IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

Comparative Analysis of Various Techniques used for Predicting Student's Performance

Yashashree K. Sonparote[1], Dr. G. R. Bamnote[2], Prof. S. V. Ahmed[3]
[3]Research Scholar,[2]Guide, [3]Co- Guide.
Department of Computer Science. & Engineering[1,2,3]
Prof Ram Meghe Institute of Technology &Research, Badnera, Amravati. Maharashtra, India

Abstract: *Research in higher education is beginning to explore the potential of data mining in analysing data to give quality service and needs of their graduates. Thus, educational data mining emerges as one tools to study academic tool data to identify patterns and help for decision making with affecting education. Data mining applications are becoming a more common tool in understanding and solving educational and administrative problems in higher education. Generally, research in educational mining focuses on modelling students' performance instead of instructors' performance. One of the common tools to evaluate instructors' performance is the course evaluation questionnaire to evaluate based on students' perceptions. In this study, four different classification techniques, decision tree algorithms, support vector machines, artificial neural networks, and discriminant analysis– are used to build classifier models. Their performances are compared over a dataset composed of responses of students to a real course evaluation questionnaire using accuracy, precision, recall, and specificity performance metrics. Although all the classifier models show comparably high classification performances, accordingly, it is shown that many of the questions in the course evaluation questionnaire appear to be irrelevant. Furthermore, the analysis shows that the instructors' success based on the students' perception mainly depends on the interest of the students in the course. The finding of the study indicate the effectiveness and expressiveness of data mining models in course evaluation and higher education mining. Moreover these findings may be used to improve measurement instruments.*

Keywords: Educational Data Mining (EDM), Machine Learning, Prediction Students' Academic Performance

I. INTRODUCTION

One of the biggest challenges of higher education institutions is the proliferation of data and how to use them to improve quality of academic programs and services and the managerial decisions. Variety of "formal and informal" procedures based on "qualitative and quantitative" methods is used by higher education institutions to solve problems, which keep them away from achieving their quality objectives. However, methods used in higher education for quality purposes are mainly based on predefined queries and charts to analyse the data. In addition, these methods lack the ability to reveal useful hidden information. Online Exam is being launched because a need for a destination that is beneficial for both college and students. With this site, institutes can register and host online exams. Students can give exams and view their results. This site is an attempt to remove the existing flaws in the manual system of conducting exams.Online Examination System fulfils the requirements of the institutes to conduct the exams online. They are do not have to go to any software developer to make a separate site for being able to conduct exams online students just have to register on the site and enter the exam details and the lists of the students which can appear in the exam. Students can give exam without the need of going to any physical destination. They can view the result at the same time. Thus the purpose of the site is to provide a system that saves the efforts and time of both the institutes and the students. Online Exams System is a web application that establishes a network between the college and the students. Institutes enter on the site the questions they want in the exam. These questions are displayed as a test to the eligible students. The answers enter by the students are then evaluated and their score is calculated and saved. This score then can be accessed by the institutes to determine the passes students or to evaluate their performance. Online Exams System provides the platform but does not directly participate in, nor is it involved in any tests conducted. Questions

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

are posted not by the site, but users of the site. The site requires an institute to register before posting the questions. The site has an administrator who keeps an eye on the overall functioning of the system. The site gets revenue by charging the institutes each time they want to conduct the exam. The growth of Information and Communication Technology has significant effects on all people around the world. With this growth, people are able to connect with each other, especially through the Internet. It could be seen as a professional level of education but with the advantages of lower time and cost. Some other advantages of e-learning include larger learner population, shortage of qualified training staff and lower cost of campus maintenance, up-to-date information and accessibility. In a typical e-learning environment the lecturers, students and information are in different geographical locations and are connected via the Internet. The e-learning promotes the construction of life-long learning opinions and learning society. E-learning is a broad concept and it consists with different types, namely Synchronous and Asynchronous e- Learning. Both methods have different characteristics and they use different methods to broadcast the learning.

