International Student Conference on Next-Gen Computing: Application of AI, Big Data, Quantum Computing, Signal Processing and Cloud Innovations (ICNGC-2026)

International Journal of Science, Engineering and Technology ISSN: 2348-4098, P-ISSN: 2395-4752

A Comprehensive Comparative Analysis of WEKAand Tanagra for Higher Education Data Mining

Dr. Rama Soni
Assistant Professor Computer Science
Maharishi University of Management and Technology, Mangla Bilaspur

Abstract- Now in days higher education is very important for development of our country as well as for the bright future of individual one . That's why the performance of students is abigdeal. Every education center is claiming that it gives quality education and best environment toits students. [1] But outcomes are not satisfying to our expectations so many researchers aretrying to research in this area to get accurate result of student's accuracy. [2] With this research we will calculated the accuracy and analyze the student performance withthe help of dif erent algorithms like K-NN, Decision Tree, Random Forest and Naive Bayes andwe will compare that which algorithm give accurate result from data set and which one is best. With this research paper students and educational institutions can predict and increase self performance as required.

Keywords: Weka, Tanagra, Higher Education, Data Mining, Students Performance.

I.INTRODUCTION

Higher education is an important part of our life. Without education life is not possible nowinday just because job, respect, knowledge, everything is depend on education. Higher educationis required for easy life and higher standard life. Higher education is also demanded for countrydevelopment and best students performance always play main role in higher education areathat's why it is necessary to increase the performance of the student so with the help of datamining tools we will analyze to the data and this data will be related to student performance.

In this research paper we will use two data mining tools Weka and Tanagra and we will alsocompare between both the tools that which tools will give more accuracy. Data mining is best technique to analyze the big data for analyzing the big data, data mining is more efficient andaffective that's why we choose data mining to analyze the data.

II. DATA COLLECTION

In this research paper we have used to secondary data and this secondary data collected fromUCI Machine Repository. We have used Portuguese data set came from UCI Machine LearningRepository. This data actually have come from an environment where ICT technology has beenused, 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 5thFuture Business Technology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7 [3].

International Journal of Science, Engineering and Technology ISSN: 2348-4098, P-ISSN: 2395-4752

In our data set there are overall 34 attributes which have several different capacities in these 34attributes we have selected one attribute as our target which is actually binary attribute i.e. it has 1 & 0 only to be used. [3]

III. OBJECTIVE

1. Talk about data mining and their two important tools that are Weka and Tanagra. 2. Find the accuracy in Weka and their different techniques.
3. Find the accuracy in Tanagra and their different techniques.
4. Compare the accuracy between Weka and Tanagra.
5. Find the best data mining tools in between Weka and Tanagra as well as best their techniques also.

IV. METHODOLOGY

First of all we have collected secondary data from UCI Machine Repository. This data is relatedto student's performance. In the next step we have applied this data set on data mining tools likeWeka and Tanagra with 10 fold cross validation. After that we have use different data model and algorithms like K-NN, Decision Tree, Random Forest and Naive Bayes and check out that which algorithm give us best accuracy and which tools of data mining is best about accuracy.

