International Journal of Computer Trends and Technology
ISSN: 2231-2803 / https://doi.org/10.14445/22312803/IJCTT-V68I2P108

Volume 68 Issue 2, 54-58, February 2020

Original Article

Classification and Prediction of Slow Learners Using Machine Learning Algorithms

Sangeeta. K[1], G.V.S.S. Naveen Babu[2], Madhuri. G[3]

[1]*Assistant Professor, Department of CSE, Aditya Institute of Technology and Management, Tekkali, India*
[2,3]*B.Tech 3rd yr student, Department of CSE, AITAM, Tekkali, Andhra Pradesh, India*

Received Date: 09 January 2020
Revised Date: 17 February 2020
Accepted Date: 23 February 2020

Abstract *- Any educational Institute's main goal is to increase the pass percentage of the students. A student's performance depends on his learning ability and is influenced by many factors. A slow learner grasps things lately requires things to be explained with many detailed resources to be successful as compared to a fast learner. As the competitive world demands more out of a student with respect to an all-around development, student classification based on learning ability is useful in predicting slow learners. The slow learners will be given appropriate training to improve /her performance and thereby achieve the institute's goal. This paper uses real-time student data of the computer science engineering department, Aditya Institute Of Technology and Management, Tekkali in Srikakulam district. The study involves experiments to understand the influence of cognitive attributes on academic performance. The classification of Students into very fast learners, fast learners, average learners, and slow learners using classification algorithms and thereby finding out the best prediction model. The proposed paper accommodates the individual differences of the learners in terms of knowledge level, learning preferences, cognitive abilities etc*

Keywords *- Classification, prediction, Slow learner.*

I. INTRODUCTION

Student's performance prediction is a tricky task as it depends on many factors such as grades, class performance, demographic data, emotional features and many more. It is important for the teachers to forecast the future performance of a student based on his past performances, identifying weak students at an early stage so that additional material and special attention can be facilitated to avoid the risk of failure. We know that identifying slow learners and their progress tracking is essential to reduce the failure rates.

II. RELATED WORK

M. Ramaswamy and R.S. Bhaskaran [1] the CHAID expectation model was valuable to break down the interrelation between factors that are utilized to anticipate the result on the presentation at higher auxiliary school instruction. The highlights like a vehicle of guidance mark gotten in auxiliary instruction, area of the school, living zone and kind of optional instruction were the most grounded markers for the understudy execution in higher auxiliary instruction. This CHAID forecast model of understudy execution was built with seven class indicator variables, though the prior models in audits were developed with restricted class indicator factors. To improve the performance of a student, Nguyen Thi-The, Andre Busche and Lars Schmidt-Thiem [2] use some machine learning techniques like Support Vector Machines (SVM), Bayesian Networks (BN) and Decision Tree (DT) on the real data sets. To resolve a set imbalance problem, first, we have to resolve for even datasets worn dedicated and cost-sensitive practice. Among all these machine learning algorithms, SVM and DT give a useful result for a short and huge data set. Praneet Kaur, Manpreet Singh, and Gurpreet Singh Josan[3] collected student's data from high schools all over the world. Among all machine learning algorithms, Multi-Layer Perception gives 75% accuracy, and thus MLP is the best-suited algorithm for classification. Comparative results for the five classifiers with the WEKA experimenter is also proved MLP to be the best with an F-measure of 82%. Cortez and Silva [4] Classification techniques are utilized as prescient devices in the information mining area and in the wake of looking at their exhibitions. From the outcomes, it is demonstrated that Multi-Layer Perception calculation is generally suitable for anticipating understudy execution. Random forest gives a 75% forecast, which is generally higher than different calculations. This examination is an endeavour to utilize grouping calculations for foreseeing the understudy execution and contrasting the exhibition of Naïve Bayes Simple, Multi-Layer Perception, SMO, J48, and REPTree were applied to the data set consisting of 788 records of students who participated in the 2006 examination. Galit proposed a case study to predict outcomes for student data and alert students to risk before their final results. V.Ramesh, P.Thenmozhi and Dr.K. Ramar [5] uses REPTree, J48, Multi-Layer Perception, SMO and Naive Bayes to identify the factors which lead to reduced student's academic performances. Among all these algorithms, MLP gives a 69.5% result.