Analyzing the large volume of data to make useful summarized information may be a complex task for human being. Grading students' academic performance is a more difficult and challenging work which will help educators to keep track of progress of performance of students. Enhancement of quality of education for improvement of student performance is more important for all educational institutes. Prior prediction of outcome of students is beneficial, knowing this in advance will affect the student performance. Data mining method plays a highly important role in the classification as well as differentiation of education data. This data is used in a properly organized way by data mining approach. This work can be done with analysis depends on the association, patterns, relations between these data so for getting useful information. The prediction in students' performance with high accuracy is more useful because it helps to finding the students with poor academic performance. In universities, enrollment system and academic performance define the student retention. Data Mining can be used within the educational area. This will be useful to increase our knowledge of learning process by recognizing the variables and evaluating them.

Mining of educational data from the educational environment called as Educational data mining [2]. In last decades, Educational data mining is becoming more popular and demanding field, which cause the exceptionally increase in count of researchers [5]. Recently lots of innovative methods are being developed by many researchers. The earlier proposed machine learning and data mining approaches are using by many researchers for deep understanding of data of educational institutes. This educational data is being used for the understanding the performance of a student.

Information exposed by using EDM approach can be used by a various kind of users like management, administrator, teachers, students and all who are connected with the educational field [1]. Basically Educational Data Mining methods includes,

- Classification and Profiling students
- Identify students learning methods
- Finding all students that are taken course together
- Predicting student's academic performance

The main objective is to get knowledge of the helpful patterns and identifying the necessary and functional information from the systems of educational data. These systems are nothing but syllabus management, course management and registration as well as admission system. These systems are more helpful to all students at various steps of educational institutes' likes universities, colleges and schools.

Nowadays, the main aim of all educational institutions is the delivery of better and higher education.

Recently, many institutes are using newly developed and innovative methods of education for developments in various education fields for making delivery of higher and better education.

II. EXISTING SYSTEM

Data mining in education is the field that allows us to make predictions about the future by examining the data obtained so far in the field of education by using machine learning techniques. There are basically three data mining methods: classification, clustering, and association rule mining.

Data mining is the process of going through the system databases and finding relevant data to analyse. For example, the predictive data mining process may use algorithm-based tools to go through a customer database to look at past transactions in order to support theories about possible future volumes of transactions

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

III. LITERATURE REVIEW

Pedro G. Espejo, Sebasti´an Ventura, and Francisco Herrera "A Survey on the Application of Genetic Programming to Classification"

Classification is one of the most researched questions in machine learning and data mining. A wide range of real problems has been stated as classification problems, for example credit scoring, bankruptcy prediction, medical diagnosis, pattern recognition, text categorization, software quality assessment, and many more. The use of evolutionary algorithms for training classifiers has been studied in the past few decades. Genetic programming (GP) is a flexible and powerful evolutionary technique with some features that can be very valuable and suitable for the evolution of classifiers. This paper surveys existing literature about the application of genetic programming to classification, to show the different ways in which this evolutionary algorithm can help in the construction of accurate and reliable classifiers.

Bin-Shyan Jong, *Member, IEEE*, Te-Yi Chan, and Yu-Lung Wu "Learning Log Explorer in E-Learning Diagnosis"

This study presents a learning behaviour diagnosis system to study students' learning status from learning portfolios. The proposed linking layer enables the proposed system to work on various e-learning platforms without reprogramming. Additionally, the use of a supervisory agent enables teachers and students to obtain their learning status or information provided by the proposed system in both Web and e-mail. Furthermore, the computer engineering curriculum operating systems was adopted to evaluate the proposed system. Evaluations of confidence between learning status and learning achievement yield positive experimental results.

Mustafa Agaoglu "Predicting Instructor Performance Using Data Mining Techniques in Higher Education"

Data mining techniques are applied in higher education more and more to give insights to educational and administrative problems in order to increase the managerial effectiveness. However, most of the educational mining research focuses on modelling students' performance. In this paper, data mining is utilized to analyze course evaluation questionnaires. Here, the most important variables that separate "satisfactory" and "not satisfactory" instructor performances based on students' perception are found. Hopefully, these can help instructors to improve their performances. .In addition, irrelevant variables that do not differentiate "satisfactory" and "not satisfactory" instructor performances are also listed. Different dimensions of course and instructor effectiveness are measured with course evaluation questionnaires in higher education institutions and these findings may be used to improve measurement instruments.

The main objective is to get knowledge of the helpful patterns and identifying the necessary and functional information from the systems of educational data. These systems are nothing but syllabus management, course management and registration as well as admission system. These systems are more helpful to all students at various steps of educational institutes' likes universities, colleges and schools.

