International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-6, August 2019

Student Learning Prediction Using Machine Learning Techniques

Anbukarasi V, A. John Martin

Abstract: ***Now a day's e-learning is smartly growing technology. This technology is more helpful for students to communicate with their professors through chats or emails. E-Learning also removes the obstacle of physical presence of an E-learner. The main aim of this paper is to predict student performance in their final exams using different machine learning techniques. Information like attendance, marks, assignments, class participation, seminar, CA, projects and semester are collected to predict student performance. This prediction helps the instructors to analyze their students based on their performance. For that we have used WEKA tool for the prediction of the student performance. WEKA (Waikato Environment for Knowledge Analysis) is one of the data mining too which is used for the classification and clustering using data mining algorithms. This prediction helps the students and the staffs to know how much effort their students need to be put in their final exams to get good marks.***

Keywords- E-learning, Data Mining, Machine Learning algorithms, Data Preprocessing

I. INTRODUCTION

The main aim of educational organizations is to provide better and excellence education to their students. The best way to achieve better quality in education is by determining the understanding of students in a particular course. At present, in educational institutions, data mining plays a vital role and it is the most essential area of research with the motive of finding significant data from knowledge keep in huge datase t. Educational DataMining (EDM) could be a main analysis space to predict helpful data from educational information to boost the standard of educational performance, to know higher and to own better assessments of students' learning. EDM may be taken mutually of the training sciences and as a field of information mining.

EDM can be applied in modelling the knowledge, behavior and experience of the user. Data Mining or knowledge discovery has emerged as an important area to analyze data from different perspectives and also helps to summarize it into useful information. Educational Data Mining uses several techniques like call trees, neural networks, naive Bayes, support vector machines and lots of others. WEKA is one of the open source software intended for data analysis and knowledge discovering. It implements a wide range of machine learning algorithms and is widely used in data mining applications. The main aim of this paper is to predict student's performance using data mining algorithms and to find the highest accuracy between the classifiers. The paper is ordered as follows. The second part of the paper represents literature review and similar research. The third part presents the datasets and description. The fourth section represents the data preprocessing, experiments and discussions. Fifth section of the paper presents conclusions and scope for the future research. The references are presented in the last section of the paper.

Revised Manuscript Received on August 30, 2019.
* Correspondence Author
Anbukarasi V*, Research Scholar, Department of Computer Applications, Sacred Heart College, Tirupattur, India.
A. John Martin, Department of Computer Applications, Sacred Heart College, Tirupattur, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

II. LITERATURE REVIEW

The researchers have referred various papers that uses different methods and methodologies to analyze and predict student results.

1. E-learning student assessment by means of evaluation strategy [1]: In this research article, they have assessed and understood the student performance using metacognitive technique. The study comprises the assembling of model and explaining mechanism. In this paper they have proposed Online Learning Stratagem. The area of perceived-skill line of attack, particularly the understanding strategies is one of the chief drawbacks of this research.
2. Analyzing student performance using data mining techniques [2]. In this paper, they have used number of parameters to predict the student performance based on their academic performance. The authors have used different assessment techniques to evaluate the student performance. The focus of this research is to use hidden Markov model to grades the students, which has used totally different learning power and substitute assessment technique. Consistent with the experimental analysis, the new evaluating strategy is considerably useful in evaluating performance of scholars.
3. Evaluation of student's performance by means of Data Mining Techniques [3], the analysis proposes technique accustomed assess the coed performance mistreatment varied prediction indicators. The most focus is to predict the students' performance. The values obtained during this analysis is showed with the intention of information mining technique available trained analyze tools for learner knowledgeable performance with 94 % success analysis.

