

# Mining of Data by WEKA Tool

Sharnjeet Kaur<sup>1</sup>

<sup>1</sup>Assistant Professor, CSE, Universal Group of Institute Lalru, India, Sharn4765@yahoo.com

\*Correspondence: Sharn4765@yahoo.com

**ABSTRACT-** Data mining is a term that is used not only for mining data but to differentiate various algorithms with each other in terms of their parameters. Sometimes we got that information about the data that we don't know. Mining of data depends on the working of different algorithms. Some algorithm gives a more accurate result than other while working in same area. The data is arranged according to their properties by different techniques such as clustering and classification technique. Various algorithms such as neural networks, decision tree and support vector machine algorithm come under the category of classification, by using WEKA tool.

**Keywords:** Classification, Clustering, Decision tree, Neural Networks.

## ARTICLE INFORMATION

**Author(s):** Sharnjeet Kaur;

**Received:** 19/03/2023; **Accepted:** 03/06/2023; **Published:** 30/06/2023;

**e-ISSN:** XXXX-XXXX;

**Paper Id:** IJCSR-020203;

**Citation:** 10.37391/IJCSR.020203



**Publisher's Note:** FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

## 1. INTRODUCTION

The two terms data and mining are defined according to requirements in different fields. Mining means extracting an important thing from a heap and data means information regarding to any field. The extraction of information according to their requirement from the heap refers to data mining. Data mining is vast field for technology. Various researching technologies are based on data mining algorithms. The clustering and classification algorithm are used according to their need. In classification the objects are assigned to predetermined classes. In clustering the objects are assigned to clusters according to their properties. Both techniques have their own application in technical field based on their evaluation criteria. Now day's data mining famous in medical field due to their applications. Various algorithms such as Genetic Algorithm, Bacterial Foraging algorithm, Particle Foraging Algorithm, Ant Colony Algorithm are used in clustering techniques as an optimization algorithms. In this research paper the classification technique is defined using various classification algorithms such as simple k-mean, decision tree etc. using WEKA tool. WEKA tool is a powerful tool used for classification and clustering of data in a proper or organized manner. In WEKA tool various prewritten algorithm are present that are used in various research studies.

## 2. LITERATURE SURVEY

Mark Hall presents the working of WEKA tool. This research paper gives the basic idea behind the WEKA tool and main functioning of WEKA tool.

Eibe Frank gives a work on bioinformatics using a WEKA tool in data mining. In this process is defined to make an analysis result on data using WEKA tool in different fields.

Micheline Kamber describes the two algorithms MedGen and MedGen Adjust the solves the scalability and efficiency issues in large database. Each algorithm generalise the data to an intermediate level of abstraction and also compress this data to use it into a better way.

Jyoti Soni presents the various parameters that detect the heart attack using different type of algorithms. In this paper the performance of decision tree based algorithm and Bayesian classification algorithm are compared with KNN and Neural Network. The decision tree and Bayesian Classification gives better result with genetic algorithm than KNN and Neural network.

Ashish Kumar Sen present the work on prediction of coronary heart disease using a neural fuzzy approach. In this paper the evaluated parameters are divided on two parts based on criticality. Then separate weights are assigned to each layer.

D. Lavanya provides the evaluation results on medical dataset by using decision tree classifiers. Parameters such as time complexity and time are also evaluated.

## 3. DATA MINING

During the mining of important data we use various data mining techniques. Classification and Clustering are data mining techniques. In classification we use predetermined classes to assign objects to classes. While Clustering is a technique classes are made on the basis of features of objects. In data mining both Classification and clustering techniques follows various algorithms to arrange the data in particular way.

## 4. CLASSIFICATION

Classification is a process of arranging the objects in predetermined classes. For example, if we consider four objects of different color then these objects are assigned to classes according on the bases of their color. In classification further various algorithm are used to arrange the objects in particular

way. The following are the algorithms that are used in classification technique:

#### 4.1 Non-Parametric or K-Nearest Neighbor

In this classification techniques the object are assigned to classes on the bases of their value. In the K-Nearest Neighbor the attributes are arranged according to their nearest feature or value into predetermined labels.