Nowadays, the main aim of all educational institutions is the delivery of better and higher education.

Recently, many institutes are using newly developed and innovative methods of education for developments in various education fields for making delivery of higher and better education [6].

Various studies are developed to improve prediction of student performance. Prediction of students' academic performance is attracted various policy makers, researchers, as well as educators for long time. K. Kasthuriarachchi and S. Liyanage [3] Designed a system which correlate with the students attendance, develop new methods for practical to encourage students to improve attendance and give the Progress of students in classroom. This will help to take decision by the academicians, educators as well as *Turkish Journal of Computer and Mathematics Education Vol.11 No.03 (2020), 1273- 1281* 1275 *Research Article*

Parents. Also management of institute gives the financial support by means of scholarships to students from families having lower income. Moreover, Students performance will degraded due to the parent's negligence, therefore parents must monitoring continuous the their children's progress. These parameters

are useful for the process of critical decision making in the educational institutes.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

Ching-Chieh Kiu [4] designed a model for predicting student academic Performance. By conducting analysis author find out the impact or effect of student background, coursework achievement and social activities on the prediction of performance. Author illustrates the significance of student background as

Well as social activities to be useful for prediction of student performance and also helpful for identifying the weak student. Early prediction by the model is helpful for the students to perform better to enhance the academic performance and teachers to make the teaching learning process more effective by invention in teaching process, which helps to the student's performance.

The research study presented by H. Muhonen et. al. [5] stated that educational discussions is naturally related with students' academic performance in the domain of language physics, chemistry and arts. By performing qualitative analysis explore the communication patterns which identified both teacher-initiated as well as student-initiated discussion learning, together with peer-centered discussions.

In the domain of arts language lessons, discussion or teaching were represented by their neutral quality, whereas in domain of the chemistry and physics lesson, is typical high-quality learning. The outcome of study suggests that to improve the student learning, there is need of improvement in both the quality as well as discussions in the classroom. R.-C. Zhang et. al. [6] done the study to on present teaching learning background through technology mediated learning (TML), The teaching learning method is more affected by the DA(dynamic assessment) approach. This study identifies the two research gaps on students' academic performance which is affected by DA. First, the previous research has more concentrates on pre- and post-test assessment based DA. The usage of Information technology and knowing the effect of DA based on computer for predication of students' performance over time is essential. Secondly, systems are designing based on TML assessment, the reason is that a many students get the support in remote TML. This system having many limitations such as class size is limited, system will not specify the others factors from outside the system which affects the result.

