

Efficient Pattern Generation using Declat Algorithm Based on Association Rule Mining in Educational Data Mining

Rajeev Sharma, Sitendra Tamrakar

Abstract: A new application called DM Educational Data Mining (EDM) involves data extraction and analysis from the classroom or area of education. In order for educators to deliver quality education to students, EDM integrates various educational information into its review. The EDM works by translating raw data from education systems invaluable information which could have a major effect on the study of education. The output of each student is measured from the database and must be sufficiently accurate to withstand changes in the academic record. Then we have transformed the overall arrangement into a modified relation for the adequacy of the Declat algorithm. The purpose of this work is to examine how prior researchers, as well as recent data mining trends in educational research, have dealt with data mining. In this paper, collected data comprised of 200 students. We define academic performance & impact of additional issues on the basis of these course's last grades, indications of attendance, class tests, and term last answer substance add up to marks and so on. Here, we compare the FP-Growth and Eclat with Declat algorithm on the bases of confidence and support value in a relation of execution time & no. of patterns generated. This paper uses a declat algorithm to create patterns or delete effective patterns. Such patterns help to illustrate a growing student's success.

Index Terms—Data Mining (DM), Association Rule Mining, EDM, FP Growth, Declat, Support, Confidence.

I. INTRODUCTION

Data mining is the processing of knowledge through huge data sets, discovering different facts which may lead to data analysis method., interpretation & prediction. DM is an automated method of discovery integrated with databases of nontrivial, previously unknown & possibly helpful patterns. DM is a technology that allows data visualization & data exploration at a high level of abstraction of very enormous databases that do not require any particular hypothesis in mind. Hided predictive data mining using big datasets has excellent potential to support businesses concentrate their data warehouses on very significant data into their data warehouses. [1]

The Association Rule Mining (ARM) is a useful and very well studied mining strategy. In recent decades, several scientists and experts from various areas have been using DM algos to

research & discover associations between both attributes of big databases in order to create correct laws. These rules equate one or more data set attributes with a different attribute, creating an if-then argument about attribute values between large database item sets. The implementation of such methods is typically not a simple process; nonetheless, five stages of the KDD process must be successively applied. ARM was used for educational content successfully [2].

EDM focuses on designing strategies for analyzing data from an educational background. In order to study education issues, EDM uses computational methods to analyze educational data. The findings can not only be used it to learn the model of learning process [3] or modeling of students[4], to evaluate or enhance e-learning systems[5] through the development of valuable learning information from portfolios of learning portfolios[6]. Statistics on education are statistics from 2 types of education systems: classroom or distance learning (web-based).

II. ADVENT OF EDUCATIONAL DATA

The creation of EDM has taken researchers into account, attempted to assess the quality, selection, quality, and volume of student data. Education institutes produce ample student data that was easily observed. However, these data are available in various formats and are often not kept on a central stage. EDM has the ability to analyze all aspects of student information. It is inferred.

Data necessary for EDM can be obtained from various sources along with historical or organizational data in educational institution databases or manuals. These data may consist of e-learning systems, intelligent tutoring systems or intelligent game systems.[7].

The descriptive data or behavioral data can be classified as two kinds of student information, often generated in the conventional education or learning system. The descriptive data includes self-reported student information. This covers sex, age, school qualification, university or geographical profile, marital status, if an international student is a member, an association or not, etc.

Competence data provide information on the topics enrolled in the program, specifics about subject assessment; points or grades, attendance records or student progress assessments and student behavior or academic practices. The patterns in education and learning have changed massively in recent years [8].

Revised Manuscript Received on February 04, 2020.

* Correspondence Author

Rajeev Sharma, Department of Computer sciences, Govt. SMS Science College, Gwalior, India. Email: sharmaraj2007@gmail.com

Sitendra Tamrakar, Department of Computer Science, Nalla Malla Reddy Engineering College, Hyderabad, India. Email: drsitendra@gmail.com

Technology has entered the education system to a considerable extent, which has changed the conventional education system. Hybrid and Web-based learning has been improved [9]. The increased usage of electronic systems in education enables learning resources to be prepared via powerful presentation instruments or advanced multimedia. Learn in pdfs and online videos from the teachers. Testing takes place in labs or online with computerized platforms. Student interactions with these systems allow huge quantities of student learning data to be collected. The theory allows for teaching to adjust assessments to minimize downtime to document interesting patterns or parameters within the learning system and thus to be analyzed. Therefore, two additional forms of student data, namely the ability to interact and attitude data, have been generated due to technological use.

