Comparative Analysis of Selected Classifiers in Mining Students' Educational Data

Ayinde A.Q
Osun State College of
Technology, Esa Oke
Osun State. Nigeria

E.O Omidiora
Ladoke Akintola University
of Technology, Ogbomoso.
Oyo State, Nigeria

A.B Adetunji, Ph.D.
Ladoke Akintola University
of Technology, Ogbomoso.
Oyo State, Nigeria

ABSTRACT
Educational data mining is concerned with developing methods that discover knowledge from educational databases. Many predictive classifiers have been applied in mining educational data with less emphasis on their performance evaluation in order to determine the most efficient. In this study, a comparative analysis of three predictive classifiers for mining educational data was conducted.

General Terms
Data Mining, Knowledge Discovery, Educational Data Mining, Students' Academic performance.

Keywords
Comparative Analysis, Selected Classifiers, Instance Based Learning, Lazy Classifier.

1. INTRODUCTION
Data Mining is a process of extracting previously unknown, valid, potentially useful and hidden patterns from large data sets. Data Mining is a technology used to describe knowledge discovery and to search for significant relationships such as patterns, associations and changes among variables in databases. The discovery of those relationships can be examined by using statistical, mathematical, artificial intelligence and machine learning techniques to enable users to extract and identify greater information and subsequent knowledge than simple query and analysis approaches [1]. Data mining techniques have been applied in many application domains such as banking, fraud detection, instruction detection, communication, education, etc.

This new emerging field called Educational Data Mining, which is concerned with developing methods that discover knowledge from data that come from educational environments. The data can be collected from historical and operational data residing in the databases of educational institutes. The student data can be personal or academic. Also it can be collected from e-learning systems which have a vast amount of information used by most institutes [2].

Educational data mining techniques includes decision trees, neural networks, nearest neighbor, bayesian network, support vector machine and many others. Using these techniques many kinds of data mining tasks can be accomplished such as estimation, description, classification, association, clustering and prediction. The discovered knowledge can be used to better understand students' performance, to assist instructors, and to improve teaching [3].

2. LITERATURE REVIEW
Data mining can be defined as the process involved in extracting interesting, interpretable, useful and novel information from data. The student data mining process allows us to have a better perspective on the student progress throughout the educational processes, and at the same time to analyze the information related to the specific programs, courses, and course assignments [4].

It has been used for many years by businesses, scientists and governments to mine volumes of data like airline passenger records, census data and the supermarket scanner data that produces market research reports [5] Educational miners compared two classifiers (decision tree and Bayesian network) to predict students GPA at the end of the third year of undergraduate and at the end of the first year of postgraduate from two different institutes. Each data set has 20,492 and 936 complete student records respectively. The results show that the decision tree outperformed Bayesian network in all classes. The accuracy was further improved by using re-sampling technique especially for decision tree in all cases of classes. In the same time it is able to reduce misclassification especially on minority class of imbalanced datasets because decision tree algorithm tends to focus on local optimum [6].

Comparative analysis of six classification methods (Naive Bayes, decision tree, feed-forward neural network, support vector machine, 3-nearest neighbour and logistic regression) were used to predict drop-outs in the middle of a course. The data set contained demographic data, results of the first writing assignments and participation to group meetings. The data set contained records of 350 students. Their best classifiers, Naive Bayes and neural network, were able to predict about 80% of drop-outs. The results also shown that simple model such as naïve bayes were able to generalize well on small data set compare to other method such as decision tree and nearest neighbour that require much larger size of data sets [7]

The comparison of six classifiers involved bayesian classifier, 1-nearest neighbours, k-nearest neighbours, Parzen window, feed-forward neural network, and decision tree to predict the course final results from a learning system log data . The data contained attributes concerning each task solved and other actions like participating in the communication mechanism and reading support material. The data set contained records of 250 students. Their best classifier, k-nearest neighbours, achieved over 80% accuracy, when the final results had only two classes which are pass or fail [8].

Researchers applied data mining classifiers as a means of analyzing and comparing the performance of students who have taken a technical course via the web. Their results show that combination of multiple classifiers leads to a significant accuracy improvement in the given data set. Often prediction performance of combining classifiers is better than a single

classifier because the decision is relying on collective output of several models. Weak classifiers are linear classifiers which less likely to suffer from over-fitting problems [8]. Combination of weak classifiers based on boosting approach was used by [9] to predict the final score. Each weak classifier used only one of 74 attributes to predict the course score. The combination achieved only 69% accuracy but the boosting revealed the most influencing factors for the course success.

