International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 1 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT195167

A Review on Various Algorithms for Student Performance Prediction

Mayuri S. Dongre[1], Neha S.Vidya[1], Vanita D. Telrandhe[1], Shweta R. Chaudhari[1], Anup A. Umrikar[1], Prof. Prachiti V. Adghulkar[2]

[1]BE Students, Department of computer Science & Engineering, Suryodaya collage of Engineering & Technology, Nagpur, Maharashtra, India

[2]Assistant Professor, Department of computer Science & Engineering, Suryodaya collage of Engineering & Technology, Nagpur, Maharashtra, India

ABSTRACT

Data in educational institutions are growing progressively along these lines there is a need of progress this tremendous data into helpful data and information utilizing data mining. Educational data mining is the zone of science where diverse techniques are being produced for looking and investigating data and this will be valuable for better comprehension of understudies and the settings they learned. Classification of data objects in view of a predefined learning of the articles is a data mining and information administration procedure utilized as a part of collection comparable data questions together. Decision Tree is a valuable and well known classification method that inductively takes in a model from a given arrangement of data. One explanation behind its prominence comes from the accessibility of existing calculations that can be utilized to assemble decision trees. In this paper we will survey the different ordinarily utilized decision tree calculations which are utilized for classification. We will likewise contemplating how these decision tree calculations are appropriate and valuable for educational data mining and which one is ideal.

Keywords : Educational data mining (EDM), Classification, Dropout Prediction, Selection Failure

I. INTRODUCTION

As of late the scientist concentrates on the new region of research, which is EDM (Educational Data mining). EDM is the information disclosure in database strategy or the data mining in instruction. The specialist concentrates on the improvement of technique to better comprehend understudies and the settings in which they learn. There are great cases of how to apply EDM systems to make models that foresee dropping out and understudy disappointment. These works have indicated promising outcomes regarding those sociological, monetary, or educational qualities that might be more pertinent in the forecast of low scholarly execution.

The data in any given educational association is growing rapidly. There is a need to change this data into helpful data and learning; henceforth we make utilization of data mining.

Distinguish and find valuable data covered up in substantial databases is a troublesome errand. An exceptionally making an assurance to answer for accomplish this goal is the utilization of learning revelation in databases techniques or data mining in training, called Educational Data Mining, EDM [15]. Educational data mining techniques drawn from an assortment of writings, including data mining and machine learning, psychometrics and different ranges of measurements, data perception, and computational displaying.

As we examined over the new territory of center for the specialists is EDM. In introduce the vast majority of the procedures was produced for the educational data mining, which foresee the disappointment and dropout understudies. Be that as it may, a large portion of the examination on the use of EDM to determine the issues of understudy disappointment and drop-outs has been connected principally to the particular instance of advanced education and all the more particularly to on the web or separation instruction. Less almost no data about particular research on basic and optional instruction has been found, and what has been discovered uses just measurable strategies, not DM methods. Along these lines, there is a need to proposed Educational data mining system that is attainable for rudimentary and optional instruction.

Classification can be depicted as a regulated learning calculation in the machine learning process. It relegates class marks to data objects in light of earlier information of class which the data records have a place. In classification a given arrangement of data records is partitioned into preparing and tests data sets. The preparation data set is utilized as a part of building the classification display, while the test data record is utilized as a part of approving the model. The model is then used to order and foresee new arrangement of data records that is not the same as both the preparation and test data sets. Decision tree calculation is a data mining enlistment methods that recursively segments a data set of records utilizing profundity first voracious approach or expansiveness initially approach until the point when every one of the data things have a place with a specific class. The tree structure is utilized as a part of arranging obscure data records. Decision tree classification system is performed in two stages: tree building and tree pruning. Tree building is done in top-down way. It is amid this stage the tree is recursively parcelled till every one of the data things have a place with a similar class name. Tree pruning is done is a base up. It is utilized to enhance the forecast and classification exactness of the calculation by limiting over-fitting. Over-fitting in decision tree calculation brings about misclassification mistake. Tree pruning is less entrusting contrasted with the tree development stage as the preparation data set is filtered just once. In this examination we will audit Decision tree calculations executed, recognize the ordinarily utilized calculations.

II. RELATED WORK

Decision tree is a vital strategy for both acceptance research and data mining, which is basically utilized for display classification and expectation. ID3 calculation is the most broadly utilized calculation in the decision tree up until this point. Through outlining on the fundamental thoughts of decision tree in data mining. ID3 calculation is a forerunner of C4.5 calculation and it was produced by a data mining software engineering scientist Ross Quinlan in 1983.It is utilized to build a decision tree by testing every hub's trait of tree in top-down way. ID calculation performs property choice component utilizing Entropy and Information Gain idea. A decision tree is an essential method for data mining and inductive realizing, which is generally used to frame classifiers and expectation models [2].