III. LITERATURE SURVEY

Gabriela Cibula et al. [2019] To predict a student's final outcome in a confident academic discipline using RARs, suggest S PRAR (Students' Performance Prediction, using Relative Association Rules). Suggest a new classification model. For different relationships between data attributes, RARs broaden the classical association laws. Three actual academic data sets are being conducted in studies from the University of Babes Bolyai in Romania. Compared with current related work, the output of the PRAR classification on studied cases is superior to the success indicator of previously proposed students [10].

E. Fernandes et al. [2019] Please present predictive student academic performance analysis in the 2015 or 2016 school cycles of public schools in the Federal District of Brazil. Next, a concise statistical analysis has been conducted to gain insight. Two datasets have been collected subsequently. The first data set includes variables that were gathered until school years as the second comprises academic variables that were collected two months later. Models were created to try to predict academic results for each data set at the end of an academic year, gradient boosting machine (GBM). Findings show attributes "grades" or "absence," which are more important to predict academic results of student performance by end of year., [11].

Sumyeha Helal et al. [2018] The study uses data collected with an Australian university to make different classification models for school performance prediction. The data included the enrolment details of students, as well as the data generated from the LMS. The inscription details contain information on student engagement (e.g. full-time vs. part-time) as a type of student attendance (e.g. by entrance exam or past experience). The LMS data report student engagement in online learning. The student heterogeneity when building the mathematical models is a major contribution to this analysis. Studies have validated the hypothesis that student subpopulation models trained only instances outpace those developed using all data. researches also showed that both registration & training action contributes to the accuracy of the identification of vulnerable students [12].

Wei Zhang and Shiming Qin [2018] introduce the fundamental ideas of BDE, EDM, & online learning platform, & then elaborate on the procedure of how EDM transform raw information into information. Finally, the main DM techniques are categorized as per their utilizes and are used in

the online education scene. the paper may offer some advice on researching & application of EDM found on online education [13].

Edona Doko and Lejla Abazi Bexheti [2018] A comprehensive research aimed at improving learning interactivity are provided in this review for videos in education technology. We saw a series of papers that showed a video field of interest. [14].

Yoga Pristyanto et al. [2018] explain imbalanced class on the processing system of multiclass EDM datasets by combining OSS & SMOTE. The OSS & SMOTE technique offers a balancing system for the allocation of datasets to improve the efficiency of the classification outcomes. The outcome demonstrates that the combination of OSS & SMOTE may improve SVM's efficiency as a technique of classification utilized in this research. This mixture of techniques produces as high as 88.637 percent precision, sensitivity, specificity, & g-mean score, 92.292 percent, 95.554 percent, 93.796 percent, respectively [15].

Sagardeep Roy and Dr. Shailendra Narayan Singh [2017] describe the recent Big Data instruments and techniques in the field of learning analysis and DM. This information is lost because standard techniques of teaching are not equipped to prepare them. Because of this, Big Data technology has now joined a field of education to mine these enormous pieces of datasets [16].

Aleksandr I. Taganov et al. [2017] considered scientific & methodological features of growth in higher education organizations of methodological, information & instrumental support to education risk control system. The proposed model for assessing educative threats in larger institutions would enable a thorough assessment of procedures in agencies, taking into account pedagogical, financial elements. The study of risk elements and roles in the instructional area was carried out. Software implementation of suggested strategies has been performed in the Postgre SQL database control system. The findings from the testing of technology proposed were shown based on open 2016 data from the University of Ryazan. [17].

IV. PROPOSED METHODOLOGY

The process has defined here:

1) Data Gathering

The generated data set provides information about students' perceptions and behaviors. In total, a total of 20 criteria, including name, age, gender, attendance, attendance marks, batch, class performance, class submission information, note-taking, class participation, seating location, punctuality, etc., have been included.

2) Data Transformation

In order to manage it in the next step, the data was collected in an appropriate format for methods of finding information.

3) Data Mining

DM approaches are used more and more to examine educational information and to discover trends for improving educational practices. EDM is used to collect interesting, useful, and unique knowledge from educational data repositories, as is the case with other DM practices [18].

This research employed 2 well-established technology for DM: pattern discovery or association rules for student data.

4) Pattern Discovery

The discovery of patterns attempts to find common trends from a multivariate regression. Specified a dataset, patterns are primarily sets of items that often occur together, making them closely related. In a given dataset, these reflect intrinsic or meaningful properties. Pattern discovery also provides a basis for many activities such as ARM, correlations or causality analysis.

A. Methodology

In this research, the Association rule mining based FP-growth, Eclat and Declat algorithms are compared with each other in terms of less duration of time taken for the execution process and the patterns generated on minimum support & minimum confidence whereas no. of generated rules are same in all the three algorithms.

