IRAQI Academic Scientific Journals

Samarra Journal of Pure and Applied Science

SJPAS

www.sjpas.com

ISSN:2663-7405

Actual Needs Criteria for Assessing Data Classification Platforms

Baydaa M. Merzah*

Information Technology Unite, College of Political Sciences, Al-Nahrain University Iraq (baydaa@ced.nahrainuniv.edu.iq)

Article Information

Received: 25/12/2020
Accepted: 10/02/2021

Keywords:

Data Classification, Data Mining Platforms, Platforms Assessment, Software Tools.

Abstract

A Software tools have an important role in different research areas. Generally they provide time and efforts saving. In computer science filed these tools can help in communications, web site development, software metrics finding, data mining, machine learning and many other fields. There are many specialized tools built to support specific purpose. Users and researchers spend a lot of time and efforts to select between the large amounts of the available platforms. Each has its own characteristics, some are open source and the other licensed with trial version to test them. In this work we will focus on some platforms related to data mining research area. The selected tools represent widely used and trusted ones with most updated version. We will study platforms from different perspectives. They have different data processing features, but they support common algorithms helps us to evaluate between them. Four data mining tools and four data set were selected. The assessment procedure done from multi-points of view as we will see in the methodology section of this article. The criteria collected from a survey done among a population of researchers interested in the field of data mining and machine learning.The
Contribution of this work is to assess the selected platforms depending on new actual needs criteria. These criteria give a clear idea for the researchers to determine the best platform according to their resources. The results highlighted the power for each platform. Orange and Weka show best performance over the rest. These results will be the guide for beginners or researchers out the computer science field to select the appropriate platform for their needs and available resources.

Introduction:

Data mining is one of the most recent used techniques. It works with different research areas. It has a major role in computer science branches. It used in classification, clustering, regression and studying the association rules. It used in software engineering life cycle [1], big data [2], image processing [3], analytics and business intelligence [4], machine learning [5], networks [6] and many other fields. Different ready tools is available to perform data mining tasks. These tools reduce the efforts for many researchers especially if they are beginners or from other scientific backgrounds. Instead of starting from the coding level for the needed algorithms they are available in a ready tool format.

There are different data mining platforms written in different programming languages. Some are open source while the others are licensed. There are no standard criteria to select the

appropriate platform for the corresponding work. Many studies published to evaluate these platforms. Prajak C. [7] compared between three data mining tools (RapidMiner, WEKA and IBM SPSS) for the regression procedure. He analysed the performance of the selected tools by analysing different dimensions like model type and functionality, system requirements, data processing tasks and model type. The limitation of this work it didn't provide information about the accuracy and total run time. Naik and Samant [8] studied some free data mining platforms such as (WEKA, Rapid miner, Tanagra, Orange and Knime) and selected some classification algorithms with one data-set. They used the accuracy results to evaluate between the mentioned tools. The accuracy and precision are not enough to evaluate between the selected tools. Triguero et al. [9] presented a comparative study between their new tool KEEL Suite 3.0 and two other tools Weka and Knime. They used common characteristics and compared depending on them. Three competitive levels were determined: Pre-processing variety, Learning Variety and Advanced Features like Post-processing Meta - Learning Statistical tests and Multi-Classifiers. The selected characteristics focused on the internal structure and algorithms and data processing with a clear neglecting to important criteria like accuracy, runtime and data format. Graczyk et al. [10] aimed to evaluate residential premises by using some data mining algorithms. They used popular tools (RapidMiner, KEEL and WEKA) with six algorithms. They carried out the procedure on actual dataset of real estate transaction record and cadastral system. This study evaluated by using Wilcoxon test for absolute percentage error and Wilcoxon signed-rank test for squared errors squared error. These statistical tests doesn't give actual needed criteria. Wahbeh et al. [11] selected four free data mining tools (Weka, Orange, Tanagra and KNIME) with nine datasets. The evaluation done depending on the accuracy results by changing some test options.

The contribution of this work is assessing data mining platforms depending on new criteria represent the actual needs of the researchers. There are wide range of researchers in the world are beginners in the field of data analysis. In addition, some of them are not from computer science background but their work needs for adopting new techniques to enhance their results. This study will be the guide to select the appropriate platform for this community. This work presents some new criteria not mentioned in previous studies.

Research Method:

This work used platforms coded in different programming languages some written in java (Weka, Knime and RapidMiner) and one platform written in python (Orange). They have various processing procedures. Some using components-based and the other group is graphical user interface while one of them mix between the previous two procedures. As shown in Fig 1.

