RESEARCH ARTICLE

SAMRIDDHI Volume 12, Special Issue 3, 2020 Print ISSN : 2229-7111 Online ISSN : 2454-5767

Weka VS Rapid Miner: Models Comparison in Higher Education with these Two Tools of Data

Sumati Pathak*[1], Pragya Bhatt[2], Rohit Raja[3], Vaibhav Sharma[4]

[1] Research Scholar, Department of CSIT, DRCVRAMAN University Bilaspur, India; e-mail : sumati.gauraha@gmail.com
[2] Pragya Bhatt mphil (CS), MCA Bilaspur(C.G.); e-mail : prt297@gmail.com
[3] Department of CSE, Sreyas Institute of Engineering and Technology, Nagole, Hyderabad, India ; e-mail : drrohitraja1982@gmail.com
[4] Department of CSIT, DRCVRAMAN University Bilaspur, India; e-mail : sharma.vaibhav11@gmail.com

ABSTRACT

Now days all higher education institutions are facing many challenging tasks, every educational center is claiming that it gives quality education and best environment to its students. However satisfactory results were not found so far. That's why many researcher research in this field. In this paper we will try to predict students' performance in higher education sectors through comparative analysis of models with two data mining tools that are: Weka and Rapid Miner. In which we will try to figure out that which tool would give accurate result among the two.

Keywords: ICT Tools, Weka, Rapid.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology, (2020); DOI : 10.18090/samriddhi.v12iS3.19

INTRODUCTION

Today Information and Communication Technology (ICT) plays main role in higher education, especially in the field of computer science last few years many changes have occurred like- Vital use of Data Mining tools.

Data Mining is nothing but a process which helps us to find a pattern of data and search useful data through large amount of data. Data mining has many tools and all tools are helpful to search data in particular pattern. Various algorithms and techniques of data mining are available today most important of them are -

1. Classification: It is a data analysis task. It is a function of data mining or we can say that classification is a process which can categories and organize of data. It is a based on machine learning.

2. Clustering: Clustering is the grouping of a particular set of objects based on their characteristics, aggregating them according to their similarities [1].

Corresponding Author : Sumati Pathak, Research Scholar, Department of CSIT, DRCVRAMAN University Bilaspur, India; e-mail : sumati.gauraha@gmail.com

How to cite this article : Pathak, S., Bhatt, P., Raja, R., Sharma, V.(2020). Weka VS Rapid Miner: Models Comparison in Higher Education with these Two Tools of Data.

SAMRIDDHI : A Journal of Physical Sciences, Engineering and Technology, Volume 12 Special Issue (3), 85-88.

Source of support : Nil

Conflict of interest : None

3. Prediction: Predictive modeling is a process that uses data mining and probability to forecast outcomes. Each model is made up of a number of predictors, which are variables that are likely to influence future results [2].

4. Association rule: Association rules are if-then statements that help to show the probability of

relationships between data items within large data sets in various types of databases [3].

5. Neural Networks: Neural networks are used to model complex relationships between inputs and outputs or to find patterns in data [4].

In our paper we will use some of the 'Classification Techniques' with the two major tools of data mining they are –

1. Weka

2. Rapid Miner

In classification we are classify and identify in particular group or class of each data from data set. For example – Use of classification can classify the students of a class on their grade (average, good, excellent) basis. Tools of classification:

Weka

Weka is a collection of machine learning algorithms for data mining tasks. It is also perfect for new machine learning schemes. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization [5].

Rapid Miner

This is very popular, it is open source software, it's give facilities to cleansing, filtering, validation of data. Rapid Miner is a data science software platform developed by the company of the same name that provides an integrated environment for data preparation, machine learning, deep learning, text mining, and predictive analytics. It is used for business and commercial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the machine learning process including data preparation, results visualization, model validation and optimization. Rapid Miner is developed on an open core model [6].