1) FP-Growth

FP Growth [19] is another basic frequent pattern basic FPM strategy, which generates visit itemset devoid of candidate age. It uses a tree-based structure. The issue of the Apriori algo was overseen, by displaying a novel, minimal data structure, known as FP tree, or FP by then in light of this structure an FP pattern fragment growth technique was made. FP tree is worked into 2 passes:

Pass 1:

- Scan data & check support to everything
- Discard irregular items
- Sort visit things in diving demand in light of their support.

Pass 2:

- Reads one exchange at any given minute & maps it for a tree.
- Fixed organize is utilized with the target that way can be shared
- Pointers are kept up among nodes comprising similar items.
- Numerous things are removed by rundown. It experiences certainly.

2) ECLAT

Eclat's first search algorithm considers the elements below as depth. The Eclat algorithm is very easy to find frequent sets of products. This algorithm uses a database vertical. The horizontal database cannot be used. If the horizontal database is available. The database does not have to be scanned repeatedly. Only once will Eclat scan the database. In this algorithm, support is counted. Trust in this algorithm is not determined.

3) DECLAT

The discrepancy in both created itemsets named different is applied to signify produced itemset rather than to use similar transaction IDs of both generated itemsets. The difference significantly reduces cardinality of sets that represent objects which contributes to a faster junction and less memory consumption. [20].

T(P) must be P's tidset. The diffset $d(P X)$ is set of IDs in $t(P)$, but not formally in $t(P X)$: $d(P X) = t(P) - t(P X) = t(P) - t(P)$. Providing the availability of $t(P X)$ & $t(P Y)$, $P X$ and $P Y$ are within the same conditional prefix P database with the objective of evaluating $\text{support}(P X Y)$. $\text{Support}(P XY) = \text{support}(P X) - XY$ by design. Thus, only $d(P XY)$ is required to measure $P XY$ support. $d(P XY) = d(P Y) - d(P X) = d(P X) - t(P X) - t(P Y)$ has been shown to allow diffset and support for item created from diffsets or tidsets or their produced objects.

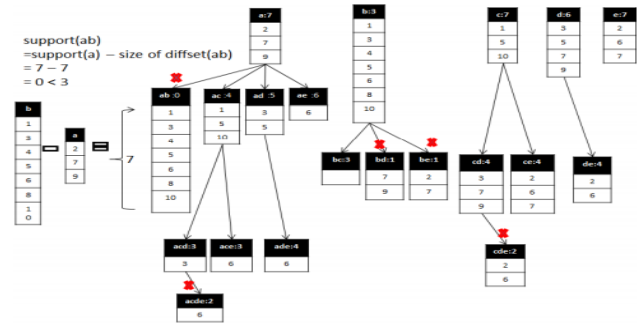


Fig. 1.dEclat algorithm: FIM Using Diffset

For diffset usage, the original database is first converted to diffset format with a set of transaction IDs whereby each object is represented by transactions that do not include the item. & then a process of diffset formation as well as a calculation of k-itemsets, in which 2— km from the first database is recursive.

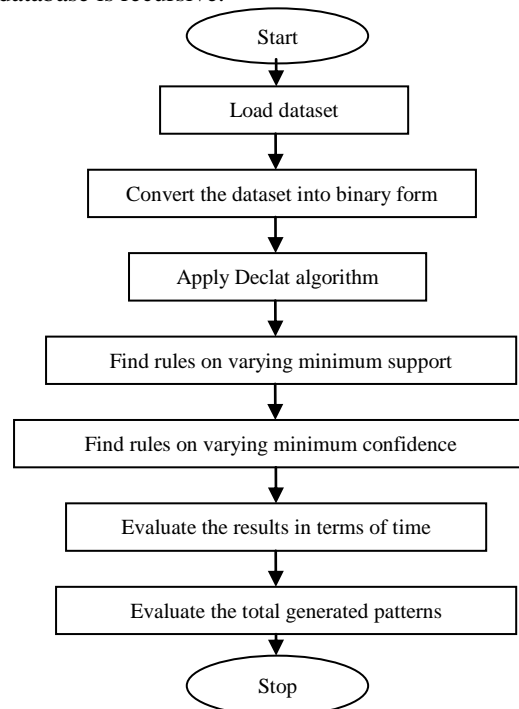


Fig. 2.Flow Diagram of Research Methodology

V. RESULTS AND DISCUSSIONS

Simulation of the proposed work and performed the implementation of the different number of parameters.