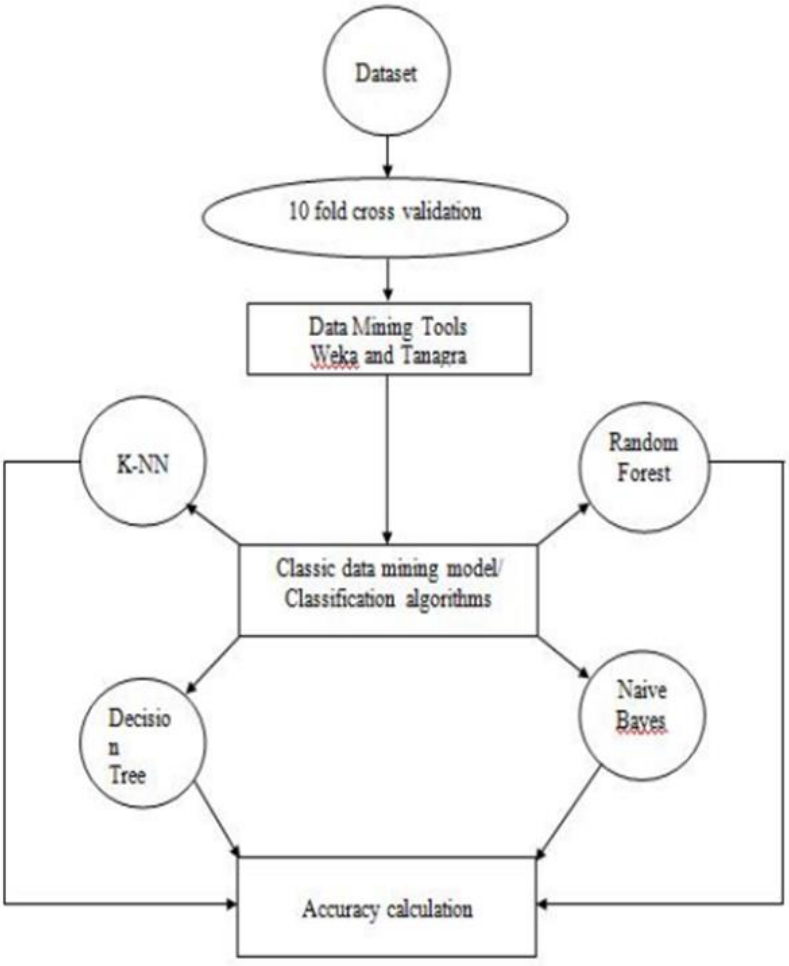

Figure1- Process of accuracy calculation

International Journal of Science, Engineering and Technology ISSN: 2348-4098, P-ISSN: 2395-4752

V. RESULT

Table1. Accuracy Comparison

USED VALIDATIONS	USED	ACCURACY(%)	
		MODELSWEKA	Tanagra
10 FOLDVALIDATION	Decision Tree	(J48) 100	90.0
	Random CROSS	Forest 98.6	NA
	K-NN	(IBK) 96.6	85.78
	NAIVE-BAYES	99.3	89.19

In this table we can see that we have used four main techniques like Decision Tree, RandomForest, K-NN and Naive Bayes in both data mining tools. And we found that the Decision Treealgorithm gave us the best results in both Weka and Tanagra data mining tools. With the helpof this table we analyzed the result like this –

1. Decision Tree of weka gave us 100% accuracy but the decision tree of Tanagra gave us 90% accuracy.
2. Random Forest of Weka gave us 98.6% accuracy but in the Tanagra this algorithmis not available.
3. K-NN of Weka gave us 96.6% accuracy but the K-NN algorithm of Tanagra gave us 85.78% accuracy
4. Naive Bayes of Weka gave us 99.3% accuracy but the Naive Bayes algorithmof Tanagra gave us 89.19% accuracy.

VI. WEKA VS TANAGRA

Now we will do a comparative study between Weka and Tanagra so that we can get an idea of the best tool and their best technique.

Weka

If we talk about Weka in the table we can see that -
1. Decision Tree gave us 100% accuracy.
2. Naive Bayes gave us 99.3% accuracy.
3. Random Forest gave us 98.6% accuracy.
4. K-NN gave us 96.6% accuracy.

So we can say that best technique of Weka is a Decision Tree because it is gave us 100%accuracy from big data set. Second best technique is Naive Bayes, after that randomforest is athird number of best technique in Weka and K-NN is a very poor algorithmabout accuracycompare than Decision Tree, Naive Bayes and Random Forest in Weka data mining tool.

Tanagra

If we talk about Tanagra in the table we can see that -
1. Decision Tree gave us 90% accuracy.
2. Naive Bayes gave us 89.19% accuracy.
3. K-NN gave us 85.78% accuracy.

International Journal of Science, Engineering and Technology ISSN: 2348-4098, P-ISSN: 2395-4752

4. Random Forest algorithm is not available in Tanagra data mining tool.

So decision tree is the best technique of Tanagra same as Weka. Naive Bayes second best technique in Tanagra, Random Forest is not available so we are unable to compare betweenWeka and Tanagra Random Forest technique but again we can say that K-NNis very poor algorithm about accuracy compare than Decision Tree and Naive Bayes in Tanagra algorithmdata mining tools.
So in this way we saw that the decision tree algorithm is the best in both Weka and Tanagra. But if we compare between Weka and Tanagra than we found that in every algorithmWeka is providing best accuracy then Tanagra so between Weka and Tanagra we have found Weka tobethe best data mining tool.

