



Data Mining Tools: A Comparative and Analytical Study for Educational Data Mining.

Het Trivedi¹, Dr.Ajay Patel²

Research Scholar¹, Research Guide²

Department of Computer Science¹,

Ganpat University, Kherva¹

Abstract: Data mining is the non-trivial process of identifying valid, potentially and understandable patterns In the form of knowledge discovery from the large volume of data. The main aim of the process is to discovering patterns and associations among preprocessed and transformed data. Data mining is used for two type of analysis: Prediction and Description. Prediction in the terms of predicts unknown or future values of selected variables. Description in terms of describes human interpretable patterns. Data mining supports supervised and unsupervised types of leaning process. This paper uses supervised learning process of data mining. For that the paper uses student's dataset which have 27 different attributes with 6000 instances. The researcher focuses on the performance mapping of different algorithm supported by different data mining tools. The different tools provides different types of association algorithm. The same data set applied for different tools. The paper shows the comparative analysis for performance of algorithm of on different data mining tools.

I. INTRODUCTION

In this research paper, researcher is going to apply data set on various of data mining tools available in the market and on the basis of those results found from those tools researcher have found best tool for their further research. Data mining, occasionally also called Knowledge Discovery in Databases(KDD), can find connections and patterns that live, but are hidden among the vast quantum of educational data. It combines machine literacy with statistical and visualization ways to discover and prize knowledge in such a way that humans can fluently comprehend. For sodalities and universities, the knowledge discovered by data mining ways would give individualized education that satisfies the demands of scholars, employers and administration. There are several of tools available which work on the data sets and it will results in some patterns, related data and many more.

II. DATA MINING

Educational associations are one of the important corridors of our society and Educational data mining is the operation of data mining. It's an arising interdisciplinary exploration area that deals with the development of styles to explore data forming in an educational environment. Educational data mining is an arising trend, designed for automatically exploring the unique types of data from large depositories of educationally related data. Relatively frequently, this data is expansive, fine - granulated, and precise [1,3]. The Educational Data Mining community defines Education data mining as follows Educational data mining is an arising discipline, concerned with developing styles for exploring the unique types of data that comes from educational settings, and using those styles to more understand scholars, and the settings in which they learn. Data mining is the field of discovering new and potentially useful information by birth huge quantum of data. We know that large quantum of data, stored in database, so in order to get the needed data and to find the retired relationship, different data mining ways are used similar as classification, prediction, association rule, outlier discovery etc. Education data mining focuses on collection, archiving, and analysis of data related to student's literacy and assessment [2]. Education data mining includes analysis(evaluation/ disquisition) of educational processes including admission, alumni relations, course selections, prognosticating drop out student, scholars success rate, course

success rate, performance evaluation of student, learning gets of scholars, list of courses taken by the student, when the student named or changed his or her academic major, chancing which tasks, courses etc are delicate easy for which scholars, chancing optional courses frequently taken by scholars etc. Data Mining in Education is an evolving field with new generalities Data Mining is the process of automatically discovering useful information in large data depositories. In other words DM is the hunt for the connections hidden among Very Large Databases(VLDB). Data mining-trapping concentrated on different ideas similar as slice, estimation, thesis, testing from statistics, searching algorithms, modeling ways, machine literacy propositions from artificial intelligence, pattern recognition and hi- performance computing. Therefore, data mining is represented as a convergence of numerous disciplines [2, 3]. Data booby-trapping ways and styles are used in the main affiliated disciplines and technologies from the ensuing are as [3].

III. PREDICTIVE DATA MINING

The thing of data mining is to produce new knowledge that the stoner can act upon. It does this by erecting a model of the real world grounded on data collected from a variety of sources which may include commercial deals, client histories and demographic information, process control data, and applicable external databases similar as credit office information or rainfall data. The result of the model structure is a description of patterns and connections in the data that can be confidently used for prediction. To avoid confusing the different aspects of data mining, it helps to fantasize a scale of the choices and opinions you need to make before you start.

