

Comprehensive Analysis of Academic Performance using Data Mining Techniques-with Special Reference to UG Class of Engineering

Ms. Juhi Arora¹, Dr. Manali Kshirsagar², Prof. Rupali Burde³

^{1, 2, 3.} Computer-Science & Engineering, Rajiv Gandhi College of Engineering, Nagpur

Abstract: *One of the main objectives of Indian educational system is evaluating or enhancing the educational organization. Data Mining is the process of searching the concealed information from a large quantity of data set. It analyzes the data from different source and it converts into meaningful information. There are a lot of advantages of data mining technique in education sector. It has been one of valuable resources to determine success and failure of an organization. In an educational system, there is a need to manage and analyze the data for organization's performance. In this paper, we have discussed the approach regarding analyzing student's result data and professor's result data, in order to monitor the progression of academic performance of students and professors for the purpose of making an effective decision by the academic planners. This leads to better results thereby increasing the profitability of the educational institutions especially in the private institutions.*

Keywords: *Data Mining (DM), Educational data mining (EDM), Educational systems, Association rule-based algorithm, FP-Growth algorithm.*

I. INTRODUCTION

Data Mining is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD. Educational data mining (EDM) is the latest advancement in education field, dealing with the development of techniques for exploring and analyzing the huge data that come from the educational database. This involves analyzing the results in depth in order to monitor the student's academic activities closely. The advent of information technology in various fields has lead the large volumes of data storage in various formats like records, files, documents, images, sound, videos, scientific data and many new data formats. 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. Techniques of data mining can be introduced into new fields of Statistics, Databases, Machine Learning, Pattern Reorganization, Artificial Intelligence and Computation capabilities etc. Quality is one of the imperative factors that advance the development of higher education, which has led to the emergence of data mining in education characterized as "Educational Data Mining". There are wide range of techniques utilized by Educational Data Mining such as classification which includes decision tree, k-nearest neighbor, naïve Bayesian, rule based induction in association, and many others. These techniques facilitate to determine various kinds of knowledge such as association, classification, clustering etc. The main objective of institutes offering courses is to impart quality education among students which in turn augment the students' academic performance and helps in better decision making. The data in education sector is growing with the increase in the number of courses in universities so managing data of student, departments and employee is a very challenging task. Data mining is also use for sorting the educational problem by using analysis techniques for measuring the student performance.

II. LITERATURE SURVEY

Nur Hani Zulkifli, Jamalayahaya, Aziz Deraman has proposed the empirical study which has been done to discover requirement analysis process in Business Intelligence and Analytics implementation. The approach involves interviewing business intelligence experts in various domains. The proposed model can be used as a guideline for practitioners in analyzing requirements for BIAPM [1].