C4.5 is a standout amongst the most exemplary classification calculations on data mining. Decision Trees a Decision Tree is a helpful and prevalent classification procedure that inductively takes in a model from a given arrangement of data. One explanation behind its fame originates from the accessibility of existing calculations that can be utilized to manufacture decision trees [3].

CART is Classification and relapse trees. It was presented by Breiman in 1984. It fabricates the two classifications and regressions trees. The classification tree development via CART. It depends on paired part of the characteristics. It is additionally in view of Hunt's model of decision tree development. It additionally can be executed serially by Breiman et al. It utilizes gini record part measure for choosing the part quality. Pruning is finished by utilizing a segment of the preparation data set in truck. This approach utilizes both numeric and clear cut properties for building the decision tree and has in-constructed highlights that arrangement with missing qualities [11].

The fundamental distinction between CART, ID3 and C4.5 is the manner by which the parceling of data is performed. Truck utilizes the Gini list to choose the part quality, though ID3 and C4.5 utilize an estimation of data pick up and the data pick up proportion.

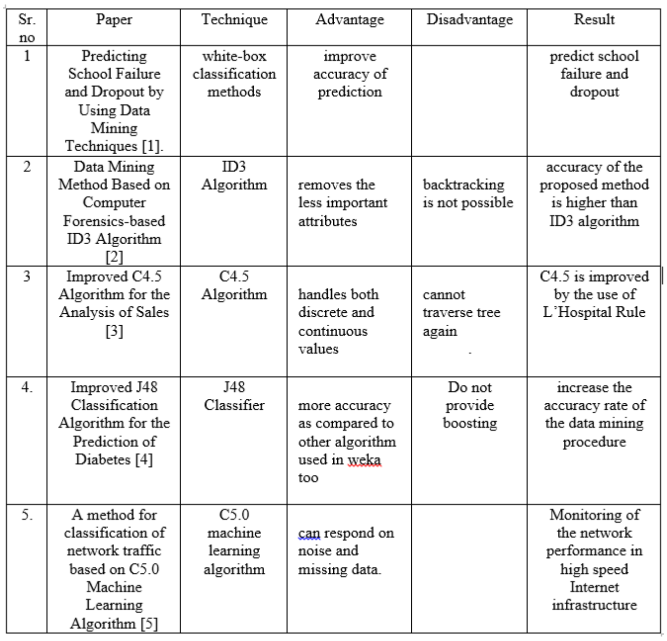
SLIQ is the Supervised Learning in Ques approach. It was presented by Mehta et al in 1996. It is a quick, adaptable decision tree calculation that can be executed in serial and parallel example. It does not depend on Hunt's calculation for decision tree classification. It segments a preparation data set recursively. It utilizes expansiveness first ravenous procedure that is coordinated with pre-sorting method amid the tree building stage [17].

Dash is Scalable Parallelizable Induction of decision Tree calculation. It was presented by Shafer et al. It is a quick and adaptable decision tree classifier. Like SLIQ it utilizes one time kind of the data things. It has no limitation on the info data estimate [18].

CHAID is a kind of decision tree method. It depends on balanced hugeness testing. This procedure was produced in South Africa. It is a comparable form to relapse examination. This adaptation of CHAID being initially known as XAID. It is valuable for classification and for recognition of connection between factors [16].

The present decision tree calculations likewise contrast in their capacity to deal with various sorts of data, and by a wide margin the most exceptional calculation of the gathering is C4.5. C4.5 depends on Quinlan's prior work with ID3 and is equipped for taking care of ceaseless data, data with missing esteems, and even has worked in ventures for rearranging the resultant decision tree. Another purpose for the ubiquity of decision trees is that they are regularly interpretable by human analyzers. The structure of a decision tree gives thinking to each statement of class esteem, and it is trusted that this thinking is anything but difficult to understand.

The classifier is tried first to group inconspicuous data and for this reason coming about decision tree is utilized. C4.5 calculation takes after the standards of ID3 calculation. Also C5.0 calculation takes after the tenets of calculation that is the vast decision tree can see as an arrangement of guidelines which is straightforward. C5.0 calculation gives the recognize on clamor and missing data. Issue of over fitting and mistake pruning is tackled by the C5 calculation. In classification procedure the C5 classifier can envision which properties are important and which are not applicable in classification [5]



