International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 17, 2017 ISSN: 2348-6600
Reference ID: IJCS-237 PAGE NO: 1480-1487
Alagappa University, Karaikudi, India 15th -16th February 2017
IT Skills Show & International Conference on Advancements in Computing Resources (SSICACR-2017)
http://aisdau.in/ssicacr ssicacr2017@gmail.com

IDENTIFYING STUDENT CAPACITY TO IMPROVE ACADEMICS PERFOMANCE USING CLASSIFICATION ALGORITHM

NANDHINI.M
PG Student
Department Of Computer Science (PG)
PSGR Krishnammal College for Women,
Coimbatore, India
nandhudass12@gmail.com

THENDRAL.T
Assistant Professor
Department Of Computer Science (PG)
PSGR Krishnammal College for Women,
Coimbatore, India
thendral@psgrkc.ac.in

***Abstract*—** The main objective of the research is to improve Students performance in academic and classify the students as slow learner and fast learner according to their marks. The research problem is identifying slow learner and fast leaner from the given dataset using classification algorithms. Data mining is the process of extracting or mining hidden knowledge from huge amounts of data. The information and knowledge gained can be used for applications ranging from Financial Data Analysis, Retail Industry, Telecommunication Industry, Biological Data Analysis, Scientific Applications and Intrusion Detection [4]. Supervised learning is the machine learning task of inferring a function from labelled training data. Classification is the data mining technique. It is applied to our real time problems. Classification is the process of classify the data According to the features of the data with predefined set of classes [16]. It is difficult to analyse the large amount of data and make the decision based on that data. To solve this problem, data mining tools and techniques can be used. The weka tool is used for this research problem. Classification technique can be used for prediction of slow learners and fast learners for improving the performance. The student dataset can be used for classification process. The research takes three algorithms or classifiers to solve the research problem. The algorithms are Naïve bayes, Multilayer perceptron and J48. Each algorithm gives the best result for this research process. The J48 algorithm gives high accuracy compared to naïve bayes and multilayer perceptron. The accuracy of the J48 algorithm is 98%.

Keywords: Slow learner, Fast learner, Classifiers, Student Performance, Accuracy, Classification, Naïve bayes, Multilayer Perceptron and J48.

I. INTRODUCTION

Large amount of data can be existing in everywhere. It can be stored in any files or databases and hardcopy. It is difficult work to predicting the some information by human. This

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 17, 2017 ISSN: 2348-6600
Reference ID: IJCS-237 PAGE NO: 1480-1487
Alagappa University, Karaikudi, India 15th -16th February 2017
IT Skills Show & International Conference on Advancements in Computing Resources (SSICACR-2017)
http://aisdau.in/ssicacr ssicacr2017@gmail.com

result is not efficient for making the decision in any applications [5].

To solve this problem, various tools and techniques are used. Data mining provide various techniques for solving various problems. It is defined as the process of finding desired information from large amount of data.

Comparing the result produced by the human and result produced by the tools and technique, the second result will be

More efficient and faster. Collecting the data is first step for our research. Collect the student academic data's such as entry test, cyclic test, assignment, seminar marks. It is useful for finding slow learner and fast learner in a class. Gain the knowledge from experiment results and make the decision for identify the student capacity. Whether the student belongs to slow learner or fast learner category and they get an idea about students learning process. Classification algorithms are naïve bayes, multilayer perceptron and J48. The accuracy of the naïve bayes is 92%. The accuracy of the multilayer perceptron is 80% and the accuracy of the j48 is 98%.

II. PROPOSED SYSTEM

The proposed system is the prediction of students learning process. If the student scored a low marks, they will put into a slow learner group. Otherwise, they will put into a fast learner group. The model will test the training dataset which refers to the student dataset and gives the result.

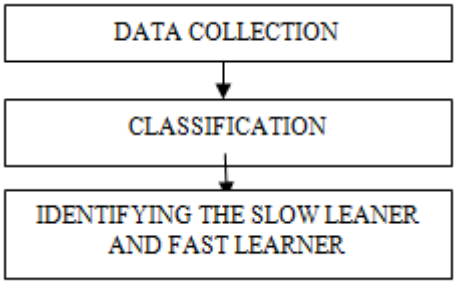

Fig. 1. Architecture of proposed system

2.1 Feature Extraction

Before doing the research, it is necessary to collect the data's regarding to our research problem. The student dataset can be used for this research process [6].

