ISSN: 2349-6002

Comparative Analysis of Tanagra and Rapid Miner in The Context of Higher Education

Dr. Rama Soni[1], Dr. Sumati Pathak[2], Dr. Indra Kumar Pandey[3], Ms. Madhavi Kaushik[4], Mr. Satanand Chandrakar[5]

[1,4,5]*Department of CSIT, Assistant Professor ,MUMT,Bilaspur(C.G.)*

[2]*Department of CS,, Assistant Ptofessor , E. Raghvendra Rao PG Science College, Bilaspur C.G*

[3]*Faculty of Computer Applications, Assistant Professor, Invertis University, Bareilly*

doi.org/10.64643/IJIRTV12I11-207105-459

***Abstract*—Data mining technology has significantly improved the ability to extract, store, and interpret enormous quantities of data, even different kinds of data samples. The prediction of students' academic performance is one of the most fascinating new pathways in the area of educational data mining. [1]. The quality of education plays a critical role in shaping an individual's future and a nation's prosperity. However, student academic performance is influenced by a wide variety of factors, not limited to classroom learning. These include individual behavior, family background, school environment, and access to learning resources. [2]. Higher education plays a crucial role in shaping the future of students and contributing to the development of a nation. Institutions aim to create an environment conducive to quality learning and academic excellence. [3]. In this research paper, we aim to evaluate the accuracy of student performance using data mining algorithms such as Decision Tree (J48), Random Forest, K-NN (IBK), Naïve Bayes, and Neural Networks or Multi-Layer Perceptron (AutoMLP). We will assess student performance and identify which algorithm provides the most accurate results, as well as determine the best-performing model from the dataset. Through this study, both students and institutions can measure their performance and also work towards improving it.**

***Index Terms*—Tanagra, Rapid Miner, Higher Education, Data Mining, Students Performance.**

I. INTRODUCTION

Higher education is performing very powerful involvement to develop our life and country. It is one of the powerful sources to shape the country. Higher education systems are associated with higher education learning institutes, college and Universities. Through Higher Education we can achieve more knowledge, good career, financial stability, analytical and logical thinking, discipline and creative skills. That's why Higher education is very beneficial for our country so to develop the student's performances is very important.

To develop the performance of students Data Mining play very powerful role. Through Data Mining we discover new method and knowledge from large data set. And also, we analyze the large data for knowing academic achievements that is very useful for student and also for the progress of our country.

Two data mining tools Tanagra and Rapid Miner will use in this research paper. We find out the accuracy of different algorithms of both tools Tanagra and Rapid Miner.

After that we compare the algorithms accuracy of Tanagra and Rapid Miner Data Mining Tool. Data Mining tools and techniques are very convenient to analyze large data set and also resolving the issues that's why we use Data Mining.

Objectives of the Study:

- Examine the data mining, Tanagra and Rapid Miner important tool of Data Mining tools.
- Examine the accuracy of Tanagra and different technique of Tanagra.
- Examine the accuracy of Rapid Miner and different technique of Rapid Miner.
- Match the accuracy of both Tanagra and Rapid Miner.
- Observe the which tool is the best data mining tools from Tanagra and Rapid Miner and also searched which is best techniques.

II. RELATED WORK

Educational Data Mining (EDM) has been widely used to analyze student data and predict academic performance. Aher and Lobo (2011) demonstrated the effectiveness of classification algorithms in educational datasets using WEKA, highlighting the potential of data mining in academic performance analysis [6].
Similarly, Ali (2013) emphasized the importance of data mining techniques for extracting valuable insights from educational databases and supporting decision-making processes [20].

Several researchers have compared different classification algorithms and data mining tools for student performance prediction. Mythili and Shanavas (2014) found that the choice of classification algorithm significantly affects prediction accuracy [22]. Khasanah et al. (2017) also reported that educational data mining techniques can effectively predict student outcomes and identify students requiring academic support [17].

Comparative studies of data mining tools have been conducted by Naika et al. (2016), who evaluated WEKA, RapidMiner, Tanagra, Orange, and KNIME, concluding that tool selection influences analytical performance and usability [10]. More recently, Soni et al. (2022) compared WEKA and Orange for higher education data analysis [4], while Soni et al. (2024) investigated the effectiveness of Orange and RapidMiner in predicting student performance [3]. Building upon these studies, the present research compares Tanagra and RapidMiner using multiple classification algorithms to identify the most suitable tool for higher education performance analysis.

