Article can be accessed online at http://www.publishingindia.com

Extraction of Actionable Knowledge to Predict Students' Academic Performance using Data Mining Technique- An Experimental Study

K. Javubar Sathick*, A. Jaya**

Abstract

Knowledge discovery in academic institution becomes more critical and crucial in terms of identifying the student's performance. In the extraction of actionable knowledge from a large database the data mining plays a vital role. The actionable knowledge extraction provides a interestingness and meaning to the mined data. This paper focuses on the prediction of the student's academic performance from the large student database. The mining algorithm like clustering and classification algorithm is revisited to predict the performance after initial mining of raw data. The main scope of this paper is to reveal the outcome of the performance analysis of a student .This work will help the university to reach betterment in providing the quality input to the student community and impart the knowledge effectively.

Keywords: Actionable Knowledge, Classification, Clustering, Prediction and Analysis.

1. Introduction

The extensive scope of growth in data creates and initializes the various new techniques, algorithms, procedures and methods to access them reliably, to identify the hidden patterns and to extract the actionable knowledge. Since the size of database is too large the data mining process is carried out with utmost concentration to reach feasible solution.

Now-a-days the data stored in educational database increases rapidly. These databases contain hidden information for improvement of students' performance. This academic performance is influenced by many factors, therefore it is essential to develop predictive data mining model for students' performance so as to identify the difference between high learners and slow learners student. In the present investigation, an experimental methodology was adopted to generate a database. The raw data was preprocessed in terms of filling up missing values, transforming values in one form into another and relevant attribute/ variable selection. Many universities are trying to fetch the information about the student. Predicting the performance of the students is a challenging task. The main goal is to improve quality and standard of education in the university. This work also helps in growth of institution as well as the student. It helps to view both academic as well as personal details of the student which would be also helpful in evaluating student's overall performance. This work illustrates graphical view of the student performance, overall mark and it helps to check whether they are eligible for placement or not.

Student evaluation has become a part of most institutions. This helps to provide a feedback for instructors to improve their teaching practices which would result in improved student performance. It also helps the administrator of the institutions to keep track of their institutions performance.

As the resource pool grows, the information from students who have multiple transactions with these resources also increases. The knowledge extraction techniques used as all data of student personal details to academic records in either in flat file or in excel form.

* Department of Computer Applications, BSA University, Chennai, Tamil Nadu, India. E-mail: javubar@gmail.com
** Professor, B.S. Abdur Rahman University, Chennai, Tamil Nadu, India. E-mail: jaya@bsauniv.ac.in

2. Literature Review

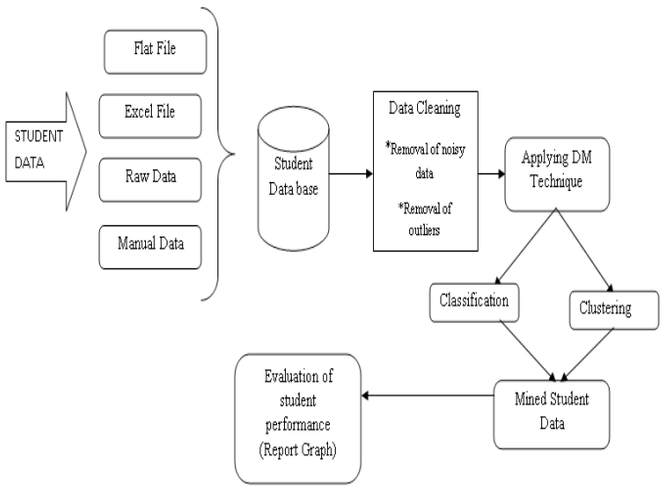
This section emphasizes some of the existing strategy for this study. In [1], the technique which is proposed is to reduce the drop put ratio to the significant level and improve the performance level of the university. Bayesian classification method is used on student data base to predict the student performance according to division on the basis of previous year database. This method is used to reduce the failing ratio of the student. In [2], student information is extracted from student database. This technique is used to identify the student failure as soon as possible before the end semester examination. The students who needs a special attention on studies and allow the teachers to provide advise/counsel them.The data mining technique used in this paper is data classification, the decision tree method. It is used to reduce the failing ratio and taking appropriate action to score good marks in next semester examination. In [3], the important features that are proposed are used to predict the performance of the faculty. It is done by gathering the feedback produced by the students using linear regression technique. The result is generated in this paper are strictly based on the student feedback.