Sr. no	Paper	Technique	Advantage	Disadvantage	Result
1	Predicting School Failure and Dropout by Using Data Mining Techniques [1].	white-box classification methods	improve accuracy of prediction		predict school failure and dropout
2	Data Mining Method Based on Computer Forensics-based ID3 Algorithm [2]	ID3 Algorithm	removes the less important attributes	backtracking is not possible	accuracy of the proposed method is higher than ID3 algorithm
3	Improved C4.5 Algorithm for the Analysis of Sales [3]	C4.5 Algorithm	handles both discrete and continuous values	cannot traverse tree again	C4.5 is improved by the use of L'Hospital Rule
4.	Improved J48 Classification Algorithm for the Prediction of Diabetes [4]	J48 Classifier	more accuracy as compared to other algorithm used in weka too	Do not provide boosting	increase the accuracy rate of the data mining procedure
5.	A method for classification of network traffic based on C5.0 Machine Learning Algorithm [5]	C5.0 machine learning algorithm	can respond on noise and missing data.		Monitoring of the network performance in high speed Internet infrastructure

Table 1. Suervy Table

III. CONCLUSION

This paper is a review of the best in class regarding EDM and reviews the most applicable work around there to date. It would be exceptionally hard to physically experience the tremendous arrangement of scholarly records to distinguish the understudy patterns and conduct and the example in which they learn. Rather, if client makes utilization of data mining strategies on the expansive measure of scholastic record, he/she can without much of a stretch gathering the understudies, distinguish shrouded designs about their learning styles, discover unfortunate understudy conduct and perform understudy profiling. In this way, data mining can unquestionably be a vital apparatus and part of mechanically progressed educational systems.

IV. REFERENCES

[1]. Carlos Márquez-Vera, Cristóbal Romero Morales, and Sebastián Ventura Soto, "Predicting School Failure and Dropout by Using Data Mining Techniques", Ieee Journal Of Latin-American Learning Technologies, Vol. 8, No. 1, February 2013 IEEE.

[2]. IU Qin, "Data Mining Method Based on Computer Forensics-based ID3 Algorithm", 978-1-4244-5265-1/10/$26.00©2010IEEE

[3]. Rong Cao, Lizhen Xu, "Improved C4.5 Algorithm for the Analysis of Sales", Sixth Web Information Systems and Applications Conference, 978-0-7695-3874-7/09 $25.00 DOI 10.1109/WISA36, © 2009 IEEE

[4]. Gaganjot Kaur , Amit Chhabra , "Improved J48 Classification Algorithm for the Prediction of Diabetes" , International Journal of Computer Applications (0975 – 8887) Volume 98 – No.22, July 2014.

[5]. Tomasz Bujlow, Tahir Riaz, Jens Myrup Pedersen, "A method for classification of network traffic based on C5.0 Machine Learning Algorithm", Workshop on Computing, Networking and Communications, 2012 IEEE.

[6]. Raisul Islam Rashu, Naheena Haq, Rashedur M Rahman, "Data Mining Approaches to Predict Final Grade by Overcoming Class Imbalance Problem", 17th International Conference on Computer and Information Technology (ICCIT) 2014.

[7]. Hina Gulati, "Predictive Analytics Using Data Mining Technique", 978-9-3805-4416-8/15/$31.00c 2015 IEEE.

[8]. Rutvija Pandya , Jayati Pandya , "C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning ", International Journal of Computer Applications (0975 – 8887) Volume 117 – No. 16, May 2015.

[9]. Alana M. de, Morais and Joseana M. F. R. Araújo, Evandro B. Costa, "Monitoring Student Performance Using Data Clustering and Predictive Modelling", 978-1-4799-3922-0/14/$31.00 ©2014 IEEE.

[10]. Kin Fun Li, David Rusk and Fred Song, "Predicting Student Academic Performance", 2013 Seventh International Conference on Complex, Intelligent, and Software Intensive Systems.

[11]. Mishra, T., Kumar, D. Gupta, S.," Mining Students' Data for Prediction Performance", Advanced Computing & Communication Technologies (ACCT), 2014 Fourth International Conference on 8-9 Feb. 2014.

[12]. B.K.Bhardwaj and S.Paul, "Mining Educational Data to Analyze Students Performance", International Journal Advanced Computer Science and application Vol. 2 No. 6, 2011.

[13]. Yohannes Kurniawan, Erwin Halim, "Use Data Warehouse and Data Mining to Predict Student Academic performance in Schools: A Case Study (Perspective Application and Benefits)".

[14]. B.K.Bhardwaj and S.Paul, "Mining Educational Data to Analyze Students Performance", International Journal Advanced Computer Science and application Vol. 2 No. 6, 2011.

[15]. Lewis, R.J. (200). An Introduction to Classification and Regression Tree (CART) Analysis. 2000 Annual Meeting of the Society for Academic Emergency Medicine, Francisco, California.

[16]. Mehta, M., Agrawal, R., and Rissanen, J. (1996). SLIQ: A fast scalable classifier for data mining. In EDBT 96, Avignon, France

[17]. Shafer, J., Agrawal, R., and Mehta, M. (1996). Sprint: A scalable parallel classifier for data mining. Proceedings of the 22nd international conference on very large data base. Mumbai (Bombay), India.

Cite this article as :