III. METHODOLOGY / DATA COLLECTION

In this research paper we have used to secondary data and this secondary data collected from UCI Machine Repository. We have used Portuguese data set came from UCI Machine Learning Repository. This data actually has come from an environment where ICT technology has been used, however there are so many ICT techniques that is not used so far but we can implement those techniques to get more better results from any Data Set.

These data set already used and upload by P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance.

In A. Brito and J. Teixeira Eds., Proceedings of 5th Future Business Technology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7 [3].

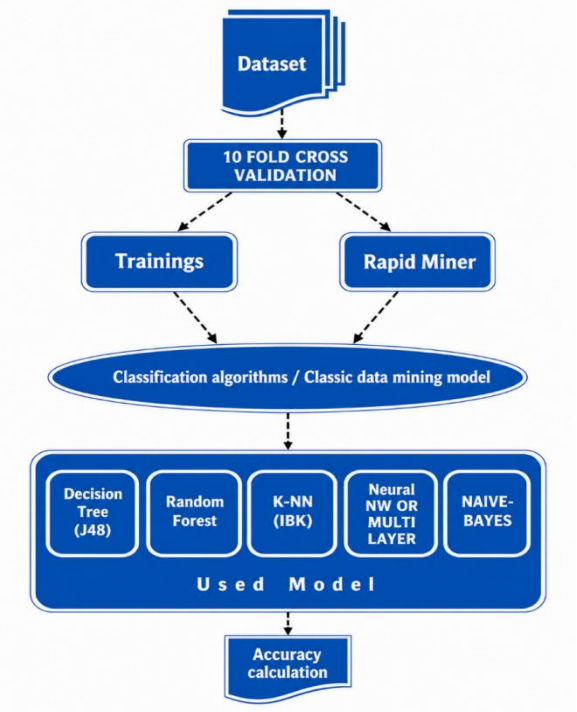

Fig.1. Performance Comparison of Classification Models in Tanagra and RapidMiner

In our data set there are overall 34 attributes which have several different capacities in these 34 attributes we have selected one attribute as our target which is actually binary attribute i.e., it has 1 & 0 only to be used.[3]
First of all, secondary data is collected from UCI Machine Repository by us and this collected data is associated with student's performance. In the next phase in this data set we applied data mining tools that are that are Tanagra and Rapid Miner with 10-fold cross validation.

After that we are applied different data models and algorithms such as Decision Tree, Random Forest, K-NN, Neural N/W, Naïve Bayes. After this we analyzed which algorithm is giving the best accuracy and which tool of Data Mining gives the best accuracy.

IV. RESULT

TABLE I. PERFORMANCE COMPARISON OF CLASSIFICATION MODELS IN TANAGRA AND RAPIDMINER

USED VALIDATION	USED MODELS	ACCURACY (%)	
		Tanagra	Rapid Miner
10-FOLD CROSS VALIDATION	Decision Tree (J48)	90.0	100
	Random Forest	NA	100
	K-NN (IBK)	85.78	98.1
	Neural N/W OR Multi-Layer Perception (AutoMLP)	92.13	98.5
	NAIVE-BAYES	89.19	99.15
	Decision Tree (J48)	90.0	100

Table I presents the comparative performance of five classification algorithms—Decision Tree (J48), Random Forest, K-Nearest Neighbors (K-NN), Neural Network, and Naïve Bayes—implemented using the Tanagra and RapidMiner data mining tools. The comparison is based on the classification accuracy achieved by each algorithm.

The results indicate that RapidMiner consistently outperformed Tanagra across all applicable algorithms. The Decision Tree (J48) and Random Forest algorithms implemented in RapidMiner achieved the highest classification accuracy of 100%, demonstrating superior predictive performance. In contrast, the Decision Tree algorithm in Tanagra achieved an accuracy of 90.00%, while the Random Forest algorithm was not available in Tanagra for evaluation.

The K-NN algorithm achieved 98.10% accuracy in RapidMiner compared with 85.78% in Tanagra. Similarly, the Neural Network algorithm produced 98.50% accuracy in RapidMiner and 92.13% in Tanagra. For the Naïve Bayes classifier, RapidMiner achieved 99.15% accuracy, whereas Tanagra recorded 89.19% accuracy.