A. Confidence and Support Representation for FP Growth.



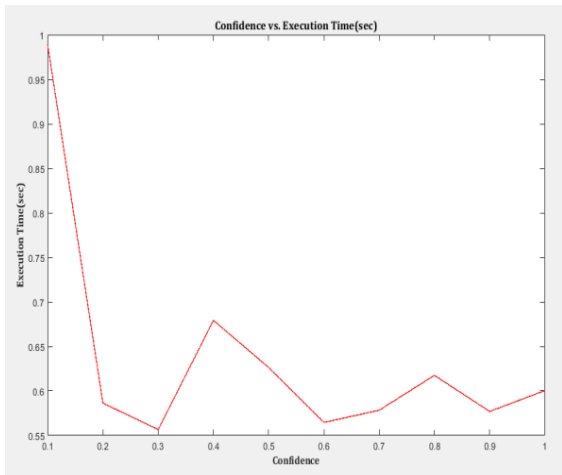


Fig. 3. Confidence vs Execution time in FP Growth

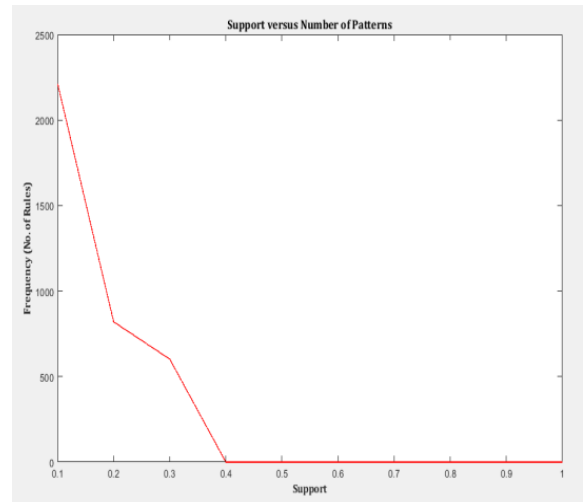


Fig. 6. Support vs no. of patterns in FP Growth

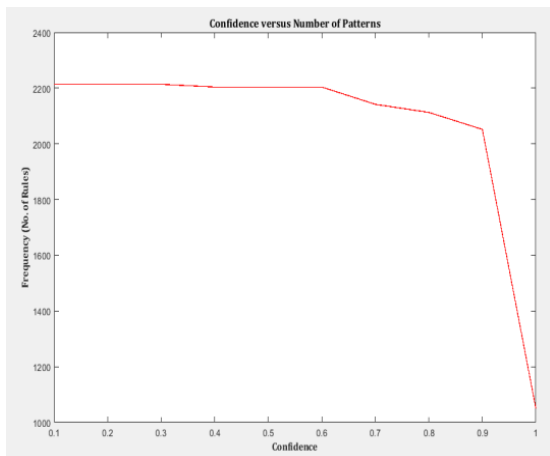


Fig. 4. Confidence vs no. of patterns in FP Growth

B. Confidence and Support Representation for Eclat.

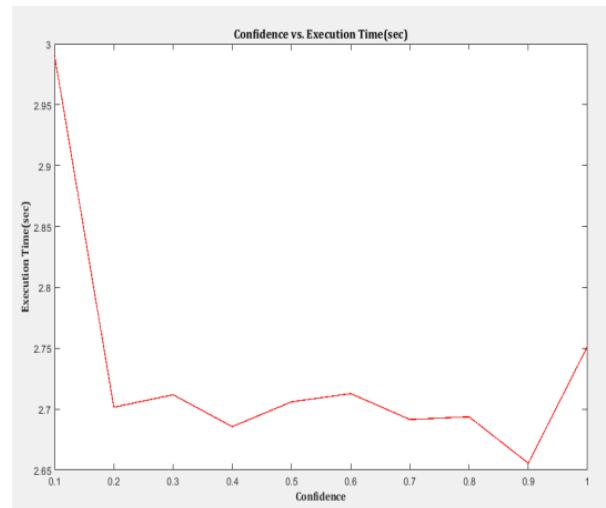


Fig. 7. Confidence vs Execution time in Eclat