In [4], the author introduces and compares some technique used to predict the performance of the student in the university. The result has been improved by comparing with data mining techniques. In [6], the methodology is proposed by using the state-of–the-art recommender system techniques for predicting student performance. The matrix factorization method is proposed to implicitly take into account two latent factors for predicting student performance. In [7], the author proposed to evaluate the faculty performance by signed and unsigned student feedback using linear regression technique. It has to be decided that how much of significance has to be associated to the student feedback in making certain decisions. Number of parameter that is concentrated for evaluated is regularity, performance, attitude, dedication, self learning initiatives of the student. In[8],apriori algorithm is used to mine the student and placement based data. The outcome of this study provides suitable analysis of student's performance and placement possibilities. It is evident that the various factors are involved in determining the overall performance of a student. Abdul fatah et al [9], briefs about the decision support system based on absorbing Markova model which targets the specific department. The target data here is student's transition enrollment data. State transition diagram is generated for student flow which enhances DSS. The semester performance and specific academic performance alone is predicted in this work. Brijesh et al [10] ensures that , Bayesian classification method is used on student database to predict the students division on the basis of previous year database. This study will help to the students and the teachers to improve the performance of the student. This study will also work to rectify those students which needed special attention to reduce failing ration and taking right action at right time. Althaf hussain et al[11] Introduced a method to predict the intra-year academic performance of the student using the historic data. The scope is to identify existing patterns in the historic data, and maintain a database for it. Comparing the current performance of a student with the existing patterns, the possible performance of the student in the future is predicted. In this process, the author identifies any new patterns that are in existence. For the purpose of identifying the interestingness of the pattern they look into the percentage of increase or decrease, appropriate methodologies are used. Mohammed M. Abu Tair et al[12] proposed a case study which describes about how to extract useful knowledge from graduate students data collected from the college of Science and Technology – Khanyounis. The data include fifteen years period [1993-2007]. After preprocessing the data, the data mining techniques is used to discover association, classification, clustering and outlier detection rules. In each of these four tasks, it represents the extracted knowledge and describes its importance in educational domain. Edin Osmanbegović et al[13] proposes, three supervised data mining algorithms which is applied on the preoperative assessment data to predict success in a course (either passed or failed) and the performance of the learning methods were evaluated based on their predictive accuracy, ease of learning and user friendly characteristics. The results indicate that the Naïve Bayes classifier outperforms in prediction decision tree and neural network methods. It has also been indicated that a good classifier model has to be both accurate and comprehensible for professors. This study was based on traditional classroom environments, since the data mining techniques were applied after the data was collected. Dorina kabakchieva et al[14] analysis the initial results from a data mining research project implemented at a Bulgarian university, aimed at revealing the high potential of data mining applications for university management. The classification technique is used in this study to measure

the performance of the student. Bayesian classifier and k-nearest neighbor classifier provides the result with the variation of 52-67%. The conclusions made from the conducted research will be used for defining the further steps and directions for the university data mining project implementation, including possible transformations of the dataset, tuning the classification algorithms' parameters, in order to achieve more accurate results and to extract more important knowledge from the available data.

3. Proposed Framework for Performance Prediction and Analysis of Student

Figure 1: Proposed Framework for Perfomance Prediction



4. Methodology

Data Mining Definition And Techniques: 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.

Classification: Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. This approach frequently employs decision tree or neural network-based classification algorithms. The data classification process involves learning and classification. In Learning the training data are analyzed by classification algorithm. In classification test data are used to estimate the accuracy of the classification rules. If the accuracy is acceptable the rules can be applied to the new data tuples. The classifier-training algorithm uses these pre-classified examples to determine the set of parameters required for proper discrimination. The algorithm then encodes these parameters into a model called a classifier.

Decision Tree: Decision tree is tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).

5. Implementation

The implementation work is based on the collected data which possess various data mining aspects. The Student data is taken into account for the performance prediction. The proposed research work is categorized into various modules. This research work is carried out with the inclusion of data mining technique and implementation software. The proposed work speculated as the useful application where, student's performance can be viewed and placement criteria for various concerns are listed efficiently. The student performance is generated as a report graph which can be used as a survey to improve the student performance in the future.

The student's and other user can login with their given credentials onto the desired application which modeled to predict the student performance and their placement criteria. The students, faculty and other user can view the past and recent performance of the respective courses to reexamine their current performance.

The figure 2. in [5] shows the statistical information about various companies' eligibility criteria which pursuits the students performance in their curriculum.

Figure 2: Placement Criteria

The actual status of the student's is viewed through the following figure 3. in [5] which represents whether a student is matching the eligibility criteria of various corporate.