A. Comparative Analysis of Tanagra

Among the evaluated algorithms in Tanagra, the Neural Network classifier achieved the highest accuracy of 92.13%, making it the best-performing technique within the Tanagra environment. The Decision Tree (J48) algorithm ranked second with 90.00% accuracy, followed by Naïve Bayes with 89.19% accuracy. The K-NN algorithm produced the lowest accuracy of 85.78% among the evaluated methods. Random Forest could not be included in the comparison because it was not available in the Tanagra tool.

The performance ranking for Tanagra is as follows:

1. Neural Network – 92.13%
2. Decision Tree (J48) – 90.00%
3. Naïve Bayes – 89.19%
4. K-NN – 85.78%

Comparative Analysis of RapidMiner

RapidMiner demonstrated consistently high classification performance across all evaluated algorithms. Both Decision Tree (J48) and Random Forest achieved the highest accuracy of 100%, indicating excellent predictive capability. Naïve Bayes ranked next with an accuracy of 99.15%, followed by Neural Network with 98.50% and K-NN with 98.10%.

The performance ranking for RapidMiner is as follows:

1. Decision Tree (J48) – 100%
2. Random Forest – 100%
3. Naïve Bayes – 99.15%
4. Neural Network – 98.50%
5. K-NN – 98.10%

The comparative evaluation demonstrates that RapidMiner achieved higher classification accuracy than Tanagra for all common algorithms. Decision Tree (J48) and Random Forest were the best-performing algorithms in RapidMiner, each achieving 100% accuracy. In Tanagra, the Neural Network algorithm performed best with an accuracy of 92.13%. Overall, the findings indicate that RapidMiner is the more effective data mining tool for the evaluated

dataset, as it consistently produced higher classification accuracy than Tanagra. Furthermore, Decision Tree (J48) and Random Forest emerged as the most effective classification techniques in RapidMiner, while the Neural Network was the best-performing technique in Tanagra.

V. CONCLUSION

We got enable to understand about the different Data Mining tools and technique like Tanagra and Rapid Miner from this research paper. We find the accuracy of different algorithms of Tanagra and Rapid Miner Data Mining Tool. After that we compare the accuracies of Tanagra and Rapid Miner Data Mining Tool. Through the comparison of accuracy of different algorithms, we tried to reach the aim of our research paper and we acquired the objective of our research paper. An ability to use Tanagra and Rapid Miner Data Mining tools through this research paper. By the use of different four technique of Tanagra, we achieve the accuracy of Tanagra. By the use of different five technique of Rapid Miner, we achieve the accuracy of Rapid Miner. We evaluate the Tanagra and Rapid Miner through comparison. Through comparison of Tanagra and Rapid Miner Data mining tool we got success to achieve the best Data Mining tools and also achieve best Data Mining technique.

VI. FUTURE WORK

Through this research paper students will be evaluating their state of progress of performance in forthcoming time. Students will also try to improve their performance progress on the basis of this evaluation. Since we are used data mining tools in this research paper that why data mining tools are very effective in research area and also very useful for every researcher.

REFERENCES

[1] R. Soni *et al*., "Educational data mining for student performance prediction: A hybrid model perspective," in *Proc. 1st Int. Conf. Research and Development in Information, Communication, and Computing Technologies (ICRDICCT)*, vol. 1, 2025, pp. 744–752, doi: 10.5220/0013872200004919.

[2] R. Soni et al., "Student performance analysis using educational and socio-economic indicators: A data-driven study," Swadeshi Research Foundation: A Monthly Journal of Multidisciplinary Research, vol. 12, no. 4, Feb. 2025.

[3] R. Soni et al., "Exploring the effectiveness of Orange and RapidMiner for higher education performance analysis," Shodhasamhita: Journal of Fundamental & Comparative Research, vol. X, no. 2, Jul.–Dec. 2024.

[4] R. Soni et al., "Comparison in higher education with the help of WEKA and ORANGE tools of data mining," in Proc. Int. E-Conf. Ancient Indian Knowledge System for Holistic Development in Engineering and Technology (CVRU-CON 2022), 2022, Paper No. CVRU-CON 2022/CSE&IT/PR/46.