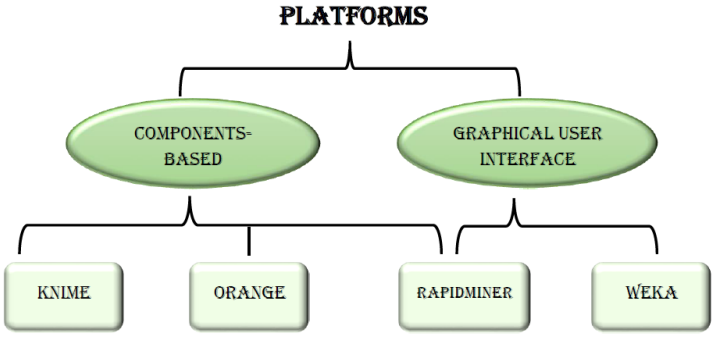

Fig 1: Platforms Types

Weka [12-16] is an open source tool written in java developed by Waikato University. It provides multiple data manipulations procedures pre-processing, classification, clustering, feature selection and visualization. It accepts datasets in ARFF (Attribute Relation File Format). The output of the experiments can be export as file and can be edited by any editor. Knime [17-20] is a components-based platform developed in java programing language and used in many researches. It has variety of data analysis tools. It accepts and exports wide range of file types. Also supports many data mining algorithms and visualization tools. Rapid Miner [21-25] is such amazing platform. It developed using java. It supports both components and GUI in processing data. It is able to build turbo model and auto model. Auto model gives the results of classification using many algorithms by one click. The turbo and auto model become limited after the trail version expire. Orange [26-29] is a python library can be imported and used as library within python 3 or installed directly from GitHub repository. The version used in this work is part of Anaconda Navigator. All the properties of the mentioned platforms is shown in table 1.

Table 1: Platforms Properties.

Tool	Version	Input File
Knime	4.1.0	ARFF
Weka	3.8	ARFF
Orange	3.23.0	CSV
RapidMiner	9.5.001	CSV

In this work five classification algorithms were selected. These algorithms are well known in the data classification research areas: Naïve Bayes (NB), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF) and Support Vector Machine (SVM). These algorithms are common in the selected platforms. The classification procedure done by using cross validation test option. All the previous mentioned algorithms supported by the selected platforms. The used criteria will evaluate between them to determine the best to use in the future work.
Data sets used in this work exported from RapidMiner platform repository and prepared to be used by all selected tools. They exported as CSV file and used directly by Orange3 tool. While we converted them to ARFF to be used by Knime and Weka tools. The characteristics of the datasets shown in table 2.

Table 2: Data Sets characteristics

No	Data Set Name	No. of Attributes	No. of Instances
1	Deals Data Set	4	1000
2	Iris Data Set	5	150
3	Ripley Data Set	3	250
4	Weighting Data Set	7	500

Assessment Criteria

The criteria used in the assessment procedure selected from a survey done on a random population of 52 researchers work in the data mining and machine learning field [30]. The questions of the survey divided into two groups depending on level of knowledge: Beginners and Experts. Most of the participants in the survey were Beginners as show in Fig 2. The selection of the criteria varied between the respondents as appeared in Fig 3. For facilitating the representation of these criteria it abbreviated as appeared in Table 3.

Many reasons for the selection of the previous criteria. First, the previous studies did not cover most of them. Second, various types of users need to work on these platforms, so we put the criteria of video tutorials availability especially for the beginners. Also, we take into account the types of computers and the required space in the hard disk. The input and exported file formats supported by the platform reflects the flexibility of it. All the experiments of this work done on laptop with the following properties: Core i5, Ram 8 GB, 1 TB HDD with Windows 10 operating system.

1. Accuracy
2. Input and output file formats
3. Time required to perform the classification process with each selected algorithm
4. Video tutorials availability and documentation
5. Tool's installation file size