3. METHODOLOGY

3.1 Overview

Prediction is an important data mining technique with broad applications. It classifies data of various kinds. Prediction is used in every field of our life. It is used to classify each item in a set of data into one of predefined set of classes or groups. In this research, students' academic record was collected from the Department of Computer Science and Engineering, LAUTECH. The performance of the algorithms was evaluated in predicting the students' final year grades. Naive Bayes algorithm is based on Bayesian Network, IBk algorithm is based on Lazy Classifier and CART algorithm is based on Decision Tree. This research did a comparative evaluation of the classifiers NAÏVE BAYES, IBk and CART in the context of educational institute dataset in term of true positive rate, false positive rate, recall and classification error using WEKA tool.

3.2 Data

The student's academic record is not available on Students' Online Portal System in Ladoke Akintola University of Technology Ogbomosho, Nigeria. In other to ensure direct statistical analysis and evaluation, this research work was streamlined towards students' academic record in the Department of Computer Science and Engineering, LAUTECH, Ogbomosho.

3.3 Research Approach

The Cross-Industry Standard Process for Data Mining (CRISP- DM) model was adopted and considered as a standard approach for this research work. It is a cyclic approach that consists of six main phases– Business understanding, Data understanding, Data preparation, Modeling, Evaluation and Deployment. The approach was applied the 385dataset that were obtain for three graduating set from the department of Computer Science. As shown in figure 1
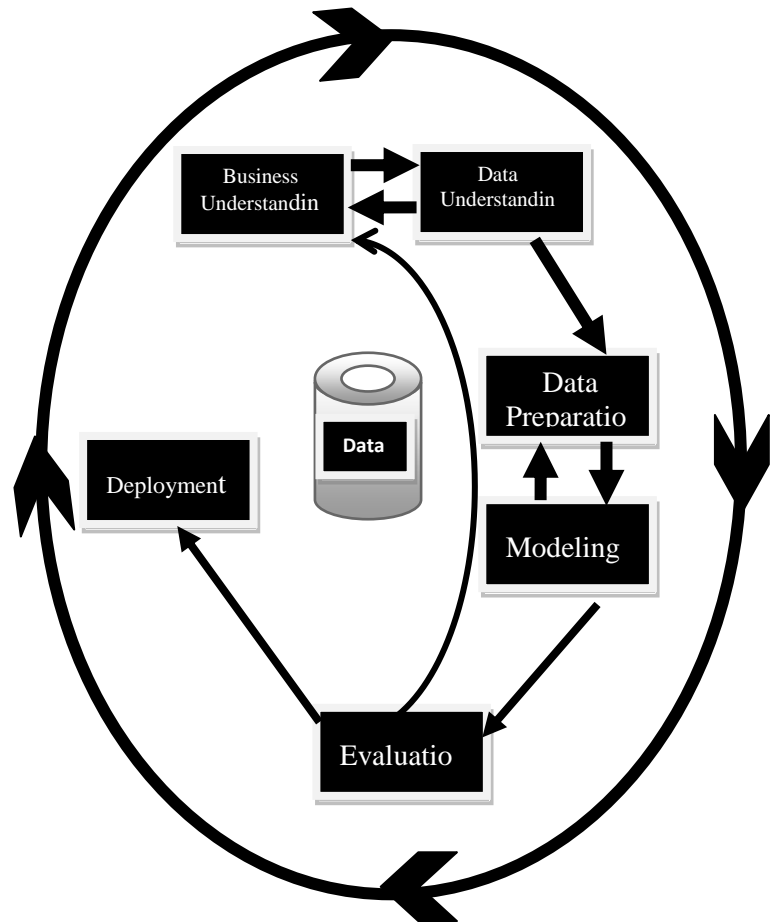

Figure 1: CRISP - DM

3.4 Equations

Naïve Bayes classifier: applies to learning tasks where each instance x is described by a conjunction of attribute values and where the target function $f(x)$ can take on any value from some finite set V. A set of training examples of tuple of attribute value $(a_1, a_2, ...a_n)$.The learner is asked to predict the target value or classification for this new instance.The Bayesian approach shall be used to classify the new instance to be assigned the most probable target, V_{map}, given the attribute value $(a_1, a_2, ...a_n)$ that describe the instance.