4. Early prediction of student performance using machine learning algorithms[4]: in this paper, they have used some undergraduate college students' details to predict the performance. The main objective of this paper is to predict weak and strong students, so that some form of remediation may be organized for them. And the future selection algorithms were applied on the datasets to reduce the number of features. Five classes of machine learning algorithm were trained and then tested.
5. Machine learning approach to predict student academic performance [5]: in the education research study various aspects like social, economic, personal, cultural, geographical, institute environment and others are evaluated. Such aspects may either help a student to shine during the academic period or halt academic program. Such failure is known as drop-out and data mining algorithm helps in finding the factors that mostly contribute to the student's performance. In this research machine learning approach or machine learning classification models are to be used, to predict students' academic performance. This is fit for Educational data mining. In this research study, the students' performance are predicted and the results of these prediction models are compared. This research also discusses how these machine learning models can help to improve an education system by considering the different factors in terms of accuracy, specificity, precision, prevalence, recall, F-Measure etc., in results. In addition to that, this research is carried on by applying Standard and Hybrid approaches as well as fuzzy logic and compared outcomes. This research also helps in predicting the performance of other state students by identifying the appropriate factors. The main aim of this research is to predict the performance of students and to identify the most appropriate model for that; so that on the basis of that we would be able to take the appropriate steps to improve their performance and to cut down the dropout rate.

III. DATASETS AND DESCRIPTION

Student Data Set consist of 1100 records of students with their Attendance, marks, assignments, class participation, seminar, CA, projects and semester. The datasets were split into two parts one is trained data and the test data, trained data consist of 1000 data's and the test data consist of 100 data in a datasets. The following table shows attribute with their descriptions and values.

Table-I: Dataset with the dataset description

ATTRIBUTE	DESCRIPTION	VALUES
Attendance	Student attendance	P>60%, F<60%
Assignment	Assignment	P>=3, F<3
Class Participation	Class participation of the students in the class.	P>=3, F<3
Seminar	Performance of the seminar.	P>=3, F<3
CA	Internal exam marks.	P>=3, F<3
Project	Project	P>=3, F<3
Semester	Semester marks	P>=35, F<35

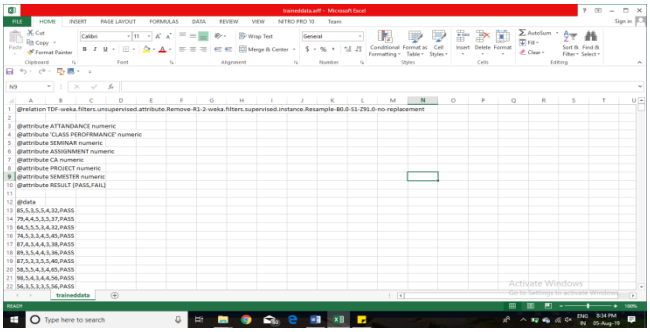

Fig. 1. Screenshot of trained dataset

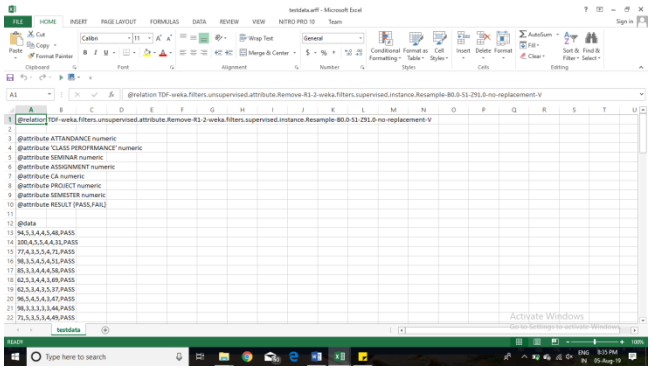

Fig. 2. Screenshot of Test dataset

IV. DATA PREPROCESSING

WEKA may be a data processing tool developed by Department of engineering, University of Waikato, New Sjaelland. It's associate degree open supply code provide out below the overall Public License (GNU). Rail is cluster of machine learning algorithms for the data mining tasks. These algorithms is applied on to the info or referred to as from the Java code. The pre-processing is an important step that is used to extract and improve the quality of data. WEKA tool import dataset from a proper file like attribute relation file format which is the preferable one. The below figures shows that the output of data pre-processor and data model visualization respectively.

