

Analyze Faculty Performance System using Data Mining Techniques

Mr. Jitendra Agrawal¹, Dr. Ravindra Tiwari²

Student, Department of Computer Application. LNCT University Bhopal, India¹

Professor, Department of Computer Application. LNCT University Bhopal, India²

Abstract: *The development of an improved and intelligent model for the evaluation of instructors' performance in higher institutions uses the efficient data mining techniques considering the drawbacks of the prior traditional techniques. This proposed system analyses the factors related with the evaluation of instructors teaching performance using predictive data mining techniques known as regression statistical model. Regression is a data mining predictive technique that is used to make statistical prediction of the variables, given a set of data. Consequently, the evaluation of instructors' performance is useful for the academic institutions as it helps to make effective managerial decisions, improve the quality, reliability and efficiency of the instructors, provides a basis for the performance improvement that will optimize students' academic outcomes and improve standard of education and contribute to successful accomplishment of the organizational goals. The knowledge is hidden among the educational data set and it is extractable through data mining techniques. Present paper is designed to justify the capabilities of data mining techniques in context of higher education by offering a data mining model for higher education system in the university. In this research, the classification task is used to evaluate student's performance and as there are many approaches that are used for data classification, the decision tree method is used here. By this task we extract knowledge that describes students' performance in end semester examination. It helps earlier in identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counseling.*

Keywords: Prediction algorithms, Statistical Regression Model, Linear Regression algorithm, Performance Evaluation, Educational Data Mining (EDM).

I. INTRODUCTION

The advent of information technology in various fields has led the large volumes of data storage in various formats like records, files, documents, images, sound, videos, scientific data and many new data formats. The data collected from different applications require proper method of extracting knowledge from large repositories for better decision making. Knowledge discovery in databases (KDD), often called data mining, aims at the discovery of useful information from large collections of data [1]. The main functions of data mining are applying various methods and algorithms in order to discover and extract patterns of stored data [2]. Data mining and knowledge discovery applications have got a rich focus due to its significance in decision making and it has become an essential component in various organizations. Data mining techniques have been introduced into new fields of Statistics, Databases, Machine Learning, Pattern Reorganization, Artificial Intelligence and Computation capabilities etc. One of the most common problems in higher educational sectors is the efficient evaluation of instructors' performances in a particular course. The most eminent tool used to evaluate the instructors' performance in a particular course is through surveying various students' responses about the course and its instructor through a questionnaire system. When evaluations of instructor performance were introduced to higher education systems, there has been a debate on the reliability, accuracy and validity of these evaluations.

Hidden information and hidden patterns in large datasets can be efficiently analysed with data mining techniques. It is also known as knowledge discovery process, which involves discovering hidden patterns, messages and knowledge within large datasets and the process of analysing outcomes or behaviours. It is a systematic process of extracting relevant knowledge and information which includes associations, changes, structures and trends that are unknown from large structured and unstructured data. It is an eminent tool in the field of academics.

Data mining techniques are used in higher education, because of its various abilities such as efficient allocation of resources, academic planning, predicting student and instructor performance, and promotion of alumni development. Educational data mining (EDM) deals with developing methods for exploring data from educational sectors with the purpose of providing quality education to students and to make effective managerial decisions. With EDM, the useful information can be gained from various entities such as lecturers, students, and alumni. This information can be used to allocate resources and assign staff more effectively, make effective decisions on educational academic activities to improve students' performance, increase students' learning behavior, increase student's retention rate, and increase students' consistency in various academic activities.

Received: 2 Nov. 2022

Revised: 18 Nov. 2022

Final Accepted for publication: 28 Nov 2022

Copyright © authors 2022

EDM can also be used in discovery and improving domain models by modification of the existing models.

II. LITERATURE SURVEY

Data mining, also popularly known as Knowledge Discovery in Database, refers to extracting or “mining” knowledge from large amounts of data. Data mining techniques are used to operate on large volumes of data to discover hidden patterns and relationships helpful in decision making. While data mining and knowledge discovery in database are frequently treated as synonyms, data mining is actually part of the knowledge discovery process. The sequences of steps identified in extracting knowledge from data are shown.

Two types of performance evaluation methods used in the literature review are formative and summative evaluations. During earlier days, instructors were evaluated based on certain typical characteristics which may or may not have been related to performance, and yet no knowledge confirms that the teaching performance is dependent on several characteristics. Hence, this evaluation technique was eliminated.

