Volume 7, Issue 5, May 2017 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Research to Improve the Quality of Education by Using Optimized Technique

Priya Babbar
Department of Computer Science Engineering, Lovely Professional University, Phagwara, (Punjab) India

Barjinder Singh
Department of Computer Science Engineering, Lovely Professional University, Phagwara, (Punjab) India

DOI: 10.23956/ijarcsse/SV7I5/0187

Abstract*— Data Mining is a process of extracting knowledge from large amount of data. It is used in EDM (Educational data mining). Educational data mining is a field for discovering knowledge from large amount of Educational data. The main purpose of EDM is to find the appropriate pattern of educational data so that there is improvement of qualification of education. Different aspects are evaluated like social, economic, personal, cultural, geographical, institute environment and other in education research study. Such aspects may either help a student in shining during academic period or halt academic program. Such failure is known as drop-out. Data mining algorithm helps in finding those factors; that are mostly contributing the student's performance. If we work on most contributing attribute better results can be achieved. In our research we are going to construct a hybrid model that can fit in Educational data mining. Hybrid approach is an approach which is combination of two or more techniques of data mining such as association, Clustering, Bayesian networks, neural network's machine learning technique, fuzzy logic, genetic algorithms etc. In this research, we discuss how a hybrid approach based data mining model can help to improve an education system by enabling better and effective teacher-student interaction.

Keywords*— EDM (Educational Data Mining), Clustering, Classification, Prediction, WEKA (Waikato Environment for Knowledge)

I. INTRODUCTION

Data mining technique is helpful for several reasons in private as well as public sectors. Many Industries use Data Mining technique to extract the valuable information from the large database to minimize costs, enhance research, and increase sales i.e. banking, medicine, insurance, retailing and EDM (Educational Data Mining). By the increase of technology of computers the collection of data, storage of data as well as manipulations of data have become straight forward. Data Mining is extensively useful in EDM (Educational Data Mining). Educational data mining is an emerging field for knowledge discovering from large amount of Educational data. The purpose of EDM is to find the pattern of educational data so that qualification of education can be improved. EDM is the educational research study of Variety of methods in which different aspects are evaluated like social, economic, personal, cultural, geographical, institute environment and other. Such aspects may either help a student in excelling during academic period or halt academic program of a student. Such failure is known as drop-out. [1] Data mining algorithm helps in finding those factors, that mostly contributing the student's performance. If we work on most contributing attribute better results can be achieved.[3]

II. PROBLEM FORMULATION

This research finds better efficiency as well as limitations of traditional algorithms. The research also finds out the best possible solution which can handle large amount of high dimensional data. Our research is about checking the performance of students with the help of decision tree and clustering algorithms [8] by applying them on different data sets using data mining tool and evaluates the outcome.

III. SCOPE OF STUDY

This research checks the effectiveness of decision tree classification as well as clustering algorithms by applying them to a large scale data set. Example: Classification methods try to find those students who are likely to fail or need more attention.[4] Focus on these kinds of students can better the quality of education and decrease the dropout rate. Clustering methods try to make cluster of students according to their knowledge of subjects. This helps the student to find job according to their taste. Experiment result will also show the best accuracy, less time taken, higher robustness and generalization ability in one of the algorithm.

IV. TYPES OF DATA MINING ALGORITHMS

A. Association rule algorithm:- It mainly deals with search statistical relations between objects in dataset. It finds how events aggregate together. [5]

B. Classification algorithm:- It can describe or classify objects related to dataset into predefined set of classes. It is supervised learning approach. It includes objects in dataset used to understand existing objects and predict behaviour of new objects.

C. Clustering algorithm:- It is collection of objects of similar type in one group. The cluster provides us better results.[28]

D. Inductive and Deductive learning:- Machine learning in mainly classify into two different types. In deductive learning, we learn something with existing knowledge and produce some new knowledge from existing knowledge. In inductive learning rules and patterns are extracted from large datasets. In clustering partition the dataset in to subsets for optimization.