PROS AND CONS OF DATA MINING

Data mining is very useful in ICT. If data mining have some pros so it has some cons also.

Pros of data mining

Data Mining has many Pros that are why it is an important part of ICT.

- Data Mining helps marketing companies build models based on historical data to predict who will respond to the new marketing campaign....etc [7].
- Data mining helps to financial institute like- bank and give help us for loan and credit score etc.
- Data Mining helps to select useful data from large amount of data set.
- Data Mining helps to divide particular class and pattern of any data from data set.
- It helps in identifying criminal suspects by law [7].

Cons of data mining

Data mining has many Cons also, let's see

- Data mining is costly just because the queries of data mining are very complex that's why collection of more data and more maintenance are required.
- Data mining have no privacy and security that's why data security is big issue in data mining.
- Data mining is not suitable for fast access of data and it cannot manage new data and can't update old data [8-13].

PROPOSED PLAN

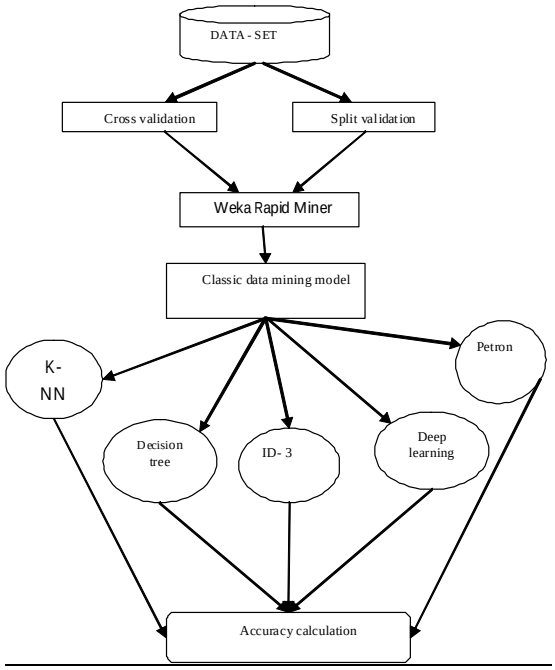
In this paper our proposed plan is as follows:-

- Data Set: we have taken Data-Set of higher education from UCI machine-learning repository [14-17]. It has one Portuguese data which has 650 records in which there are 34 attributes which record is calculated as per many properties such as student ID, Romantic, health, address, sex, age etc.These data set already used and upload by P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance.
- Method of Validation: In this paper we have used two types of validation method which one is 10 fold Cross Validation method and another one is Split Validation method. Both methods are ideal for Validation process of data.

Accuracy Table

We have worked on two tools of data mining Weka and Rapid Miner. We have got the following Accuracy:

USED VALIDATION	USED MODELS	ACCURACY(%)	
		WEKA	RAPID MINER
10 FOLD CROSS VALIDATION	Decision Tree (J48)	100	100
	Random Forest	98.6	100
	K-NN (IBK)	96.6	98.1
	MULTI LAYER Perception (AutoMLP)	97.9	98.5
	NAIVE-BAYES	99.3	99.15
SPLIT VALIDATION	Decision Tree (J48)	99.6	94.67
	Random Forest	97.0	94.67
	K-NN (IBK)	92.9	94.67
	MULTI LAYER Perception (AutoMLP)	94.4	97.22
	NAIVE-BAYES	97.7	94.67



CONCLUSION

Here in this research as we can see in the above table some algorithms give best accuracy in Rapid Miner where as some give best accuracy in weka but for a precise conclusion we can say that Rapid Miner somehow best in class because of more algorithm, better accuracy and greater functionality it has. However any research doesn't claims that this is the last result, there are so many chance of changes in result or many more algorithms enhancements are possible.