- Business thing
- Type of prediction
- Model type
- Algorithm
- Product

For illustration, seeking patterns in your data to help you retain good guests, you might make one model to prognosticate client profitability and a alternate model to identify guests likely to leave (waste) [7]. Your knowledge of your association's requirements and objects will guide you in formulating the thing of your models. The coming step is deciding on the type of prediction that's most applicable

- 1) Classification: predicting into what order or class a case falls
- 2) Regression: predicting what number value a variable will have, if it's a variable that varies with time.

In predictive models, a data mining tool harbors it from the complications of statistical ways; it requires you to understand the workings of the tools you choose and the algorithms on which they're grounded. The choices you make in setting up your data mining tool and the optimizations you choose will affect the delicacy and speed of your models.

IV. TOOLS OF DATA MINING

Now days, wide range of operations similar as client profiling, targeted marketing, work inflow operation, store layout, and fraud discovery are advantaged with Data mining ways. Lately, there has been an adding trend that data mining ways were used to ameliorate the effectiveness of educational academy/ institutions [7, 8, 9,]. Mining ways can be used in educational terrain. In similar case, mining is called Educational Data Mining, concerned with developing new styles to discover knowledge from educational databases [7] Data mining-trapping ways and tools can help in bridging knowledge gaps in advanced education system.

3.1 WEKA

index. Weka(Waikato Environment for Knowledge Analysis) is a package that offers collection of tools that they can use for data mining. Weka algorithms can be used to classify the available data, used to filter the data contents and use it for data conniving. Weka includes tools for data clustering and Regression, association rules and attributes annotator. New machine learning schemes

can be developed by Weka. It's an open source and freely available. It's platform-independent and fairly easier to use. It provides flexible installations for scripting trials [10].

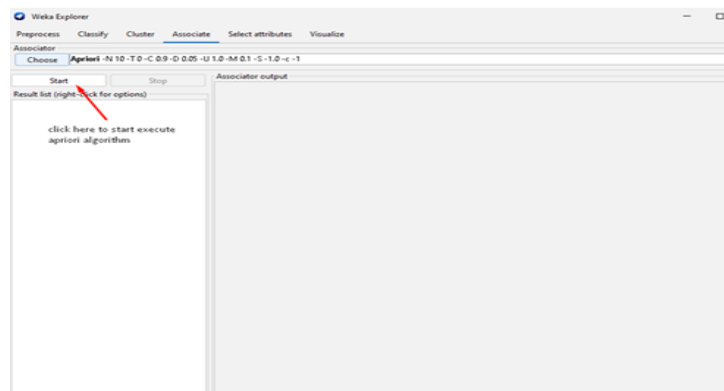


Figure 1.1- apriori in weka

3.2 TANAGRA

Tanagra is also freely available data mining tool. It provides various statically, non-parametric test, Spv learning techniques association and clustering. Tanagra works with .arff and other file formats specified by Tanagra. It is component based visualize tool [11].