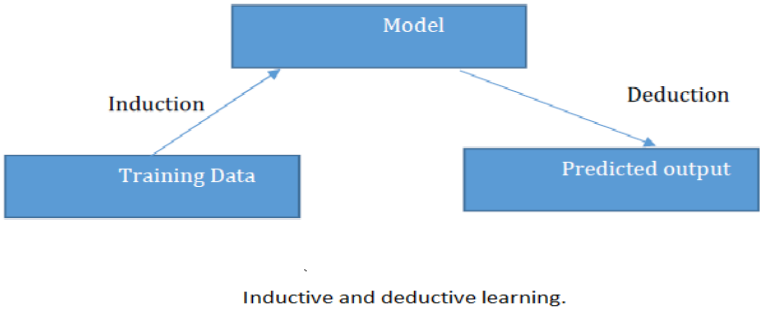


Fig 1: Inductive and deductive learning.

V. METHODOLOGY

Student's performance is a great concern for academic institutions. Classification and clustering methods like decision trees, Bayesian network, k-means etc can be applied on the educational data for predicting the student's performance in examination. These classification methods will be useful to identify the weak students and help them to score better marks. Various decision tree and clustering algorithms like C4.5, ID3 (Iterative Dichotomiser 3), k-means and CART (Classification and Regression Trees) can be applied to the research.

In this study, C4.5 algorithm is applied on Students of different colleges to predict their performance in the final exam. The outcome of the clustering is to group the similar types of students and analysis with inter cluster students. The outcome of the decision tree predicts the number of students who are likely to pass, fail or promoted to next year. The results provide steps to improve the performance of the students who were predicted to fail or promoted.

In our research, the data is firstly converted to an optimal dataset by applying various Expert rules [7] and then on this dataset feature reduction is performed. Finally from this optimal dataset decision tree has been constructed. Following steps are performed in our research:

- Primary data has been collected from 1300 students from two different schools through questionnaires and interviews.
- After that training dataset has been created from this primary data by randomly selecting 650 rows.
- Knowledge elicitation has performed from domain experts.
- From this knowledge of experts, rules have been created in Java Language.
- These rules have been applied to the training data set.
- Then feature reduction is performed to this dataset by calculating gain ratio of every attribute and the attributes having minimum gain ratio have been deducted. After deduction we obtain an optimal dataset.
- The optimal decision tree has been constructed using c4.5 algorithm.

Finally, comparison has been made between existing approach and our approach to find the outcome.