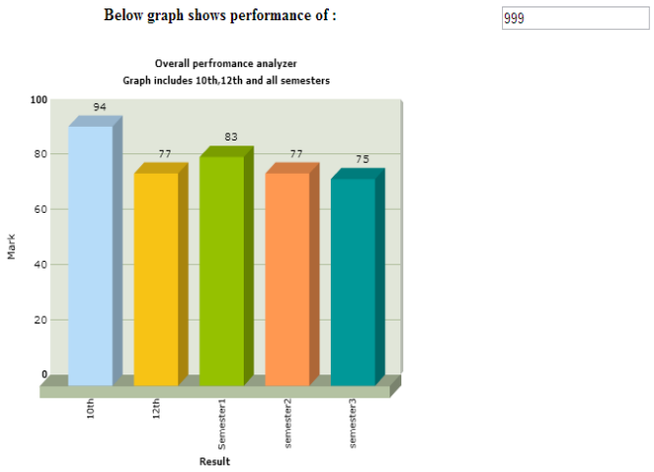
Figure 3: Student Eligibility Criteria

The figure 4. in [5] reveals the optimal performance ranking of a student, after considering the various factors such as academic performances, extracurricular activity and placement criteria. The ranking is based on their accumulation of all the performances. The performer can be of various classes such as optimal performer, intermediate performer so on..,

Figure 4: Performance Ranking

The overall performance is represented as performance chart or performance graph which reveals the effective data which can be used by the institution for the efficient improvement in the performance of the student in the future. Figure 5. in [5] illustrates the overall performance of the student.

Figure 5: Overall Performance Graph



6. Conclusion

In this study, student's performance is predicted and graph is generated to improve efficiency and make the student apt for the better placement. This study will also work to identify those students who in need of special attention to increase their performance. Predicting students' academic performance is great concern to the higher education. With the help of classification and clustering technique the performance of student is identified to a maximum extent, and the result obtain through this research work reveals the positive outcome of student involvement in improving university quality. Classification technique is used to classify the student according to their academic results. They classified as average performer, intermediate performer and better performer. This experimental study can be further expanded which meets lot more academic constraints which creates effective impact in the overall outcome of the student and institution.

References

1. Pandey, U. K. & Pal, S. (2011). Data Mining: A prediction of performer or underperformer using classi-

fication. *International Journal of Computer Science and Information Technologies*, 2(2), 686-690.

2. Lakshmi, T. M., Martin, A., Begum, R. M. & Venkatesan, V. P. (2013). An analysis on performance of decision tree algorithms using student's Qualitative data. *International Journal of Modern Education and Computer Science*, June, 5(5), 18-27.
3. Singh, C., Gopal, A. & Mishra, S. (2011). Management faculty performance evaluation with signed and unsigned student feedback using linear regression technique. *International Journal of Information Technology and Knowledge Management*, 4(2), 591-594.
4. Thai-Nghe, N. Busche, A. & Schmidt-Thieme, L. (2009). *Improving Academic Performance Prediction by Dealing with Class Imbalance*. International Swaps and Derivatives Association, (pp. 878-883).
5. Kumutha, S. & Sathick, K. J. (2013). *Performance Prediction and Analysis of a University using Data Mining Technique*. National Conference on Advanced Computing Technology.
6. Thai-Nghe, N., Drumond, L. & Horv´ath, T. (2011). Matrix and Tensor Factorization for Predicting Student Performance. 3rd International Conference on Computer Supported Education.
7. Sinhgad, C. S., Gopal, A. & Mishra, S. (2012). Faculty performance prediction from student feedback using linear regression technique. *International Journal on Computational Sciences and Applications*, 2(4), 1-4.
8. Shreenath, A. & Madhu, N. (2012). Discovery of students' academic patterns using data mining techniques. *International Journal on Computer Science and Engineering*, 4(6), 1054-1062.
9. Mashat, A. F. S. & Khedra, A. M. (2012). *Decision Support System Based Markov Model for Performance Evaluation of Students Flow in FCIT-KAU*. International Conference on Communication and Information Technology.
10. Bhardwaj, B. K. & Pal, S. (2011). Data Mining: A prediction for performance improvement using classification, *International Journal of Computer Science and Information Security*, 9(4), 136-140.
11. Basha, S. K. A. H. & Kumar, Y. R. R. (2012). Predicting student academic performance using temporal association mining. *International Journal of Information Science and Education*, 2(1), 21-41.
14. Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61-72.
13. Osmanbegović, E. & Suljić, M. (2012). Data mining approach for predicting student performance. *Economic Review - Journal of Economics and Business*, 10(1), 3-12.
12. Tair, M. M. T. & El-Halees, A. M. (2012). Mining educational data to Improve student's performance: A case study. *International Journal of Information and Communication Technology Research*, 2(2).