Table 1. shows the summary of the research review.

Table 1. Research Review

S.NO	Author	Algorithm used	Result
1	M. Ramaswamy and R.S. Bhaskaran	CHAID	59.4%
2	Nguyen Thi-Nhe, Andre Busche and Lars Schmidt-Thiem	SVM	80%
3	Praneet Kaur,Manpreet Singh and Gurpreet Singh Josan	Multilayer Perception	75%
4	Cortez and Silva	RandomForest	75%
5	V.Ramesh	REPTree, J48, Multi-Layer Perception, SMO, and Naive Bayes	54.4%,65.8%,69.5%,58.5% and 55.8%,

III. PROPOSED MODEL

In this research, we prepared a set of questionnaires that analyses the cognitive abilities of engineering students. The questionnaires train and test on analytical thinking, errors and misconceptions, decision making, knowledge level etc., of a student and constitutes the raw data. The static information such as academic performance, gender, demographic data, learning preferences and class performance is collected and converted into *.CSV format for preprocessing and then converted to AIFF format. A total of 314 records were collected to analyse the student's performance and classified into very fast learners, fast learners, average learners, and slow learners. We used the weka tool for data preprocessing and statistical analysis of various machine learning algorithms. The preprocessed data is inputted to algorithms such as decision tree, Naïve Bayes, k-NN, SVM, C4.5, algorithms for the statistical output. The results of various algorithms are compared using parameters precision, recall, score and accuracy and performance analysis is displayed using the ROC curve.

A. System Architecture

It consists of the following components:

- Raw Data collection: Real-time Data is collected from engineering students through questionnaires.
- Data Preprocessing: Data is normalised and converted into AIFF format.
- Statistical analysis of algorithms: The statistical output of various algorithms are analysed.
- Comparisons of algorithms: The results of various algorithms are compared on parameters precision, recall, score etc.

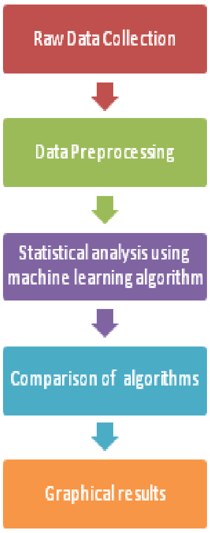


Fig. 1 System Architecture

B. Feature Set

The features of a student considered for performance analysis are demographic, physical, cognitive and learning preferences.

Table 2. Feature set and description

Features	Description
Gender	Male/Female
class_performance	Student's active performances in the Class
CGPA	Academic performance
urban/rural	Demographic feature
Analytical Thinking	Analytical ability
Knowledge Level	Understanding and remembering concepts
Problem Solving Skills	Programming knowledge
Decision Making	Deciding given the information
errors	Error identification/misconception

IV. RESULTS

We experimented with various algorithms and compared the statistical results. The Weka tool supports various machine learning algorithms such as Naïve Bayes, Support Vector Machine, J48, KStar, Random Forest and Random Tree and the results thus obtained are compared. The parameters used for comparison are FPRate, TPRate, Recall, F-Measure, Precision, MCC and ROC Area. The confusion matrix and ROC curves also prove to be useful for comparison.

A. Naïve Bayes model for prediction of student's performance

Out of 313 instances,266 instances(84.984%) are correctly classified and 47 instances are incorrectly classified(15.016%). Table 3 shows the accuracy achieved for each Class.

Table 3.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	Class
0.852	0.096	0.838	0.852	0.845	0.754	0.949	fast
0.476	0.007	0.833	0.476	0.606	0.612	0.983	very_fast
0.937	0.140	0.821	0.937	0.875	0.785	0.951	avg
0.780	0.000	1.000	0.780	0.876	0.865	0.995	slow
0.850	0.092	0.850	0.846	0.846	0.775	0.959	Weighted Avg.