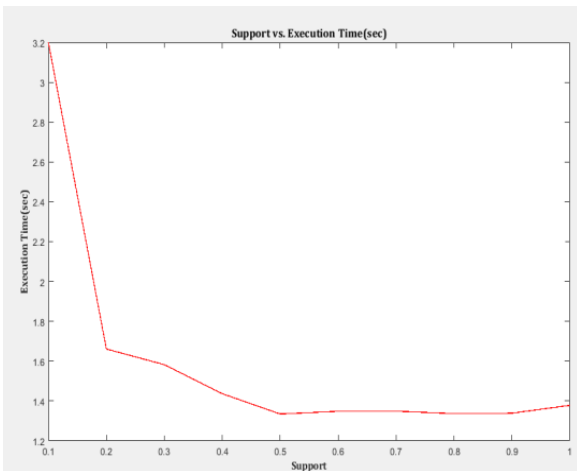


Fig. 5. Support vs execution time in FP Growth

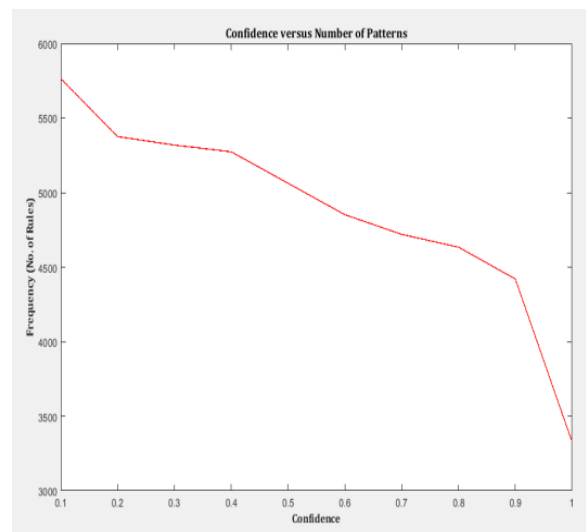


Fig. 8. Confidence vs no. of patterns in Eclat

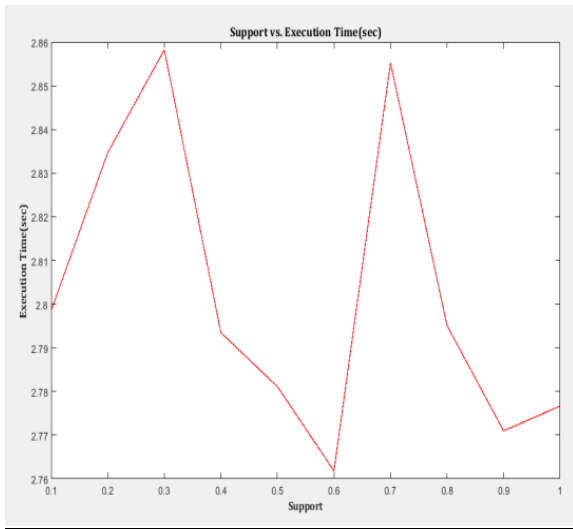


Fig. 9. Support vs execution time in Eclat

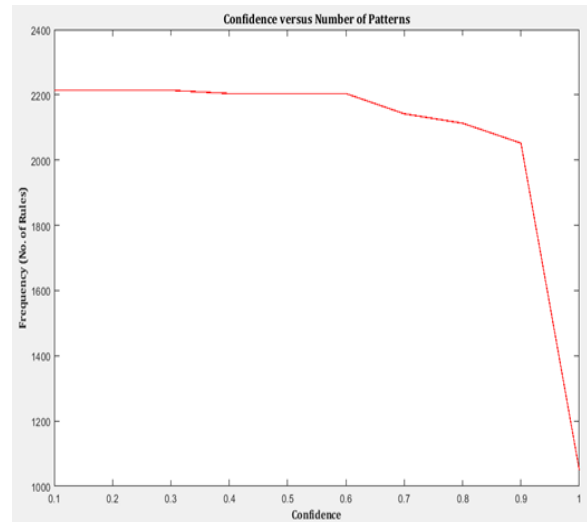


Fig. 12. Confidence vs no. of patterns in Declat

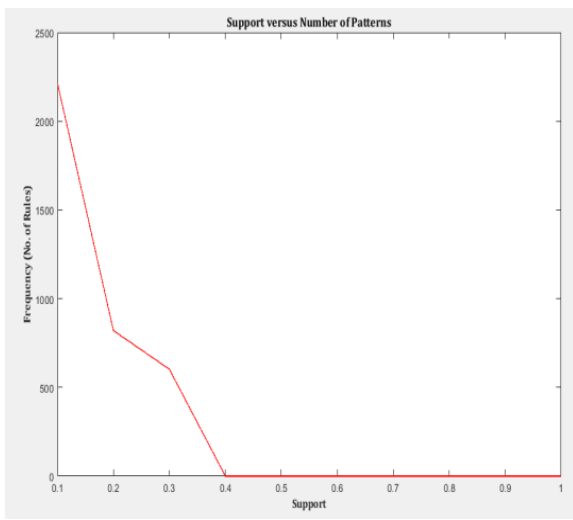


Fig. 10. Support vs no. of patterns in Eclat

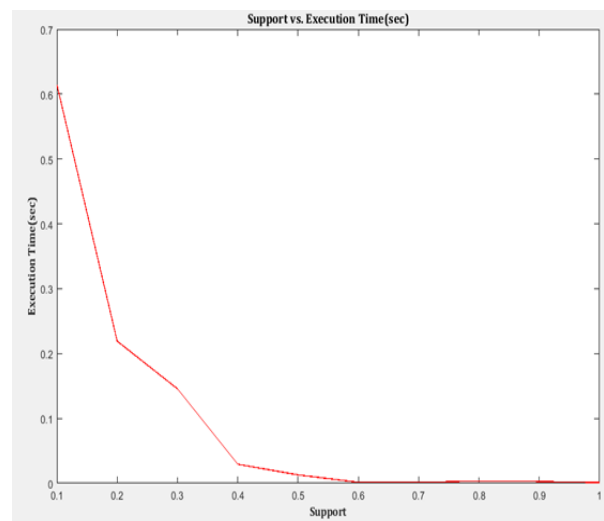