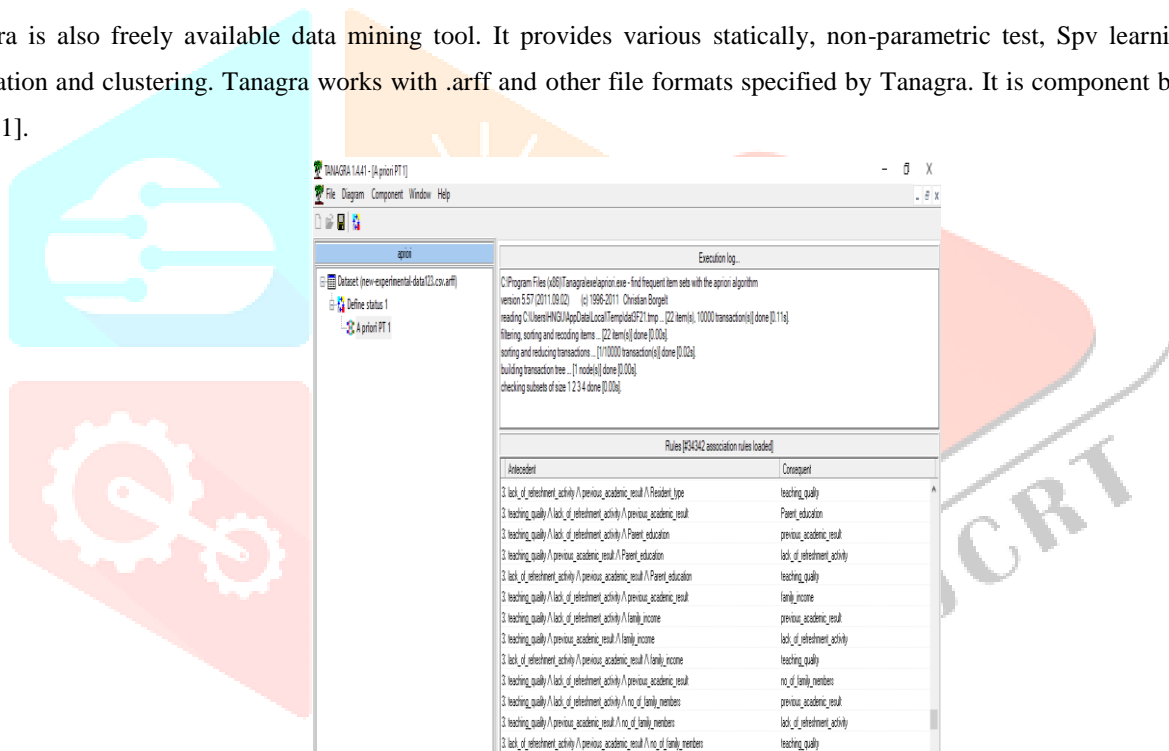


Figure 1.2- apriori in Tanagara

3.3 ORANGE

Orange is a perfect software suite for machine literacy & data mining. It stylish aids the data visualization and is a element-grounded software. As its software, the factors of orange are called “widgets” [11].

- Widgets offer major functionalities like
- Showing data table and allowing to elect features
- Reading the data
- Training predictors and to compare literacy
- Algorithms Imaging data elements etc.

Also, it brings a more interactive and delightful vibe to the dull logical tools. It's relatively intriguing to operate.

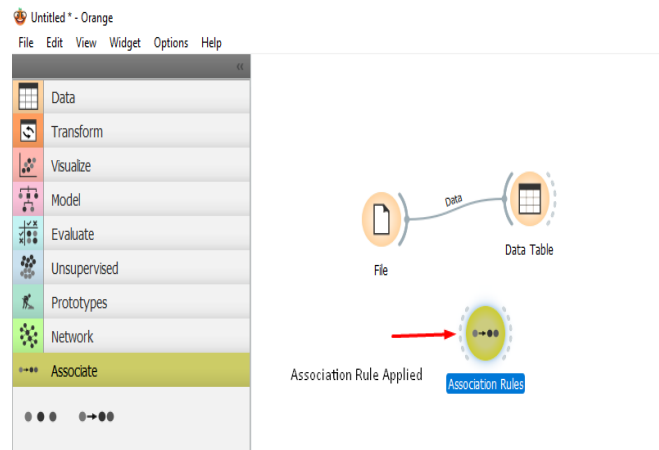


Figure 1.3- apriori in Orange

IV. RESULTS AND DISCUSSION

4.1 Results of TANAGRA

In Tanagra tool it will perform various association rule algorithms, it have apriori algorithm, pt-apriori algorithm, FP-growth algorithm as a part of association algorithm. In this research, researcher has found various of patterns after applying apriori algorithm on his dataset. The highest confidence value after applying algorithm it will provide 0.92. The following decision table shows the confidence values of particular patterns.