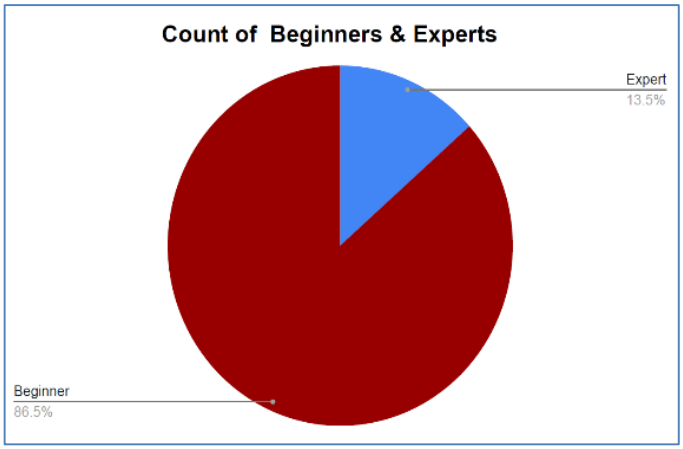

Fig 2

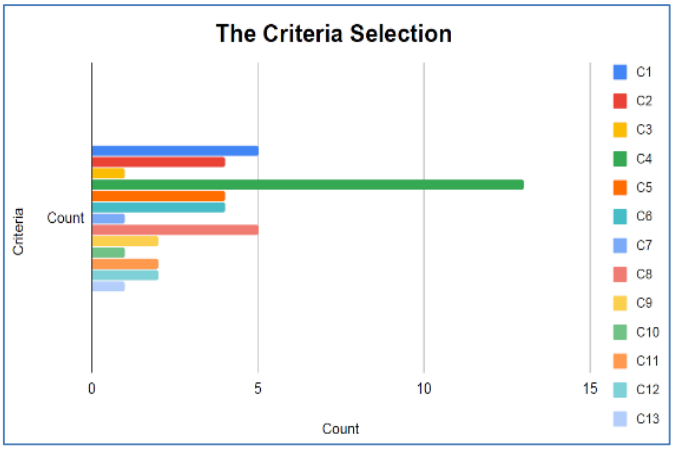

Fig 3

Table 3: Criteria's Name Abbreviation

Criteria	Favourite
C1	Accuracy
C2	Accuracy, Availability of Video tutorials & Documentation
C3	Accuracy, Input and output file formats Availability of Video tutorials & Documentation
C4	Accuracy, Input and output file formats, Run Time Availability of Video tutorials & Documentation
C5	Accuracy, Run Time, Availability of Video tutorials & Documentation
C6	Accuracy, Run Time
C7	Accuracy, Input and output file formats, Run Time
C8	Availability of Video tutorials & Documentation Real time projects
C9	Availability of Video tutorials & Documentation
C10	Input and output file formats Availability of Video tutorials & Documentation
C11	Input and output file formats, Run Time Availability of Video tutorials & Documentation
C12	Input and output file formats
C13	Run Time, Availability of Video tutorials & Documentation

.

The steps of assessment procedure summarized in Fig 4.

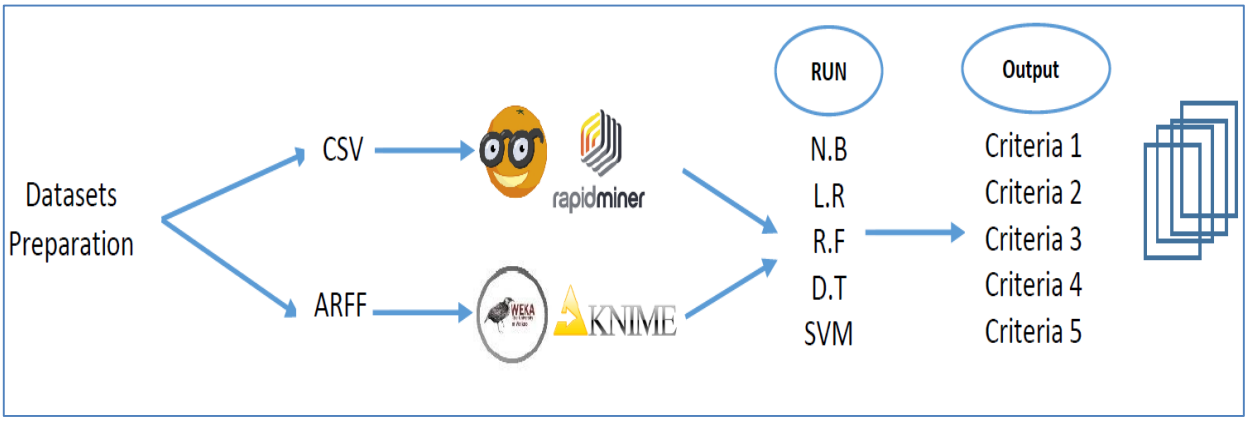

Fig 4. Methodology Scheme

Results and Discussion

The proposed assessment criteria aim to help beginners in data mining and machine learning to select the appropriate platform depending on their actual needs and available resources. This help in save efforts and time without trying randomly to work on any tool and then find a problem of incompatibility with their needs. In this work, the selected criteria tested on four platforms with five classification algorithms and four data-sets.

