



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

Analysis of Student Performance Using Data Mining Technique

Ankita A Nichat, Dr. Anjali B Raut

M.E Student, Department of Computer Science & Engineering, HVPM's College of Engineering & Technology,
Amravati, India

HOD, Department of Computer Science & Engineering, HVPM's College of Engineering & Technology,
Amravati, India

ABSTRACT: Data mining is the analysis step of the "knowledge discovery in databases" process. Data mining is the process of analyzing data from different perspectives and summarizing it into useful information. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining is also use for sorting the educational problem by using analysis techniques for measuring the student performance, instructor performance. In this paper, measuring student performance using classification technique - decision tree, artificial neuralnetwork areused to build classifier models... The work processed based on the several attributes to predict the performance of the student. Analyzing the weakness and strength of student which may be helpful to improve the performance in future.

KEYWORDS: Data mining, Classification algorithms, decision tree.

I. INTRODUCTION

Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills. Data mining use in many areas for analysing large amount of data. Generally, data mining (sometimes called data or knowledge discovery) is the process of analysing data from different perspectives and summarizing it into useful information. Data mining software is one of a number of analytical tools for analysing data. It allows users to analyse data from many different dimensions or angles, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.

The process of discovering "hidden messages," patterns and knowledge within large amounts of data and process of making predictions for outcomes. Data mining is an automated process of extracting useful knowledge and information including patterns, associations, changes, trends, anomalies, and significant structures that are unknown from large or complex datasets.

There is an increased popularity of using data mining techniques in higher education. New field called educational data mining has emerged. Educational data mining (EDM) is describes a research field with the application of data mining, machine learning and statistics to information generated from educational settings e.g., universities and intelligent tutoring systems. 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 do not have to go to any software developer to make a separate site for being able to conduct exams online. They just have to register on the site and enter the exam details and the lists of the students which can appear in



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 1, January 2017

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 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 this study measures the student performance by using data mining technique like classification, decision tree algorithm using to build the classifier model on base on dataset composed of responses of students to courses evaluation questions.

II. LITERATURE REVIEW

Mustafa Agaoglu [1] research in educational mining focuses on modeling student's performance instead of instructors' performance. One of the common tools to evaluate instructors' performance is the course evaluation questionnaire to evaluate based on students' perception. 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, C5.0 classifier is the best with respect to accuracy, precision, and specificity. In addition, an analysis of the variable importance for each classifier model is done. 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.

Tripti Mishra, Dr. Dharminder Kumar, Dr. Sangeeta Gupta [2] use different classification techniques to build performance prediction model based on students' social integration, academic integration, and various emotional skills which have not been considered so far. Two algorithms J48 (Implementation of C4.5) and Random Tree have been applied to the records of MCA students of colleges affiliated to Guru Gobind Singh Indraprastha University to predict third semester performance. Random Tree is found to be more accurate in predicting performance than J48 algorithm.

Keno C. Peadar, Menchita Dumlaio, Melvin A. Ballera, Shaneth C. Ambat [3] predicts the employability of IT graduates using nine variables. First, different classification algorithms in data mining were tested making logistic regression with accuracy of 78.4 is implemented. Based on logistic regression analysis, three academic variables directly affect; IT_Core, IT_Professional and Gender identified as significant predictors for employability. The data were collected based on the five year profiles of 515 students randomly selected at the placement office tracer study.

Bipin Bihari Jayasingh [4] initiates a sample study that is taken for a particular institution, in the particular environment, for the particular batch and particular set of students. The sample data are collected from a classroom by distributing the questionnaire attempted by two different batches of student having questions pertaining to Inquiry based and deductive learning. The system is developed and tested twice after teaching the content using inductive method and implemented using attribute relevance, discriminant rules of class discrimination mining. The results are visualized through bar charts and shows that the two batches of learners of different years have different learning characteristics.

S. M. Merchán [5] presents and analyzes the experience of applying certain data mining methods and techniques on 932 Systems Engineering students' data, from El Bosque University in Bogotá, Colombia; effort which has been pursued in order to construct a predictive model for students' academic performance. As an iterative discovery and learning process, the experience is analyzed according to the results obtained in each of the process' iterations. Each obtained result is evaluated regarding the results that are expected, the data's input and output characterization, what



ISSN(Online): 2320-9801
ISSN(Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 1, January 2017

theory dictates and the pertinence of the model obtained in terms of prediction accuracy. Said pertinence is evaluated taking into account particular details about the population studied, and the specific needs manifested by the institution.