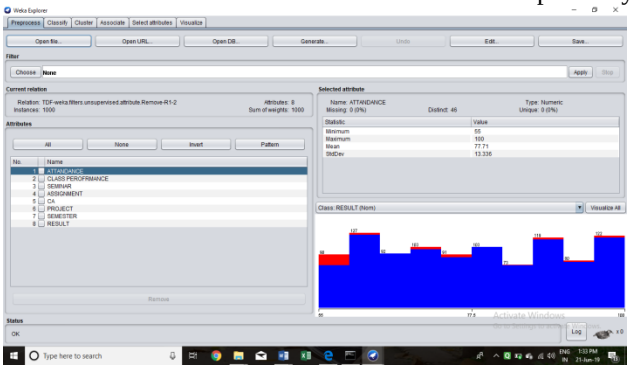

Fig. 3. Model Visualization of a data

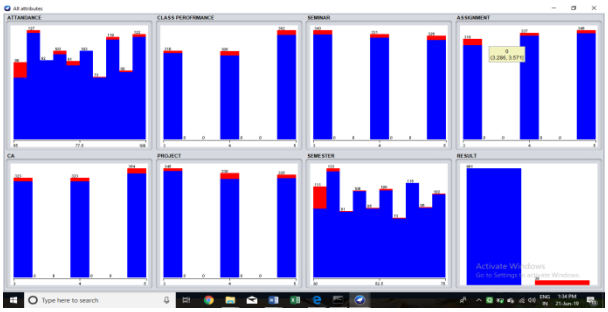

Fig. 4. Data Pre-Processor Output

a. EXPERIMENTAL RESULTS AND DICUSSION

A comparative analysis has been done on set of specified dataset using some machine learning algorithms. Those algorithms are BayesNet, J48, Random Forest, Navie Bayes, Random Tree, and Decision Stump, Hoeffding Tree and LMT. Each algorithms has its own characteristics to classify the dataset. Following table shows the predicted results of all classifiers.

Table-II: classifiers with the results

Criteria	Classifiers								
	BayesNet	***J48***	***Random Forest***	***Navie Bayes***	***Random Tree***	***Decision Stump***	***REPTree***	***Hoeffding Tree***	***LMT***
Accuracy %	100	100	100	94	100	96	100	94	100
Correctly Classified Instance	100	100	100	94	100	96	100	94	100
Incorrectly Classified Instance	0	0	0	6	0	4	0	6	0

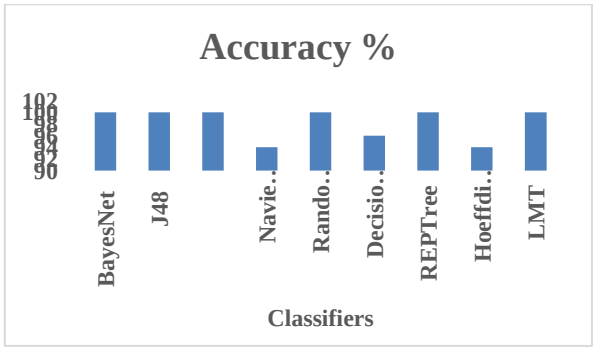

Fig .5.Accuracy of the classifiers results

The below two graphs shows the analysis of all classifiers based on the RMSE And RRSE It shows that the J48 and REPTree has the minimum error values among other classifiers.

