Journal of the Nigerian Association of Mathematical Physics
Volume34,(*March, 2016*), pp 249 – 262

Predicting Academic Performance of Students in Higher Institutions with k-NN Classifier

[1]O.M Omisore and [2]N.A Azeez

[1]Centre for Information Technology and Systems,University of Lagos,Akoka, Yaba, Lagos, Nigeria.
[2] Department of Computer Science, University of Lagos, Akoka, Yaba, Lagos, Nigeria.

Abstract

Educational Information is the pre-processed data of profile and characters in any institution of learning. The volume of such data depends on the level of learning in the institution. As a result, in several institutions of higher learning, the massive information hosted bulges out of their database and thereby making it very difficult to establish presence of common consistent interesting patterns needed for decision making. In this paper, k-Nearest Neighborhood (k-NN) classifier is adopted for predicting academic performance of students in higher institution. Case study of returning students in department of computer science, University of Lagos, Nigeria is observed. Experimental result shows the classifier can predict performance of students who can be distinctive, hapless or intermediate in their studies.

Keywords: Education, Data Mining, Data Classification, Predictive Model, Nearest Neighborhood, Performance Prediction
Categories and Subject Descriptors
H.1.2 [**User/Machine Systems**]: Human information processing;H.2.8 [**Database Applications]**: Data miningI.5.2 [**Design Methodology**]: Design and EvaluationI.6.1 [**Simulation Theory]:** Model classification
General Terms: Experimentation, Human Factors, and Performance.

1.0 Introduction

Educational information is the pre-processed data about current or prospective students of academic institutes at any level. This information has shown to exhibit hidden knowledge that can be used to predict performance of student in different educational level. Prediction of students' success is crucial for higher institutions because an objective of teaching is an ability to develop and produce self sufficient students who can improve to meet up with academic expectations. As observed in [1], one of the several issues faced in academic institution today is predicting the success of students in an institution. This educational information can be if great importance to admission system of any institution and for placement of faculty programmes. For instance, if stakeholders in admission system of an academic institution are aware of students that are likely to perform better if enrolled in a particular programme or students that will need assistance in order to complete their programme, then end products of the institution are likely to be excellent in their studies and as well, they will be of great influence for the community at large.

Research efforts of pervious times have proved data mining as veritable technique that can be applied to study available data in educational field to bring out hidden knowledge [2]. Such knowledge can strengthen actors of academic institutes to identify weak students and help them to improve their performance in their course assessments [3]. One way to address these challenges is data analysis, discovery and presentation, a viable tool termed data mining. Knowledge discovery enables organizations to use their current reporting capabilities to uncover and understand hidden patterns in vast databases. These patterns are then built into data mining models and to predict individual behavior with good accuracy [4].

Data mining is a computational process used to extract implicit and interesting samples, trends and information from data in order to predict unknown formal outcomes as useful knowledge. Data mining techniques are used to operate on large amount of data to discover hidden patterns and relationships helpful in decision making. These techniques can be of great assistant for primary stakeholders of any academic institution, as both students and teachers needs to understand, track, and improve teaching and learning performance in the environment. The technique helps stakeholder in academia to discover and master new relationships and dependencies of features in educational phenomenon [5].

Corresponding author: O.M Omisore, E-mail:ootsorewilly@gmail.com, Tel.: +234703-196-7847& 7066838551(NAA)

The rate of poor students in tertiary institutions of developing countries had being on an increase as greater number of candidates enroll into the institutions. Techniques to assess performance have generally relied on subjective evaluations which are, often, carried out by same person responsible for teaching the students. These subjective gateways had in no way assisted institutions of higher learning because it does not offer a way to filter inferior students in the institutions; thereby students produce woeful results [6]. If such institutions know in advance the weak students that is, those that are likely to fail, they will take the necessary actions to determine students that will be of best performance before placing them on admission or faculty programmes. Also preventive measures like increasing study hours per week can be adopted to improve the weak students.

The academic performance of students in educational fields depends on diverse physiological and psychological factors such that the success or failure of a student can be pre-determined. The ability to predict success can, hence, be deemed as a very important method of improving end products in educational environments. In [7], learning behavior of students were analyzed to predict their final results, and to improve students who were at risk of failing their final exams.

Different mining techniques can be applied in education systems like clustering, classification, outlier detection, association rule mining and sequential pattern mining [8]. With these techniques, large amount of records and information can be traversed to discover interestingness that will assist in making optimal decision [9]. Some tasks of knowledge discovery are to find important pattern, regularity, and relationship or association, between attributes of educational dependences. Many information-theoretic measures have been proposed and applied to quantify the importance of attributes and relationships that features amongst them. Beyond these, other frequently used techniques include Genetic Optimization, Market Basket Analysis, Neural and Kohonen Networks, Link Analysis, Time Sequence, and Text Mining algorithms [10]. Genetic optimization and neural networks are artificial intelligence techniques that provide human supports in varying fields. Artificial Intelligence is a coding method applied to machines to enhance decision making from complex imprecise information that requires human experts. In the last two decades, researches in this field have improved performance of many service systems [11]. This paper presents a k-Nearest Neighborhood (k-NN) model for predicting academic performance of students in institution of higher learning.

The remaining part of this paper is organized with literature review presented in Section 2. A k-NN classification model designed to relate educational attributes for predicting students' performance is described in Section 3 while the report of experimental study and evaluation carried observed to measure system's performance is presented in Section 4. Finally, conclusion and area of future research are highlighted in Section 6.

2.0 Literature Review

Data mining is a discovery and predictive technique employed to determine the essential quality of relationships that exits between different attributes of an entity when there are incomplete a priori expectations as to the nature of other entities. In this section, an overview of mining techniques and their applications in performance prediction are presented. This section reviews related works that are pertinent and prominent to the subject domain.

2.1 Overview of Data Mining System

Data mining system employs several tasks and techniques toward extracting useful knowledge from large amount of data. Muqasqas et. al, 2014 emphasized taxonomy of such systems on mining techniques [12] while Han et al [34] identified type of data to be mined, kind of knowledge discovered, and type of application to be adapted as major criteria for selecting a mining technique. In essence, these criteria dictate the type of technique that could be utilized.

Classification is a mining technique that focuses information extraction as a way to describing important features of a class. Applications of this mining technique have been utilized in fraud detection, target marketing, and medical diagnosis [13]. Data Classification can be viewed as a two-step process which consists of learning phase and actual classification. In the learning (training) phase, a general model is constructed to walk through to study patterns found available in a sample dataset to understand the relationships that exists amongst the data. Information gained from inter-class and intra-class relationships can then be used in the second phase to label and classify entities in test data. Accuracy of classification models can be evaluated on a certain test data as the amount of instances that are correctly classified by the model.

The use of data mining in educational field is being adopted to enhance humans' understanding of learning process. This has afforded human race to focus on identifying, extracting and evaluating variables related to the learning process of students [14]. Although data mining has been found critical for business intelligence and web search engines and its use in higher education is still relatively new [15]. For instance, data mining techniques can be applied on student demographics combined with other information to find some correlations capable of predicting the students' performance in their academic engagements. Also, Kovačić in 2010 applied data classification to explore the socio-demographic variables of students from an open polytechnic in New Zealand. The objective is to explore factors that impact study outcome in the Polytechnic[16]. Some factors identified as been important for students' success were used to develop typical profiles of likely successful and unsuccessful students. This can enable colleges and institutions of higher learning to improve their understanding about student capabilities.