IV. SYSTEM DESIGN

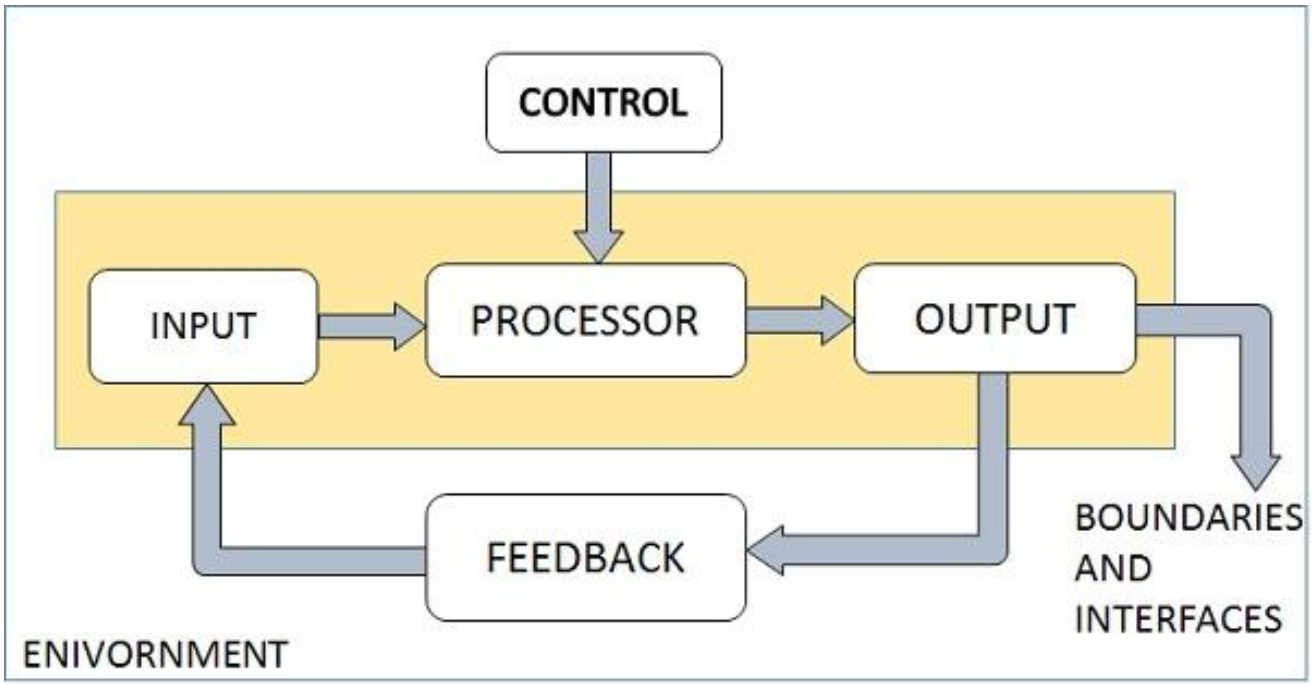
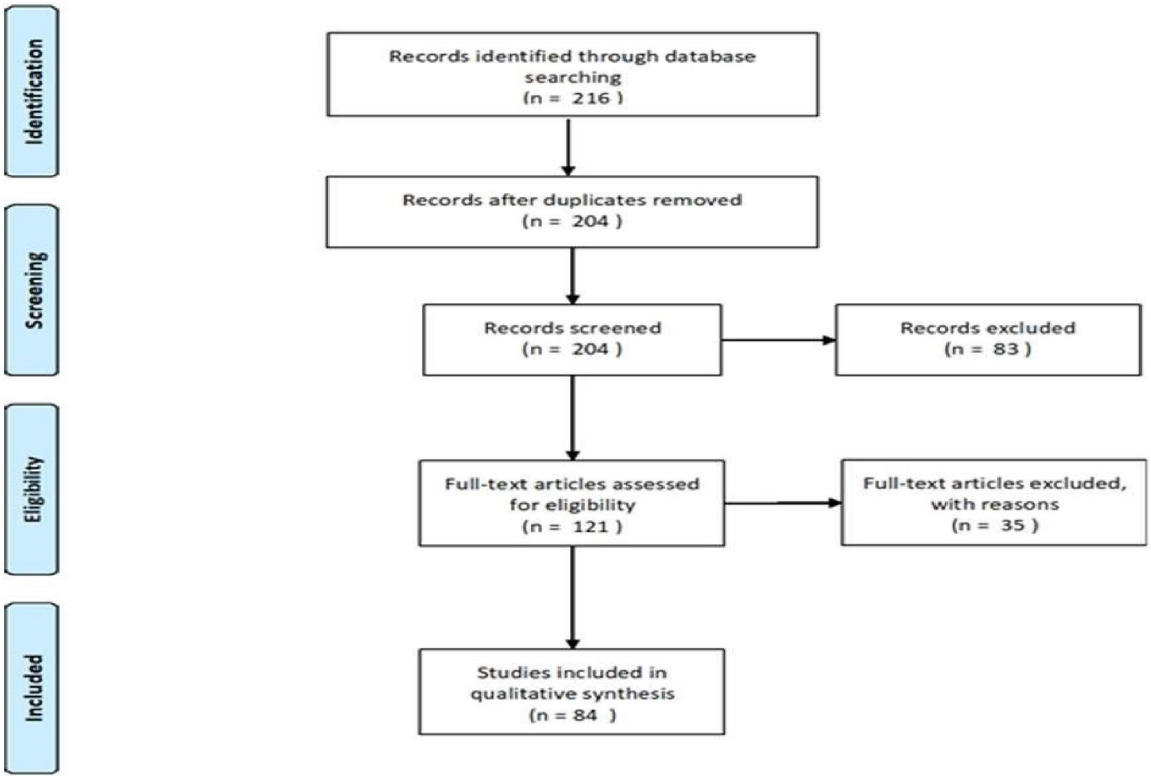