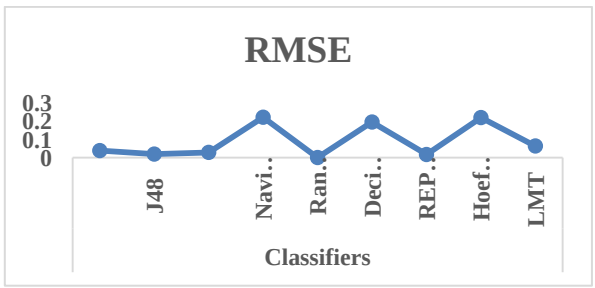

Fig.6. Root mean square error value

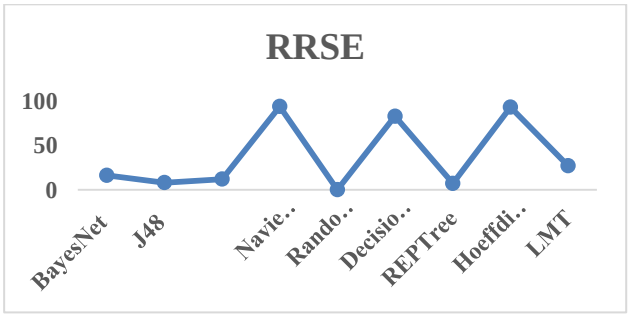

Fig. 7. Root relative square error value

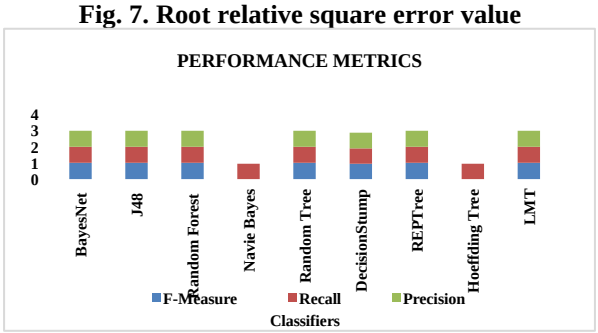

Fig.8. value of performance metrics

The above chart shows that the performance accuracy of the five classifiers based on different classification metrics.

These metrics are; Precision, Recall and F-measure measure are very important to determine the classifiers based on the accuracy. These metrics shows that Bayesian Network classifier performs better than other classifiers.

a. COMBINATION OF TWO ALGORITHMS USED FOR THE PREDICTION

Here we have combined two algorithms for the prediction of student's performance. And the below table shows that the J48 with REPTree, BayesNet with LMT and LMT with Hoeffding Tree produced best result compared to others and the result shows that these combination of algorithms are the best classifiers**.**

Table-III Result Of Combined Classifiers

Criteria	Classifiers						
	BayesNet with NavieBayes	***J48 with REPTree***	***Random Forest with Random Tree***	***Random Forest with Navie Bayes***	***J48 With Random Forest***	***BayesNet with LMT***	***LMT with Hoeffding Tree***
Accuracy %	97	100	100	97	100	100	96
Correctly Classified Instance	97	100	100	97	100	100	96
Incorrectly Classified Instance	3	0	0	3	0	0	4