- **The Results of 1st Criterion:**

Table 4: The Accuracy Results for the Deals Data Set

Algorithm	**Knime Accuracy**	**Orange Accuracy**	**RapidMiner Accuracy**	**Weka Accuracy**
Naïve	96,14	70,00	95,80	92,20
Logistic Regression	91,14	76,20	97,20	99,70
Decision Tree	98,57	75,40	93,00	99,30
Random Forest	Not Applied	74,00	96,20	99,60
SVM	70,86	66,20	96,50	98,70

Table 5: The Accuracy Results for the Iris Dataset

Algorithm	**Knime Accuracy**	**Orange Accuracy**	**RapidMiner Accuracy**	**Weka Accuracy**
Naïve	96,19	98,10	97,80	96,00
Logistic Regression	95,24	98,80	97,80	96,00
Decision Tree	88,57	96,50	97,80	96,00
Random Forest	Not Applied	98,80	97,80	95,33
SVM	94,29	99,70	100,00	96,00

Table 6: The Accuracy Results for the Ripley Dataset

Algorithm	**Knime Accuracy**	**Orange Accuracy**	**RapidMiner Accuracy**	**Weka Accuracy**
Naive	82,86	93,40	82,00	83,20
Logistic Regression	82,29	93,10	83,40	85,60
Decision Tree	83,43	83,20	77,70	84,80
Random Forest	Not Applied	91,80	84,80	82,00
SVM	77,14	93,60	90,20	86,00

Table 7: The Accuracy Results for the Weighting Dataset

Algorithm	**Knime Accuracy**	**Orange Accuracy**	**RapidMiner Accuracy**	**Weka Accuracy**
Naïve	90,26	96	92,30	89,80
Logistic Regression	97,71	100,00	99,30	99,60
Decision Tree	82,29	91,30	99,30	92,20
Random Forest	Not Applied	97,50	84,70	94,20
SVM	96,86	99,00	96,50	94,20

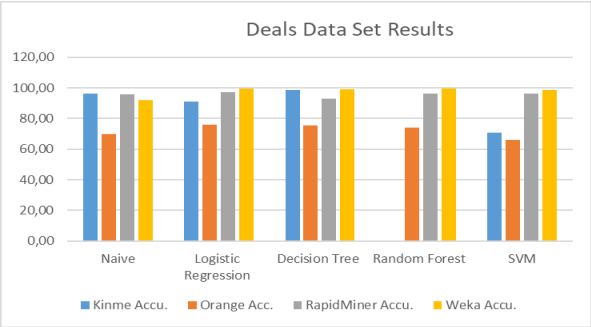

Fig 5

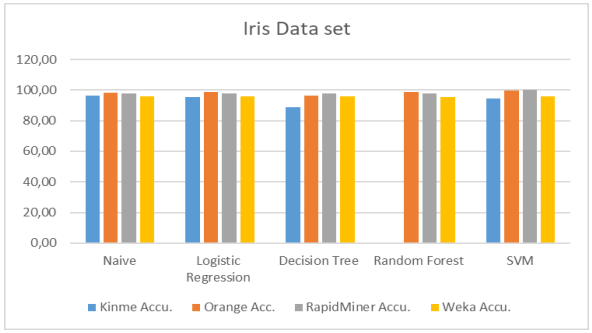

Fig 6

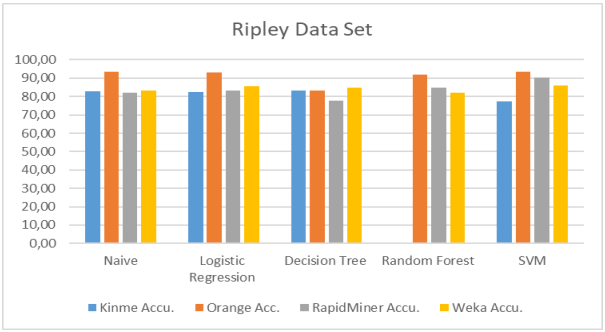

Fig 7

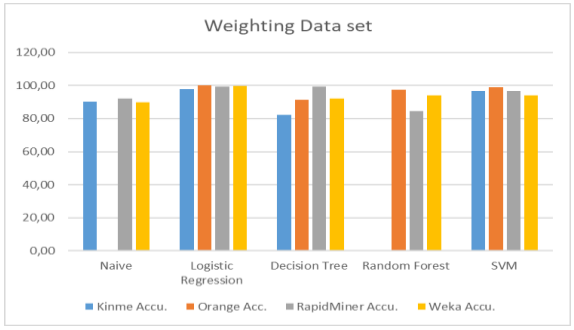

Fig 8

All the accuracy results in this section calculated directly by the paltforms after applying the classification procedure for each algorithm. The accuracy represents the number of correctly classified instances to the total number of instances. So, it gives a clear view of the actual performance of the classifier and the platform. The results varied from minimum value equal to 66.20% in applying SVM algorithm using Orange platform. The maximum accuracy value was 100% in two platforms RapidMiner with Iris data-set and SVM and Orange with Weighting by applying LR.