$$V_{map} = \arg_{max} \ P(V_j | a_1, a_2, ...a_n)$$

The Bayes theorem in equation above shall be used to rewrite this expression as

$$V_{map} = \arg_{\substack{max \\ Vj \in V}} \ P(V_j | a_1, a_2, ...a_n | v_j) P(V_j)$$

CART Classifier: The classification and regression tree (CART) method was suggested by [10]. The decision trees produced by CART are strictly binary, containing exactly two branches for each decision node. CART recursively partitions the records in the training data set into subsets of records with similar values for the target attribute. The CART algorithm grows the tree by conducting for each decision node, an exhaustive search of all available variables and all possible splitting values, selecting the optimal split (Kennedy, Yunchun, Benjamin, Christopher and Richard, 1995).

$$\phi\left(\frac{s}{t}\right) = 2P_L P_R \sum_{J=1}^{\#classes} P\left(\frac{j}{t_L}\right) - P\left(\frac{j}{t_R}\right)$$

$$P_L = \frac{number\ of\ records\ at\ t_L}{number\ of\ records\ in\ training\ set}$$

$$P_R = \frac{number\ of\ records\ at\ t_R}{number\ of\ records\ in\ training\ set}$$

$$P\left(\frac{j}{t_L}\right) = P_L = \frac{number\ of\ class\ j\ records\ at\ t_L}{number\ of\ records\ at\ t}$$

$$P\left(\frac{j}{t_R}\right) = P_L = \frac{number\ of\ class\ j\ records\ at\ t_R}{number\ of\ records\ at\ t}$$

Then the optimal split is whichever split maximizes this measure $\phi\left(\frac{s}{t}\right)$ over all possible splits at node *t*.

IBK Classifier: IBk algorithm is a non-parametric method for classifying objects based on closest training examples in the feature space. IBk is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. The IBk algorithm is amongst the simplest of all machine learning algorithms. It is used in classifying object based on a majority vote of its neighbors, with the object being assigned to the class most common amongst its nearest. The beauty of the IBk is that, it can handle incremental data or data in batches.

IBK equation

It stored the training set by classifying a new example *x* by finding example (*x*, *x_i*) that is nearest to *x* according to

Euclidean distance: $\left\| x - x_i \right\| = \sum_j \left(x_j - x_{ij} \right)^2$

Where; *x* denotes training data sets and x_i denotes target variables The equation predicts the target class x_i.

4. ANALYSIS OF RESULT
The results were evaluated based on the following parameters: True Positive Rate, False Positive Rate, Correctly Classified instances using Naïve Bayes, CART and IBk classifiers.

Each classifier was applied for cross validation testing option (using 2/3 of the dataset for training set and the rest 1/3 for testing). The results obtained in the evaluation carried out using the classifiers are as presented in Tables below. This was based on five classes of grades (First Class, Second Class Upper, Second Class Lower, Third Class and Pass). Also, shows the analysis of variance conducted on the results obtained in the Tables. Where PA = Performance Accuracy

Table 1 First Class Grade

PERFORMANCE	Naive Bayes	CART	IBK
FP Rate	0.746 (3.7)	0.547	0.918
TP Rate	0.574	0.598	0.663
PA	0.453	0.675	0.798

(3.8)

Table 2 Second Class Upper Grade

PERFORMANCE	Naive Bayes	CART	IBK
TP Rate	0.519 (3.9)	0.613	0.924
FP Rate	0.611	0.815	0.921
PA	0.643 (3.10)	0.874	0.612

Table 3 Second Class Lower Grade

PERFORMANCE	Naive Bayes	CART	IBK
TP Rate	0.598	0.741	0.818
FP Rate	0.711	0.410	0.312
PA	0.612	0.785	0.909

Table 4 Third Class Grade

PERFORMANCE	Naïve Bayes	CART	IBK
FP Rate	0.481	0.611	0.752
TP Rate	0.517	0.910	0.321
PA	0.643	0.815	0.762

Table 5 Pass Class Grade

PERFORMANCE	Naïve Bayes	CART	IBK
FP Rate	0.781	0.652	0.923
TP Rate	0.813	0.652	0.313
PA	0.853	0.706	0.951