Table 4.1: Tanagra Tool rules

Rule No.	Rule Name	Confidence Value
1	Family Guidance, Class Attendance, Previous Semester Result, Gender	0.92
2	Family Guidance, Faculty Guidance, Previous Semester Result, Gender	0.90
3	Gender, Infrastructure, Family Income, Syllabus	0.87
4	No of Family Members, Previous Semester Result, No. of Hours Spend for Study	0.86
5	Gender, Parents Education, Family Guidance	0.84

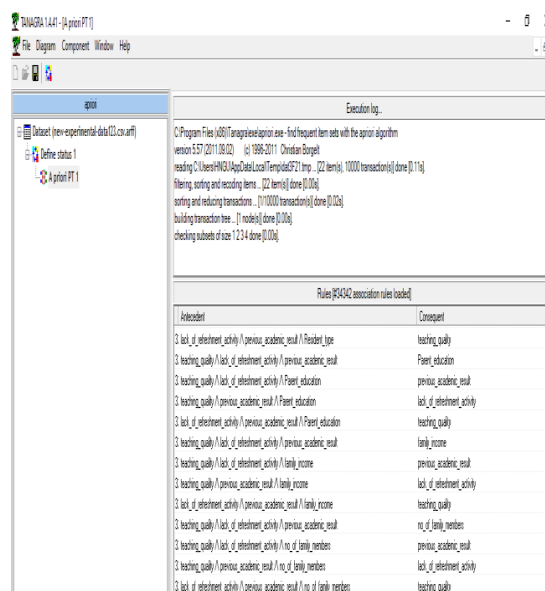


Figure 1.4- apriori result in Tanagara

4.2 Results of ORANGE

In Orange tool, it will perform associate rule algorithm for finding different patterns of given data. Again in this tool, it will provide different confidence level value of given data set. Here, using this tool it will have maximum of 0.89 confidence level. And all the rules with their confidence values will be show below.

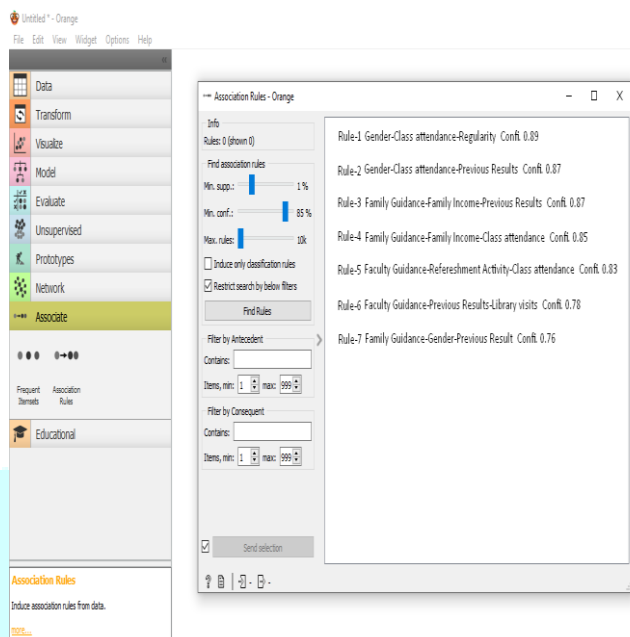


Figure 1.5- apriori result in Orange

Table 4.2: Orange Tool rules

Rule No.	Rule Name	Confidence Value
1	Gender, Class Attendance, Regularity	0.89
2	Gender, Class Attendance, Previous Semester Result	0.87
3	Family guidance, Family Income, Previous Semester Result	0.87
4	Family guidance, Family Income, Class Attendance	0.85
5	Faculty guidance, Refreshment Activity, Class Attendance	0.83
6	Faculty guidance, Previous Semester Result, Library Visits	0.78
7	Family Guidance, Gender, Previous Semester Result	0.76

4.3 Results of WEKA

In Weka tool, it will have different association rule mining algorithms, like, Apriori, FP-Growth algorithm. In this research paper, researcher applies apriori algorithm for finding different patterns. All rules with their confidence value displayed below. In below figure, we are clearly seen that the maximum confidence level value is 0.96.