2.2 Common Techniques for Data Mining

Naïve Bayes, Decision trees, Genetic algorithms, and Neural Networks are few of several techniques that have been proposed and used for mining data in recent times. Naïve Bayes, popularly known as Bayesian Network and decision trees are common classifiers applied in many works. For instance, Edin & Mirza in 2012applied decision trees modeled from Naïve Bayes on preoperative assessment data to predict students' performance in a summercourse[17]. Genetic algorithms have been applied to predict their students' final grade [18]. However, neural network deems less suitable for data mining purpose due to lack of comprehensibility because such models are built on black-box mechanisms. Genetic Algorithm and NEURAL network have been hybridized in [19], and neural evolutionary programming is proposed in [20]. Review of these operational methods is given herein.

2.2.1 Naïve Bayes

Naïve Bayes is a data mining classification technique that applies Bayes' theorem with strong independence assumptions between the features using simple probability. This classifier is scalable in that it requires a number of parameters linear with number of attributes in problem domain and maximum likelihood training only takes a linear time because evaluation involves a closed-form expression rather than by expensive iterative approximation as used by many other classifiers [21].

To classify an item x_i as an instance of X using Naïve Bayes, the classification model computes a posterior likelihood of x_i as:

$$P(C_i|X) = P(C_i)\prod_{i=1}^{n} P(x_i|C_i) \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots \ldots (1)$$

Where C_i is a classifier label and $P(C_i|X)$ is the priori probability that an instance of X is tagged (labeled) in C_i.

Naïve Bayes is widely used for learning data relationships unlike other complex classification techniques [22]. Despite its simplicity, computational efficiency and performance for real-life problems, results from decision tree are not that accurate making its potential for probabilistic modeling unexploited [23, 24].

2.2.2 Decision Trees

Decision tree is a support tool with tree-like model used in taking decisions and showing relative consequences, chance of event outcomes, resource costs, and utility. This approach, commonly used in operation research helps to identify a strategy most likely to reach a goal. In a flowchart-like tree structure, the internal nodes denoted by rectangles are used to test values of expressions on given attributes while leaf nodes denoted by ovals depict the class label an entity can be categorized.

Decision trees can adopt classification or regression tree analyses. In the classification tree analysis, predicted outcome is a class to which data belongs while in regression analysis outcome can be a real number. Most decision tree algorithms such as ID3, C4.5, and CART adopt top-bottom, and divide-conquer procedures to learn from training dataset [25]. These algorithms starts by assuming an empty tree then split on next best attribute, and hold on to recursion on each leaf. Choosing the best attribute upon which splitting is done is a great challenge. This could be addressed by quantifying the predictive feature of an attribute since outcome at current nodes depends on this. In CART, variance reduction is often employed in cases with continuous target variable [26].

Other recent algorithms include Chi-squared Automatic Interaction Detector which performs multi-level splits when computing classification trees [27], multivariate adaptive regression splines extends decision trees to handle numerical data better. The regression technique is a non-parametric extension of linear models that automatically models interactions between variables. In general, there might be multiple independent variables but their relationship to dependent variable(s) is unclear by manual analysis and neither is it visible in plot. Hence, regression techniques are used to discover non-linear relationships in multiple variables.

Decision trees represent rules in format that can be easily read, understood and interpreted by users. Over time, decision trees had been built with conditional inference. Conditional inferential is a statistical technique that takes non-parametric tests as splitting criteria for multiple testing to achieve unbiased prediction without pruning nor over-fitting. ID3 and C4.5 were older techniques used to generate decision trees from a dataset. C4.5 has realized a number of improvements over its successor ID3 [28]. An example is its ability to handle both discrete and continuous attributes.

2.2.3 Neural Network

Neural Network is an artificial intelligent technique that hosts a group of interconnected artificial neurons to mimic the properties of biological neurons in human. The technique follows analog and parallel computing system made up of simple processing elements that communicate through a rich set of interconnections with varying contributory weights [11]. Over a period of time, neural network has been used to solve pattern recognition problems such as text, image and medical bioinformatics. In Kim et. al, 2005, a predictive model based on neural network was applied to estimate the compressive strength of concrete mixture from constituent's proportion [30]and the prediction accuracy was improved with an iteration method in [31].

Since neural network is an adaptive artificial intelligent technique with self-tunning used that adjusts its structure during learning process, its usage in students' academic performance prediction can be described as being probabilistic. Probabilistic

neural network, implemented based on Parzen non-parametric probability density function estimation and Bayes classification, use feed-forward network to effectively solve prediction problems. The approach was found to build its architecture and train the network in fewer seconds, thence provides probabilistic viewpoints from all input with deterministic classification results. The Probabilistic network usually adopts Radial Basis Function to map any input pattern for classification.

Neural networks have shown success in areas of performance prediction, estimation and approximation, and pattern recognition problems [32]. Probabilistic network model was proven to be more time efficient when compared to conventional back-propagation based models. The network type is recognized as a way to solving real-time classification problems as observed in [31]. However, it works better if a desired output is expressed as one of several pre-defined classes [33]. Finally, NN and some statistical methods, like decision trees built on conditional inference, are deemed less suitable for data mining purposes [8].

2.2.4 k-Nearest Neighborhood Classification

k-Nearest Neighborhood (k-NN) classification is a method adoptable for classifying entities based on closest training examples in a feature space. k-NN is a lazy learning classifiersthat adopts instance-base learning hence having prediction done in two stages. Firstly, it undergoes minimal operations of analyzing the attribute values of individual instances in training dataset.

Loads of works come in the second stage where classification and prediction processes are carried out. Despite greater expenses incurred at the later stage, k-NN classifier is a simple learning algorithm. Objects are classified by a majority vote of neighbors before been assigned a class label of most common labels in its k-nearest neighbors [34]. k is a positive integer, typically small with best choice value depending on dataset used in the study. However, effects of noise on the classification are reduced if a big value of k is chosen, and the boundaries between classes are less distinct. The accuracy of the k-NN algorithm can be severely degraded by the presence of noisy or irrelevant features, or if the feature scales are not consistent with their importance.

K-NN algorithm has been adopted by statisticians as a machine learning approach for some decades now [35]. This algorithm has capability to classify new instance of entities with minimal training sets because it only does more work during classification and thereby achieves prediction with good likelihood and relatively inexpensive computational resources. Also, its error rate is minimal compared to the Naïve Bayes and decision trees [36]. A set of continuous dependent variable and categorical predictors are use in [32] to perform a three-way comparison of prediction accuracy. Non-linear regression, neural network and CART models were taken in the study. Results show that neural models CART produced better prediction accuracy than non-linear regression model due to the categorical predictor variables.

