ITJ | The International Tax Journal

Comparative study of data mining algorithm between orange and Weka: Theoretical study

Professor. Fatiha Bouhrine [1], **Professor. Salima Tebaibia** [2] and **Doctor. Bilel Merabet** [3]

[1] Faculty of science economic, commercial and science management, University of constantine2 – Abdelhamid Mehri (Algeria), Email: fatiha.bouhrine@univ-constantine2.dz
[2] Faculty of science economic, commercial and science management, University of 8 mai 1945 – Guelma (Algeria), Email: salima.tebaibia@univ-guelma.dz
[3] Faculty of science economic, commercial and science management, University of Algiers 3 (Algeria), Email: merabet.bilel@univ-alger3.dz

Abstract---This study aims to conduct a comparison between two important programs, namely Weka and Orange, in terms of several aspects, the most important of which are the algorithms used in both. The study reached several results, the most important of which is that both programs are open source.

Keywords---data mining; algorithm; weka; orange; comparative.
JEL Classification Codes: c70, c88.

I- Introduction

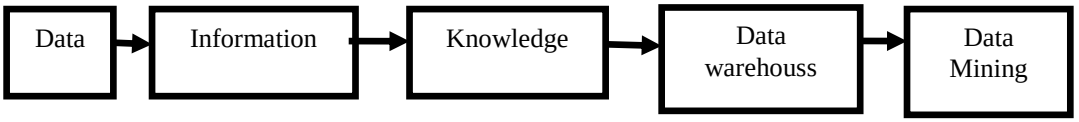
Data mining is one of the important topics as it given en new view of the data. The decision- magink process at the level of institutions of all kinds depends on accurate information. Hence, it becomes clear to us that data is a raw material for building a data society, after processing it to benefit from it in decision- making, it's called the information , latter is stored and collected, and here it's culled the term knowledge, through the above, we can define data mining as the discovery of knowledge within databases.

How to Cite:
Bouhrine, F., Tebaibia, S., & Merabet, B. (2025). Comparative study of data mining algorithm between orange and Weka: Theoretical study. *The International Tax Journal*, *52*(1), 79–84. Retrieved from https://internationaltaxjournal.online/index.php/itj/article/view/42

The International tax journal ISSN: 0097-7314 E-ISSN: 3066-2370 © 2025

Submitted: 9 Sept 2024 | Revised: 18 Nov 2024 | Accepted: 16 Dec 2024

Figer 1: defined data mining



There are many methods that can be used for data mining including algorithms, accordingly, this paper came to compare the algorithms related to data mining, which exist in several programs, including orange &weka. Therefore, the aim of this paper is to compare the data mining algorithms in orange and weka.

I.1 Literature review:

There is significant theoretical debate among researchers retarding appropriate algorithms for knowledge discovery. As ether were many and different studies including in table 1.

Table1: summary of literature review contributions according to area of research

year	author	Study title	Main contribution
October8, 2007	- Xindony wu - Vipin kumar	Top 10 algoritms in data mining	The study dealt with the ten algorithms identified by the IEEE international conference on data mining in December 2006. These 10 algorithms are all among the most impotent topics in data mining research and development.
October 2012	- Fartash. Haghanikhameneh - Payam. Hassany Shariat Panahy - Nasim. Khanahmadliravi - Seyed Ahmad. Mousav	A Comparison Study between Data Mining Algorithms over Classification Techniques in Squid Dataset	This study deal with algorithms related to classification which is considered the most important supervised technique in data mining
September 2015	- Pankaj Singh - Sudhakar Singh - Rakhi Garg - Devisha Singh	Comparative Study of Data Mining Algorithms through Weka	In this paper the authors exposed the algorithms used within the weka program, which is one of the data mining tools.
January 2021	- Madhu Kahar - Manisha Kahar*	Comparative analysis of data mining algorithms and techniques	This paper help us to know the information about the data mining technologies, data mining algorithms and some data of corporation which have changed their technologies of data mining to midgrade or enhance their profits and got impressive output from it.

I. 2. Data mining and algorithms:

Data mining is a way to extract useful data and patterns from big data. It is also known as the process of obtaining information, extracting information from data, extracting information or analyzing

information. Data mining is a sensible process used for search by using a large amount of information to obtain useful data (Madhu Kahar, Manisha Kahar: 2022, p 969). Data mining depends on two pillars. The first is data mining strategies such as classification and prediction, the second is scientific method for data mining and detection of new patterns and relationships, such as decision trees and genetic algorithms (shama.A & panigraph PK: February 2012, p 37-48). In two pillars of data mining is used algorithms.

A data mining algorithm is a set of heuristics and calculations that creates a data mining model from data (Sethunya R Joseph, Hlomani Hlomani, Keletso Letsholo: april 2016, p6806).

Data mining technically consists of applying a data-mining algorithm to the cleansed dataset to discover hidden patterns and relationships. A data- mining algorithm is a set of heuristics and calculations (Liane Colonna: December 2017, p 331). We have many data mining algorithms including in table 2.

Table2: summary of data mining algorithms

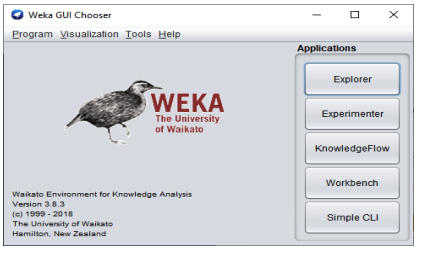
Name of algorithm	year
chaid	1980
cart	1984
ID3	1986
C4.5	1993
SLIQ	1996
SPRINT	1996
J48	1996
SPRINT	1996
Learn++	2001
SSDM	2005

Whatever the type of algorithm, its use of several programs and in this paper we will focus on Orange and weka.