Formative performance evaluation method is referred as qualitative evaluation on the instructors’ teaching assessment, which mainly deals with identifying strengths, weaknesses and providing appropriate professional development opportunities for particular areas which need significant improvements. Formative performance evaluation technique is based on the criteria such as classroom observations and student assessment report as important tools to measure the staff performance. The overall intention of this method is to provide useful informative feedback about a particular faculty for assisting them in improving the effectiveness and efficiency of their teaching methodologies.

Summative performance evaluation technique is illustrated as a source of documentation and a reliable way for the evaluation of instructors’ performance by providing a set of statements regarding instructor’s capabilities through regular inspection, examination and various informative interviews, in order to measure aptitude skills, effectiveness and knowledge to ensure that the required pre-defined standards are met, and to promote the level of performance. This method is used to determine the worth and career advancement of an instructor, and ensures that instructors adopt the appropriate actions and best practices which improves student performance. Hence, summative evaluation method is a source of useful documentation. In many institutions, annual performance evaluation report (APER) and informative interviews are used as effective tools to measure their quality, performance and effectiveness.

Han and Kamber [3] describes data mining software that allow the users to analyze data from different dimensions, categorize it and summarize the relationships which are identified during the mining process.

Pandey and Pal [5] conducted study on the student performance based by selecting 600 students from different colleges of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayes Classification on category, language and background qualification, it was found that whether new comer students will performer or not.

Hijazi and Naqvi [6] conducted as study on the student performance by selecting a sample of 300 students (225 males, 75 females) from a group of colleges affiliated to Punjab university of Pakistan. The hypothesis that was stated as "Student's attitude towards attendance in class, hours spent in study on daily basis after college, students' family income, students' mother's age and mother's education are significantly related with student performance" was framed. By means of simple linear regression analysis, it was found that the factors like mother’s education and student’s family income were highly correlated with the student academic performance.

Khan [7] conducted a performance study on 400 students comprising 200 boys and 200 girls selected from the senior secondary school of Aligarh Muslim University, Aligarh, India with a main objective to establish the prognostic value of different measures of cognition, personality and demographic variables for success at higher secondary level in science stream. The selection was based on cluster sampling technique in which the entire population of interest was divided into groups, or clusters, and a random sample of these clusters was selected for further analyses. It was found that girls with high socio-economic status had relatively higher academic achievement in science stream and boys with low socioeconomic status had relatively higher academic achievement in general.

Galit [8] gave a case study that use students data to analyze their learning behavior to predict the results and to warn students at risk before their final exams.

Al-Radaideh, et al [9] applied a decision tree model to predict the final grade of students who studied the C++ course in Yarmouk University, Jordan in the year 2005. Three different classification methods namely ID3, C4.5, and the

NaïveBayes were used. The outcome of their results indicated that Decision Tree model had better prediction than other models.

Pandey and Pal [10] conducted study on the student performance based by selecting 60 students from a degree college of Dr. R. M. L. Awadh University, Faizabad, India. By means of association rule they find the interestingness of student in opting class teaching language.

Ayesha, Mustafa, Sattar and Khan [11] describes the use of k-means clustering algorithm to predict student’s learning activities. The information generated after the implementation of data mining technique may be helpful for instructor as well as for students.

Bray [12], in his study on private tutoring and its implications, observed that the percentage of students receiving private tutoring in India was relatively higher than in Malaysia, Singapore, Japan, China and Sri Lanka. It was also observed that there was an enhancement of academic performance with the intensity of private tutoring and this variation of intensity of private tutoring depends on the collective factor namely socioeconomic conditions.

Bhardwaj and Pal [13] conducted study on the student performance based by selecting 300 students from 5 different degree college conducting BCA (Bachelor of Computer Application) course of Dr. R. M. L. Awadh University, Faizabad, India. By means of Bayesian classification method on 17 attribute, it was found that the factors like students’ grade in senior secondary exam, living location, medium of teaching, mother’s qualification, students other habit, family annual income and student’s family status were highly correlated with the student academic performance..

In recent years, teacher evaluation is used to measure the pedagogical effectiveness of teaching and the quality of education. During the earlier days, educational sectors used the evaluation results to assess the performance and helping teachers improve their pedagogical effectiveness. These results are also used to implement other ideas by providing the required information to other entities in educational sec. Thus the admin uses these ratings to make important decisions like, hiring new staff, making promotions, rewarding awards, assigning the appropriate staff, and other managerial decisions. Thus the instructors use these results to improve their pedagogical effectiveness and to monitor their performance Students use these dimensions to select courses and to choose teachers for rewards. Though various studies have been carried out, there is still no guarantee about the reliability, validity, usefulness and dependability of these techniques. It is subjected to debatable issues regarding the accuracy of these methods.