- **The Results of 2nd Criterion:**

Table 8: Input and Output File Format

Tool	Input file Formats	Exported File Formats
RapidMiner	EXCEL(.xls,.xlsx), CSV(.csv, .tsv) Data Base ,URL ,SPSS, Stata, Sparse, ARFF, XRFF, C4.5, BibTex, DasyLab, XML, Access	EXCEL , CSV, ARFF XRFF, Access
Knime	Excel(xls), ARFF, CSV, Line Table, PMML, Images, Database	Excel(xls), ARFF, CSV, Line, Table, PMML, Images, DataBase
Orange	.tab, CSV, Table, Data Base SQL Table	.ows (Orange Workflows),HTML PDF, .report, Images as (.png , .svg, .pdf)
	ARFF, CSV ,URL	

Weka	Data Base	All Files

The acceptance of different file format gives the platform more flexibility. It enable the users from dealing with data-set from different repositories and various types. Table 7 show all the accepted file formats by each tool. Exported files from any tool or program enable the user to save his work and integrate it with another program. So, the researchers need platform provides many types of exported files manage their results. For this reason, we added this criterion to the assessment procedure. According to that RapidMiner is the most flexible, while Knime platform has more flexibility in exporting different file formats.

- **The Results of 3rd Criterion:**

The time required to complete the classification procedure of one data-set by one algorithm via the selected platform are shown in Fig 9, 10, 11 and 12. Fastest platform was Weka in the run time required to complete the classification procedure. Whereas, RapidMiner took longer time than other platforms. The availability of learning videos and documentation in the official web site gives a great support for the platform. This indicated as important criteria as appeared in the responses of the survey.