Fig. 13. Support vs Execution time in Declat

C. Confidence And Support Representation For Eclat.

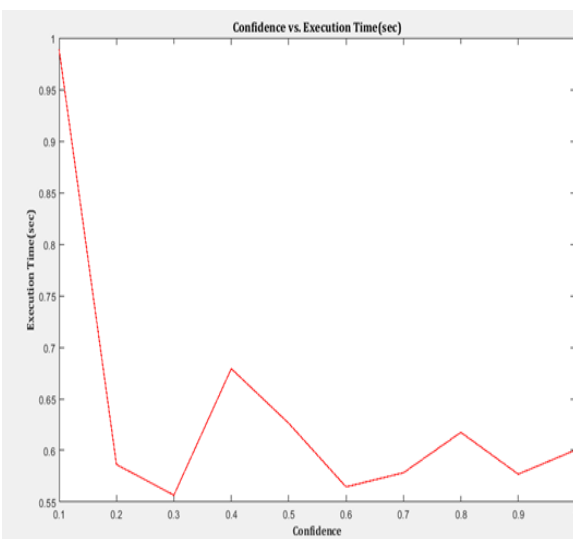


Fig. 11. Confidence vs Execution time in Declat

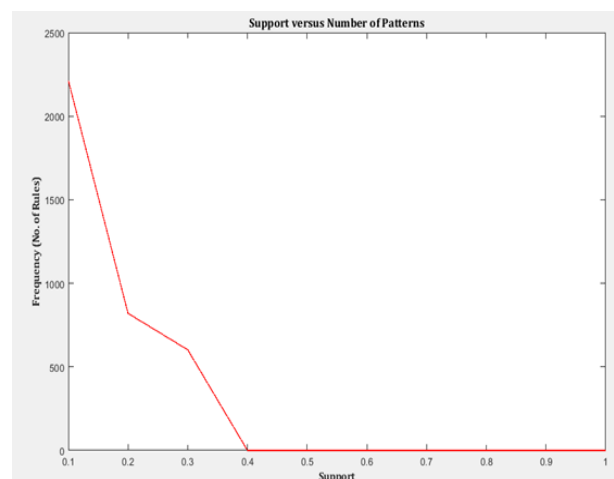


Fig. 14. Support vs no. of Patterns in Declat

In the result section, the impact of various parameters is shown with their generated rules. Different rules are generated in this section by considering minimum support and confidence value.