Konstantina Chrysaftadi and Maria Virvou [6] considered a novel approach of web-based education that performs individualized instruction on the domain of programming languages is presented. This approach is fully implemented and evaluated in an educational application module, which is called Fuzzy Knowledge State Definer (FuzKSD). In particular, FuzKSD performs user modeling by dynamically identifying and updating the student's knowledge level for all the concepts of the domain knowledge. The operation of FuzKSD is based on Fuzzy Cognitive Maps (FCMs) that are used to represent the dependencies among the domain concepts. FuzKSD uses fuzzy sets to represent the students' knowledge level as a subset of the domain knowledge.

M. Mayilvaganan, D. Kalpanadevi [7] research, the paper have been focused the improvement of Prediction/classification techniques which are used to analyze the skill expertise based on their academic performance by the scope of knowledge. Also the paper shows the comparative performance of C4.5 algorithm, AODE, Naïve Bayesian classifier algorithm, Multi Label K-Nearest Neighbor algorithm to find the well suited accuracy of classification algorithm and decision tree algorithm to analysis the performance of the students which can be experimented in Weka tool.

Cristóbal Romero [8] Educational data mining (EDM) is an emerging interdisciplinary research area that deals with the development of methods to explore data originating in an educational context. EDM uses computational approaches to analyze educational data in order to study educational questions. This paper surveys the most relevant studies carried out in this field to date. First, it introduces EDM and describes the different groups of user, types of educational environments, and the data they provide. It then goes on to list the most typical/common tasks in the educational environment that have been resolved through data-mining techniques, and finally, some of the most promising future lines of research are discussed.

III. ANALYSIS OF PROBLEM

The existing system is not user friendly. The retrieval and storing of data is not fast and data is not maintained efficiently. Moreover the graphical user interface is provided in proposed system which provide user to deal with the system very easily. Report cannot be easily generated in the existing system. The existing system requires less paper work. All the data is feed into the computer immediately and report can be generated through computer. Moreover works become very easy because there is no need to keep on paper.

The manual examination system leads to errors, more time consumption, inefficient and wastage of valuable resources. There is repetition of work in the existing system, the same data is written again and again by different branches. In existing system, managing of student's record is very tedious. Searching of students' records in manual registers, maintaining of records and reconciliation etc. are very time consuming.

Standardization of data and models. Current tools for mining data pertaining to a specific course/framework may be useful to their developers only. There are no general tools or reusing tools that can be applied to any educational system. Therefore, a standardization of input data and output model are needed.

Traditional mining algorithms need to be tuned to take into account the educational context. Data mining techniques must use semantic information when applied to educational data. The need for more effective mining tools that integrate educational domain knowledge into data mining algorithms.

IV. PROPOSED SYSTEM

We 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. We 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



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 1, January 2017

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. In the era of Information Technology, there is hardly any department which is not adopting this technology. To bring, efficiency, transparency and reliability, universities should also adopt this new technology for managing the examination system. Some of the benefits of integrated automated examination system are defined as under.

Proposed Algorithm

Decision Tree:

A decision tree is a structure that includes a root node, branches, and leaf nodes as show in figure. 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.

Decision tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy). Leaf node (e.g., Play) represents a classification or decision. The topmost decision node in a tree which corresponds to the best predictor called root node. Decision trees can handle both categorical and numerical data.

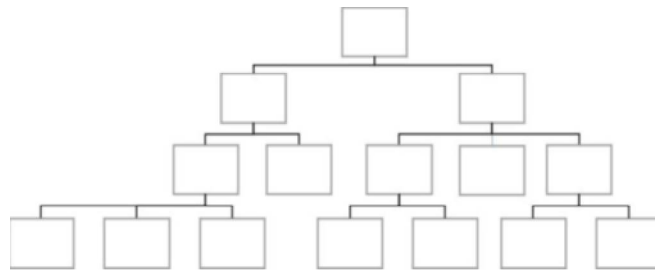


Fig: Decision Tree

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.

V. CONCLUSION

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 student's performance. In this paper, data mining is utilized to analyse course evaluation questionnaires. Here, the most important variables that separate "satisfactory" and "not satisfactory" student performances and there weakness' in particular subject or field. Hopefully, these can help instructors to improve their performances.