The main issue encountered is the relationship between assessments and learning claims that an average relationship exists between assessments and learning. But this relation is situational and is not applicable to all cases of academic sectors and all the educational levels. It is claimed that, the level of measurement of learning, affects the relation between learning and assessment.

III. SYSTEM DESIGN

Student’s feedback evaluations are used to improve the course result and increase the level of teaching quality. It is also used for evaluation process for staff appraisal systems. Several prior research have identified that student evaluation offers a reliable and valid assessment of the instructors. Students’ evaluation is collected and stored in the database, and this huge dataset is analysed using various data mining techniques and further used for managerial purposes. Figure 1 depicts the overall dataflow of the performance evaluation system.

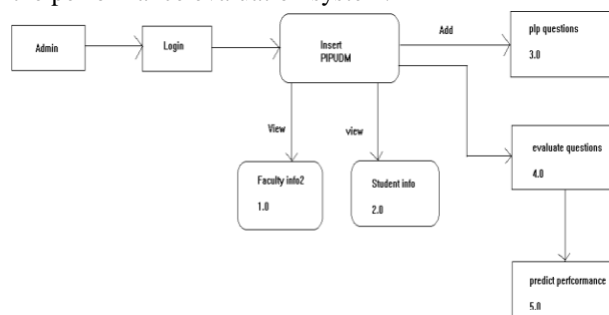


Figure 1: Overall dataflow diagram

A. Data collection and variables:

Received: 2 Novt. 2022

Revised: 18 Nov. 2022

Final Accepted for publication: 28 Nov 2022

Copyright © authors 2022

The performance evaluation system consists of data with 26 variables. The responses are measured on an interval scale. Response option values of these 26 questions are "Never", "Rarely", "Sometimes", "Often", "Always" that have different weightage such as 1, 2, 3, 4, 5 respectively for the questions 1 to 4 questions. The responses 1, 2, 3, 4, 5 Represents "Strongly disagree", "Disagree", "Neutral", "Agree", and "Strongly agree" respectively for the questions 5 to 25. The last variable is measured on a nominal scale in the form of {1, 2} where 1 stands for "Not satisfactory" and 2 stands for "Satisfactory". Details of the variables are as follows:

- 1) How frequently you attended this course?
- 2) How frequently did the instructor attend this course?
- 3) Did the instructor come to class on time?
- 4) Did the instructor use the class hours fully?
- 5) The instructor's knowledge of the subject was adequate?
- 6) The instructor came to class well prepared.
- 7) Lectures were clear and well structured?
- 8) The syllabus was followed and fully covered
- 9) There was a good balance between theory and application.
- 10) The instructor was effective in presenting the material in lectures and discussions.
- 11) Subjects were not well organized and did not follow each other.
- 12) The instructor answered questions satisfactorily.
- 13) The instructor was enthusiastic about this course.
- 14) The instructor motivated the students to ask questions and to get involved in the discussions.
- 15) The instructor was approachable, nice and easy to communicate with?
- 16) The instructor was available to give help outside the class.
- 17) The course materials (textbooks, handouts etc) were satisfactory.
- 18) The assignments (case studies, homework, projects, presentations etc) were relevant and helpful.
- 19) This was an easy AA course?
- 20) The assignments/exams increased the ability to think creatively.
- 21) The assignments/exams adequately tested the knowledge taught.
- 22) I had to spend a lot of time to get prepared for assignments/exams.
- 23) This course was helpful in increasing my knowledge and interest in the subject matter.
- 24) I learned a lot in this course.
- 25) The instructor's grading (assignments/exams) was fair.
- 26) The overall performance of the instructor was satisfactory.

B. Data Pre-processing:

Pre-processing of the data is an important phase in the data mining process. The raw data is more prone to inconsistent and inaccurate data. Hence the raw data is pre-processed to increase the quality of data. It includes tasks such as data cleaning, selection, transformation and integration. It is the precise collection of data being used for appropriate reason from different origin. It has been recognized, registered and classified. The scope of collecting the data is to retrieve information, to preserve data, to form decisions regarding crucial problems and transmission of data to the residuals. The considered dataset consisting of 600 records and students academic details is uploaded to the database by the faculty. It is pre-processed manually by checking the attribute entries in the Microsoft excel format. This data is pre-processed to remove the noisy and inconsistent data which are not required by our analysis. This data is then stored in the WAMP server for further processing. It is often used for development of web interface and testing purposes.