Several studies have discussed the use of data mining techniques for mining educational data and those studies present different attributes for measuring performance. However, evaluating metrics used and showing how such metrics dictates student's performances is lacking. K-NN is a CART algorithm for prediction models but its usage has is uncommon when compared to other methods.Ribeiro, 2013juxtaposed the adoption of ten common data mining techniques presented in Figure 1 and reported decision trees and NN as most used ones [37]. The authenticity of this analysis could not be validated since most of the research works in this domain lack of a baseline dataset.

k-NN can be been trained for online and real time analysis of data to identify interestingness inherent to data stream, match particular user group for classification, or recommend exhaustive options that meet specific users' needs. Though the technique requires expensive resources during computation, but it is transparent, consistent, straightforward, simple and easy to implement with high tendency to possess desirable qualities than most other data mining techniques, specifically when there is little or no prior knowledge about data distribution. Popular techniques of K-NN classifiers are Euclidean distance or cosine similarity between training and test datasets. In both techniques, the entity to be predicted is assigned a common class among its k-nearest neighbors, weights are assigned to selected variables, and the noisy data are pruned.

2.3 Application of Data Mining in Academic Performance Prediction

The explosive growth of educational data in higher institutions of learning has imitated several studies in performance prediction. Objectively, these studies mostly focus improving the quality of decisions necessary to impart delivery of quality education. Survey in [8] describes educational data mining as a leading way to determine academic performance in learning institutions. This can be backed by an expert system developed in [38] to predict students' success in gaining admission into higher education institution through Greece national exams. In the study, prediction is made at three points with different variables considered at each point. An initial

prediction is observed after the second year of studentship with specialization, sex, age, grades of students as input variables. This prediction gives a first indication of student's possibility to pass his/her exams otherwise specifying the effort necessary for student to pass. The prediction system demonstrates accuracy and sensitivity potencies of 75% and 88% respectively on experimentation. Students' performance has some form of non-linear relationship with various factors that make up their socio-demographic data [16]. This indicates that success of a student depends on patterns that exist in physiological and psychological features possessed by the student. Lori et al, [39] attempted to justify some variables that may be related to predicting

success among students who enrolled for a distant educational programme. The study concludes that learners' characteristics have a major impact on the manner in which online courses are designed and the pedagogy employed to deliver them.

Oladokun et al [40] traced the poor quality of graduates from Nigerian Universities to inadequacies admission systems in the nation. The study employed artifacts of neural network to predict the performance of candidates considered for admission into University of Ibadan, Nigeria. In the study, 10 variables from demographic and educational backgrounds information of prospective students were transformed into a format suitable for neural analysis. An output variable representing the students' final Grade Point Average on graduation is then predicted. The model operates with a promising accuracy rate of 74% but no psychometric factors were considered in the study. Also, Stamos and Andreas [41] utilized the promising behaviour of neural network to predict students' final results. The network feature a three-layered perceptron trained with back-propagation. Experimentation was conducted with a case study of 1,407 profiles of students at Waubonsee Community College, Chicago, USA. The study concluded an average predictability rate of 68%.

Pandey and Pal [42] predicted the academic performance of 600 students from different colleges of Awadh University in India by means of Bayes Classification. The study considered variables from the background qualification of students and language used in teaching them. From conclusions, new comer students will likely perform low. Furthermore, Sajadin et al.[43] applied kernel method as mining techniques to analyze relationships between students' behavior and their success. The model is developed with using smooth support vector machine to cluster final year students at Universiti Malaysia Pahang, Malaysia. Clustering observed on k-means and results expressed a strong correlation between mental conditions and academic performance of students. Hence, psychometric information can be taken as an important factor while predicting students' academic performance. In another recent study, Ahmed and Elaraby [44] used ID3 classification technique to predict final grade of students in department of Management Information System, American University, Cairo, Egypt. Student information from data mart of1547 records were used to train the model. Decision rules were used to find interestingness in the training data to predict and improve students' performance. Also, the model can help identify students that might need special attention to reduce fail rate.

Several variables have been used for establishing students' performance in prediction problems. For instance, Edin and Mirza [17] proposed predicting the performance of students with twelve variables taken as input. Conclusion can be drawn from close observations on these variables.Dorina in 2013 proposed to find patterns in the student data that could be useful for predicting students' performance at the university based on their personal and pre-university characteristics [45]. After a pretty overview of works in this domain, it can be concluded that an attribute with conditional or prior probability of 0.5 in any training set is likely an high potential variable in predicting performance with high probability [46]. In this research work, our interest is to study the robustness of K-NN in predicting students' academic performance in terms of their attributes variability.

3.0 Methodology

A proper view of the research problem and state of arts in this domain shows we have a classification problem at hand. A model conceptualized for predicting academic performance of students in higher institution is presented as Table 2

.

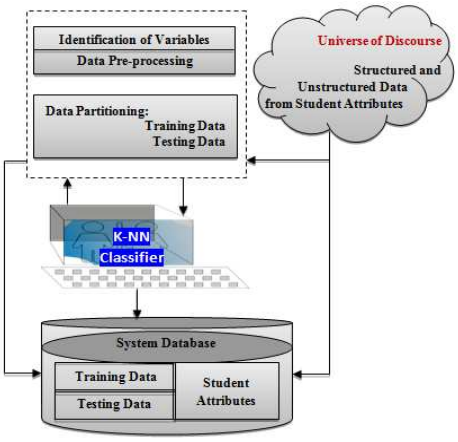

Figure 1: Conceptualized model for prediction of students' academic performance

The core of the model is to create mappings between dataset of a given group of known students to some class labels. k-NN classifier is adopted to determine academic classes of students.

3.1 Data Pre-processing

Data pre-processing is applied to improve the quality of input data to the prediction system so as to ensure an accurate, consistent and complete predictive model. The current state-of-the-art has proven having a clear understanding of the

problem domain as a great factor considered to proffer a suitable solution. The model conceptualized as Figure 1 intuits identification of variables as a major step in data pre-processing. Since there is no linear relationship that always exists between the data values of student attributes that can serve as input variables and corresponding output in educational data, it is important to observe some technical variables upon which the academic performance of students can be linearized.

Educational universe of discourse is bulged with diverse attributes from psychometric and physiological factors thereby making it very difficult to analyse. In this study, questionnaire is used to acquire student data from different variables covering demography, personal, academic, and family information. Broad overview of this subject shows that each attributes in Table 1 have useful magnitude in performance prediction of students.

Table 1: Categorization of Performance Prediction

Group	Category	SN	Attribute
A	**Demography Information**	**1**	Gender
		2	Age
		3	Category
		4	Campus Location
B	**Personal Information**	**1**	Department Management
		2	Academic Advisor
		3	Curriculum
		4	Courses
C	**Academic Information**	**1**	Level
		2	Secondary School Grade
	Academic Information	**3**	Entrance Score Average
		4	Grade Point Average
D	**Family Information**	**1**	Family Location
		2	Family Size
		3	Annual Income
		4	Fathers Qualification
		5	Mother's Qualification
		6	Father's Occupation
		7	Mother's Occupation

In another view, Pandey and Pal [42] emphasized students' grade in senior secondary school and their living location as high potential attributes in predictions process. Hence, two major steps taken during data pre-processing are data cleaning and data reduction. The first step is data cleaning which comprises of operations like filling missing values, smoothening noisy data, identifying remove outliers, and resolving inconsistencies in crude dataset. For instance, data full of impurity will cause confusion when mined, and its results cannot very reliable. For this purpose, use of central tendency is employed to fill in missing value for different attribute groups and binning is utilized for smoothening in the case of noise. In the second step, data reduction is used to obtain a simplified representation of dataset with reduced dimension while integrity of the original data is still firmed. During this process, the number of attributes under consideration is reduced to produce a more accurate and complete analytical result with integrity of the data maintained. The resulting dataset is smaller and thereby consuming lesser computing resources.