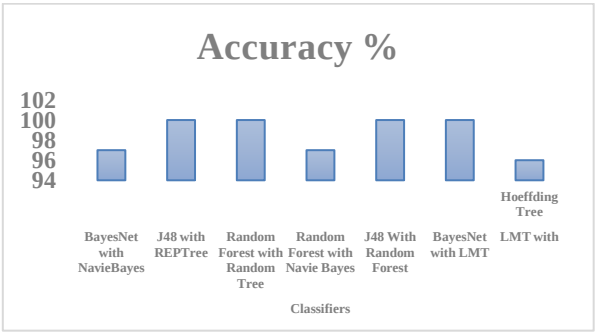

Fig. 9. Accuracy of the classifiers results

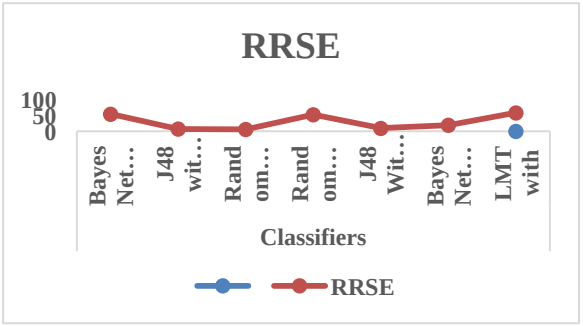

Fig.11. Root mean square error value

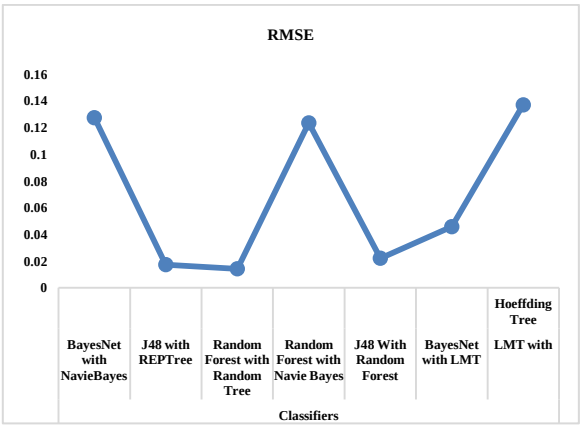

Fig.10. Root relative square error value

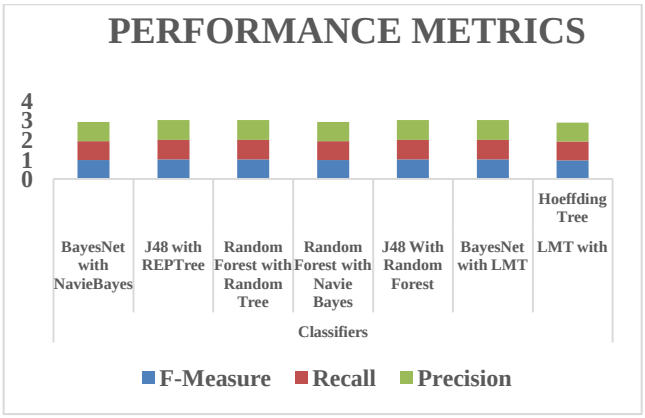

Fig.12. shows the performance metrics of the classifiers

The above graph shows the accuracy of the combined classification algorithms. And the below two graphs shows the analysis of all classifiers based on the RMSE and RRSE. It shows that the Random Forest with Random Tree combination has the minimum error values among other classifiers. discovers the likelihood to forecast the success rate of the scholars within the specific course mistreatment machine learning algorithms. From the above discussion, we can conclude that Bayesian Network, J48, Random Forest, REPTree and LMT algorithms for predictive modelling gave us the best result as compared to the Decision Stump,

Naive Bayes, and Hoeffding Tree using single algorithms. And also we have used combination of two algorithms to predict student academic performances, in that we have got better results. J48 with REPTree, Random Forest with Random Tree, J48 with Random Forest, and Bayes Net with LMT these combination of the classifiers gives 100% result. With this paper, we try to find out the best result with only selected attributes. This paper helps other researchers to find out which are the algorithms are best to predict student performance. In our future work, we try to implement our work on the different dataset with different classification algorithm. By doing this we can able to find the best algorithm which is going to perform almost perfect on every dataset in hand.

V. CONCLUSION

Student evaluation is one of the important task in instructional data processing. This prediction discovers the likelihood to forecast the success rate of the scholars within the specific course mistreatment machine learning algorithms. From the above discussion, we can conclude that Bayesian Network, J48, Random Forest, REPTree and LMT algorithms for predictive modelling gave us the best result as compared to the Decision Stump, Naive Bayes, and Hoeffding Tree using single algorithms. And also we have used combination of two algorithms to predict student academic performances, in that we have got better results. J48 with REPTree, Random Forest with Random Tree, J48 with Random Forest, and BayesNet with LMT these combination of the classifiers gives 100% result. With this paper, we try to find out the best result with only selected attributes. This paper helps other researchers to find out which are the algorithms are best to predict student performance. In our future work, we try to implement our work on the different dataset with different classification algorithm. By doing this we can able to find the best algorithm which is going to perform almost perfect on every dataset in hand.