Tree-based methods classify instances by sorting the instances down the tree from the root to some leaf node, which provides the classification of a particular instance. Each node in the tree specifies a test of some attribute of the instance and each branch descending from that node corresponds to one of the possible values for this attribute.

Decision tree algorithms, support vector machines, artificial neural networks, and discriminant analysis– different classifiers are used for performance measures. This finding indicates the effectiveness of using data mining techniques



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 1, January 2017

in course evaluation data and higher education mining. It is hoped this study can contribute to the literature in two major areas: data mining, and higher education.

REFERENCES

1. Mustafa Agaoglu, "Predicting Instructor Performance Using Data Mining Techniques in Higher Education," *IEEE Access*, Volume: 4, 2016.
2. Tripti Mishra, Dr. Dharminder Kumar, Dr. Sangeeta Gupta, "Mining Students' Data for Performance Prediction," in fourth International Conference on Advanced Computing & Communication Technologies, 2014.
3. Keno C. Piad, Menchita Dumlaio, Melvin A. Ballera, Shaneth C. Ambat, "Predicting IT Employability Using Data Mining Techniques," in third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC), 2016.
4. Bipin Bihari Jayasingh, "A Data Mining Approach to Inquiry Based Inductive Learning Practice In Engineering Education," in IEEE 6th International Conference on Advanced Computing, 2016.
5. S. M. Merchán, "Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic," *IEEE Latin America Transactions*, vol. 14, no. 6, June 2016.
6. Konstantina Chrysafiadi and Maria Virvou, "Fuzzy Logic for adaptive instruction in an e-learning environment for computer programming," *IEEE Transactions on Fuzzy Systems*, Volume: 23, Issue: 1, Feb. 2015.
7. M. Mayilvaganan, D. Kalpanadevi, "Comparison of Classification Techniques for predicting the performance of Students Academic Environment," in International Conference on Communication and Network Technologies (ICCNT), 2014.
8. Cristóbal Romero, "Educational Data Mining: A Review of the State of the Art," *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews*, Vol. 40, No. 6, November 2010.
9. Behrouz Minaei-Bidgoli, Deborah A. Kashy, Gerd Kortemeyer, William F. Punch, "Predicting Student Performance: An Application Of Data Mining Methods With An Educational Web-Based System," in 33'd ASEE/IEEE Frontiers in Education Conference, T2A-13, November 5-4, 2003.
10. Peter Brusilovsky, Sibel Somyürek, Julio Guerra, Roya Hosseini, Vladimir Zadorozhny, Paula J. Durlach, "Open Social Student Modeling for Personalized Learning," *IEEE Transactions on Emerging Topics in Computing*, Volume: 4, Issue: 3, July-Sept. 2016.
11. Pedro G. Espejo, Sebastián Ventura, and Francisco Herrera, "A Survey on the Application of Genetic Programming to Classification," *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews*, Vol. 40, No. 2, March 2010.
12. Brent Harrison, Stephen G. Ware, Matthew W. Fendt, and David L. Roberts, "A Survey and Analysis of Techniques for Player Behavior Prediction in Massively Multiplayer Online Role-Playing Games," *IEEE Transactions on Emerging Topics in Computing* Volume: 3, Issue: 2, June 2015.
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)," in IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), 2013.
14. Chia-Yin Ko, Fang-Yie Leu, "Applying data mining to explore students' self-regulation in learning contexts," in IEEE 30th International Conference on Advanced Information Networking and Applications (AINA), 2016.
15. Kin Chew Lim, Low, S, Samir, A, Cheang, P., "A Model For Teaching, Assessment And Learning In Engineering Education For Working Adults," in IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE), 20-23 Aug. 2012, pp. T1A-15 - T1A-17.
16. Priyanka Anandrao Patil, R. V. Mane, "Prediction of Students Performance Using Frequent Pattern Tree," in sixth International Conference on Computational Intelligence and Communication Networks, 2014.

BIOGRAPHY

Ankita A Nichat is a Student of M.E Computer Science & Engineering Department, H.V.P.M'S College of Engineering & Technology, Amravati, Maharashtra, India. She received Bachelor of Engineering Degree in 2015 from SGBAU Amravati, Maharashtra, India. Her research interests are Education technology and Data Mining.

Dr. Anjali B Raut is a HOD of Department of Computer Science & Engineering, H.V.P.M'S College of Engineering & Technology, Amravati, Maharashtra, India.