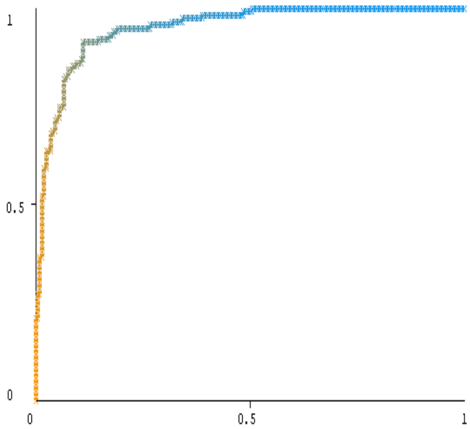
The confusion matrix is given in Table 4. Out of 115 fast instances,98 are classified as fast, and 2,15 are wrongly classified as very_fast and avg, respectively. Out of 21 very_fast instances,10 are classified as very_fast, and 11 are wrongly classified as fast. Out of 127 avg instances,119 are classified as avg, and 8 are wrongly classified as fast. Out of 50 slow instances,39 are classified as slow, and 11 are wrongly classified as avg, thereby giving an accuracy of 85%.

Table 4. Confusion Matrix

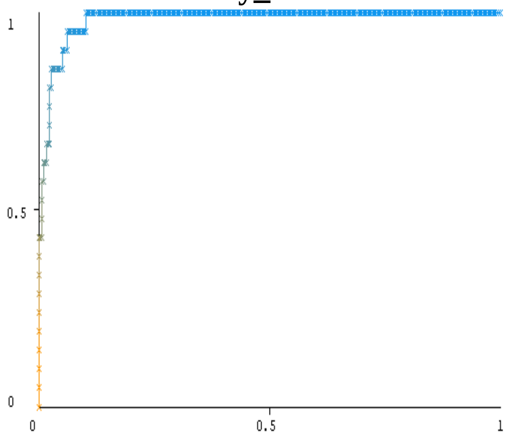
	fast	**very_fast**	**avg**	**slow**
fast	98	2	15	0
very_fast	11	10	0	0
avg	8	0	119	0
slow	0	0	11	39

The ROC curve is given below :

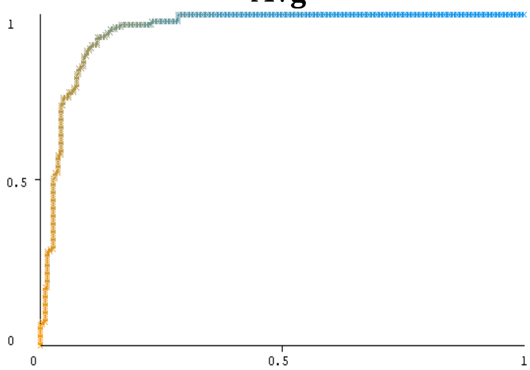
Fast



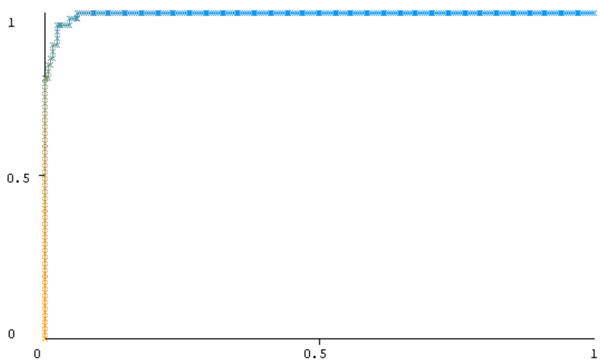
very_fast



Avg



slow



A. J48 model for prediction of student's performance

Out of 313 instances,280 instances(89.4569%)are correctly classified and 33 instances are incorrectly classified(10.5431%). Table 5 shows the accuracy achieved for each Class

Table 5.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	Class
0.904	0.066	0.889	0.904	0.897	0.836	0.965	fast
0.619	0.003	0.929	0.619	0.743	0.745	0.972	very_fast
0.961	0.102	0.865	0.961	0.910	0.847	0.955	avg
0.820	0.000	1.000	0.820	0.901	0.890	0.977	slow
0.895	0.066	0.900	0.895	0.893	0.843	0.963	Weighted Avg.