It is also called as training dataset. It contains the set of features with predefined class labels. The features are entry test, CT1(cyclic test), observation, group discussion , decision making , weekly test, home work, oral presentation, seminar, class participation, extra activity, attendance, assignment, quiz, library , sum , average, class label. Based on this information, the result will be produced whether the student is a slow learner or fast learner.

2.2 Training Dataset

The training dataset contains set of attributes with class labels. Here, the training set is student dataset. The class labels will be denoted as a and b. Binary classification is used. It contains 0's and 1's. The one will refers to a and the zero will refers to b. The "a" refers to the fast learner the "b" refers to the slow learner. The training dataset contains two fifty tuples or records. Each tuple contains

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 17, 2017 ISSN: 2348-6600
Reference ID: IJCS-237 PAGE NO: 1480-1487
Alagappa University, Karaikudi, India 15th -16th February 2017
IT Skills Show & International Conference on Advancements in Computing Resources (SSICACR-2017)
http://aisdau.in/ssicacr ssicacr2017@gmail.com

student information. It contains 21 attributes or features. The data set have fifty tuples. Each tuple contains individual student information. Based on this information, class label will be assigned. The classifier is used for classify the tuples and predict the class labels correctly [4].

III. SUPERVISED LEARNING ALGORITHM

Supervised learning is the machine learning task of inferring a function from labelled training data. In supervised learning, each example is a pair consisting of an input object and desired output value. A supervised learning algorithm analyses the training data and produces an inferred function, which can be used for mapping new examples. Classification is the data mining technique. It is applied to our real time problems. Classification is the process of classify the data according to the features of the data with predefined set of classes [16]. Training set contains set of tuples associated with known class labels. It is based on the analysis of a set training data (data objects whose class label is known).

Algorithms

Naïve bayes, Multilayer perceptron, J48 are the classification algorithms. The classification is done by using above three data mining algorithms or classifiers.

In this research, we perform classification using these three algorithms with same training dataset (student dataset).

Each algorithm gives the output with various parameters.

Naïve bayes

The naïve bayes algorithm is a simple probabilistic classifier that calculates a set of probabilities by counting the frequency and combinations of values in a given dataset. The algorithm uses bayes theorem and assumes all attributes to be independent given the value of the class variable.

$$P(H|X) = \frac{P(X|H)\,P(H)}{P(X)}$$

Bayesian classifiers are statistical classifiers. They can predict class membership probabilities such as the probability that a given tuple belongs to a particular class [4].

Multilayer perceptron

It is a Feed forward artificial neural network model .A neural network consists of a set of nodes. Such nodes are input nodes, output nodes and intermediate nodes. Input nodes is used for receive the input signals. Output node gives the output signals. It has unlimited number of intermediate layers contain the intermediate nodes [9].

MLP utilizes a supervised learning technique called back propagation for training the network. A back propagation network has input and output nodes, set of intermediate layers with hidden nodes.

J48

J48 classifier is a simple c4.5 decision tree for classification. And also useful for creating a binary tree. J48 allows classification via either decision trees or rules generated from decision trees [16].

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 17, 2017 ISSN: 2348-6600
Reference ID: IJCS-237 PAGE NO: 1480-1487
Alagappa University, Karaikudi, India 15th -16th February 2017
IT Skills Show & International Conference on Advancements in Computing Resources (SSICACR-2017)
http://aisdau.in/ssicacr ssicacr2017@gmail.com

IV. EXPERIMENT AND RESULTS

The classification was performed using Naïve bayes, Multilayer perceptron and J48 algorithm on student dataset in weka tool.

a. Results of classification using Naïve bayes algorithm

The naïve bayes algorithm is applied to the student dataset. It produces the results. It is called as parameters. The confusion matrix is generated for class label. It has two values (i.e.) yes and no.

The "yes" will refers to "a". The "no" will refer to "b". The fast learner is referred as "a" and the slow learner is referred as "b".