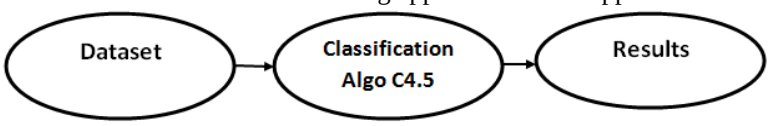


Fig. 2:The diagram represents existing approach

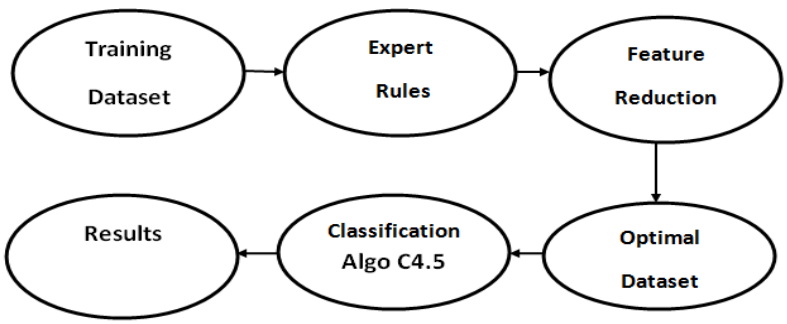
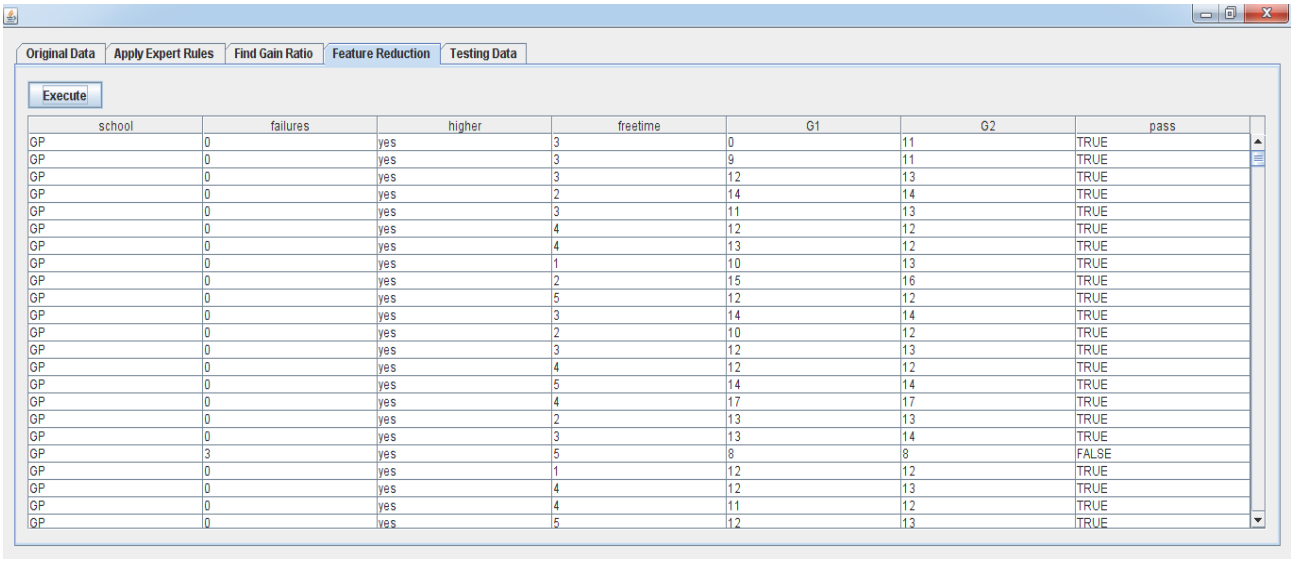


Fig 3: Representing our approach

A. *Attributes used:* The attributes considered for the analysis of students performance are: [6] Sex, Age,Address, Parents status, Mother's education, Father's education etc. Attribute 'Sex' has domain values {Male, Female}, Age{15-20} etc.

B. *C4.5 Algo:-* These are developed by Quinlan for inducing Classification Models from data that are also called decision trees. We are given a set of accounts. Each record has the same construction, consisting of a number of quality/value pairs. These attributes shows the group of the record. The problem is to decide a decision tree. This decision is done on the basis of answers to questions. These questions are about the non-category attributes predicts correctly the value of the category attribute. Usually the category attribute takes only the values {true, false}, or {success, failure}, or something equivalent. In any case, one of its values will mean failure.[2]

C. *Finding gain ratio:-* After applying expert rules on training dataset we will find the gain ratio of all the attributes by using WEKA tool. In our research the algorithm C4.5 will reduce the parameters with less gain ratio. Our approach will use only 6 attributes that we got after finding gain ratio. At last we will gain an optimal dataset with only 6 attributes.



school	failures	higher	freetime	G1	G2	pass
GP	0	yes	3	0	11	TRUE
GP	0	yes	3	9	11	TRUE
GP	0	yes	3	12	13	TRUE
GP	0	yes	2	14	14	TRUE
GP	0	yes	3	11	13	TRUE
GP	0	yes	4	12	12	TRUE
GP	0	yes	4	13	12	TRUE
GP	0	yes	1	10	13	TRUE
GP	0	yes	2	15	16	TRUE
GP	0	yes	5	12	12	TRUE
GP	0	yes	3	14	14	TRUE
GP	0	yes	2	10	12	TRUE
GP	0	yes	3	12	13	TRUE
GP	0	yes	4	12	12	TRUE
GP	0	yes	5	14	14	TRUE
GP	0	yes	4	17	17	TRUE
GP	0	yes	2	13	13	TRUE
GP	0	yes	3	13	14	TRUE
GP	3	yes	5	8	8	FALSE
GP	0	yes	1	12	12	TRUE
GP	0	yes	4	12	13	TRUE
GP	0	yes	4	11	12	TRUE
GP	0	yes	5	12	13	TRUE