3.2 System Database

Another major component of the model conceptualized for prediction of students' academic performance in institutions of learning is the System Database. This component is a mart used to house or preserve data values that qualifies attributes of students in the universe of discourse. Basically, the system database stores both numeric and nominal information of students in any institution of learning. The information can be segmented as training and test dataset.

The training dataset is a group of data analyzed for the purpose of learning any form of interesting pattern inherent within the independent and dependent variables in a dataset. This knowledge is for the purpose of labeling newly introduced (unknown) data object to classes that best describes them. However, test dataset is used to estimate and evaluate accuracy of rules upon which class label prediction is based. Such accuracy becomes acceptable if the rules can be applied to classify new and unknown data tuples with higher precision.

Pattern learning is an interesting field whose concentric effort can be based on any learning technique described in [34]. Supervised learning technique designed on formal procedures is considered for predicting students' performance. The machine learning tactic is described as a simple K-NN classification technique that analyses and infer from available training data for the purpose of classifying new students with accuracy.

3.3 K-NN Classification

K-NN classification is a labour intensive algorithm best adopted in situations of large training dataset. The algorithm is found to conform to Euclidean distance measure in terms of distance metrics. The algorithm assumes a student, entity of our discourse herein, can be described as a set of values representing their attributes. Suppose a student (Si) is represented by the attributes described in Table 1, then relationship between the students and a set of other students is defined by their Euclidean distance.

This describes closeness that exists between the set of known students $(S_1, S_2, S_3, \dots, S_n)$ and an unknown student (S_u) in terms of distance metrics. For instance, if (S_k) represents a known student with attributes given as $(S_{k1}, S_{k2}, S_{k3}, \dots, S_{km})$, then the Euclidean distance between unknown student (S_u) and a known student (S_k) is given as:

$$E(S_u, S_k) = \sqrt{\sum_{i=1,j=1}^{n,m} \left(S_{uj} - S_{ij}\right)^2} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

If the tuples have numeric values on attributes $i = 1,2,3, \dots, m$; then V-1 the cumulative aggregation of squares of each attribute differences is normalized before Eq (1) is applied. To normalize the square of the differences, min-max normalization is applied. This function, given as Eq (2), prevents attributes with large initial ranges from outweighing attributes with smaller ranges. It transforms numeric value V of attribute A to V^{-1} in range of $[0, 1]$.

$$V^{-1} = \frac{V - \min_A}{\max_A - \min_A} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (3)$$

Where $\max_A$ and $\max_A$ are the minimum and maximum values of attribute A.

In the case of nominal attributes where values are non-numeric, the distance metric can be deduced by correlating the corresponding values of attributes in tuples $S_k$and S_u. If such values are similar, the difference is taken to be zero (0); otherwise the difference is one (1) irrespective of their ordering. If the value of A is missing in tuple S_k and/or in tuple S_u, the maximum possible difference is assumed. For instance, in nominal categorical variables, difference value of 0 is assigned if the corresponding values of the attributes are similar otherwise 1 is assigned. However, if either one or both of the corresponding values of A are missing, then the maximum possible value isassigned. For instance, if A is non-numeric and missing from both tuples $S_k$and S_u, then the difference is taken to be 1.

On a Final note, determining the number of neighboring tuples for prediction should be taken care of. A better value for k denoting the number of nearest neighbors to be considered is deduced experimentally. In reality, an odd value of k usually performs diligent especially to avoid ties. Meanwhile, a value of k=1 gives the best nearest neighborhood classification yet other values are tested to select a k value with minimum error rate.

Table 2: Attributes and Possible Values Considered during Data Preparation of the Study

Sex	Age	Category	Level	Campus Location	Department Management	Academic Advisor	Curriculum	Courses	Secondary School Grade					Entrance Score	GPA	Family Location	Family Size	Annual Income	Fathers Qualification	Mother's Qualification	Father's Occupation	Mother's Occupation
									S1	S2	S3	S4	S5									
M, F	14 … 60	UTME, DE	100, 200, 300, 400	Own Hostel Space, Share Hostel Space, Off Campus	Poor, Average, Good, Best	Poor, Average, Good, Best	Poor, Average, Good, Best	Poor, Average, Good, Best	A, B2, B3, C4, C5, C6, O	A, B2, B3, C4, C5, C6, O	A, B2, B3, C4, C5, C6, O	A, B2, B3, C4, C5, C6, O	A, B2, B3, C4, C5, C6, O	50 … 100	0.00 … 5.00	Anything	1-20	10000 … 200000	NA Elementary Secondary Bachelor Masters PhD	NA Elementary Secondary Bachelor Masters PhD	NA Personal Private Civil Public None	NA Personal Private Civil Public None
2	**R**	**2**	**4**	**4**	**4**	**4**	**4**	**4**	**7**	**7**	**7**	**7**	**7**	**R**	**R**	**A**	**R**	**R**	**6**	**6**	**5**	**5**

using a instance based learning algorithm, the experimental study is observed in a simulation under Waikato Environment for Knowledge Analysis (WEKA) software. WEKA is a free software built at the University of Waikato, New Zealand and made available under GNU General Public License with popular learning algorithms pre-designed in Java [13].

4.1 Data Preparation

Psychometric and physiological data of students were elicited by means of questionnaire presented in Appendix A, and direct link with educational database of students in department of Computer Science, University of Lagos, Nigeria. The dataset comprises of information about 310 students from all levels in 2013/2014 academic session. The data sources provide information about students' demographics, current and previous academic standings, departmental structure, and family backgrounds as in Table 2.

Subsequently, data captured into different tables were joined and attributes with lesser entropy were removed. Entropy test is a simple probability test used to reduce the dimension of attribute fields under consideration. For instance in this step, values from derived attributes, such as age, were determined while some attributes, like secondary school grades, were combined to form just one field. In the former case, list of valid values are indicated in second row, while the fields from nominal attributes have range of valid values instead.

Information gotten from data sources were compiled in Microsoft Excel 2007 and saved as .arff for usage in WEKA. A result of this is records of 310 students used for proposed system learning and classification.

Students' academic standings were stratified in five different groups based on scale in Table 3. Hence, a new attribute is considered with derived values. This nominal attribute has instance values depending on student'sGPA value.