Mohiuddin Ali Khan, Waje Gharibi, Sateesh Kumar Pradhan has proposed the analysis of real data collected from several institutions, a sample study has been conducted to show how the Apriori Algorithm can be used in educational field and the results have been observed. Data mining can be used effectively in educational institutes for leading education activities in an effective way, for watching students' performances continuously and directing students in course and profession choosing. Thus, the level of student's success can be raised, and we can concentrate on students' weakness at selected courses thus improving the standard of students [2]. Leena Khanna, Dr. Shailendra Narayan Singh, Dr. Mansaf Alam has proposed the systematic review on Educational Data Mining technique which helps in predicting the future and changing the future. It contains application areas of EDM, various methods of EDM and the factors affecting the Student's academic performance and the teaching learning process [3]. Oswaldo Moscoso-Zea1, Andres-Sampedro, and Sergio Luján-Mora has proposed the differences between existing methodologies for DW design. Looking at differences between a business DW design and an educational DW design we found out that the value chain is the most important characteristic that changes the vision of the implementation project. Kimball and Inmon are the two most popular and recommended methodologies in the literature [4]. Lida Shams, Hassan Rashidi has proposed the BI architecture based on the service oriented concept (i.e. the flexible services that can be accessed on demand) is used for analyzing the current situation of educational courses and predicting their near future in order to support decision making. Thus the proposed SOA for BI considerably improves the decision making process [5]. Mohammed I.A.I. Twijri, Amin Y. Noaman has proposed a new Data Mining model to be applied in higher education institutions. The Suggested model assists in decision making process in the strategic levels of higher institutions as well as regulates the disciplines of student's admission [6]. Ms. Tismy Devasia ,Ms. Vinushree T P, Mr. Vinayak Hegde has proposed the web based application which makes use of the Navie Bayesian mining technique for the extraction of useful information. The system aims at increasing the success graph of students using Naive Bayesian and the system which maintains all student admission details, course details, subject details, student marks details, attendance details, etc. [7]. C. Romero and S. Ventura, presents the peer review of 15 years from 1993 to 2009 which focuses on EDM an upcoming research area which take account of various fields like data mining, learning analytics, e-learning, adaptive hypermedia, intelligent tutoring systems, web mining etc[11]. Pooja Thakar, Anil Mehta and Manish, describes a Comprehensive literature review of researches in EDM from 2002 to 2014 with main focus on papers published in Educational Data Mining, Predicting Academic Performance with Pre/Post Enrolment Factors, DM techniques to predict learning performance, association between Pre/Post Enrolment Factors and Employability, DM applications in supplementary areas of education like Faculty Evaluation, Analysis of the Curriculum Chosen by Students etc. Need of Integrated Models/Frameworks for all the stakeholders to ensure Sustainable Growth [12]. Ahmed, A.B.E.D. and Elaraby, conducted a similar research that mainly focuses on generating classification rules and predicting students' performance in a selected course program based on previously recorded students' behavior and activities. The authors processed and analyzed previously enrolled students' data in a specific course program across 6 years (2005–10), with multiple attributes collected from the university database. As a result, this study was able to predict, to a certain extent, the students' final grades in the selected course program, as well as, "help the student's to improve the student's performance, to identify those students which needed special attention to reduce failing ration and taking appropriate action at right time" [23]. Dr. P. Nithya, B. Uma maheswari, A. Uma devi, presented the survey on Educational data mining. This paper describes the goals of EDM, methods used in EDM, trends of EDM and application area of Educational data mining. These algorithms have shown a remarkable improvement in strategies like course outline formation, teacher student understanding and high output and turn out ratio [27].

III.BLOCK DIAGRAM OF THIS PROJECT

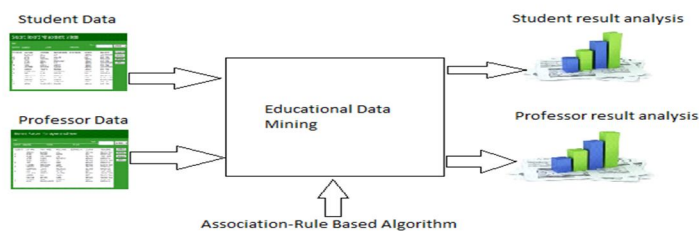


Fig: Above diagram shows the block diagram of our project

IV. PROPOSED WORK

In the project we collect the student’s data on regular basis of their daily activities to analyze the performance by using the concept of Educational data mining. With the help of Association algorithm we do the analysis of data. On the basis of student’s analysis we are going to analyze the performance of professor. After the analysis of student’s and professor’s performance we will give the graphical presentation of their results. We are establishing one-to- many and many-to-many relationship between student and professor so as to find the correlation between them. And for the proposed work we are using FP-Growth Algorithm.

A. FP-Growth Algorithm:

FP-Growth is an algorithm for discovering frequent item sets in a transaction database. FP-Growth is a very fast and memory efficient algorithm. It uses a special internal structure called an FP-Tree.

B. Input of the FP-Growth algorithm:

The input of FP-Growth is a transaction database and a threshold named *minsup* (a value between 0 and 100 %).

A transaction database is a set of transactions. Each transaction is a set of items. For example, consider the following transaction database. It contains 5 transactions (t1, t2, ..., t5) and 5 items (1,2, 3, 4, 5).

For example, the first transaction represents the set of items 1, 3 and 4. It is important to note that an item is not allowed to appear twice in the same transaction and that items are assumed to be sorted by lexicographical order in a transaction.

Transaction id	Items
t1	{1, 3, 4}
t2	{2, 3, 5}
t3	{1, 2, 3, 5}
t4	{2, 5}
t5	{1, 2, 3, 5}

Output of the FP-Growth algorithm;

FP-Growth is an algorithm for discovering itemsets (group of items) occurring frequently in a transaction database (frequent itemsets). A frequent itemset is an itemset appearing in at least *minsup* transactions from the transaction database, where *minsup* is a parameter given by the user.