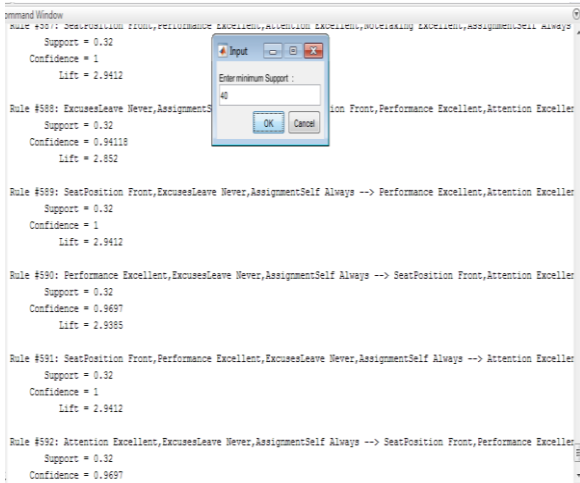


Fig. 15. Rules generated by FP Growth on minimum Support 0.4



Fig. 18. Rules generated by Eclat on minimum Confidence 1.0

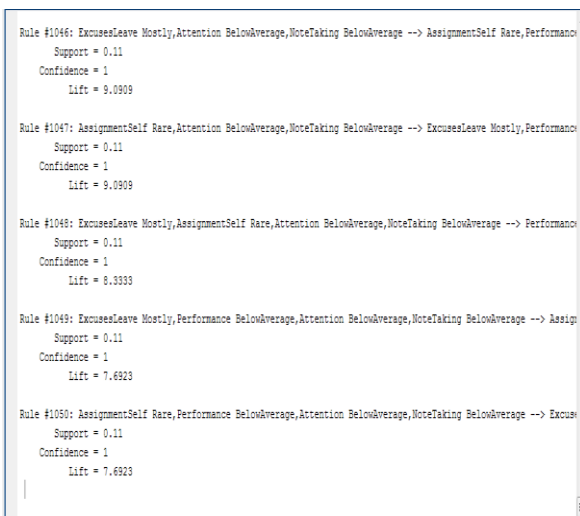


Fig. 16. Rules generated by FP Growth on minimum Support 1.0

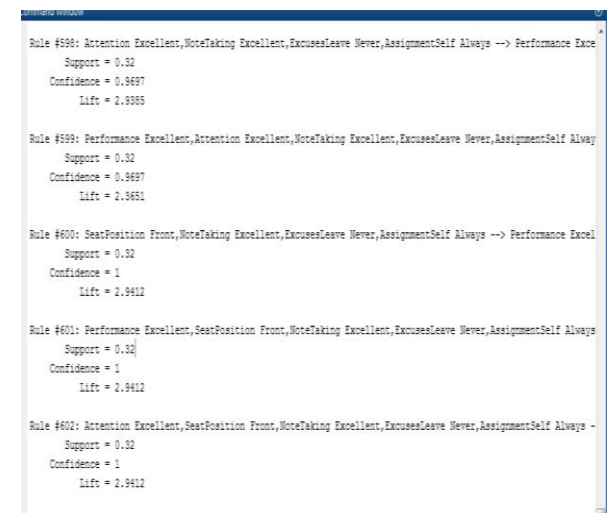


Fig. 19. Rules generated by Declat on minimum Support 0.3

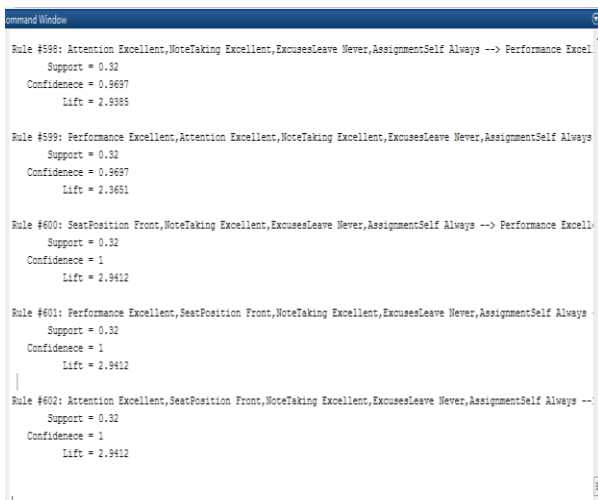


Fig. 17. Rules generated by Eclat on minimum Support 0.3

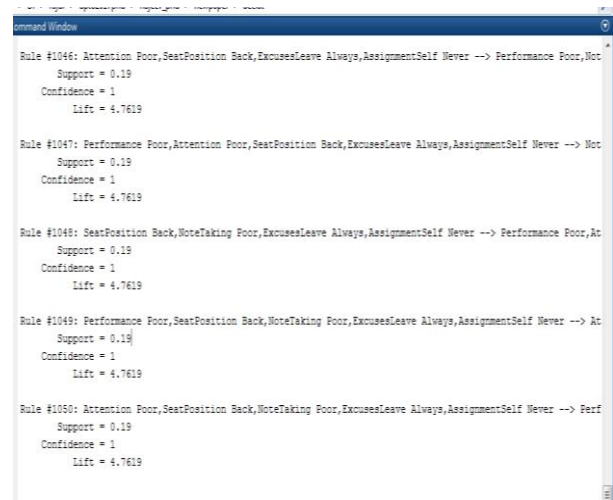


Fig. 20. Rules generated by Declat on minimum Support 1.0

VI. CONCLUSION

EDM is a DM sub-domain that deals with academic database information used to create different methods & identify distinctive trends. The knowledge acquired may then be utilized for giving suggestions to educational organizers in higher education institutions for improve their decision-making procedure, increase the academic achievement of learners, reduce failure rates, better comprehend student behavior, help teachers, improve learning, & build regression models & decision trees (DTs) to predict learners perfectly. Retention and abandonment are two main causes for the performance of the students. DM techniques are applied to mine useful data & generate the most influential rules to detect the main causes. FP Growth, Eclat and Declat algorithms are used in this paper for the generation of rules over the pre-processed data and make it more useful for the institution. It was found that the Declat algorithm proved to be most efficient in terms of time execution & no. of patterns generated.