Table 3: Classification of Students Base on Academic Standings

S/N	GPA Range	Class
1	4.50 - 5.00	Distinctive
2	3.50 - 4.49	Good
3	2.40 - 3.49	Average
4	1.50 - 2.39	Weak
5	0.00 – 1.49	Hapless

Table 4: View of Student Dataset after Data Preparation Step

No.	Gender Nominal	Age Numeric	Category Nominal	Level Numeric	Campus Location	Department Management	Academic Advisor	Curriculum Nominal	Courses Nominal	Secondary School Grade	Entrance Score Average	FamilyLocation Nominal	FamilySize Numeric	Annual Income	Fathers Qualification	Mother's Qualification	Father's Occupation	Mother's Occupation	**Class** Nominal
1	M	20.0	UTME	400.0	Share Hostel	Good	Poor	Average	Good	598916.0	58.0	ABIA	3.0	3.0	PhD	Elementary	Civil	Civil	Average
2	M	29.0	DE	400.0	Share Hostel	Average	Best	Average	Good	236794.0	94.0	ABIA	2.0	4.0	Secondary	Bachelor	Civil	Private	Distinctive
3	M	23.0	DE	400.0	Off Campus	Poor	Best	Best	Best	565756.0	65.0	ABIA	2.0	2.0	PhD	NA	Public	Public	Average
4	F	28.0	UTME	300.0	Share Hostel	Good	Poor	Good	Best	353347.0	58.0	ADAMAWA	5.0	1.0	Secondary	NA	Civil	NA	Average
5	F	26.0	UTME	100.0	Share Hostel	Average	Average	Good	Best	279863.0	53.0	ADAMAWA	4.0	2.0	PhD	PhD	Private	NA	Weak
6	F	15.0	UTME	400.0	Share Hostel	Poor	Best	Best	Good	498574.0	79.0	ADAMAWA	3.0	2.0	Bachelor	PhD	Civil	NA	Good
7	F	20.0	DE	100.0	Share Hostel	Poor	Good	Average	Best	550306.0	73.0	ADAMAWA	1.0	3.0	NA	Bachelor	NA	Private	Good
8	F	23.0	UTME	400.0	Own Hostel	Best	Good	Best	Poor	512380.0	65.0	AKWA IBOM	1.0	4.0	Secondary	Masters	Private	NA	Average
9	M	14.0	DE	200.0	Own Hostel	Good	Good	Average	Best	210856.0	100.0	AKWA IBOM	5.0	2.0	Secondary	Masters	Personal	Personal	Distinctive
10	M	21.0	UTME	300.0	Off Campus	Good	Poor	Good	Average	553127.0	64.0	AKWA IBOM	2.0	4.0	Secondary	PhD	NA	Civil	Average
11	F	11.0	UTME	400.0	Own Hostel	Good	Average	Average	Poor	663552.0	53.0	ANAMBRA	3.0	3.0	NA	Elementary	NA	Personal	Hapless
12	F	28.0	UTME	300.0	Off Campus	Poor	Best	Average	Good	518504.0	97.0	ANAMBRA	1.0	3.0	PhD	Bachelor	NA	Civil	Distinctive
13	F	16.0	DE	400.0	Own Hostel	Poor	Good	Average	Average	361448.0	61.0	BAUCHI	2.0	3.0	Bachelor	NA	NA	Personal	Average
14	M	27.0	DE	400.0	Own Hostel	Average	Average	Average	Good	240489.0	69.0	BAUCHI	1.0	1.0	Secondary	Elementary	Personal	NA	Good
15	F	13.0	UTME	100.0	Own Hostel	Average	Average	Poor	Poor	604867.0	77.0	BAUCHI	6.0	3.0	Bachelor	NA	Public	Personal	Good
16	M	20.0	DE	100.0	Off Campus	Best	Average	Average	Good	133476.0	91.0	BAYELSA	1.0	1.0	Secondary	Secondary	Public	Civil	Good

GPA value of a student depends on other variables hence, prediction in this case study will forecast the class a student will likely fall on graduation. Grouping of studentsis based on students' grade point class which is the academic standard in every university system.Table 4 shows a view of student records after preparation. The precision '.0' in some columns of the figure was added by WEKA but this does not have effect in prediction process. Table 4 presents a subset of data used to train k-NN classifier of the model in WEKA.

4.2 Experimental Setting

In WEKA, K-NN classification is applied on the dataset during the experimental study. Well, before choosing classifier, dataset was loaded from the Preprocess tab and this activates other tabs in WEKA explorer panel. Apart from this, the tab also gives a statistical visual preview of the data showing histograms for attributes in the data in separate cubicle as displayed in Figure 2.

The next step is to select and configure the classifier to be used for the experiment. The study proposed using k-NN which chooses majority class from several k-neighbors.

WEKA workbench is suite of visualization techniques pre-programmed for data analysis and pattern learning for the purpose of knowledge discovery. The techniques are aided with graphical user interfaces for easy access [29]. In this experiment, instance based classifier of the suite is adapted for prediction.

Examining a value of K upon which the classifier predicts student's performance with best accuracy, that is a k value with higher predictive accuracy, becomes an issue. However, since the value of k to be used strongly relates to number of training tuples and relationships that exist in their attribute values, we can easily predict the test tuples with 1-NN. On another note it might not be completely correct to base prediction on such shallow specific closest tuple, so we tested different values of k (3-9) and it was discovered that the experiment works better with k = 3 but with best accuracy and minimum error rate at k=5. We thereby chose 5-NN for prediction.Furthermore, the classifier is evaluated on how well it predicts a test data. For this purpose, experiment is conducted on 96% split training thereby leaving 12 instances of the dataset for validation and evaluation of the proposed model.

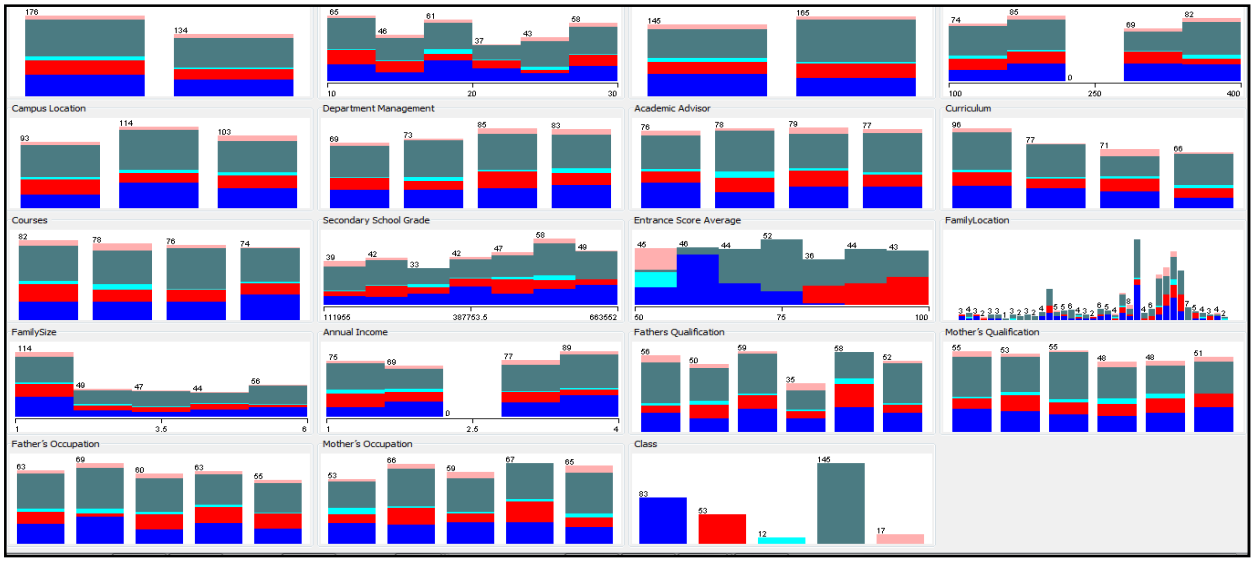

Figure 2: Statistical Visual Preview of the Experiment Dataset

4.3 Results and Evaluation

A total of 4% of the split data resorted to 12 instances used for classifier validation. The result observed from experiment with this data, is summarized in Table 5. Summary of result shows the classifier is 58.33% to have predicted classes of 7 from 12 instances correctly with Mean Absolute Error (MAE) of 0.217.