The confusion matrix is given in Table 6. Out of 115 fast instances,104 are classified as fast, and 1,10 are wrongly classified as very_fast and avg, respectively. Out of 21 very_fast instances,13 are classified as very_fast, and 8 are wrongly classified as fast. Out of 127 avg instances,122 are classified as avg, and 5 are wrongly classified as fast. Out of 50 slow instances,41 are classified as slow, and 9 are wrongly classified as avg, thereby giving an accuracy of 89.5%.

Table 6. Confusion matrix

	fast	very_fast	avg	slow
fast	104	1	10	0
very_fast	8	13	0	0
avg	5	0	122	0
slow	0	0	9	41

B. Support Vector Machine model for prediction of student's performance

Out of 313 instances,280 instances(89.4569%)are correctly classified and 33 instances are incorrectly classified(10.5431%). Table 7 shows the accuracy achieved for each Class

Table 7.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	Class
0.913	0.086	0.861	0.913	0.886	0.818	0.936	fast
0.286	0.000	1.000	0.286	0.444	0.521	0.938	very_fast
0.984	0.086	0.887	0.984	0.933	0.886	0.952	avg
0.880	0.000	1.000	0.880	0.936	0.928	0.989	slow
0.895	0.066	0.903	0.895	0.883	0.843	0.951	Weighted Avg.

The confusion matrix is given in Table 8..Out of 115 fast instances,105 are classified as fast, and 10 are wrongly classified as avg. Out of 21 very_fast instances,6 are classified as very_fast, and 15 are wrongly classified as fast. Out of 127 avg instances,125 are classified as avg, and 2 are wrongly classified as fast. Out of 50 slow instances,44 are classified as slow, and 6 are wrongly classified as avg, thereby giving an accuracy of 89.5%.

Table 8. Confusion matrix

	fast	very_fast	avg	slow
fast	105	0	10	0
very_fast	15	6	0	0
avg	2	0	125	0
slow	0	0	6	44

C. *k-NN model for prediction of student's performance*

Out of 313 instances,all instances(100%)are correctly classified and no instances are incorrectly classified(0%).Table 9 shows the accuracy achieved for each Class.

Table 9.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	fast
1.000	0.000	1.000	1.000	1.000	1.000	1.000	very_fast
1.000	0.000	1.000	1.000	1.000	1.000	1.000	avg
1.000	0.000	1.000	1.000	1.000	1.000	1.000	slow
1.000	0.000	1.000	1.000	1.000	1.000	1.000	Weighted Avg.

The confusion matrix is given in Table .10.

Table 10. Confusion matrix

	fast	very_fast	avg	slow
fast	115	0	0	0
very_fast	0	21	0	0
avg	0	0	127	0
slow	0	0	0	50

V. CONCLUSION

The k-NN model for student's learning has 100%accuracy compared to other algorithms and can be used to predict the performance of students for the quick action and remedial measures. This will help in reducing the failure rates.

REFERENCES

[1] M.Ramaswamy and R.S.Bhaskaran, A CHAID Based Performance Model in Educational Data Mining, International Journal of Computer Science Issues(IJCSI), 7(1) (1) (2010).

[2] Praneet Kaur, Manpreet Singh and Gurpreet Singh Josan, Classification and Prediction based data mining algorithms to predict slow learners in the educational sector 3rd International Conference on Recent Trends in Computing (ICRTC). (2015).

[3] Cortez and Silva, Using data mining to predict secondary school student 5th Future Business Technology Conference(FUBUTEC). (2008).

[4] Ramesh, P.Thenmozhi and Dr.K. Ramar, Study of influencing factors of academic performance of students: A data mining Approach. International Journal of Scientific & Engineering Research. 3(7) (2012).

[5] R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, High-speed digital-to-RF converter, U.S. Patent. 5(16) (1997) 668-842.