REFERENCES

1. AartiKaushal, Manshi Shukla, "Comparative Analysis to Highlight Pros and Cons of Data Mining Techniques-Clustering, Neural Network and Decision Tree", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 2014, 651-656.
2. Ougiaroglou, S., & Paschalis, G. (2012). *Association Rules Mining from the Educational Data of ESOG Web-Based Application. Artificial Intelligence Applications and Innovations*, 105–114. DOI:10.1007/978-3-642-33412-2_11.
3. Hamalainen, W., Suhonen, J., Sutinen, E., Toivonen, H.: Data mining in personalizing distance education courses. In: World Conference on Open Learning and Distance Education, Hong Kong (2004)
4. Tang, T., McCalla, G.: Student modeling for a web-based learning environment: A data mining approach. In: Eighteenth National Conference on Artificial Intelligence, Menlo Park, CA, USA, pp. 967–968 (2002).
5. Zaïane, O., Luo, J.: Web usage mining for a better web-based learning environment. In: Proceedings of the Conference on Advanced Technology for Education, Banff, Alberta, pp. 60–64 (2001)
6. Hwang, W., Chang, C., Chen, G.: The relationship between learning traits, motivation, and performance-learning response dynamics. *Computers & Education Journal* 42(3), 267–287 (2004).
7. <https://files.eric.ed.gov/fulltext/ED566873.pdf>
8. Graham Hall, "The Routledge Handbook of English Language Teaching", book.
9. Jennifer Groff, "Technology-Rich Innovative Learning Environments", February 2013, pp. 1-30.
10. Gabriela Czibula, Andrei Mihai, Liana Maria Crivei, "S PRAR: A novel relational association rule mining classification model applied for academic performance prediction", *Procedia Computer Science* 159 (2019) 20–29, doi: 10.1016/j.procs.2019.09.156.
11. Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Erven, G. V. (2018). *Educational data mining: Predictive analysis of the academic performance of public school students in the capital of Brazil. Journal of Business Research*. DOI:10.1016/j.jbusres.2018.02.012.
12. Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D. J., & Long, Q. (2018). *Predicting academic performance by considering student heterogeneity. Knowledge-Based Systems*. DOI:10.1016/j.knsys.2018.07.042.
13. Wei Zhang and Shiming Qin, "A brief analysis of the key technologies and applications of educational data mining on online learning platform", *IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, 2018, pp. 83-86.
14. EdonaDoko and LejlaAbaziBexheti, "A Systematic Mapping Study of Educational Technologies based on Educational Data Mining and Learning Analytics", 7th Mediterranean Conference On Embedded Computing (MECO), 2018, pp. 1-4.
15. Yoga Prityanto et al, "Data Level Approach for Imbalanced Class Handling on Educational Data Mining Multiclass Classification",

- International Conference on Information and Communications Technology (ICOIACT), 2018, pp. 310-314.
16. Sagardeep Roy and Dr. Shaileendra Narayan Singh, "Emerging Trends in Applications of Big Data in Educational Data Mining and Learning Analytics", *7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, 2017, pp. 193-198.
17. Aleksandr I. Taganov et al, "Management of educational risk on the basis of data mining in GIS", *International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)*, 2017, pp. 577-580.
18. Yoga Prityanto et al., "Data Level Approach for Imbalanced Class Handling on Educational Data Mining Multiclass Classification", International Conference on Information and Communications Technology (ICOIACT), 2018, pp. 310-314.
19. Han J., Pei H., and Yin. Y., Mining Frequent Patterns without Candidate Generation, In Proc. Conf. on the Management of Data (2000).
20. Mustafa Man, Wan Aezwani Wan Abu Bakar, Masita Masila Abd. Jalil, Julaily Aida Jusoh, "Postdiffset Algorithm in Rare Pattern: An Implementation via Benchmark Case Study", *International Journal of Electrical and Computer Engineering (IJECE)* Vol.8, No.6, December 2018, pp. 4477-4485. DOI: 10.11591/ijece.v8i6.pp4477-4485.

AUTHORS PROFILE



Rajeev Sharma, is working in Dept. of Computer Science at, Govt. SMS Science College Gwalior as a visiting faculty. He obtained MCA from RGPV and M.Tech from RGPV Bhopal. He is currently doing research work in Educational data mining.



Dr. Sitendra Tamrakar, is working in Dept. of Computer Science & Engineering, at, Nalla Malla Reddy Engineering College, Hyderabad as Associate Professor. He has guided several scholars at Ph.D., M.Tech level. He has authored total of 67 research papers & book chapters, which have been published in national and International journal as well as in proceedings of conferences in India and abroad.