Fig: Predicting Students Performance

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

4.1 Flowchart



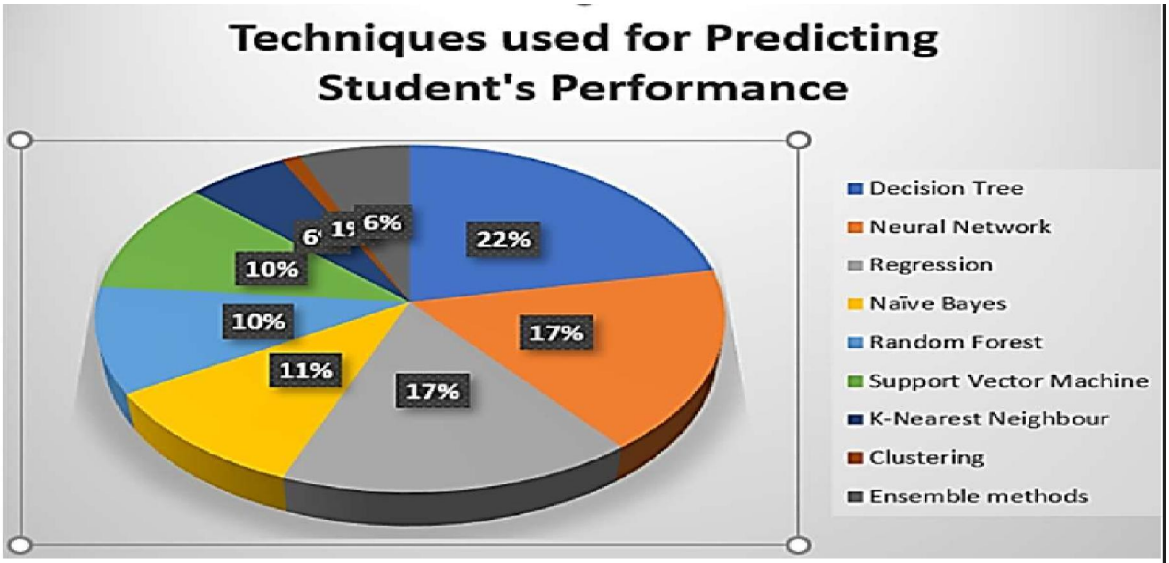


Figure 1: Distribution of techniques used for Predicting students' performance

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

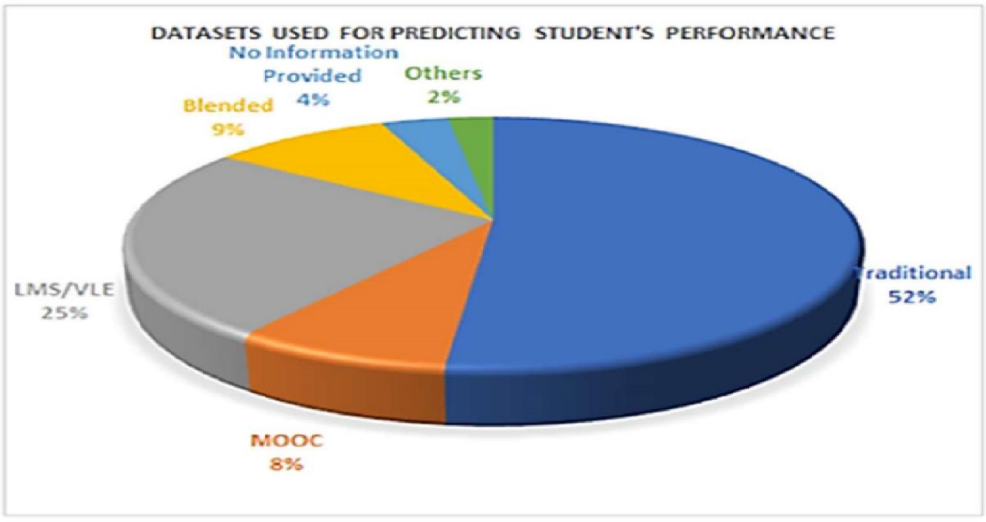


Figure 2: Distribution of datasets used for predicting student's performance

V. PROPOSED SYSTEM

They are going to propose the system by using which the user can give a test on specific educational or subject categories. When he / she complete the test, system will calculate the performance of the user by using the algorithm decision tree. The system will suggest to the teacher that on which topics the user is weak or need to study again.