Fig 4: Optimal dataset

VI. INTERPRETATION OF RESULTS

When C4.5 algorithm is run on an optimal dataset which is generated by applying expert rules and feature reduction technique, the percentage of accuracy has been increased by 12%. It has been observed that, correctly classified instances in this approach is 91.1051% and 7.8647% instances are in correctly classified. Computational time has also been reduced by using this approach that is 0.09 sec to 0.02 sec. Above analysis shows the decrease in computation time by applying C4.5 algorithm on datasets. The implemented technique has been applied to the dataset of 650 rows to make it an optimal dataset and then comparison has been done between computational time of existing and our technique.

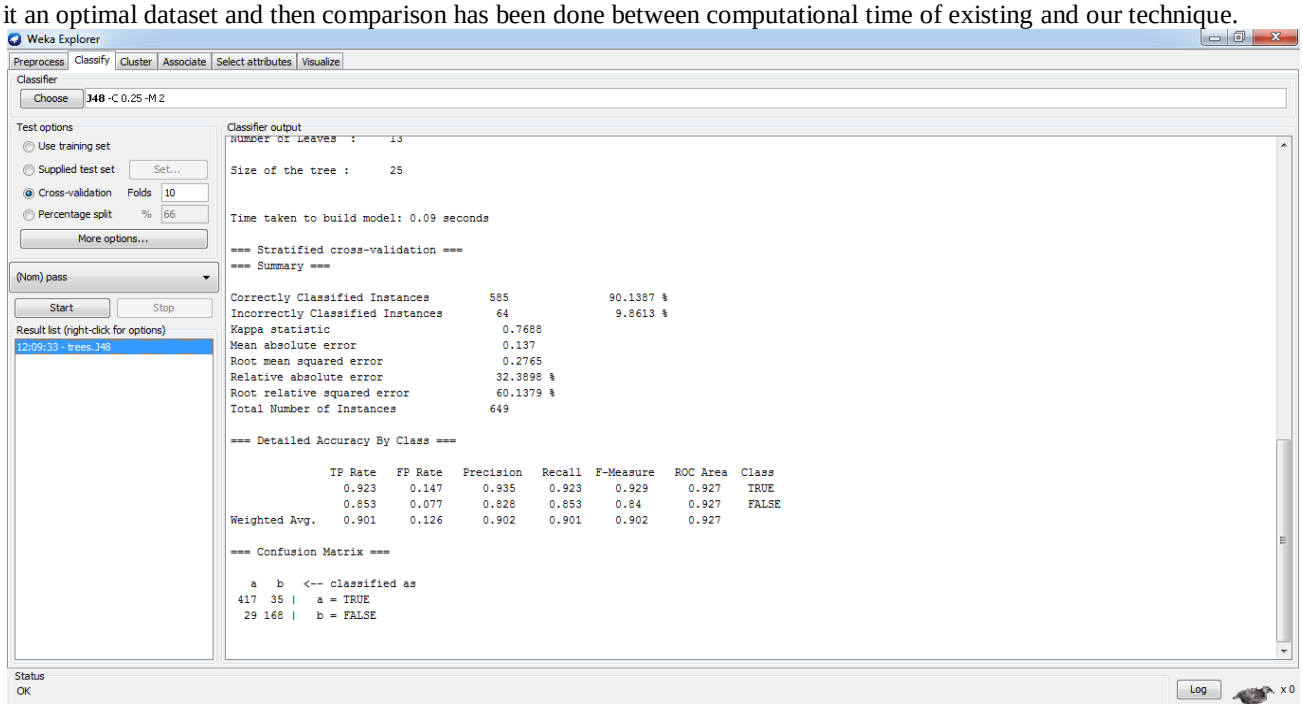


Figure 5: Representing accuracy of an existing Approach

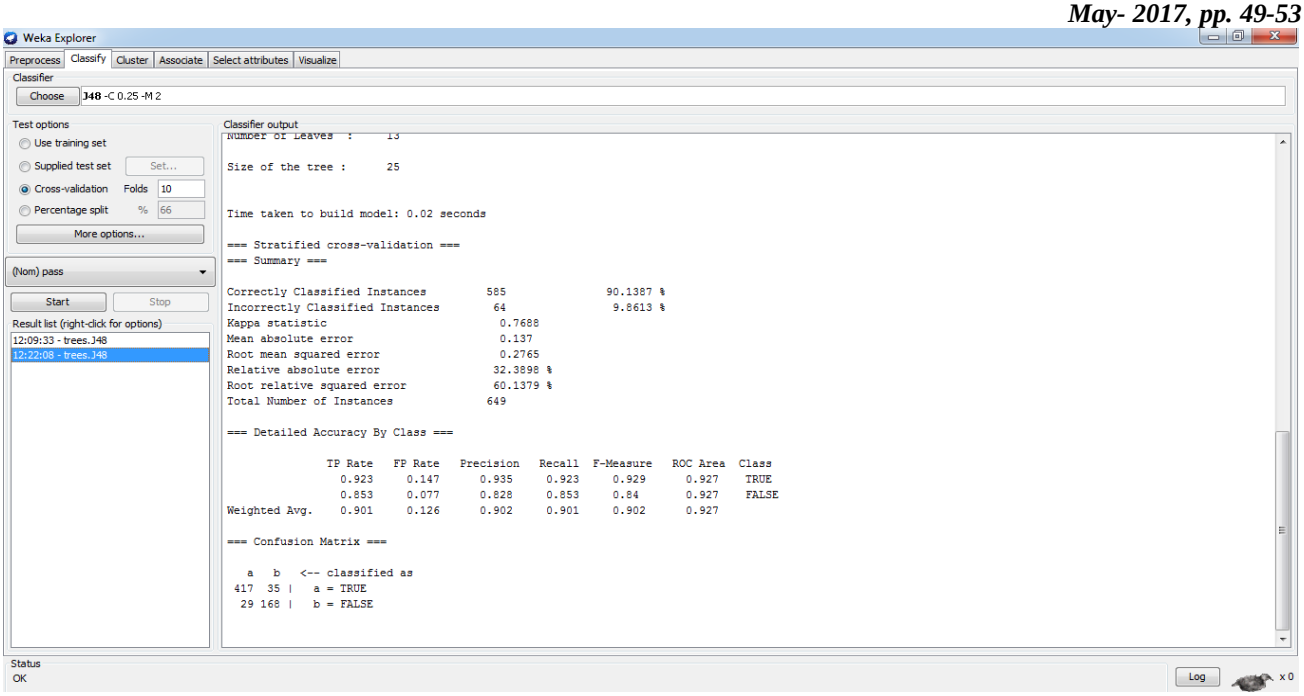


Figure 6: Representing accuracy of an optimal Approach

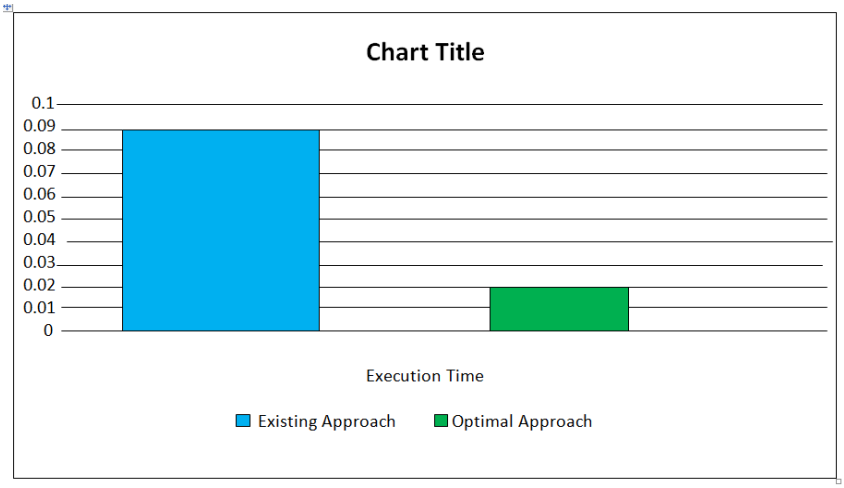


Figure 7: Representing time comparison between Existing Approach and Optimal Approach graphically

VII. CONCLUSION

The optimal algorithm is based on the C4.5 methods. This method introduces new characteristics such as implementation of Expert rules and technique of feature reduction on large dataset. Several attempts have been made to design and develop the specific data mining system but no system is found completely specific or generic. Thus the domain expert's support is compulsory for every domain. The domain experts shall be guided by the system