#### 4.2 Rule Based or Decision Tree

Decision tree algorithm divides the dataset in subsets by arranging them into tree like structure. In Tree the top most nodes is called root node and root node is best predictor. Tree gives a decision node decision nodes and leaf nodes. In decision tree the nodes represent attributes. Branches represent the values of the attribute.

#### 4.3 Neural Networks

Neural Network is a system that contains programs and databases that work with human brain. A neural network contains large number of processor that work parallel with each other. A neural network is trained and fed with data inputs to get an appropriate result.

### 5. WEKA TOOLS

WEKA tool is the collection of machine learning algorithms such as K-MEAN, CJ algorithm, Decision tree, Clustering, Association rules and Visualization WEKA is open source tool. Its name came from a bird.



Figure 1: Shows the WEKA Tool

### 6. METHODOLOGY

The heading of a section should be in Times New Roman 12-point bold in all-capitals flush left with an additional 6-points of white space above the section head. Sections and subsequent sub- sections should be numbered and flush left. For a section head and a subsection head together (such as *Section 3* and *subsection 3.1*), use no additional space above the subsection head.

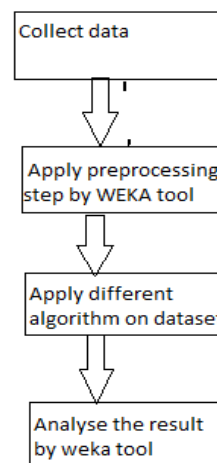


Figure 2: Shows the Methodology

#### 6.1 Dataset

This radar data was collected by a system in Goose Bay, Labrador. This system consists of a phased array of 16 high-frequency antennas with a total transmitted power on the order 6.4 kilowatts. See the paper for more details. The target were free electrons in the ionosphere. “Good” radar returns are those showing evidence of some type of structure in the ionosphere. “Bad” returns are those that do not; their signals pass through the ionosphere. Received signals were processed using an autocorrelation function whose arguments are the time of the pulse and pulse. There are 17 pulse members of the goose bay system. Instances in this database are described by 2 attributes per pulse number, corresponding to the complex values returned by the function resulting from complex electromagnetic signals.

