the various classifiers can be compared in a report interface for the best classification technique to be selected by the user.

ACKNOWLEDGMENT

The authors would like to take this opportunity to show our appreciation to all our colleagues for their support throughout the period of this research. Without such help, this work would not be completed or published in one of the international journals around the word.

REFERENCES

[1] Liautaud, B., & Hammond, M. (2002). E-business intelligence. Turning information into knowledge into profit. New York: McGraw-Hill.

[2] Hugh J. Watson, Barbara H. Wixom, "The Current State of Business Intelligence", Computer, vol.40, no. 9, pp. 96-99, September 2007, doi:10.1109/MC.2007.331

[3] Carey, J. C., Dimmit, C., Hatch, T., Lapan, R. T., & Whiston, S. C. (2008). Report of the National Panel for Evidenced-Based School Counseling: Outcome research coding protocol and evaluation of student success skills and second step. Professional School Counseling, 11(3), 197-206.

[4] Sav, G. T. (2012). Four-Stage DEA Efficiency Evaluations: Financial Reforms in Public University Funding. International Journal of Economics and Finance, 5(1). doi:10.5539/ijef.v5n1p24

[5] Brunsden, V., Davies, M. Shelvin, M. & Bracken, M. (2000). Why do HE Students Drop Out? A test of Tinto's Model. Journal of Further and Higher Education.

[6] Tinto,V. (1975) "Dropout from Higher Education: A Theoretical Synthesis of Recent Research" Review of Educational Research vol.45, pp.89-125.

[7] Tinto, V. (1986). Theories of Student Departure Revisited. In J.C. Smart (ed), Higher Education: Handbook of Theory and Research, Vol. 11, New York: Agathon Press: 359-384

[8] Astin, Alexander W., "Personal and Environmental Factors Associated with College Dropouts Among High Aptitude Students," Journal of Educational Psychology, VoL 56, 1964, pp. 219-227

[9] Ozga,J & Sukhnandan,L. (1998) "Undergraduate non-completion: Developing an explanatory model" Higher Education Quarterly vol.52 no.3 pp.316-333

[10] Bean, J. P. (1982). Student attrition, intentions, and confidence: Interactions effects in a path model. Research in Higher Education, 17,291–319

[11] Chickering, A.W., and William Hannah, "The Process of Withdrawal," Liberal Education, Vol. 66, 1969, pp. 661-668.

[12] Olszak, C. M., & Ziemba, E. (2006). Business intelligence systems in the holistic infrastructure development supporting decision-making in organizations. Interdisciplinary Journal of Information, Knowledge and Management, 1, 47-58. Retrieved December 1, 2006fromhttp://ijikm.org/Volume1/IJIKMv1p047-58Olszak19.pdf

[13] Chaudhuri, S. (1998). Data Mining and Database Systems: Where is the Intersection? InIEEE Bulletin of the Technical Committee on Data Engineering, 21(1), (pp. 4-8).

[14] Hämäläinen, W. and Vinni, M. (2010). Classifiers for educational technology. In C. Romero, S. Ventura, M. Pechenizkiy, R.S.J.d. Baker (eds.), Handbook of Educational Data Mining,(pp. 54-74). CRC Press.

[15] Rumelhart (1986) D.E. Rumelhart, G.E. Hinton, R.J. Williams Learning internal representations by error propagation Parallel distributed processing, MIT Press, Cambridge (1986), pp. 318–362

[16] Zhang, G. (2000), Neural networks for Mateo, CA: Morgan Kaufmann. classification: a survey. IEEE Transactions on [115] Wettschereck, D., Aha, D. W. & Mohri, T. Systems, Man, and Cybernetics, Part C 30(4): 451- (1997). A Review and Empirical Evaluation of 462. Feature Weighting Methods for a Class of Lazy

[17] Good, I.J., 1950. Probability and the Weighing of Evidence.Charles Griffin, London

[18] Nilsson, N.J. (1965). Learning machines. New York: McGraw-Hill.

[19] Cestnik, B., Kononenko, I., Bratko, I., (1987).Assistant 86: A knowledge elicitation tool for sophisticated users. In: Proceedings of the Second European Working Session on Learning. pp. 31-45.

[20] Cestnik, B. (1990), Estimating probabilities: A crucial task in machine learning. In Proceedings of the European Conference on Artificial Intelligence, pages 147-149.

[21] V Vapnik The Nature of Statistical Learning Theory Springer NY, 1995

[22] N.Cristianini, C Campbell and J ShaweTaylor Dynamically adapting kernels in support vector machines, Advances in Neural Information Processing Systems 11, ed. M. Kearns, S .A. Solla, and D. Cohn, MIT Press, p .204-210, 1999.

[23] Shim, J. P., Warkentin, M., Courtney, J. F., Power,D. J., Sharda, R., & Carlsson, C. (2002). Past,Present, and Future of Decision Support Technology.Decision Support Systems, 32(1), 111–126.doi:10.1016/S0167-9236(01)00139-7

[24] B. Boehm, "Spiral Development: Experience, Principles and Refinements,"Proc. Software Engineering Institute Spiral Development Workshop, p.49, 2000.