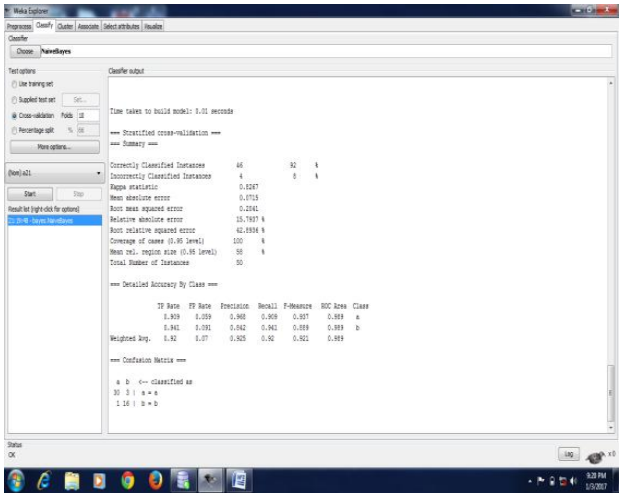

Fig. 2. Naive Bayes

Confusion matrix A confusion matrix illustrates the accuracy of the solution to a classification problem [16] or it is a useful tool for analysing how well your classification can recognize tuples of different classes [4]. Dividing the true-positive by sum of true-positive and false-positive is called as precision. Dividing true-positive by sum of true-positive and false-negative is called as recall. Product of precision and recall divide by sum of precision and recall is called as F-Measures [16].

Precision= true-positive/true-positive +false-positive
Recall=true-positive/true-positive + false-negative
F-Measures = 2* precision *recall / (precision +recall)

True-positive refers to positive tuples that were correctly labelled by the classifier. True-negative refers to negative tuples that were correctly labelled by the classifier. False-positive refers to positive tuples that were incorrectly labelled by the classifier. False-negative refers to negative tuples that were incorrectly labelled by the classifier [4].

Table 1. Class Label Prediction

True-Positive	Yes	Yes
True-Negative	No	No
False-Positive	Yes	No
False-Negative	No	Yes

a. Results of Classification Using Multilayer Perceptron Algorithm

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 17, 2017 ISSN: 2348-6600
Reference ID: IJCS-237 PAGE NO: 1480-1487
Alagappa University, Karaikudi, India 15th -16th February 2017
IT Skills Show & International Conference on Advancements in Computing Resources (SSICACR-2017)
http://aisdau.in/ssicacr ssicacr2017@gmail.com

The Multilayer Perceptron algorithm is applied to student dataset. It produces the results. The confusion matrix is generated for class label. It has two values a and b.

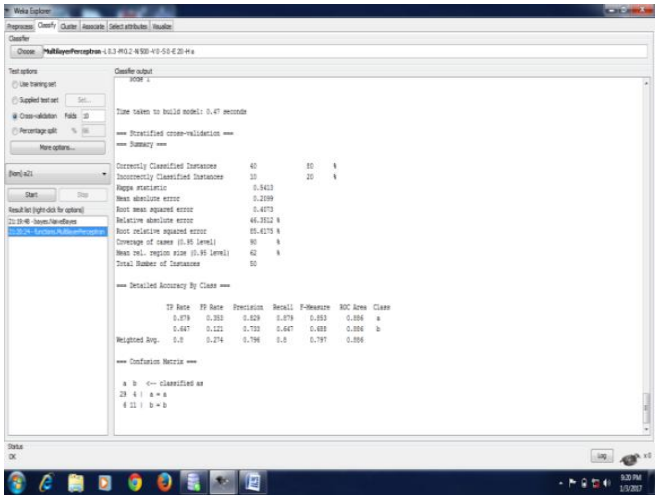

Fig. 3. Multilayer Perceptron

a. Results of classification using J48 algorithm

The J48 algorithm is applied to the student dataset. It produces the results. The confusion matrix is generated for class label. It has two values a and b.