REFERENCES

[1] Y Vidya, B Shemimol Secured Friending in Proximity based Mobile Social Network. Journal of Excellence in Computer Science and Engineering, Vol. 1(2) 2015, pp. 1-10

[2] Bøe, T., Gulbrandsen, B. and Sørebø, Ø., 2015. How to stimulate the continued use of ICT in higher education: Integrating information systems continuance theory and agency theory. Computers in Human Behavior, 50, pp.375-384.

[3] Sørebø, Ø., Halvari, H., Gulli, V. F., & Kristiansen, R. (2009). The role of selfdetermination theory in explaining teachers' motivation to continue to use elearning technology. Computers an d Education, 53, 1177–1187.

[4] Jasperson, J.S., Carter, P.E. and Zmud, R.W., 2005. A comprehensive conceptualization of post-adoptive behaviors associated with information technology enabled work systems. MIS quarterly, 29(3), pp.525-557.

[5] Osei-Bryson, K.-M. (2010). Towards supporting expert evaluation of clustering results using a data mining process model. Information Sciences, 180(3), pp.414–431

[6] Sinha, A. P., & Zhao, H. (2008). Incorporating domain knowledge into data mining classifiers: An application in indirect lending. Decision Support System, 46(1), pp.287–299.

[7] Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy,32, 1761–1768

[8] A K Shriwas & Pragya Tiwari "Comparative analysis of models for student performance with data mining tools" Department of IT, Dr. C V. Raman University (C.G.) India.

[9] Khushbu N.Shah1, Misika R. Patel2, Neha V. Trivedi3, Priyanka N. Gadariya4, Richa H. Shah5, Ms. Nehal Adhvaryu6 Department of Computer Science and Technology, Uka Tarsadia University Bardoli, Gujarat, India.

[10] O. F. Noah and B. Barida, "Evaluation of Student Performance Using Data Mining over a Given Data Space", International Journal of Recent Technology and Engineering (IJRTE), Volume-2, Issue-4, pp.101-104, 2013.

[11] 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 978-9077381-39-3.

[12] Mrs. Bharti M. Ravageri Modern Institute of Information Technology and research, Department of Computer application, Pune Maharashtra, India "Data Mining Techniques and Applications.

[13] Sumati Pathak, Rohit Raja, Vaibhav Sharma, and K. Ramya Laxmi, A Framework Of ICT Implementation On Higher Educational Institution With Data Mining Approach, European Journal of Engineering Research and Science, ISSN (Online) : 2506-8016, Aprl 2019.

[14] Sumati Pathak, Rohit Raja, Vaibhav Sharma The Impact of ICT in Higher Education. Published in IJRECE Vol. 7 Issues 1 January-March, 2019. ISSN: 2393-9028 (PRINT) ISSN: 2348-2281 (ONLINE) ISSN: 2393-9028 (PRINT). Vol 7, Issue 1, pp 1650-1656. (UGC Approved)

[15] Rohit Raja, Tilendra Shishir Sinha, Raj Kumar Patra and Shrikant Tiwari(2018), Physiological Trait Based Biometrical Authentication of Human-Face Using LGXP and ANN Techniques, Int. J. of Information and Computer Security, Vol. 10, Nos. 2/3, pp. 303- 320.

[16] A. C. Bhensle and Rohit Raja (2014), An efficient face recognition using PCA and Euclidean Distance classification, An efficient face recognition using PCA and Euclidean Distance classification, Vol. 3 Issue.6, pp. 407-413. ISSN: 2320–088X.

[17] Rakesh Kumar Lenka, Amiya Kumar Rath, Zhiyuan Tan, Suraj Sharma, Deepak Puthal, N V R Simha, Rohit Raja, Shankar Sharan Tripathi, and Mukesh Prasad Building Scalable Cyber-Physical-Social Networking Infrastructure Using IoT and Low Power Sensors, , Vol. 6, Iss. 1, pp. 30162-30173, Print ISSN: 2169-3536, Online ISSN: 2169-3536, Digital Object Identifier: 10.1109/ACCESS.2018.2842760. (SCI Index)