C. Data Mining Model:

After performing the required pre processing on the data, the transformed data is fed into the suitable data mining model which will carry out the analysis. In this case, the prediction technique is used to predict the staff performance by computing the analysis using the regression statistical model. This model identifies the related variables and forms a suitable equation. The evaluation of the data is performed and the final result is represented in the graphical manner.

D. Predictive analytics technique:

Predictive analysis deals with prediction of future trends and frequently occurring behavior. It comes under statistics. The main idea lies in identifying the relation between the relevant variables previously occurred. This relation is then exploited to predict the final outcome. The assumptions made and the analysis level greatly impacts the accuracy or the

result. It is used in credit card fraud, business applications, forecasting.

It deals with generating the probabilities for each entity at a very coarse and detailed level so that all the aspects are covered. It is useful in driving better decisions in the future since it learns from the past patterns. It is useful in optimization and prevents various issues that might occur in the future. The final data can also be used for improvement of the loop life cycle. It is also known as decision modeling or descriptive modeling or optimization. It is widely used in business applications to make effective decisions.

The models deal with the relation between the capacity of the units and characteristics of the units. The main aim here is to identify that the units exhibit different behavior in every different sample. It is used widely in applications that require live transactions such as customer relation management to make decisions. With latest innovations, predictive analysis is also used to capture the behavior of the human in specific scenarios.

The training example consists of the known information such as attributes and performances. The other samples consisting of unknown performances are known as out of training examples. The predictive technique used in this evaluation system is the regression statistical model.

E. Regression Statistical model:

Regression is a data mining predictive technique that is used to make statistical prediction of the variables, given a set of data. Used to predict numerical or a continuous value. It is used in multiple areas such as business marketing and planning, applications that need forecasting, environmental modelling and analysis of trends.

Regression analysis is a step-wise statistical process. It deals with identifying the relationship between dependent and independent variables. It provides an outlook of how the values of the variables change, when any one of the variables is varied. It calculates the average mean value of the dependent variable when the independent variables are constant. The variation of values of the dependent variable observed in the regression function is known as probability distribution. It is widely used for applications that use forecasting, prediction and machine learning.

The regression function consists of finite unknown parameters. Nonparametric regression consists of infinite parameters. The data generating process and its relationship with the type of regression approach being used is responsible for the final performance of the regression analysis. Simple linear regression falls under the category of parametric regression.

F. Linear regression algorithm:

Linear regression is used for predictive analysis. The main things to be observed are:

1. The capability of the set of variables in the prediction of a result variable.
2. Analysing the variables that are dependent variable predictors and observing their influence on that particular dependent variable. These observations are used to find the relationship between the dependent and independent variables.
3. The prediction of the final outcome variable is determined by the regression equation which consists of predictor variables. The simplest equation with a dependent and an independent variable is given by: $y = c + b * x$ where $y =$ dependent variable, $c =$ constant, $b =$ regression coefficient and $x =$ independent variable.

IV. RESULT AND ANALYSIS

The sequences of events under the staff module are:

1. The staff can login and add their branch, course, and subject details.
2. All the informative data about the staff will be stored in the database for further evaluation by the regression analysis.
3. The staff can view their individual performance and percentage in a graphical representation.

The sequences of events under the staff student are:

1. The students can login with their login-id and password.
2. The students can select the particular staff and give the feedback according to the questionnaire.
3. The feedback is fed to the regression analysis and performance evaluation is computed.
4. The students can view the feedback of the staff in a graphical representation.

The admin can login and search through the subjects and view the details such as:

1. The names of the students who have submitted the feedback according to the questionnaire in the feedback form.
2. The names of the staff for whom the feedback has been submitted by the students according to the feedback form.
3. The total points obtained by each staff after performing the regression analysis.

Figure 2 represents the individual performance evaluation graph. Figure 3 represents the performance evaluation graph by questions.