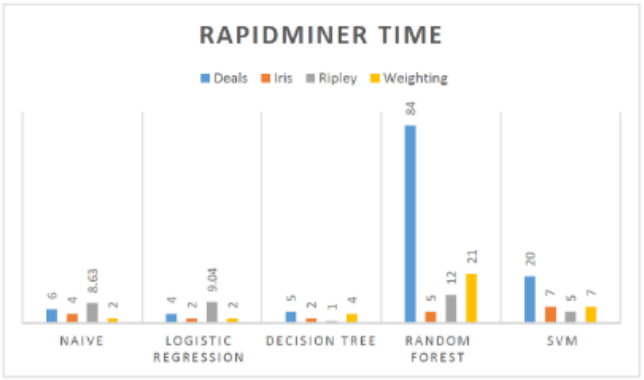

Fig 9

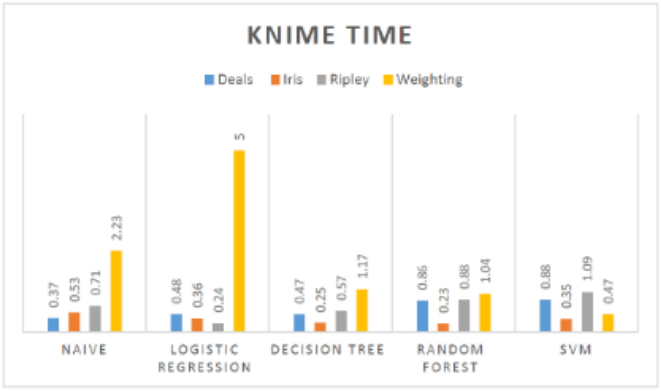

Fig 10

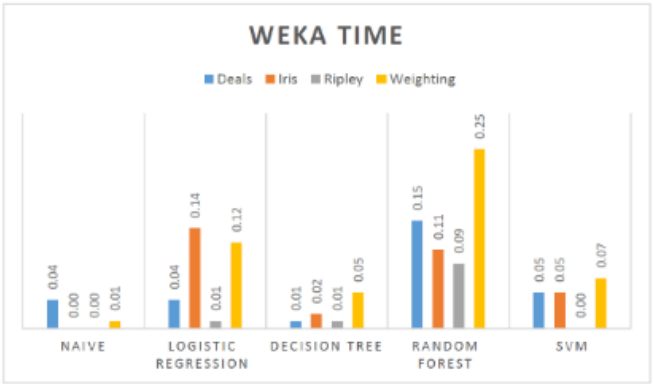

Fig 11

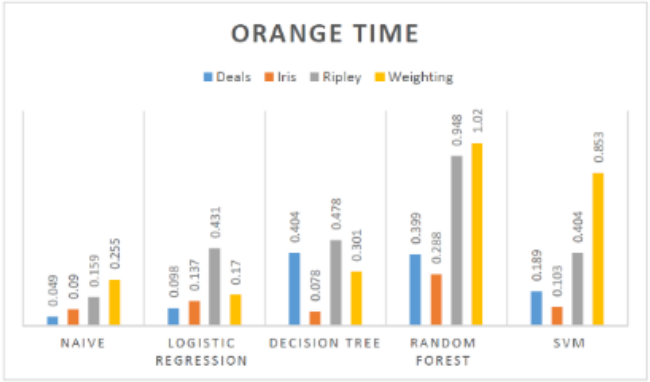

Fig 12

- **The Results of 4th Criterion:**

Video tutorials is important criteria for beginners as shown in the survey results. This depended on the videos and documentation available in the official sites as shown in Table 9. Video Tutorials for any product gives strong support for it. These tutorials enables the beginners to start from the early level towards proficiency. For this reason we put this as a criteria to evaluate the selected platforms. The results in Table 9 show that Weka has the most numerous tutorials. But this is not dependable because in some cases there is one video of 30 minutes equal to three videos of ten minutes. The importance is to give a complete guide.

Table 9: Number of Video Tutorials

Tool	No. Video Tutorials	Documentation
Knime	27	✓
Orange	31	✓
RapidMiner	28	✓
Weka	34	✓

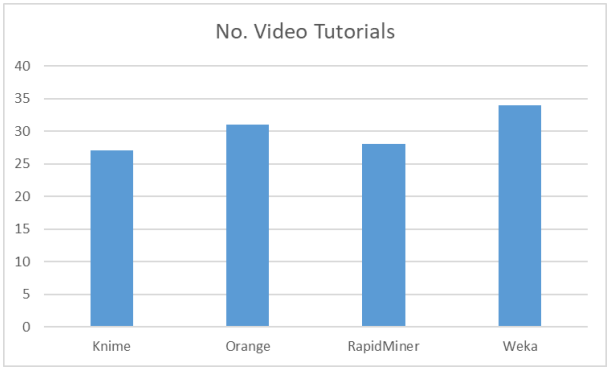

Fig 13

- **The Results of 5th Criterion:**

Usually, most people check the file size of a specific program before installing it on the computer. Table 10 shows the space required to install each platform on the computer. This helps researchers with limited resources working in this field. The installation file size represents the space should available in the hard disk to install the platform. This criterion gives a clear view for the users to select platform suitable for their computer. This directly proportional to available space in the hard disk. This can be a good indication for users with limited resources. As presented in Table 10 all the platforms require normal space in the hard disk to be installed.

Table 10: File Size

Tool	Installation file size	Availability
Knime	475.367 KB	Free
Orange	67900 KB	Free
RapidMiner	272.147 KB	License with free trial version for one month
Weka	118000 MB	Free

Table 11 views the summary for all criteria with the corresponding platform. As we can see, there is no specific platform distinguished as the best. Each one has its specific good

properties and disadvantages. The user can select the appropriate platform depending on the needed requirements.

Table 11: Evaluation Summary

Tool	Criteria 1	Criteria 2	Criteria 3	Criteria 4	Criteria 5
Knime		✓			
Orange	✓				✓
RapidMiner		✓			
Weka			✓	✓	

Based on the results of this work, we can underline the following points:

- Orange show better performance among other platforms in criteria 1.
- RapidMiner and Knime have the best performance in criteria 2.
- Weka has absolute good execution in criteria 3.
- Weka performs efficiently in criteria 4.
- Orange show good feature in criteria 5.

Conclusion

This work experimented assessment procedure to determine the best platform depending on actual needs criteria. It was shown that each platform has its own strength points. As soon as the actual needs of the researcher determined, the appropriate platform can be selected. During the assessment procedure some notes can be registered. Not all datasets compatible to work with all platform.

Future Work

Prior work has documented evaluation procedure of the selected platforms depending on different criteria as mentioned in section 1. However, these studies focused on specific important criteria but not enough. In this study, another valuable criteria have been experimented. Most of the selected criteria not mentioned in the previous studies. Limitation of current work, it just evaluated the platforms from classification perspective, but all of them have another characteristics like clustering and data visualization. This will be a plan for future work. Also, there are some algorithms have equal accuracy value. In this case additional evaluation option will be chosen to evaluate between them. So, another test options will be investigated to evaluate between too classifiers have the same accuracy value as a plan for future work.

References

1. Merzah, B. M. (2019). Software quality prediction using data mining techniques. 2019 International Conference on Information and Communications Technology, ICOIACT 2019. https://doi.org/10.1109/ICOIACT46704.2019.8938487.