Table 5: Analysis on Actual and Predicted Classes in Test Data

Id	Actual	Predicted	Error	Id	Actual	Predicted	Error
1	Distinct	Good	0.200	7	Distinct	Distinct	0.399
2	Distinct	Distinct	0.001	8	Average	Good	0.200
3	Good	Good	0.001	9	Good	Good	0.200
4	Distinct	Average	0.399	10	Average	Good	0.399
5	Average	Average	0.399	11	Average	Average	0.399
6	Good	Good	0.001	12	Average	Good	0.001

Table 6 shows the model can predict academic performance of students in terms of their class with percentage harmonic measures of: 72.7% assurance for **Good** students, 57.1% for **Distinctive** students, and 33.3% for **Average** students. However, the model does not show tendency to predict students that are weak or hapless in their performance. Also, the Receiver Operating Characteristics (ROC) values are noted.

Table 6: Performance of k-NN on using Percentage Split

No	TP Rate	FR Rate	Precision	Re-call	F1- measure	ROC	Class
1	0.333	0.222	0.333	0.333	0.333	0.611	Average
2	0.4	NA	1	0.4	0.571	0.7	Distinctive
3	0	NA	0	0	0	?	Weak
4	1	0.375	0.571	1	0.727	0.813	Good
5	0	NA	0	0	0	?	Hapless

However, result obtained on another run round with cross validation as the test option shows the predictive model can be improved to predict academic performance of students that are hapless. For instance, the result of a 10-fold cross validation built in few micro-seconds is presented in Table 7.

Table 7: Performance of k-NN using 10-Fold Cross Validation

No	TP-Rate	FR-Rate	Precision	Re-call	F1-measure	ROC	Class
1	0.41	0.357	0.296	0.41	0.343	0.502	Average
2	0.226	0.101	0.316	0.226	0.264	0.611	Distinctive
3	0	0.007	0	0	0	0.539	Weak
4	0.572	0.43	0.539	0.572	0.555	0.548	Goo
5	0.059	0	1	0.059	0.111	0.702	Hapless

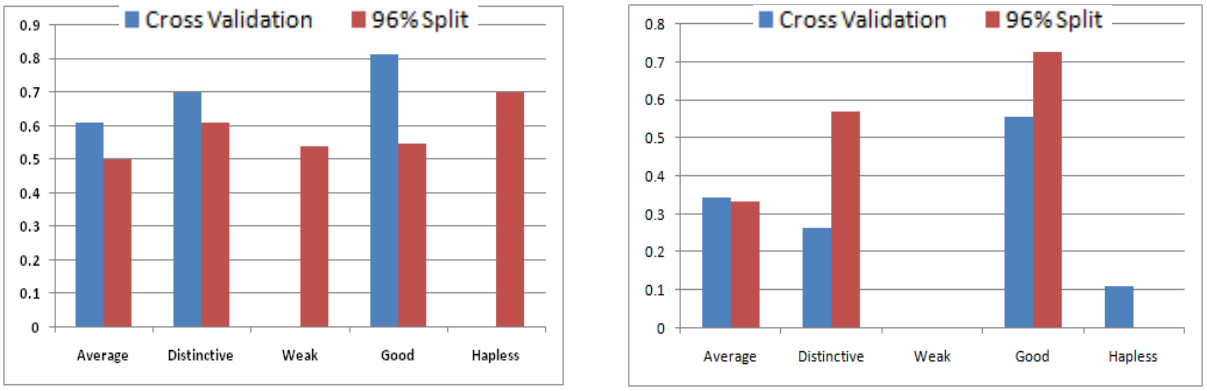

Figure 3: Basic graph for Harmonic (Left) and ROC (Right) Measures on the Test Options

The effect of changing the test option to cross-validation indicates that the k-NN (with K=5) can determine students that will be hapless in their study with percentage harmonic measure of 11.1%. Therefore, the classifier can predict performance of students that are Weak, Average, Good, and Distinctive.

Summarily, the classifier can predict the performance of student from all classes except the weak ones, and this effect can be accorded to variations in dataset used for the experiment. Moreover, this study is aimed at improving quality of teaching in higher educational institutions hence major focus is on Hapless students. It is better such students are not taken into the educational system since the proposed model shows they cannot complete their studies. A basic graph for the harmonic and ROC measures is presented as Figure 3.

Performance of learning techniques is strongly dependent on the nature of training data used. To verify the predictive accuracy of the k-NN classifier using benchmark observation, the model is compared with four other learning algorithms. Table 8 shows the results observed from five predictive models with each experimented on 96% split training dataset. The remaining 4% dataset has 12 record instances chosen from blind randomization.

Table 8: Evaluation of k-NN and Four Other Classifiers using 96% Split Training Dataset

Evaluation Metric	**k-NN**	**Naïve Bayes**	**ZeroR**	**CART**	**OneR**
Correctly Classified	7	11	4	7	7
Time Taken to Build (Seconds)	0.09	0.51	0.10	1.36	0.23
Accuracy	58.33%	91.67%	33.33%	58.33%	58.33%
MAE	0.2166	0.2306	0.2838	0.2888	0.2667
Kappa Statistics	0.1025	0.8737	0	0.3939	0.3939

Table 8 shows that Naïve Bayes is an optimal classifier with an accuracy of 91.67% followed by k-NN (with K=5) having 58.33%. However, the time taken to build Naïve Bayes classifier is more than that of k-NN and therefore the former will not be optimal in cases of very large training data. Also, the prediction error associated with k-NN is smaller meaning the classifier's predictions are more dependable. Clearly, the kappa statistics shows the measures of agreements between the actual and predicted values for all the classifiers. Unlike MAE and other metrics, kappa statistics shows that prediction made by k-NN agrees very well with the actual data and this is done in minimal time when compared with other classifiers. Lastly, Analysis of the classifiers' accuracy and time taken to build them are visually presented in Figure 4 and Figure 5 respectively.