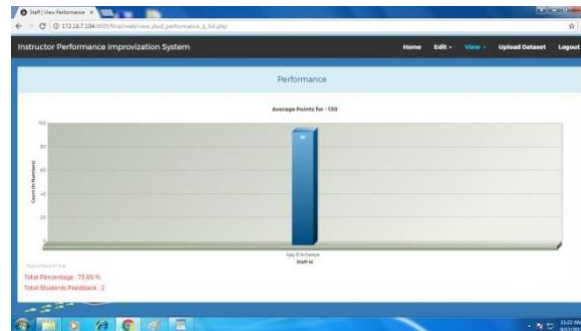


Figure 2: Individual Performance Evaluation graph

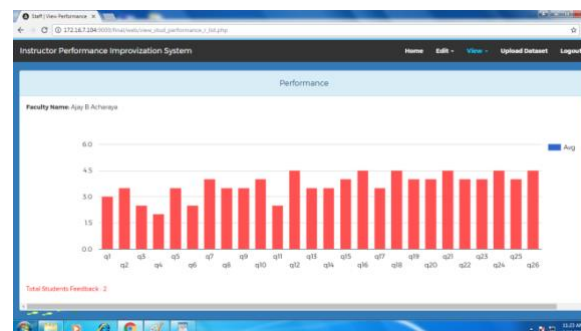


Figure 3: Performance Evaluation graph by question.

V. CONCLUSION

Data mining techniques are applied in higher education to give appropriate solutions to educational and administrative problems in order to increase capability and consistency. The need for an intelligent statistical model for performance evaluation of instructors in higher institutions all over the world has become necessary in order to find an efficient solution to the limitations of the old traditional methodologies. Using factors and resources obtained from course evaluation questionnaire, a system framework for the instructors' performance evaluation system is presented. The framework is designed with some basic components which satisfy the reliability and efficiency. This method examines the factors associated with the assessment of teaching performance and promotes them to reach the highest level of performance quality. The proposed system will assist the higher administrators in making appropriate managerial decisions, which will optimize students' academic outcomes and improves the standard of education. Consequently, this will lead to the accomplishment of objectives and goals.

REFERENCES

- [1] C. Guarini, and B. Stacy, Review of Gathering Feedback for Teachers. Boulder, CO: National Education Policy Center, 2012. Retrieved from http://nepc.colorado.edu/files/pb-options-1-teval_1.pdf.
- [2] E.D. Nakpodi, "A Critique of the Methods of Evaluating the Competency of Lecturers in Nigerian Tertiary Institutions"- African Journal of Education and Technology, volume 1 Number 1, April 2011; pp. 53-59.
- [3] J.L Steele, L.S. Hamilton, and B.M. Stecher, Incorporating Student Performance Measures into Teacher Evaluation Systems. Palo Alto, CA: RAND. ISBN: 978-0-8330-5250-6, 2010. Retrieved from: http://www.rand.org/content/dam/rand/pubs/technical_reports/2010/RAND_TR9
- [4] Kulikk, J.A. (2001). Student ratings: validity, utility, and controversy. In: Theall M., Abrami, P.C. Mets, L.A. (Eds.), New Directions for

- Institutional Research. Jossey-Bass, 109, 9–24.17
- [5] Claayson, D.E. (2009). Student evaluations of teaching: are they related to what students learn? A meta-analysis and review of the literature. *Journal of Marketing Education*, 31(1), 16–30. *Higher Education*, *Int. J. Adv. Compu. Sci. Appl.*, Vol. 2, no. 3, pp. 80_84 2011.
- [6] Heikki, Mannila, *Data mining: machine learning, statistics, and databases*, IEEE, 1996.
- [7] U. Fayadd, Piatiesky, G. Shapiro, and P. Smyth, *From data mining to knowledge discovery in databases*, AAAI Press / The MIT Press, Massachusetts Institute Of Technology. ISBN 0–262 56097–6, 1996.
- [8] J. Han and M. Kamber, “Data Mining: Concepts and Techniques,” Morgan Kaufmann, 2000. [4] Alaa el-Halees, “Mining students data to analyze e-Learning behavior: A Case Study”, 2009..
- [9] U . K. Pandey, and S. Pal, “Data Mining: A prediction of performer or underperformer using classification”, (IJCSIT) *International Journal of Computer Science and Information Technology*, Vol. 2(2), pp.686-690, ISSN:0975-9646, 2011.
- [10] S. T. Hijazi, and R. S. M. M. Naqvi, “Factors affecting student”’s performance: A Case of Private Colleges”, *Bangladesh e-Journal of Sociology*, Vol. 3, No. 1, 2006