2. Wu, X., Zhu, X., & Member, S. (2014). Data Mining with Big Data. 26(1), 97–107.

3. Dey, S., & Rahman, M. (2019). Application of image processing and data mining techniques for traffic density estimation and prediction. 2019 2nd International Conference on Advanced Computational and Communication Paradigms, ICACCP 2019. https://doi.org/10.1109/ICACCP.2019.8882878.

4. Kollwitz, C., Dinter, B., & Krawatzeck, R. (2018). Tools for Academic Business Intelligence and Analytics Teaching: Results of an Evaluation. https://doi.org/10.1007/978-3-319-58097-5_17.

5. Karandikar, J. (2019). Machine learning classification for tool life modeling using production shop-floor tool wear data. Procedia Manufacturing, 34. https://doi.org/10.1016/j.promfg.2019.06.192.

6. Madi, M., Jarghon, F., Fazea, Y., Almomani, O., & Saaidah, A. (2020). Comparative analysis of classification techniques for network fault management. Turkish Journal of Electrical Engineering and Computer Sciences, 28(3). https://doi.org/10.3906/elk-1907-84.

7. Wahbeh, A. H., Al-radaideh, Q. A., Al-kabi, M. N., & Al-shawakfa, E. M. (2011). A Comparison Study between Data Mining Tools over some Classification Methods. (September). https://doi.org/10.14569/SpecialIssue.2011.010304.

8. Naik, A., & Samant, L. (2016). Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime. Procedia Computer Science, 85. https://doi.org/10.1016/j.procs.2016.05.251.

9. Triguero, I., Gonz, S., Moyano, J. M., Garc, S., Luciano, S., Fern, A., & Jos, M. (2017). KEEL 3. 0: An Open Source Software for Multi-Stage Analysis in Data Mining. 10, 1238–1249.

10. M. Graczyk, T. Lasota, and B. Trawiński, “Comparative analysis of premises valuation models using KEEL, rapidminer, and WEKA,” in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 2009, vol. 5796 LNAI.

11. Graczyk, M., Lasota, T., & Trawiński, B. (2009). Comparative analysis of premises valuation models using KEEL, rapidminer, and WEKA. Lecture Notes in Computer

Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 5796 LNAI. https://doi.org/10.1007/978-3-642-04441-0-70.

12. Witten, I. H., Frank, E., & Geller, J. (2002). Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. SIGMOD Record, 31(1). https://doi.org/10.1145/507338.507355.

13. Kumar, V., & Kumar, A. (2020). Classification of Comparison-Based Analysis between Rattle and Weka Tool on the Mental Health Disorder Problem Dataset Using Random Forest Tree Classifier. In Lecture Notes in Networks and Systems (Vol. 121). https://doi.org/10.1007/978-981-15-3369-3_47.

14. Ramotra, A. K., Mahajan, A., Kumar, R., & Mansotra, V. (2020). Comparative Analysis of Data Mining Classification Techniques for Prediction of Heart Disease Using the Weka and SPSS Modeler Tools. Smart Innovation, Systems and Technologies, 165. https://doi.org/10.1007/978-981-15-0077-0_10.

15. Berthold, M. R., Cebron, N., Dill, F., Di Fatta, G., Gabriel, T. R., Georg, F., Wiswedel, B. (2006). KNIME: The Konstanz information miner. 4th International Industrial Simulation Conference 2006, ISC 2006. https://doi.org/10.1007/978-3-540-78246-9_38.

16. Karthikeyan, V. G., Thangaraj, P., & Karthik, S. (2020). Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation. Soft Computing. https://doi.org/10.1007/s00500-020-05075-4.

17. Qundus, J. Al, Peikert, S., & Paschke, A. (2020). AI supported Topic Modeling using KNIME-Workflows. CEUR Workshop Proceedings, 2535.

18. Hany Salman, R., Zaki, M., & A. Shiltag, N. (2020). A STUDYING OF WEB CONTENT MINING TOOLS. Al-Qadisiyah Journal of Pure Science, 25(2). https://doi.org/10.29350/qjps.2020.25.2.1067.

19. RapidMiner Studio Code Free Workflow Designer for Data Scientists. RapidMiner Studio website. [Online]. Available: https://rapidminer.com/products/studio/.

20. Ye, D. (2020). Association Analysis of Supermarket Products Based on RapidMiner. https://doi.org/10.2991/assehr.k.200331.054.