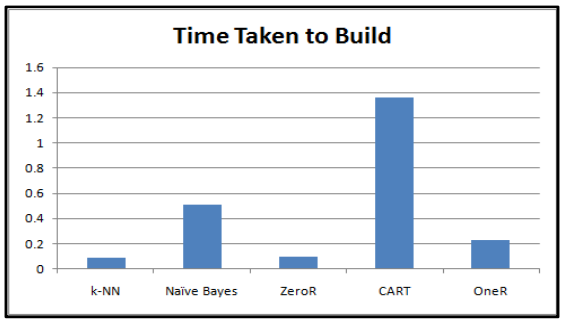

Figure 4: Chart of Built Time of the Classifiers

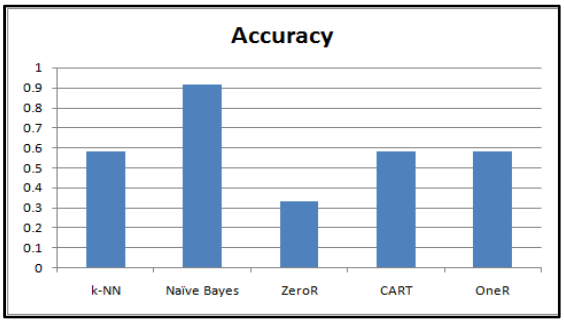

Figure 5: Chart of Accuracy of the Five Classifiers

5.0 Conclusions

Data mining is a computational approach adopted to discover hidden patterns and relationships that are helpful in decision making. Techniques such as clustering, outlier detection, and association rule mining have shown advances in educational sector. In this study, a neighborhood classification approach (k-NN) is applied to predict the academic performance of students in institutions of higher learning. The use of data mining in educational field is being adopted to enhance humans' understanding of learning process. In this study, a model capable of predicting academic performance of students in higher institutions is proposed. For a real-life development, the model is conceptualized as three-tier architecture with k-NN classification technique adopted as the prediction procedure at the second tier.

To validate, the model is simulated in WEKA with 5-NN IBK classifier chosen after many brute-force attempts. Results and evaluation of the model when trained and tested with a vast amount of dataset shows the predictive algorithm can determine the educational status of a student on a 5-likert GPA scale: Distinctive, Good, Average, Weak, and Hapless. However, experiment with 96% split shows the model's inability to foresee weak students, and this was confirmed on a further test option with 10-fold cross validation. Results from simulation shows that Naïve Bayes has better accuracy over other classifiers, though an exception is the longer time needed to build. This makes it sub-optimalfor experiments with large dataset.

The prediction scheme can be adopted in different institutions to promote their standards because quality of students in a school has direct relationship with the school's standard. Therefore, it is good to note that the model can only predict success likelihood, and educational status of students that are already admitted into an institution of higher learning. However, it is also good to examine cases of students who are just seeking admission or to predict alternative programme (course) for those who prefer to change from their existing department because of their poor performance.

Knowledge models obtained under these paradigms are usually considered to be black-box mechanisms, able to attain very good accuracy but they are difficult for people to understand. This conventional modeling approach makes use of mathematical models to connect upstream information to downstream consequences. Typically, such models are constrained by the variables utilized, and usually they have limited or no ability to add, reject or change these variables during prediction processes.

Lastly, the expression of certainty in outcome of predictive models is a major factor in accepting the predictions made by such model. It is therefore recommended that future works look into evaluating the performance of predictive models with test of certainty. This can be dealt with through sensitivity analysis wherein the accuracy of outcomes from a model responds to changes in the number, types and values of variables through simulation of potential ranges of values for those variables. Also, importance of some distinct variables should be put into consideration using weighting predictive approaches.

APPENDIX A: Questionnaire

Please complete this questionnaire by selecting the best group you fall. This is for strict academic use hence, no identifying information is needed.

SECTION A: Demography Information (*Thick the appropriate box*)

Gender	☐ Male	**Category**	☐ UTME
	☐ Female		☐ DE
Age on Admission	☐ Below 14	**Campus Accommodation Type**	☐ Own Hostel Space
	☐ Between 14-18		☐ Share Hostel Space
	☐ Above 18		☐ Campus Accommodation

SECTION B: Personal Information (*Rate the following indices with the factors below*)

Department Management	☐ Poor	**Curriculum Outfit**	☐ Poor
	☐ Average		☐ Average
	☐ Good		☐ Good
	☐ Best		☐ Best
Teaching Style	☐ Poor	**Academic Advisor**	☐ Poor
	☐ Average		☐ Average
	☐ Good		☐ Good
	☐ Best		☐ Best

Course Content	☐ Poor	☐ Average	☐ Good	☐ Best

SECTION C: Academic Information (*Thick or Shade the following boxes as appropriate*)

Secondary School Grades in Five Core Subjects	**English**						**UTME/DE Entrance Score Average**	☐ % and Above
	A1	B2	B3	C4	C5	C6		☐ % - 89%
	Mathematics							☐ % - 79%
	A1	B2	B3	C4	C5	C6		☐ % - 69%
	Physics							☐ Below 60%
	A1	B2	B3	C4	C5	C6	**Cumulative Grade Point Average (CGPA)**	☐ 0 and Above
	Chemistry							☐ 0 - 4.49
	A1	B2	B3	C4	C5	C6		☐ 0 - 3.49
	Biology							☐ 1.50 - 2.39
	A1	B2	B3	C4	C5	C6		☐ Below 1.50

SECTION D: Family Information (*Rate the indices in the following boxes*)

Annual Income	☐ low 50, 000	☐ 50,000-100,000	☐ 101,000 - 200,000	☐ Above 200,000
Father's Qualification	☐ No Education		**Mother's Qualification**	☐ No Education
	☐ Elementary			☐ Elementary
	☐ Secondary			☐ Secondary
	☐ Graduate			☐ Graduate
	☐ Post Graduate			☐ Post Graduate
	☐ Post Graduate (Ph.D)			☐ Post Graduate (Ph.D)
	☐ Not Applicable			☐ Not Applicable
Father's Occupation	☐ Personal Service		**Mother's Occupation**	☐ Personal Service
	☐ vate Service			☐ vate Service
	☐ Civil Service			☐ Civil Service
	☐ Public Service			☐ Public Service
	☐ Not Applicable			☐ Not Applicable

6.0 References

[1] Tawarish M. & Waseemuddin F. (2013), Analysis of Data Mining Concepts in Higher Education with Needs to Najran University, International Journal of English and Education, 2(2), 590-97

[2] Kumar A. & Vijayalakshmi M. (2011), Implication of Classification Techniques in Predicting Student's Recitals, International Journal of Data Mining & Knowledge Management Process, 1(5), 41-51

[3] Bienkowski M., Feng M., & Means B. (2012), Enhancing teaching and learning through educational data mining and learning analytics: An Issue Brief, Technical Report, Office of Educational Technology, U.S.

[4] Akinola O., Akinkunmi B. & Alo T. (2012), A Data Mining Model for Predicting Computer Programming Proficiency of Computer Science Undergraduate Students, African Journal of Computer & ICTs, 5(1), 43-52

[5] Yao Y. Y. (2003), Information-Theoretic Measures for Knowledge Discovery and Data Mining, Maximum Entropy Principle and Emerging Applications, Studies in Fuzziness and Soft Computing, 119, 115-136

[6] Alexander, H. (1996), Physiotherapy Student Clinical Education: The Influence of Subjective Judgments on Observational Assessment, Assessment and Evaluation in Higher Education, 21(4), 357-366.

[7] Galit (2007), Examining online learning processes based on logfiles analysis: a case study, Research, Reflection and Innovations in Integrating ICT in Education.

[8] Romero C., & Ventura S. (2007), Educational data mining: A survey from 1995 to 2005, Expert Systems with Applications 33 (2007), 135-146

[9] Samuel O., Omisore M., & Atajeromavwo E. (2014), Real Time Fuzzy Based System for Human Resource Performance Appraisal, Journal of International Measurement Confederation, 55, 452 - 461.