For example, if FP-Growth is run on the previous transaction database with a *minsup* of 40 % (2 transactions), FP-Growth produces the following result:

Itemsets	support
{1}	3
{2}	4
{3}	4
{5}	4
{1, 2}	2
{1, 3}	3
{1, 5}	2
{2, 3}	3
{2, 5}	4
{3, 5}	3
{1, 2, 3}	2
{1, 2, 5}	2
{1, 3, 5}	2
{2, 3, 5}	3
{1, 2, 3, 5}	2

C. How should I interpret the results?

In the results, each itemset is annotated with its support. The support of an itemset is how many times the itemset appears in the transaction database. For example, the itemset {2, 3 5} has a support of 3 because it appears in transactions t2, t3 and t5. It is a frequent itemset because its support is higher or equal to the *minsup* parameter.

D. Advantages of FP-Growth algorithm

- 1) The biggest advantage found in FP-Growth is the fact that the algorithm only needs to read the file twice, as opposed to apriori who reads it once for every iteration.
- 2) FP-Growth is more scalable because of its linear running time. This makes it O(n) which is much faster than apriori.
- 3) The FP-Growth algorithm stores in memory a compact version of the database.

V. WORKING

In this project, we are analyzing student as well as professor performance. In Student analysis, we are analyzing their performance according to their 4 month attendance, PUT marks, Midterm marks and University results. We have collected their data (attendance, PUT marks and Midterm marks) from college and on this data we applied FP-growth algorithm. By using this FP-growth algorithm we got particular students attendance and marks individually after that we calculate overall analysis of student. And we represent the university marks of students in graph.

In Professor Analysis, we are analyzing them according to students feedback related to the subject and project, total number of lecture taken by Professor and how many students had pass and fail in particular subject. We also established one-to-many and many-to-many relationship between students' and professors'. Lastly we generate the summary chart which shows the overall correlation between students' and professors'.

A. Parameters used in Project

Students' → attendance, PUT marks, Midterms marks and university results.

Professor → subject feedback, project feedback, total no of lecture taken, number of pass and fail students, number of swapping lectures and university results.

VI.RESULT

This project has three different login profiles that is student, professor and Head of Department (HOD).The below graph shows the result of this project.

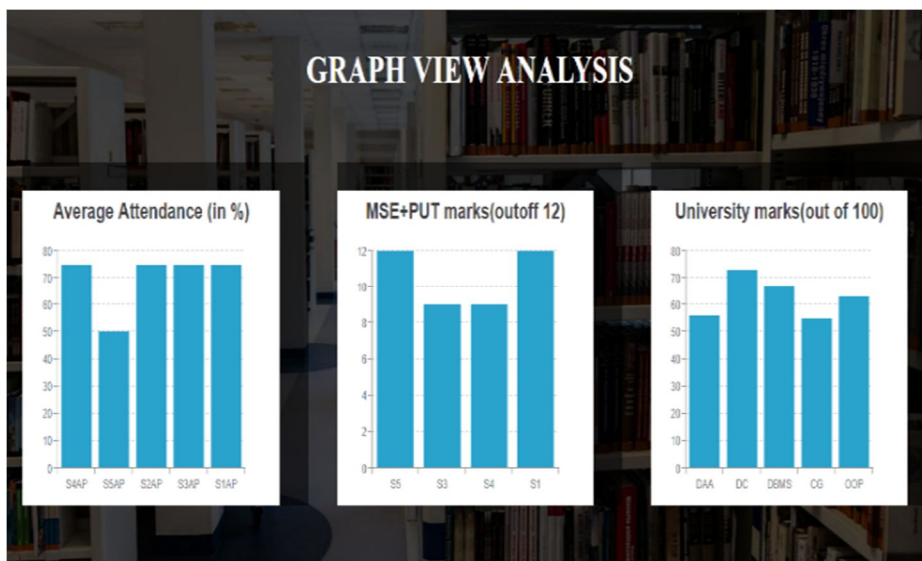


Fig1: When login through student

When login through student, the above three graphs are generated where the first graph describe the average attendance of particular student in all subjects (throughout the semester), the second graph describe the mid semester exam(mse) plus pre-university exam(put) exam marks achieved out off 12 in all subjects and the last graph describe the university marks.