To solve the problems faced with manual examination writing, there is need for a computerized system to handle all the works. I propose a web based application that will provide a working environment that will be flexible and will provide ease of work and will reduce the time for report generation and other paper works.

These examinations are conducted as open-book or open table type examinations. Today many organizations are conducting online examinations worldwide successfully and issue results online but they are not measuring the performance of the student and teacher not know about the weak points of the students and we are focusing on this issue.

The main advantage is that it can be conducted for remote candidates and evaluation of answers can be fully automated for all questions and other essay type questions can be evaluated manually or through automated system, depending on the nature of the question's and the requirements. Also online examinations can be conducted at any time. When comparing with traditional exam scenario the cost for an online examination will be almost zero after the online exam system is establishment and if maintenance cost is not considered.

5.1 Advantages of the Proposed System

- The system can automatically create exams by randomly choosing questions from your exam question database.
- The system will randomly order the questions for each student.
- The system will provide the performance details about the student to the admin.
- It will help the admin to find out on which concept the student is weak.
- It will improve the efficiency of the workforce who in turn carries out many operations manually.
- With the adoption of new system, there will centralized, secure and robust database of the candidates appearing in the examination without duplication of records.
- In a networked environment system different units of the examination wing are connected so that all these are able to share up-to-date information without redundancy of database at various stages.
- The complete Computerization of Examination System will lead to effective monitoring of Examination Process by the concerned authorities.
- The Computerized Statistical Analysis of Data will facilitate the management to see trends in result at various levels.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

- Level Management to make strategic decision in favor of University and students.
- Once the students' data is in database form, lots of academic benefits can be drawn in terms of the statistical analysis of the results belonging to different faculties, courses, social groups, colleges (private/government) and different area of the state and the country.

5.2 Proposed Algorithm

A. Decision Tree Algorithm

A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node.

B. Advantage:

- Decision trees implicitly perform variable screening or feature selection.
- Decision trees require relatively little effort from users for data preparation.
- Nonlinear relationships between parameters do not affect tree performance.
- The best feature of using trees for analytics - easy to interpret and implement

VI. SUMMARY

Literature survey is done by studying various papers, using the different approaches performed on educational data to predict student's academic performance. Form [1] to [7] shows the generalized study on EDM process. Whereas, Big data Mining and Generalized data mining techniques are studied, [8] to [11]. For machine learning approach to predict the students' performance referred [12] to [16]. For prediction of student's performance proposed models using hybrid approaches referred the existing models of [7] to [25].

VII. RESEARCH GAP

We study the various research will states the effects of data pre-processing and fine tuning algorithms on the effectiveness of these approaches. Many researchers performed experiment to provide proof for efficiency of existing EDM methods to in identifying weak students who may fail. It is important to be aware of some threats. All the studied systems work on specific educational data that means system is not producing general result. Only fine-tuning of Naive Bayes techniques and SVM was done automatically. Fine tuning task of algorithm is done manually may affect the efficiency. This task is main threat for systems. [3].

To predict performance of student researchers applied various classification techniques. Some are uses the multi-class classification gives best accuracy and requires a less execution time. For predicting student performance only a few of them consider choosing associated attributes as an important step and parameters for fine tuning of algorithm [5].

VIII. CONCLUSION

A model for predicting students' academic performance using Decision tree algorithm has an improved accuracy and easily be implement at institutions for higher education for predicting students' performance and also mine interesting features pertaining academics of students.

For improving the prediction find out enhanced prediction model by reviewing and compared the many existing techniques (final). This paper has presented a exhaustive survey of research works on Educational data mining (EDM).This paper reviews several existing researches and identifies other future pathways based on EDM insights. Clustering techniques help identify key variables such as students

Behavior in class, group learning, time need to spend learning a particular module, the classroom environment, and student motivation, etc. Clustering on EDM provides various useful factors and it can be multilevel nonhierarchical and hence the researchers must carefully choose the algorithm and the variables that result in better and accurate clusters and hence provide useful information.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 3, Issue 1, January 2023
Impact Factor: 6.252

IX. ACKNOWLEDGEMENT

First and foremost, I would like to express my sincere gratitude to my **Dr. G. R. Bamnote, & Prof. S. V. Ahmed** who has in the literal sense, guided and supervised me. I am indebted with a deep sense of gratitude for the constant inspiration and valuable guidance throughout the work.