21. Manjunath Kumar, T., Murugeswari, R., Devaraj, D., & Hemalatha, J. (2020). Comparison of Deep Learning and Random Forest for Rumor Identification in Social Networks. Advances in Intelligent Systems and Computing, 1087. https://doi.org/10.1007/978-981-15-1286-5_12

22. Jain, S., Jain, S. C., & Vishwakarma, S. (2020). Analysis and Prediction of Customers' Reviews with Amazon Dataset on Products. Advances in Intelligent Systems and Computing, 1077. https://doi.org/10.1007/978-981-15-0936-0_48.

23. Ramos, A. L. A., Dadiz, B. G., & Santos, A. B. G. (2020). Classifying Emotion based on Facial Expression Analysis using Gabor Filter: A Basis for Adaptive Effective Teaching Strategy. Lecture Notes in Electrical Engineering, 603. https://doi.org/10.1007/978-981-15-0058-9_45.

24. Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Zupan, B. (2013). Orange: Data mining toolbox in python. Journal of Machine Learning Research, 14.

25. Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U., & Sattar, M. U. (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. Applied Sciences (Switzerland), 10(11). https://doi.org/10.3390/app10113894.

26. Shastri, D. J. (2020). Machine Learning for Non-Programmers. https://doi.org/10.1145/3334480.3375051.

27. Lakhtaria, K. I., & Iyer, S. S. (2020). Data analytics using satellite remote sensing in healthcare applications. In Studies in Computational Intelligence (Vol. 836). https://doi.org/10.1007/978-3-030-20212-5_7.

28. Pejović, V., Majhen, I., Janež, M., & Zupan, B. (2020). RICERCANDO: Data mining toolkit for mobile broadband measurements. Computer Networks, 177. https://doi.org/10.1016/j.comnet.2020.107294.

29. Tougui, I., Jilbab, A., & El Mhamdi, J. (2020). Heart disease classification using data mining tools and machine learning techniques. Health and Technology. https://doi.org/10.1007/s12553-020-00438-1.

30. The link of The Survey: https://forms.gle/WaHJWEJ3urBWEkcJ6 .

IRAQI Academic Scientific Journals

Samarra Journal of Pure and Applied Science

SJPAS

ISSN:2663-7405

www.sjpas.com

تقييم منصات تصنيف البيانات وفقاً لمعايير الاحتياجات الفعلية

بيداء محمد مرزة *

وحدة تقنية المعلومات، كلية العلوم السياسية، جامعة النهرين (baydaa@ced.nahrainuniv.edu.iq)

معلومات البحث:

تأريخ الاستلام: 2020/12/25
تأريخ القبــول: 2021/02/10

الكلمات المفتاحية:

تصنيف البيانات، منصات التنقيب عن البيانات، تقييم المنصات، الأدوات البرمجية

الخلاصة:

للأدوات البرمجية دور مهم في مختلف المجالات البحثية. بشكل عام توفر الوقت والجهد. في مجال علوم الحاسوب، يمكن أن تدعم هذه الأدوات في مجال الاتصالات، تطوير مواقع الويب، ايجاد مقاييس البرامج، التنقيب عن البيانات، والتعلم الآلي والعديد من المجالات الأخرى. هناك العديد من الأدوات المتخصصة المصممة لدعم غرض معين. يقضي المستخدمون والباحثون الكثير من الوقت والجهود للاختيار بين الأعداد الكبيرة من المنصات المتاحة. لكل منها خصائصه الخاصة، بعضها مفتوح المصدر والآخر مرخص بإصدار تجريبي لاختبارها. سنركز في هذا العمل على بعض الأدوات الأساسية المتعلقة بمجال استخراج البيانات. تمثل الأدوات المختارة أدوات مستخدمة على نطاق واسع وموثوقة وبأحدث إصدار. سوف ندرس المنصات من جهات متعددة. لديهم ميزات معالجة بيانات مختلفة، لكنهم يدعمون الخوارزميات الشائعة التي تساعدنا على التقييم فيما بينها. تم اختيار أربع أدوات للتنقيب عن البيانات وأربع مجموعات بيانات. يتم إجراء التقييم من وجهات نظر متعددة كما سنرى في قسم المنهجية في هذا البحث. تم جمع عدة معايير من استطلاع تم إجراؤه بين مجموعة من الباحثين المهتمين بمجال التنقيب عن البيانات والتعلم الآلي. تتمثل مساهمة هذا العمل في تقييم المنصات المختارة اعتمادًا على معايير الاحتياجات الفعلية الجديدة لتعطي فكرة واضحة للباحثين لتحديد أفضل منصة وفقًا لمواردهم. . ستكون النتائج بمثابة دليل للمبتدئين أو الباحثين خارج المجال لاختيار المنصة المناسبة لاحتياجاتهم والموارد المتاحة.