5. RESULT DISCUSSION
Educational database classifiers such as: Naive Bayes (NB), Instance Based Learning (IBk) and Classification and Regression Tree (CART) were selected for comparative analysis. An educational database of 385 university

undergraduate students' bio-educational data was created, each with 7 attributes. The attributes were sex, age, mode of admission, state of origin, religion, 200 level cumulative grade point average (CGPA) and students' final grade. Six of the attributes served as the input (sex, age, mode of admission, state of origin, religion and 200 level CGPA) in predicting the output attribute (students' final year grade).These classifiers were tested using a percentage split evaluation methodology (2/3 of the dataset for training and 1/3 for testing) using the Knowledge flow platform of the Wekaito Environment for Knowledge Analysis (WEKA). The performance of the classifiers was evaluated based on: True Positive Rate (TPR), False Positive Rate (FPR) and Performance Accuracy (PA). The Tables revealed that the lazy classifier (IBk) was able to predict for all the classes of grade with and overall performance rate of 86%, the Bayesian network (Naive Bayes) has the performance rate of 64% and the decision tree (CART) performance rate was of 62%.

The outstanding performance of the lazy classifier (IBk) was because it stored all of the training datasets and do not build a model until a new datasets need to be predicted (incremental learning) which differs from Bayesian network (Naive Bayes) and decision tree (CART) which build the model before receiving the new datasets (batch learning) in mining the students' educational data. The lazy classifier is faster at the training time compared to others but slower at predicting time since all computation was delayed till it receive the last datasets based on true positive rate, false positive rate and correctly classified instance.

A one-way analysis of variance (ANOVA) was conducted on the performance evaluation results. The performance evaluation results of the classifiers show that NB, IBk and CART had average TPR of 55.4%, 79.4% and 58.7% with their corresponding FPR of 58.4%, 42.3% and 65.8%, respectively. NB, IBk and CART had PA values of 59.3%, 81.6% and 68.9%, respectively. The ANOVA results of NB, IBk and CART in predicting the students' final year grade are 0.37, 0.03 and 0.29, respectively based on null hypothesis probability.

6. CONCLUSION

The comparative analysis of the three predictive classifiers showed that IBk gave the best PA with the lowest null hypothesis value. Also, TPR of NB and CART decreases with decreasing value of the PA while TPR of IBk increases with increasing value of PA, which implies that IBk performs best with increasing value of TPR.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Han and Kamber 2000.Data Mining Concepts and Techniques.

[2] Witten I.H., Frank E.,(2005) "Data Mining: Practical Machine Learning Tools and Techniques", 2nd Edition. San Francisco: Morgan Kaufmann, 2005.

[3] Daniel Larose (2005): Knowledge Discovery In Data. An Introduction to Data Mining. First Edition ISBN 0-471-66657-2 pg 5,2005

[4] Michalski.K and Michalski.R (2004).Educational Data Mining and Reporting: Analyzing student Data in Order To Improve Educational Process.In L. Cantoni and C.McLoughlin(Eds).Proceedings of World Conference on Educational Multimedia,Hypermedia and Telecommunications 2004(pp. 1088 1094) Chesapeake,VA: AAC

[5] Han and Kamber (2006):Data Mining Concept and Techniques.2nd Edition ISBN 1-55860-901-6

[6] Nguyen Thai Nghe, P. Janecek, and P. Haddawy (2007)"A comparative analysis of techniques for predicting academic performance", ASEE/IEEE Frontiers in Education Conference,pp. T2G7-T2G12, 2007.

[7] Kotsiantis .S ,Pierrakeas .C, and Pintelas. P (2004), "Prediction of Student's Performance in Distance Learning Using Machine Learning Techniques", Applied Artificial Intelligence, Vol. 18, No. 5, 2004, pp. 411-426

[8] B. Minaei-Bidgoli, G. Kortemeyer, and W. F. Punch, (2004) "Enhancing Online Learning Performance: An Application of Data Mining Method", In proceedings of The 7th IASTED International Conference on Computers and Advanced Technology in Education (CATE2004), Kauai, Hawaii, USA, pp. 173-8, August 2004.

[9] W. Zang, and F. Lin,(2003) "Investigation of web-based teaching and learning by boostingalgorithms". In *Proceedings of IEEE International Conference on Information Technology: Research and Education (ITRE 2003)*, pp. 445–449, 2003.