[10] Luan J. (2002), Data Mining and Its Applications in Higher Education, New Directions for Institutional Research, Chapter Two, Wiley Periodicals, No. 113,17-36.

[11] Atajeromavwo E., Omisore O., & Samuel O. (2015), A Genetic-Neuro-Fuzzy Inferential Technique for Diagnosis of Tuberculosis, 5th MobiHoc Workshop on Pervasive Wireless Healthcare, Hangzhou, China, June 22-26.

[12] Muqasqas S. A., Al Radaideh Q. A., & Abul-Huda B. A. (2014), A Hybrid Classification Approach Based on Decision Tree and Naïve Bays Methods, International Journal of Information Retrieval Research, 4(4), 61-72

[13] Omisore M. O., Fayemiwo M. A., Ofoegbu E. O., & Adeniyi S. A. (2015), Simulation of a Stock Prediction System using Artificial Neural Networks, International Journal of Business Applications.

[14] Alaa el-Halees, Mining Students Data to Analyze e-Learning Behavior: A Case Study, 2009.

[15] Zeljko P. (2012), The Evolution of Business Intelligence: From Historical Data Mining to Mobile and Location-based Intelligence, Recent Researches in Business and Economics, ISBN: 978-161804-102-9

[16] Kovačić Z. (2010), Early Prediction of Student Success: Mining Students Enrolment Data, In: Proceedings of Informing Science and IT Education Conference, Cassino, Italy, June 21-24, 647-665.

[17] Edin O. & Mirza S. (2012), Data Mining Approach for Predicting Student Performance, Journal of Economics and Business, 10(1), 3-12

[18] Minaei-Bidgoli B. & Punch, W. (2003), Using Genetic Algorithms for Data Mining Optimization in an Educational Web-based System, Genetic and Evolutionary Computation, Part 2, 2252-2263

[19] Yao, X. (1999), Evolving Artificial Neural Networks, Proceedings of the IEEE, 87(9), 1423-1447

[20] Martínez F., Hervás C., Gutiérrez P., Martínez, A. & Ventura S. (2006), Evolutionary Product-Unit Neural Networks for Classification, Conference on Intelligent Data Engineering and Automated Learning,1320 - 28.

[21] Russell S. & Norvig P. (2003), Artificial Intelligence: A Modern Approach, Second Edition, Prentice Hall, ISBN 978-0137903955.

[22] John G. & Langley, P (1995), Estimating Continuous Distributions in Bayesian Classifiers, 11th Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann Publishers, SM, 338-345

[23] Nghe N., Janecek P., & Haddawy P. (2007), A Comparative Analysis of Techniques for Predicting Academic Performance, 37th ASEE/IEEE Frontiers in Education Conference, October 10-13, Wisconsin, USA, p. 7-12.

[24] Lowd D. & Domingos P. (2005), Naive Bayes Models for Probability Estimation, Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany, 2005

[25] Kotsiantis S. (2007), Supervised Machine Learning: A Review of Classification Techniques, Informatica Vol. 3 (2007), 1249-268.

[26] Yi-Horng L. (2015), Applying Data Mining Technology on the Using of Traditional Chinese Medicine in Taiwan, International Journal of Computer and Information Technology, 4(1), 50-57

[27] Satya G., Somayajula S., & Karthika P. (2011), Fuzzy SLIQ Decision Tree for Quantitative length Test Data-sets, International Journal of Computer Science & Technology, 2(1), 61-64

[28] Quinlan, J. (1993), C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers.

[29] Sudhir B. J. & Kodge B. G. (2013), "Census Data Mining and Data Analysis using WEKA", International Conference in Emerging Trends in Science, Technology and Management, Singapore, 35-40

[30] Kim D., Lee J., Lee J., & Chang S. (2005), Application of Prediction of Probabilistic Neural Networks of Concrete Strength, Journal of Materials in Civil Engineering, 17(3), 353-362, 2005

[31] Kim D., Kim D. & Chang S. (2007), Modified Probabilistic Neural Network Considering Heterogeneous Probabilistic Density Functions in the Design of Breakwater, Journal of Civil Engineering, 11, 65-71

[32] Razi M., & Athappilly K. (2005), A Comparative Predictive Analysis of Neural Networks, Nonlinear Regression and Classification and Regression Tree Models, Expert Systems with Applications, 29, 65-74

[33] Cowell R., Lauritzen S., Spiegelhater D., & David P. (2003), Probabilistic networks and expert systems. New York, NY: Springer.

[34] Han J., Kamber M., & Pei J (2012), Data mining concept and Techniques, Third Edition, Morgan Kaufmann Publishers, Elsevier inc., 225 Wyman Street, Waltham, MA 02451, USA, 285-350

[35] Zdravko M., Daniel T. (2007), Data mining the Web, Uncovering patterns in Web Content, Structure and Usage, John Wiley & sons Inc., New Jersey, USA, 115-132.

[36] Amartya S. & Kundan K (2007), Application of Data mining Techniques in Bioinformatics, Bachelor Thesis, Department of Computer Science and Engineering, National Institute of Technology, Rourkela.

[37] Ribeiro P. S. (2013), Strategies to thwart students' attrition in higher education: a data mining approach, Doctoral Programme in Informatics Engineering, Faculty of Engineering, University of Porto, Portugal.

[38] Hatzilygeroudis I., Karatrantou A., & Pierrakeas C. (2004), PASS: An Expert System with Certainty Factors for Predicting Student Success in Negoita et al. (Eds.): KES 2004, LNAI 3213, 292-298, 2004.

[39] Lori B., Frederick B., Scott W. (2004), Student Traits and Attributes Contributing to Success in Online Courses: Evaluation of University Online Courses, Journal of Interactive Online Learning, 2(3), 1-17

[40] Oladokun V., Adebanjo A., & Charles-Owaba O. (2008), Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course, The Pacific Journal of Science and Technology 9(1), 72-79.

[41] Stamos T. & Andreas V (2008), Artificial Neural Network for Predicting Student Graduation Outcomes, Proceedings of the World Congress on Engineering and Computer Science, San Francisco, USA

[42] Pandey U. & Pal S. (2011), Data Mining: A Prediction of Performer or Underperformer using Classification, International Journal of Computer Science and Information Technology, 2(2), 686-690

[43] Sajadin S., Zarlis M., Dedy H., Ramliana S., Elvi W. (2011), Prediction of Student Academic Performance By An Application Of Data Mining Techniques, International Conference on Management and Artificial Intelligence, IACSIT Press, Bali, Indonesia, 6, 110-114

[44] Ahmed A, & Elaraby I (2014), Data Mining: A prediction for Student's Performance Using Classification Method, World Journal of Computer Application and Technology, 2(2), 43-47.

[45] Dorina K. (2013), Predicting Student Performance by Using Data Mining Methods for Classification, Cybernetics and Information Technologies, 13(1), 61-72.

[46] Hijazi S., & Naqvi R. (2006), Factors Affecting Student's Performance: A Case of Private Colleges, Bangladesh e-Journal of Sociology, 3(1), 67-74.