REFERENCES

[1]. Han, J. 2012. Micheline Kamber, & Jian Pei."Data Mining: Concepts and Techniques".
[2]. Aziz, A. A., Ismail, N. H., & Ahmad, F. 2013." Mining Students' Academic Performance" inJournal of Theoretical & Applied Information Technology,53(3).
[3]. Ahmad, F., Ismail, N. H., & Aziz, A. A. 2015. "The Prediction of Students' Academic Performance Using Classification Data Mining Techniques." inApplied Mathematical Sciences,9(129), 6415-6426.
[4]. El-Halees, A. 2009. Mining Student's Data to Analyze ELearning Behavior: A Case Study.
[5]. Tair, M. M. A., & El-Halees, A. M. 2012." Mining Educational Data to Improve Students' Performance: A Case Study" in International Journal of Information,2(2), 140-146.
[6]. Garc´ıa, E. P. I., & Mora, P. M. 2011." Model Prediction of Academic Performance for First Year Students". In Artificial Intelligence (Micai), 2011 10th Mexican International Conference (pp. 169-174).
[7]. Bhardwaj, B. K., & Pal, S.2012." Data Mining: A Prediction for Performance Improvement Using Classification".Arxiv Preprint Arxiv:1201.3418.
[8]. Al-Radaideh, Q. A., Al-Shawakfa, E. M., & Al-Najjar, M. I. 2006." Mining Student Data Using Decision Trees." In International Arab Conference on Information Technology, Yarmouk University, Jordan.
[9]. Lakshmi, D., Arundathi, S., & Jagadeesh, D. 2014." Data Mining: A Prediction for Student's Performance Using Decision Tree Id3 Method".
[10]. Ali, M. M. 2013. "Role of Data Mining in Education Sector" in International Journal of Computer Science and Mobile Computing,2(4), 374-383.
[11]. Sembiring, S. 2012."An Application of Predicting Student Performance Using Kernel K-Means and Smooth Support Vector Machine" in(Doctoral Dissertation, Ump).
[12]. Bhullar, M. S., & Kaur, A. 2012." Use of Data Mining in Education Sector." InProceedings of The World Congress on Engineering and Computer Science(Vol. 1, pp. 24-26).
[13]. Sumitha, R.,& Vinothkumar, E. S. 2016." Prediction of Students Outcome Using Data Mining Techniques." inInternational Journal of Scientific Engineering and Applied Science,2(6)8.
[14]. Saa, A. A. 2016." Educational Data Mining & Students' Performance Prediction." inInternational Journal of Advanced Computer Science and Applications,7(5), 212-220.
[15]. Asif, R., Merceron, A., Ali, S. A., & Haider, N. G. 2017. "Analyzing Undergraduate Students' Performance Using Educational Data Mining." inComputers & Education,113, 177-194.
[16]. Hussain, S., Dahan, N. A., Ba-Alwib, F. M., & Ribata, N. 2018." Educational Data Mining and Analysis of Students' Academic Performance Using Weka." inIndonesian Journal of Electrical Engineering and Computer Science,9(2).
[17]. Chen, T. Y., Kuo, F. C., & Merkel, R. 2004." On the Statistical Properties of The F-Measure." InQuality Software, 2004. Qsic 2004. Proceedings. Fourth International Conference on (pp. 146-153). IEEE.
[18]. Arora, R. 2012." Comparative Analysis of Classification Algorithms on Different Datasets Using Weka." in International Journal of Computer Applications,54(13).
[19]. Samrat Singh, Dr. Vikesh Kumar , "Performance Analysis of Engineering Students for Recruitment Using Classification Data Mining Techniques ",IJCSET February 2013. [20] M. Goyal and R. Vohra, "Applications of Data Mining in Higher Education", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue2, No 1, March 2012.
[20]. Jason Brownlee ,"How to Save Your Machine Learning Model and Make Predictions in Weka", August 3, 2016.