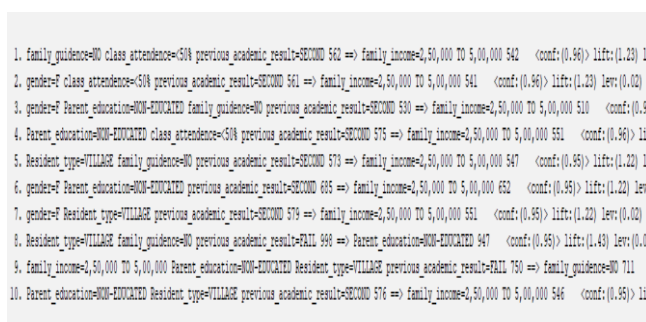


Figure 1.6- apriori result in Weka

V. CONCLUSION

Recent research suggests that data mining is the best technique for working on different datasets. The association rule mining is the best suited algorithm because it deals with different rules. In association rule, apriori algorithm is the best suited algorithm which provides best rule of given datasets. For applying the association rules, there are various of data mining tools available in market to found the best rule, in this research, researcher have uses 3 different tools with same dataset. And researcher can clearly found that WEKA have provides the best rule with highest confidence level values. Other tools also provide different rules but the confidence value is not higher as expected with researcher's demands. WEKA is also user friendly and open source tool which is widely available in market with free of cost.

REFERENCES

- [1] Amruta R. Joshil , Prof. A. Thomas 2017 Educational Data Mining for Classification of Students based on their Performance International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 5 Issue IV, April 2017, DOI: 10.22214/ijraset.2017.4163
- [2] Rakesh Agrawal and Ramakrishnan Srikant Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases, VLDB, pages 487-499, Santiago, Chile, September 1994.
- [3] The data science behind IP address matching Published by deductive.com, September 6, 2018, retrieved September 7, 2018
- [4] Bayardo Jr, Roberto J. (1998). "Efficiently mining long patterns from databases"
- [5] "Data Mining Techniques: Top 5 to Consider". Precisely. 2021-11-08. Retrieved 2021-12-10.
- [6] "16 Data Mining Techniques: The Complete List - Talend". Talend - A Leader in Data Integration & Data Integrity. Retrieved 2021-12-10
- [7] "Drawbacks and solutions of applying association rule mining in learning management systems". ResearchGate. Retrieved 2021-12-10.
- [8] Ankita Kalale, "Outlier Detection Methods and the Challenges for their Implementation with Streaming Data" Publication 25 November 2020.
- [9] P.K.Srimani & Annapurna S Kamath 2012, Data Mining Techniques for the Performance Analysis of a Learning Model A Case Study International Journal of Computer Applications Volume 53–No.5, September, 0975 – 8887
- [10] Varun Kumar, Anupama Chadha, 2012 "Mining Association Rules in Student Assessment Data", IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 3, September , p211-215
- [11] Manuel Fernandez-Delgado, Manuel Mucientes, Borja V 'azquez-Barreiros and Manuel Lama 2014 Learning Analytics for the Prediction of the Educational Objectives Achievement 2014 IEEE 7th International Conference on Engineering Education, ISSN: 0190-5848 DOI: 10.1109/FIE.2014.7044402
- [12] Pensri Amornsriphachai 2016 Efficiency of data mining models to predict academic performance and a cooperative learning model Faculty of Science and Technology, Nakhon Ratchasima Rajabhat University, Thailand, 978-1-4799-3922-0/14/\$31.00 ©2016 IEEE.
- [13] Karan Sukhija, Dr. Manish Jindal , Dr. Naveen Aggarwal 2015 The Recent State of Educational Data Mining: A Survey and Future Visions 2015, IEEE international Conference on Engineering Education DOI: 10.1109/MITE.2015.7375344.
- [14] Norlida Buniyamin, Usamah bin Mat, Pauziah Mohd Arshad 2015 Educational Data Mining for Prediction and Classification of Engineering Students Achievement 2015 IEEE 7th International Conference on Engineering Education (ICEED) ISBN: 978-1-47998810-5 DOI: 10.1109/ICEED.2015.7451491