### 7. RESULTS

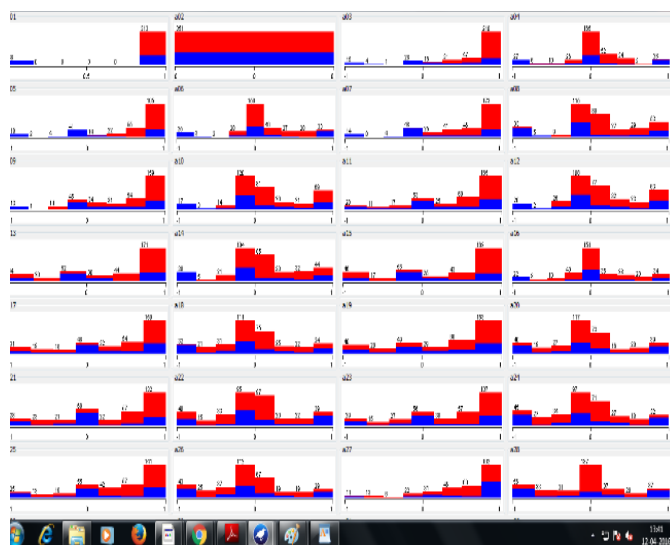


Figure 3: Shows the preprocessing step by WEKA tool

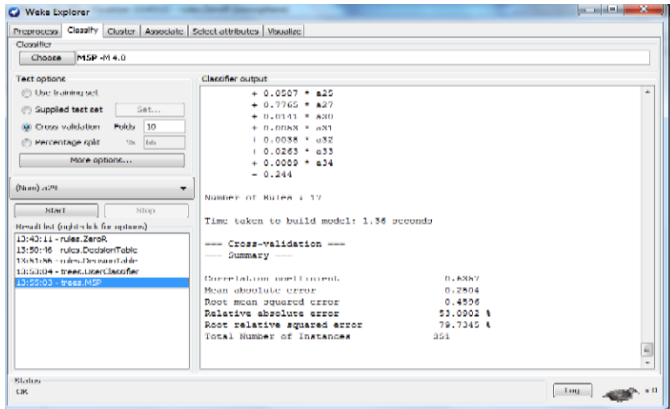


Figure 4: Shows the result of M5J algorithm

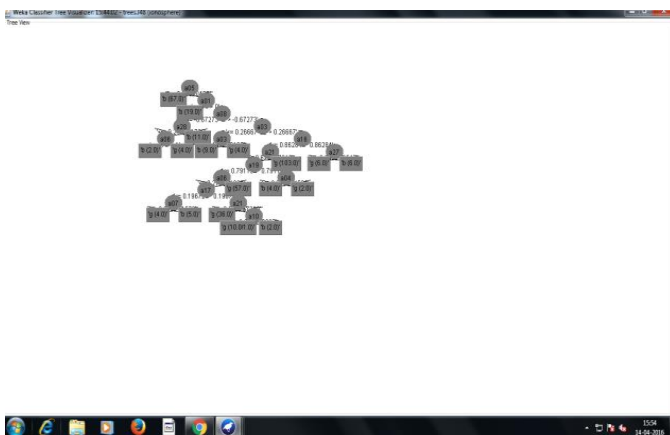


Figure 5: Shows the decision tree algorithm

Table 1: Shows the result of Naïve Bayes algorithm

Total Number of instances	351
Correctly Classified Instances	290
Incorrectly classified Instances	61
Root mean Squared error	0.3935

## 8. CONCLUSION

This research paper shows the classification of data by WEKA tool. In this paper various algorithms are defined to show the result of WEKA tool. WEKA tool is a powerful tool that works with various algorithms to show results in proper and fast way. We can use this tool to analyze different algorithms with their result

## REFERENCES

[1] Kamber, Micheline, et al. "Generalization and decision tree induction: efficient classification in data mining." Research Issues in Data Engineering, 1997. Proceedings. Seventh International Workshop on. IEEE, 1997.

[2] Mobasher, Bamshad, Robert Cooley, and Jaideep Srivastava. "Automatic personalization based on web usage mining." Communications of the ACM 43.8 (2000): 142-151.

[3] Antonie, Maria-Luiza, Osmar R. Zaiane, and Alexandru Coman. "Application of Data Mining Techniques for Medical Image Classification." MDM/KDD 2001 (2001): 94-101.

[4] Kantardzic, Mehmed. Data mining: concepts, models, methods, and algorithms. John Wiley & Sons, 2011.

[5] KamHan, Jiawei, Micheline Kamber, and Jian Pei. Data mining: concepts and techniques. Elsevier, 2011.1997.

[6] Mark A., and Geoffrey Holmes. "Benchmarking attribute selection techniques for discrete class data mining." Knowledge and Data Engineering, IEEE Transactions on 15.6 (2003):

[7] Frank, Eibe, et al. "Data mining in bio informatics using weka" Bioinformatics 20.15 (2004): 2479-2481

[8] Y.T. Yu, M.F. Lau, "A comparison of MC/DC, MUMCUT and several other coverage criteria for logical decisions", Journal of Systems and Software, 2005, in press.

[9] Berkhin, Pavel. "A survey of clustering data mining techniques." Grouping multidimensional data. Springer Berlin Heidelberg, 2006. 25-71.

[10] Wu, Xindong, et al. "Top 10 algorithms in data mining." Knowledge and information systems 14.1 (2008): 1-37.

[11] Hall, Mark, et al. "The WEKA data mining software: an update." ACM SIGKDD explorations newsletter 11.1 (2009): 10-18.

[12] Soni, Jyoti, et al. "Predictive data mining for medical diagnosis: An overview of heart disease prediction." International Journal of Computer Applications 17.8 (2011): 43-48.

[13] Lavanya, D., and K. Usha Rani. "Performance evaluation of decision tree classifiers on medical datasets." IJCA International Journal of Computer Applications 26.4 (2011).

[14] Sen, Ashish Kumar, Shamsher Bahadur Patel, and D. P. Shukla. "A data mining technique for prediction of coronary heart disease using neuro-fuzzy integrated approach two level." International Journal of Engineering and Computer Science ISSN (2013): 2319-724.



© 2023 by the Sharnjeet Kaur. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).