to effectively apply their knowledge for the use of data mining systems to create required knowledge. The domain experts are required because they determine the category of data that should be composed of in the specific problem domain, selection of particular type of data for data mining, cleaning of data, transformation of data, extracting various patterns for generation of knowledge and finally interpretation of the patterns and knowledge generation. By applying this optimal approach on student's data, we can find out the best features or attributes that will influence student's performance.

ACKNOWLEDGMENT

All praise in the name of almighty God, who give us in the darkness and help in difficulties. The research is the result of full semester of work whereby I have been accompanied and supported by many people. It is a pleasant aspect to that I have the opportunity to express my gratitude for all of them. I am also extremely indebted to my guide Mr. Barjinder Singh. I am very much thankful to him for picking me as a student at the critical stage of my masters. I warmly thank him for his valuable advice, constructive criticism and his extensive discussions around my work. I expand my thanks to my friends and family who always kept my spirits up with their extended love, affection and support at the time of my project work. At last but not the least, I would like to pay high regards to the authors whose work I have consulted very often during my project work. And I would like to thank Lovely Professional University that provided me the road for the completion of my degree in this particular field.

REFERENCES

[1] Apurva A Mehta, Niyati J Buch "Dept and Breadth of Education Data Mining Researchers' Point of View"

[2] Md. Fahim Sikder, Md. Jamal Uddin and Sajal Halder 'Predicting student's yearly performance using Neural Network' 5th International Conference on Informatics, Electronics and Vision (ICIEV), 2016 pp. 524-529

[3] Hind Almayan, Waheeda Al Mayyan, 'Improving Accuracy of Students' Final Grade Prediction Model Using PSO' 6th International Conference on Information Communication and Management, 2016 pp. 35-39

[4] Norlida Buniyamin, Usamah bin Mat, Pauziah Mohd Arshad 'Educational Data Mining for Prediction and Classification of Engineering Students Achievement' IEEE 7th International Conference on Engineering Education (ICEED), 2015, pp. 49-53

[5] Data Mining: Concepts and Techniques Second Edition Jiawei Han *University of Illinois at Urbana-Champaign* Micheline Kamber

[6] John Jacob, Kavya Jha, Paarth Kotak, Shubha Puthran 'Educational Data Mining Techniques and their Applications' International Conference on Green Computing and Internet of Things (ICGCloT), 2015, pp. 1344-1348

[7] "Data Mining Concepts." [Online]. Available: https://technet.microsoft.com/en-us/library/ms174949.aspx. [Accessed: 21-Jan-2016].

[8] Ashish Dutt, Saeed Aghabozrgi, Maizatul Akmal Binti Ismail, and Hamidreza Mahroeian 'Clustering Algorithms Applied in Educational Data Mining' International Journal of Information and Electronics Engineering, 2015, pp. 112-116