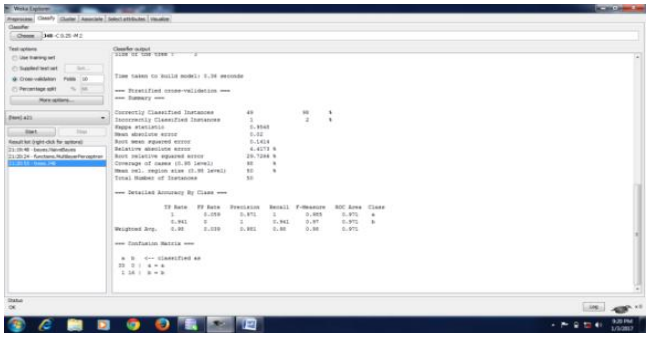

Fig. 4 .J48

TABLE 2. PERFORMANCE OF THE CLASSIFIERS

Parameters	Naïve bayes	Multilayer perceptron	J48
Correctly classified instances (%)	92	80	98
Incorrectly classified instances (%)	8	20	2
Kappa statistic	0.826	0.541	0.954
Mean absolute error	0.071	0.209	0.02
Root mean squared error	0.204	0.407	0.141
Relative absolute error (%)	15.793	46.351	4.417
Root mean squared error (%)	0.204	0.407	0.141
Relative absolute error	15.793	46.351	4.417

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 17, 2017 ISSN: 2348-6600
Reference ID: IJCS-237 PAGE NO: 1480-1487
Alagappa University, Karaikudi, India 15th -16th February 2017
IT Skills Show & International Conference on Advancements in Computing Resources (SSICACR-2017)
http://aisdau.in/ssicacr ssicacr2017@gmail.com

(%)			
Root relative squared error	42.893	85.617	29.726
Coverage of cases (0.95 level)(%)	100	90	98
Mean rel.region Size (0.95 level) (%)	58	62	50
Total number of instances	50	50	50
Time taken to build model(Second)	0.01	0.49	0.01

Correctly a classified instance is the set of tuples that are correctly classified by the classifier. Incorrectly classified instances are the set of tuples that are incorrectly classified by the classifier. Mean absolute error, root mean squared error, relative absolute error and root relative squared error- these are the measures of predictor error. Total number of instances is the set of tuples or records in a dataset[4].

TABLE 3. AVERAGE OF THE CLASSIFIERS

Algorithms	Accuracy
Naïve bayes	92%
Multilayer perceptron	80%
J48	98%

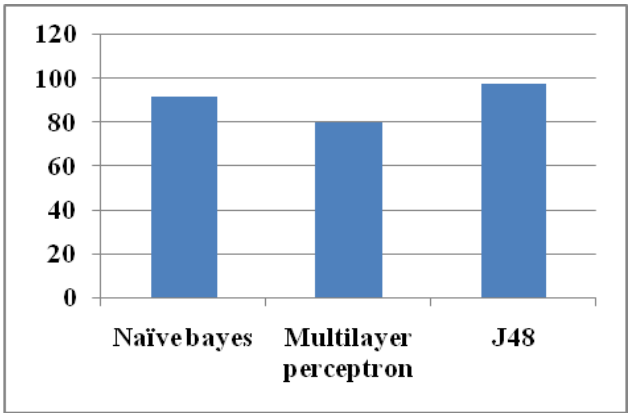

Fig. 5. Prediction Accuracy

V. CONCLUSION

The student dataset has been correctly classified by naïve bayes, multilayer perceptron and J48 algorithms. The classifier was predict the class label a and b. The correctly classified instances will be taken as accuracy of research. After study the experimental results, conclude that each classifier gives best result for our research process. The naïve bayes algorithm gives 92% accuracy and multilayer perceptron gives 80% and The J48

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 17, 2017 ISSN: 2348-6600
Reference ID: IJCS-237 PAGE NO: 1480-1487
Alagappa University, Karaikudi, India 15th -16th February 2017
IT Skills Show & International Conference on Advancements in Computing Resources (SSICACR-2017)
http://aisdau.in/ssicacr ssicacr2017@gmail.com

algorithm gives 98% accuracy. The J48 classifier gives more accurate result for our research. So J48 algorithm gives best result compared to naïve bayes and multilayer perceptron. The analysis of result will help to teachers to identify or category the student learning process for knowing whether they are slow learner or fast learner. And also teachers can take extra effort for improving the academic performance of slow learners. In future different classification techniques can be used for providing efficient result and high accuracy.