I. 3. Data mining algorithm in weka & orange:

In this part of the paper, we will address the algorithms that can be applied to both Weka and Orange.

A- Weka:



is an abbreviation for Waikato Environment for Knowledge Analysis. Development of weka begun in 1997. It has many advantages they are:

- ✓ Open source
- ✓ The program contains a set of data processing and modeling tools
- ✓ The program features an easy-to-use graphical interface

Can handle other programs Available functions in weka are:

- ➢ Filter and Preprocess on data
- ➢ Classification
- ➢ Clustering
- ➢ Association rules
- ➢ Selecting attributes
- ➢ Viewer

The reason why WEKA is chosen for the experiments is that the Factor Analysis or feature selection algorithms and especially information gain are almost identically. Implemented when compared with the original algorithms. These implementations exist under the classification and regression capabilities of the too(Amal Al-Hatali: 2018,p 12).

a- Classification: Weka implements a very large number of classifiers (based on rules, trees, Bayesian networks, etc.). These CLASSIFICATION ALGORITHMS include:

- ZeroR: majority class rule;
- J48: decision tree;
- NaiveBayes: Naive Bayes;
- IBk: KNN

b- Clustering : CLOPE, Cobweb, OPTICS, farthest first, EM, DBSCAN, Simple K-mean and Hierarchical are various Clustering algorithms and we have applied Weka tool on Simple K-mean, Hierarchical, DBSCAN and EM on the diabetes data set (Pankaj Singh, Sudhakar Singh, Rakhi Garg, Devisha Singh: September 2015, p8).

B- **Orange**:

Orange is an open source tool dedicated to data mining and machin6 e learning.

Orange is very effective when the concept of innovation, reliability or quality is involved.

It has many advantages they are:

- ✓ Open source ;
- ✓ Ability to perform multiple operations on data;
- ✓ Rich and interactive graphical visualization ;
- ✓ Visual programming ;
- ✓ Orange as a Python module.

It gives a platform for experiment selection, predictive modeling, and recommendation systems and can be used of genomic research, biomedicine, bioinformatics, and teaching. Orange is always preferred when the factor of innovation, quality, or reliability is involved.(Rajeev Shishodia, Prabhakar Anand, Viswanathan.R :2019, p55).

In orange we use many algorithms for classification , regression and the decision tree excembal:

- ID3
- J48
- C4.5
- CART
- kNN

I- 4. A comparative study between Weka and Orange:

We have many deferent between weka and orange we give theme in the table

Table3: deference between weka and orange

statement	weka	orange
Organization / Country	University of Waikato (New Zealand)	University of Ljubljana (Slovenia)
Programming Language	java	Python, C++

Price	free	
OS Platform	Windows, Mac OS X, Linux	
Availability	Open source	
Compatibility with database	ARFF,CSV, C4.5	CSV and Excel

The past table explained the deference between weka and orange in general, but the table N 4 and 5 we have Shawn Technical comparison of WEKA and Orange.

Table4: Comparative study of WEKA and Orange tool Precision Metric

Classifier	WEKA(%)	Orange(%)
Naïve bays	83.7	82.4
Random Forest	81.8	77.9
k-nearest	75.3	58.0

Source; Ritu Ratra , Preeti Gulia ; Experimental Evaluation of Open Source Data Mining Tools (WEKA and Orange), International Journal of Engineering Trends and Technology (IJETT) – Volume 68 Issue 8 - Aug 2020, p 33

Table5: Comparative study of WEKA and Orange tool Recall Metric

Classifier	WEKA(%)	Orange(%)
Naïve bays	83.7	80.6
Random Forest	81.9	73.4
k-nearest	75.2	54.7

Source; idem, p 34

II- Conclusion:

According to this Theoretical analysis, a number of results were reached, which are:

1- They are many program about data mining
2- They are many algorithm used in weka and orange
3- In weka and orange the program features an easy to use graphical interface
4- Algorithm in weka and orange are easy to use because they don't require writing code, as those who aren't skilled in the field can create with ease by relying on the features of the drab 'and –drop interface

References

1) Madhu Kahar, Manisha Kahar: (January2021) , A COMPARATIVE ANALYSIS OF DATA MINING-ALGORITHMS AND TECHNIQUES, International Research Journal of Modernization in Engineering Technology and Science, Volume03, Issue 01,.
2) shama.A & panigraph PK; (February2012) areview of financial accounting fraud detection based on data mining techniques, international journal of computer applications, volume39 , number 1,
3) Sethunya R Joseph, Hlomani Hlomani, Keletso Letsholo:(april 2016) Data Mining Algorithms: An Overview, international journal of computers and technology, volume 15, number 6, .
4) Liane Colonna : (December 2017), A Taxonomy and Classification of Data Mining, the SMU *Science & Technology Law Review* , volume 16, number 2, , https://scholar.smu.edu/scitech .
5) Amal Al-Hatali: (2018) A Comparative Study of the Efficient Data Mining Algorithm to find the most influenced factor on price variation in Oman Fish Markets, Scholar Journal of Applied Sciences and Research, Volume 5, number 1.
6) Pankaj Singh, Sudhakar Singh, Rakhi Garg, Devisha Singh:(September 2015), Comparative Study of Data Mining Algorithms through Weka, International Journal of Emerging Research in Management &Technology, Volume 4, Issue9.

7) Ritu Ratra , Preeti Gulia ;(Aug 2020) ,Experimental Evaluation of Open Source Data Mining Tools (WEKA and Orange), International Journal of Engineering Trends and Technology (IJETT) , Volume 68 ,Issue 8 .