VII. CONCLUSION

In this research paper we learned about Weka and Tanagra data mining tools. Also triedtocheck the accuracy of all their algorithms, at the same time we tried to achieve our objectivein this research paper, in which we have been completely successful
1. In this research paper we learned about data mining as well as learned about their twoimportant tools Weka and Tanagra.
2. We achieve the accuracy of Weka through four important techniques.
3. We achieve the accuracy of Tanagra through four important techniques.
4. We did the comparative study between Weka and Tanagra.
5. And through this comparative study we found the best data mining tools along withthis we also got success in getting the best technique of those data mining tools.

FUTURE WORK:–
In future through this research students will be able to easily predict their overall performance and the data mining tools used in this research paper will prove useful for everyresearcher.

REFERENCES

1. Sumati Pathak et. al. (2019).”Weka vs Rapid Miner: Model Comparison In Higher Education WithThese Two Tools Of Data Mining”, paper present in the track on MICS of the ICIECE-2019 at Guru Nanak Institute Hydrabad, India.
2. Rama Soni et. al. (2022), “ Comparison in Higher Education with the help of WEKA and ORANGETools of Data Mining”, International E-Conference [CVRU-CON 2022] on The Ancient IndianKnowledge System For Holistic Development In Engineering And Technology, CVRU-CON2022/CSE&IT/PR/46.
3. Sumati Pathak et. al. (2020), ”Application of ICT on student performance: an analysis based uponORANGE and TANAGRA”, National conference on emerging trends and technologies in the area of IT science, commerce and management SAICON- 2020, NCSCCS- 021, ISBN: 978-81-937540-8-5.
4. Sunita B Aher et. al.(2011), “Data Mining in Educational System using WEKA”, International Conference on Emerging Technology Trends (ICETT), 2011 Proceedings published by International Journal of Computer Applications® (IJCA).
5. Kanwal Preet Singh Attwal et. al. (2020), “Exploring Data Mining Tool - Weka And Using Weka ToBuild And Evaluate Predictive Models”, Advances and Applications in Mathematical Sciences Volume 19, Issue 6, April 2020, © 2020 Mili Publications.

6. Eshwari Girish Kulkarni et.al. (2016), "WEKA Powerful Tool in Data Mining", International Journal of Computer Applications (0975 – 8887) National Seminar on Recent Trends in Data Mining(RTDM 2016).
7. Satish Kumar David et. al. (2013), "Comparative Analysis of Data Mining Tools and ClassificationTechniques using WEKA in Medical Bioinformatics", Computer Engineering and Intelligent Systems www.iiste.org, ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol.4, No.13, 2013.
8. Amrita Naika et. al. (2016), "Correlation review of classification algorithm using data mining tool: WEKA, Rapid miner, Tanagra, Orange and Knime", International Conference on Computational Modeling and Security (CMS 2016).
9. Pushpa Devi et.al. (2020) "Association Rule Mining on Spam base Dataset using Tanagra", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278- 3075 (Online), Volume-9 Issue-6, April 2020.
10. R. Kishore Kumar et. al. (2012), "Comparative Study on Email Spam Classifier using Data MiningTechniques", Proceedings of the International Multi Conference of Engineers and Computer Scientists 2012 Vol I, IMECS 2012, March 14 - 16, 2012, Hong Kong.
11. H. Abdullah et.al. (2011), "A Comparison Study between Data Mining Tools over some Classification Methods", (IJACSA) International Journal of Advanced Computer Science andApplications, Special Issue on Artificial Intelligence, September 2011.
12. S. Tanmayee et. al. (2014), "A Survey and Comparative Study of Different Data Mining Techniques for Implementation of Intrusion Detection System", International Journal of Current Engineeringand Technology, E-ISSN 2277 – 4106, P-ISSN 2347 - 5161 ©2014.
13. A. O. Ameen et. al. (2018), "Performance Evaluation Of Select Data Mining Software Tools For Data Clustering", Department of Computer Science, University of Ilorin, Kwara State, Nigeria, 2018.