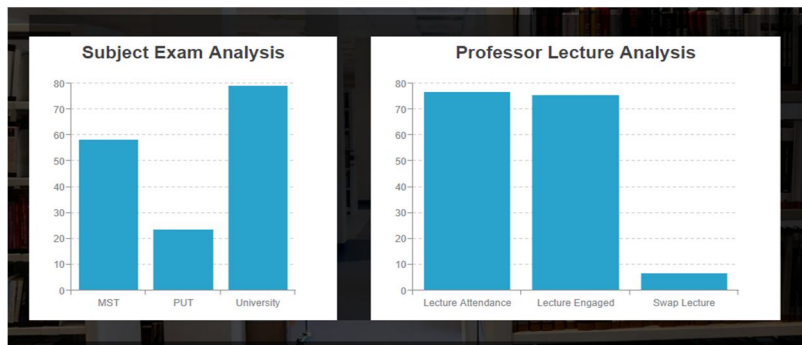


Fig2: When login through professor

When login through Professor, the above two graphs are generated. The first graph describe all exam result analysis according to their particular subject and the other graph describe the lecture analysis of that particular faculty.

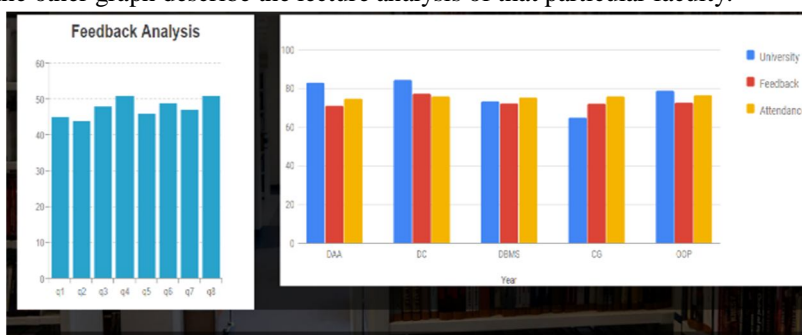


Fig3: When login through Head of Department

When login through HOD, we can see the analysis of both student and professor. In professor analysis, we can also view the feedback and project analysis of all faculty. In this profile, summary chart considering all important parameters can also be viewed.

VII. CONCLUSIONS

In this project, we are finding the correlation between the student performance and professor performance. This paper conclude the result of student on the basis of their put result, mid-term result as well as the attendance in lectures. We analyze this result by using the FP-Growth algorithm it shows the performance graph of the student for the further prediction. This paper also analyze the professor performance using the following parameters such as number of lectures engaged, average attendance of student’s in particular subject, number of swapped lectures, university result, mse(mid semester exam) and put(pre-university exam) result and feedback. In the current study, it was slightly found that the student’s performance is not totally dependent on their academic efforts, in spite, there are many other factors that is equal to greater influences as well.

In conclusion, this study can motivate and help colleges to perform data mining tasks on their students’ data and professors’ data regularly to find out interesting results and patterns which can help both the colleges as well as the students in many ways.

VIII. FUTURE SCOPE

Using the same dataset, it would be possible to do more data mining tasks on it, as well as, apply more algorithms. For the time being, it would be interesting to apply association rules mining to find out interesting rules in the students data. Similarly, clustering would be another data mining task that could be interesting to apply. This project can also be used for different datasets such as various departments of colleges.

REFERENCES

- [1] Nur Han iZulkifli, Jamalayahaya, Aziz Deraman (Article August2016), “Business Intelligence and Analytics in Managing Organizational Performance:The Requirement Analysis”,Journal of Advances in Information Technology Vol. 7, No.3,August 2016
- [2] Mohiuddin Ali Khan, Wajeb Gharibi, Sateesh Kumar Pradhan (2014) “Data Mining Techniques for Business Intelligence in Educational System: A Case Mining”978-14799-3351-8/14/\$31.00,2014 IEEE.
- [3] Leena Khanna, Dr.Shailendra Narayan Singh, Dr.Mansaf Alam,“Educational Data Mining and its Role In Determining Factors Affecti Students AcademicPerformance: A SystematicReview, 978-1-4673-69848/16/\$31.00 © 2016IEEE