REFERENCES

1. Meng-Jung Tsai, “The Model of Strategic e-Learning: Understanding and Evaluating Student eLearning from Metacognitive Perspectives”, 2009.
2. Sandeep Roy, Anchal Garg “Analyzing performance of students by using data mining techniques” 2017, 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)
3. Mukesh Kumar, Prof.A.J. Singh “Evaluation of Data Mining Techniques for Predicting Student’s Performance”, I.J. Modern Education and Computer Science, 2017, 8, 25-31 Published Online August 2017 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijmecs.2017.08.04
4. Anal Acharya, Devadatta Sinha, “Early Prediction of student performance using machine learning technique”, International Journal of Computer Applications, December 2014.
5. Amandeep Kaur, Nitin Umesh, Barjinder Singh, “Machine learning approach to predict students’ academic performance”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), April 2018.
6. S. B. Kotsiantis, “Use of machine learning techniques for educational proposes: a decision support system for forecasting students grades”,Artificial Intelligence Review, 37(4):331–344, 2012.
7. C. Vera, A.Cano, C. Romero,S. Ventura, “Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data”, DOI 10.1007/s10489-0120374-8
8. P. Cortez and A. Silva, “Using Data Mining to Predict Secondary School Student Performance”, In A. Brito and J. Teixeira Eds., Proceedings of 5th FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 512, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7.
9. A. Azevedo and M. F. Santos, “KDD, SEMMA and CRISP-DM: a parallel overview”, In Proceedings of the IADIS European Conference on Data Mining 2008, pp 182-185. Archived January 9, 2013, at the Wayback Machine.
10. S. S. Rohanizadeh, M.B. Moghadam, “A Proposed Data Mining Methodology and its Application to Industrial Procedures”, Journal of Industrial Engineering 4 (2009) pp 37-50
11. S. Gabrilson, D. Fabro, P. Valduriez, Towards the efficient development of model transformations using model weaving and matching transformations, Software and Systems Modeling 2003. Data Mining with CRCT Scores, Office of information technology, Geogia Department of Education.
12. K. Umamaheswari, S. Niraimathi “A study on student data analysis using data mining techniques”, International Journal of Advanced Research in Computer Science and Software Engineering, vol. 3, Issue 8, August 2013, pp. 117-120.
13. O. Oyelade, O. Oladipupo, I. Obagbuwa, “Application of k-Means clustering algorithm for prediction of students’ academic performance” (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, num. 1, 2010, pp. 292-295.
14. S. Aher, L.M.R.J. Lobo, “Data mining in educational system using weka”, International Conference on Emerging Technology Trends, 2011, pp. 20-25.
15. R. Kirkby, E. Frank, Weka explorer user guide for version 3-4-3, The University of Waikato, 2004, pp. 1-13.
16. Edin Osmanbegovic, Mirza Suljic, 2012. Data mining approach for predicting student performance, Economic Review – Journal of Economics and Business, vol. 10, issue 1.
17. Neelam Naik, Seema Purohit, 2012. Prediction of Final Result and Placement of Students using Classification Algorithm, International Journal of Computer Applications, vol. 56, no. 12, pp. 35-40.

AUTHORS PROFILE:

Anbukarasi V has completed her M.C.A and is perusing her M.Phil in Computer Applications. Presently she is working as an Assistant Professor in Arignar Anna College of Arts and Science, Krishnagiri. Her research area includes Machine Learning and Block chain technologies. Her passion is to further her research in Block chain technology and contribute to the advancement of humanity.

A. John Martin holds M.C.A., M.Phil. and B.Ed in Computer Science. He has great passion towards teaching profession. Presently he is working for MCA department as Assistant Professor. He is a part-time Ph.D Scholar and he is currently in the second year of his PhD in Computer Science at Sacred Heart College, Tirupattur. His current research area is in Artificial Intelligence using Machine Learning Techniques. He has published research articles in National and International Journals.