[5] S. Pathak et al., "Application of ICT on student performance: An analysis based upon ORANGE and TANAGRA," in Proc. Nat. Conf. Emerging Trends and Technologies in the Area of IT, Science, Commerce and Management (SAICON 2020), 2020, ISBN: 978-81-937540-8-5.

[6] S. B. Aher et al., "Data mining in educational system using WEKA," in Proc. Int. Conf. Emerging Technology Trends (ICETT), 2011, published in International Journal of Computer Applications (IJCA).

[7] K. P. S. Attwal et al., "Exploring data mining tool—WEKA and using WEKA to build and evaluate predictive models," Advances and Applications in Mathematical Sciences, vol. 19, no. 6, Apr. 2020.

[8] E. G. Kulkarni et al., "WEKA: Powerful tool in data mining," International Journal of Computer Applications, National Seminar on Recent Trends in Data Mining (RTDM), 2016.

[9] S. K. David et al., "Comparative analysis of data mining tools and classification techniques using WEKA in medical bioinformatics," Computer Engineering and Intelligent Systems, vol. 4, no. 13, 2013.

[10] A. Naika et al., "Correlation review of classification algorithm using data mining tools: WEKA, RapidMiner, Tanagra, Orange and KNIME," in Proc. Int. Conf. Computational Modeling and Security (CMS), 2016.

[11] P. Devi et al., "Association rule mining on Spambase dataset using Tanagra," International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 9, no. 6, 2020.
[12] R. K. Kumar et al., "Comparative study on email spam classifier using data mining techniques," in Proc. Int. MultiConference of Engineers and Computer Scientists (IMECS), vol. I, Hong Kong, Mar. 14–16, 2012.
[13] H. Abdullah et al., "A comparison study between data mining tools over some classification methods," International Journal of Advanced Computer Science and Applications (IJACSA), Special Issue on Artificial Intelligence, 2011.
[14] S. Tanmayee et al., "A survey and comparative study of different data mining techniques for implementation of intrusion detection system," International Journal of Current Engineering and Technology, vol. 4, 2014.
[15] A. O. Ameen et al., "Performance evaluation of selected data mining software tools for data clustering," Dept. Comput. Sci., University of Ilorin, Kwara State, Nigeria, 2018.
[16] A. H. Wahbeh et al., "A comparison study between data mining tools over some classification methods," International Journal of Advanced Computer Science and Applications (IJACSA), Special Issue on Artificial Intelligence.
[17] A. U. Khasanah et al., "A comparative study to predict students' performance using educational data mining techniques," IOP Conference Series: Materials Science and Engineering, vol. 215, Art. no. 012036, 2017.
[18] A. H. Wahbeh et al., "A comparison study between data mining tools over some classification methods," International Journal of Advanced Computer Science and Applications (IJACSA), Special Issue on Artificial Intelligence, 2011.
[19] K. Rangra et al., "Comparative study of data mining tools," International Journal of Advanced Research in Computer Science and Software Engineering.
[20] M. Ali, "Role of data mining in education sector," International Journal of Computer Science and Mobile Computing, vol. 2, no. 4, pp. 374–383, 2013.
[21] A. Tekin et al., "Early prediction of students' grade point averages at graduation: A data mining approach," Eurasian Journal of Educational Research, no. 54, pp. 207–226, 2014.
[22] M. S. Mythili et al., "An analysis of students' performance using classification algorithms," IOSR Journal of Computer Engineering (IOSR-JCE), vol. 16, no. 1, ver. III, pp. 63–69, 2014.
[23] A. K. Shriwas et al., "Comparative analysis of models for student performance with data mining tools," Dept. Inf. Technol., Dr. C. V. Raman University, Chhattisgarh, India.
[24] D. Koech, "List of ICT tools for teaching and learning: Tech devices for teachers and students," Kenyayote, 2018.
[25] O. F. Noah et al., "Evaluation of student performance using data mining over a given data space," International Journal of Recent Technology and Engineering (IJRTE), vol. 2, no. 4, pp. 101–104, 2013.
[26] K. Singh, I. K. Pandey et al., "The power of automation: Streamlining processes and enhancing efficiency," Digital Transformation in Business: Integrating Technology with Management and Co*mmerce*, vol. 2, no. 4, pp. 32–37, 2025.