- [4] Oswaldo Moscoso-Zea I., Andres-Sampedro, and Sergio Luján-Mora (2016), "Data warehouse design for Educational Data Mining", International Conference on Enterprise Information Systems (ICEIS), 2016, pp. p. 244–249
- [5] Lida Shams, Hassan Rashidi (2016), "Factors affecting the Increase and Decrease student achievement in Primary School with Business Intelligence Approach", International Journal of Computer & Information Technologies (IJOCIT).
- [6] Mohammed I.AI-Twijri, Amin Y. Noaman, "A New Data Mining Model Adopted for Higher Institutions", Procedia
- [7] Amjad Abu Saa, "Educational Data Mining & Students' Performance Prediction", International Journal of Advanced Computer Science and Applications, Vol.7, No.5, 2016
- [8] S. M. Merchan and J. A. Duarte, "Analysis of Data Mining Techniques for Constructing a Predictive Model for Academic Performance", IEEE Latin America Transactions, Vol. 14, No. 6, June 2016
- [9] Trupti A. Kumbhare, Prof. Santosh V. Chobe, "An Overview of Association Rule Mining Algorithms", IJCSIT, Vol. 5(1), 2014, 927-930
- [10] C. Romero and S. Ventura, "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, pp. 601–618, Nov. 2010.
- [11] Pooja Thakar, Anil Mehta and Manisha, "Performance Analysis and Prediction in Educational Data Mining: A Research Travelogue." arXiv preprint arXiv: 1509.05176 (2015)
- [12] B. Kumar and S. Pal, "Mining educational data to analyse students' performance," International Journal of Advanced Computer Science and Applications, vol. 2, no. 6, 2011
- [13] A. M. Shahiri, W. Husain, and N. A. Rashid, "A review on Predicting student's performance using data mining techniques," Procedia Computer Science, vol. 72, pp. 414–422, 2015
- [14] P. Baepler and C. Murdoch, "Academic Analytics and data mining in higher education," International Journal for the Scholarship of Teaching and Learning, vol. 4, no. 2, Jul. 2010
- [15] S. Hari kumar, "A study on educational data mining," International Journal of Computer Trends and Technology, vol. 8, no. 2, pp. 90–95, Feb. 2014
- [16] M. Berland, R. S. Baker, and P. Blikstein, "Educational data mining and learning Analytics: Applications to Constructionist research," Technology, Knowledge and Learning, vol. 19, no. 1-2, pp. 205–220, May 2014
- [17] M. S. Bhullar and A. Kaur. "Use of data mining in education sector". Lecture Notes in Engineering and Computer Science 2200(1), pp. 513-516. 2012
- [18] 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
- [19] BipinBihariJayasingh, "A Data Mining Approach to Inquiry Based Inductive Learning Practice In Engineering Education", in IEEE 6th International Conference on Advanced Computing 2016
- [20] K. Jothi and K. Venkatalakshmi, "Intellectual performance Analysis of students by using data mining techniques", International Journal of Innovative Research in Science, Engineering and Technology, vol 3, Special iss 3, March 2014
- [21] M. Jayakameswaraiah, S. Ramakrishna, "A Study on Prediction Performance of Some Data Mining Algorithms", International Journal of Advance Research in Computer Science and Management Studies, ISSN: 2321-7782 (Online) Volume 2, Issue 10, October 201
- [22] Ahmed, A.B.E.D. and Elaraby, I.S., "Data Mining: A prediction for Student's Performance Using Classification Method." World Journal of Computer Application and Technology, 2(2), pp.43-47, 2014
- [23] Yadav, S.K. and Pal, S., 2012. Data mining: A prediction for performance improvement of engineering students using classification. World of Computer Science and Information Technology Journal (WCSIT). (ISSN: 2221-0741), Vol. 2, No. 2, 51-56, 201
- [24] Dina Abdulaziz Alhammedi, "Data Mining in Education- An Experimental Study", International Journal of Computer Applications Volume 62, No.15, January 2013
- [25] Baker, R.S.; Yacef, K (2009). "The state of educational data mining in 2009: A review and future visions". JEDM-Journal of Educational Data Mining 1 (1): 2017
- [26] Dr. P. Nithya, B. Umamaheswari, A. Umadevi "A Survey on Educational Data Mining in Field of Education", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 1, January 201
- [27] Hashmia Hamsa, Simi Indiradevi and Jubilant J. Kizhakkethottam, "Student academic performance prediction model using decision tree and fuzzy genetic algorithm", Procedia Technology 25 (2016) 326 – 33
- [28] M. Anoop kumar and Dr. A. M. J. Md. Zubair Rahma, "A Comprehensive Survey on Educational Data Mining and Use of Data Mining Techniques for Improving Teaching and Predicting Student Performance", Advances in Innovative Engineering and Technologies ISBN: 978-0-9948937-1-0
- [29] O.S. Akinola, B.O. Akinkunmi, T.S. Alo, "A Data Mining Model for Predicting Computer Programming Proficiency of Computer Science Undergraduate Students", African Journal of Computing & ICT January, 2012, Vol 5. No. 1 - ISSN 2006-1781