References

[1] Azwa Abdul Aziz, Nur Hafieza Ismail, Fadhilah Ahmad, "Mining Students' Academic Performance", Journal of Theoretical and Applied Information Technology, Vol. 53 No.3,31 st July 2013 , ISSN: 1992-8645

[2] Cristóbal Romero, Sebastián Ventura, Pedro G. Espejo and César Hervás, "Data Mining Algorithms to Classify Students"

[3] A.A. Aziz, N. H. Ismail, & F. Ahmad, "Mining Students' Academic Performance", Journal of Theoretical and Applied Information Technology, vol. 53(2013), no. 3, 485–495.

[4] Han, J. and Kamber, M., "Data Mining: Concepts and Techniques", 2nd edition. The Morgan Kaufmann Publishers, 2006.

[5] Jigna Ashish Patel , "Classification Algorithms and Comparison in Data Mining" , International Journal of Innovations & Advancement in Computer Science, IJIACS, Volume 4, May 2015, ISSN 2347 –8616

[6] Ms. Priti S. Patel, Dr.S.G.Desai,"Various Data Mining Techniques used to Study Student's Academic Performance", International Journal of Computer Science and Mobile Applications, Vol.3 Issue. 6, June-2015, pg. 55-58 ISSN: 2321-8363

[7] Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz, "The Prediction of Students' Academic Performance Using Classification Data Mining Techniques", Applied Mathematical Sciences, Vol. 9, 2015, no. 129, 6415-6426

[8] Brijesh kumar bhardwaj and saurabh pal , "Data mining : A prediction for performance improvement using classification", (IJCSIS) International Journal of Computer Science and Information Security, vol. 9 , no. 4, April 2011

[9] Pieter Adriaans and Dolf Zantinge, Data Mining, 1999, ISBN: 981-235-966-4.

[10] Pooja Thakar, Anil Mehta, Manisha, "Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue", International Journal of Computer Applications (0975 – 8887), Volume 110 – No. 15, January 2015.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 17, 2017 ISSN: 2348-6600
Reference ID: IJCS-237 PAGE NO: 1480-1487
Alagappa University, Karaikudi, India 15th -16th February 2017
IT Skills Show & International Conference on Advancements in Computing Resources (SSICACR-2017)
http://aisdau.in/ssicacr ssicacr2017@gmail.com

[11] DR. A. PADMAPRIYA, "PREDICTION OF HIGHER EDUCATION ADMISSIBILITY USING CLASSIFICATION ALGORITHMS" ,VOLUME 2, ISSUE 11,NOVEMBER 2012, ISSN: 2277 128X

[12] JASVINDER KUMAR, "A COMPREHENSIVE STUDY OF EDUCATIONAL DATA MINING", INTERNATIONAL JOURNAL OF ELECTRICAL ELECTRONICS & COMPUTER SCIENCE ENGINEERING ,SPECIAL ISSUE –TELMISR, 2015, ISSN : 2348-2273

[13] MRS. M.S. MYTHILI, DR. A.R.MOHAMED SHANAVAS,"AN ANALYSIS OF STUDENTS' PERFORMANCE USING CLASSIFICATION ALGORITHMS", IOSR JOURNAL OF COMPUTER ENGINEERING (IOSR-JCE), VOLUME 16, VER. III (JAN.2014), E-ISSN: 2278-0661, P-ISSN: 2278-8727, PP 63-69

[14] SONALI AGARWAL, G. N. PANDEY, AND M. D. TIWARI,"DATA MINING IN EDUCATION: DATA CLASSIFICATION AND DECISION TREE APPROACH", INTERNATIONAL JOURNAL OF E-EDUCATION, E-BUSINESS, E-MANAGEMENT AND E-LEARNING, VOL. 2, NO. 2, APRIL 2012

[15] DORINA KABAKCHIEVA,"STUDENT PERFORMANCE PREDICTION BY USING DATA MINING CLASSIFICATION ALGORITHMS ",INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND MANAGEMENT RESEARCH,VOL 1 ,ISSUE 4 NOVEMBER 2012 , ISSN : 2278-733X

[16] TINA R. PATIL, MRS.S. S. SHEREKAR,"PEFORMANCE ANALYSIS OF NAÏVE BAYES AND J48 CLASSIFICATION ALGORITHM FOR DATA CLASSIFICATION", INTERNATIONAL JOURNAL OF COMPUTER SCIENCE AND APPLICATIONS, VOL.6, NO.2, APRIL 2013, ISSN: 0974-1011