A Comparison Study between Data Mining Tools over some Classification Methods

Abdullah H. Wahbeh, Qasem A. Al-Radaideh, Mohammed N. Al-Kabi, and Emad M. Al-Shawakfa
Department of Computer Information Systems
Faculty of Information Technology, Yarmouk University
Irbid 21163, Jordan

***Abstract*- Nowadays, huge amount of data and information are available for everyone, Data can now be stored in many different kinds of databases and information repositories, besides being available on the Internet or in printed form. With such amount of data, there is a need for powerful techniques for better interpretation of these data that exceeds the human's ability for comprehension and making decision in a better way. In order to reveal the best tools for dealing with the classification task that helps in decision making, this paper has conducted a comparative study between a number of some of the free available data mining and knowledge discovery tools and software packages. Results have showed that the performance of the tools for the classification task is affected by the kind of dataset used and by the way the classification algorithms were implemented within the toolkits. For the applicability issue, the WEKA toolkit has achieved the highest applicability followed by Orange, Tanagra, and KNIME respectively. Finally; WEKA toolkit has achieved the highest improvement in classification performance; when moving from the percentage split test mode to the Cross Validation test mode, followed by Orange, KNIME and finally Tanagra respectively.**

Keywords-component; data mining tools; data classification; Wekak; Orange; Tanagra; KNIME.

I. INTRODUCTION

Today's databases and data repositories contain so much data and information that it becomes almost impossible to manually analyze them for valuable decision-making. Therefore, humans need assistance in their analysis capacity; humans need data mining and its applications [1]. Such requirement has generated an urgent need for automated tools that can assist us in transforming those vast amounts of data into useful information and knowledge.

Data mining is the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information repositories. Data mining involves an integration of techniques from multiple disciplines such as database and data warehousing technology, statistics, machine learning, high-performance computing, pattern recognition, neural networks, data visualization, information retrieval, image and signal processing, and spatial or temporal data analysis [2]. Data mining has many application fields such as marketing, business, science and engineering, economics, games and bioinformatics.

Currently, many data mining and knowledge discovery tools and software are available for every one and different usage such as the Waikato Environment for Knowledge Analysis (WEKA) [3] [4], RapidMiner [5][6], Clementine [6], Rosetta, Intelligent Miner [1] etc. These tools and software provide a set of methods and algorithms that help in better utilization of data and information available to users; including methods and algorithms for data analysis, cluster analysis, Genetic algorithms, Nearest neighbor, data visualization, regression analysis, Decision trees, Predictive analytics, Text mining, etc.

This research has conducted a comparison study between a number of available data mining software and tools depending on their ability for classifying data correctly and accurately. The accuracy measure; which represents the percentage of correctly classified instances, is used for judging the performance of the selected tools and software.

The rest of the paper is organized as follows: Section 2 summaries related works on data mining, mining tools and data classification. Section 3 gives a general description on the methodology followed and provides a general description of the tools and software under test. Section 4 reports our experimental results of the proposed methodology and compares the results of the different software and tools used. Finally, we close this paper with a summary and an outlook for some future work.

II. RELATED WORKS

King and Elder [7] have conducted an evaluation of fourteen data mining tools ranging in price from $75 to $25,000. The evaluation process was performed by three kinds of user groups: (1) four undergraduates; who are inexperienced users in data mining, (2) a relatively experienced graduate student, and (3) a professional data mining consultant. Tests were performed using four data sets. To test tools flexibility and capability, their output types have varied: two binary classifications (one with missing data), a multi-class set, and a noiseless estimation set. A random two-thirds of the cases in each have served as training data; the remaining one-third was test data. Authors have developed a list of 20 criteria, plus a standardized procedure, for evaluating data mining tools. The tools ran under Microsoft Windows 95, NT, or Macintosh 7.5 operating systems, and have employed Decision Trees, Rule

Induction, Neural Networks, or Polynomial Networks to solve two binary classification problems, a multi-class classification problem, and a noiseless estimation problem. Results have provided a technical report that details the evaluation procedure and the scoring of all component criteria. Authors also showed that the choice of a tool depends on a weighted score of several categories such as software budget and user experience. Finally, authors have showed that the tools' price is related to quality.

Carrier and Povel [8] have described a general schema for the characterization of data mining software tools. Authors have described a template for the characterization of DM software along a number of complementary dimensions, together with a dynamic database of 41 of the most popular data mining tools. The business-oriented proposal for the characterization of data mining tools is defined depending on the business goal, model type, process-dependent features, user interface features, system requirements and vendor information. Using these characteristics, authors had characterized 41 popular DM tools. Finally; authors have concluded that with the help of a standard schema and a corresponding database, users are able to select a data mining software package, with respect to its ability, to meet high-level business objectives.

Collier et al. [9] have presented a framework for evaluating data mining tools and described a methodology for applying this framework. This methodology is based on firsthand experiences in data mining using commercial data sets from a variety of industries. Experience has suggested four categories of criteria for evaluating data mining tools: performance, functionality, usability, and support of ancillary activities. Authors have demonstrated that the assessment methodology takes advantage of decision matrix concepts to objectify an inherently subjective process. Furthermore, using a standard spreadsheet application, the proposed framework by [9] is easily automatable, and thus easy to be rendered and feasible to employ. Authors have showed that there is no single best tool for all data mining applications. Furthermore, there are a several data mining software tools that share the market leadership.

Abbott et al. [10] have compared five of the most highly acclaimed data mining tools on a fraud detection application. Authors have employed a two stage selection phase preceded by an in-depth evaluation. For the first stage, more than 40 data mining tools/vendors were rated depending on six qualities. The top 10 tools continued to the second stage of the selection phase and these tools were further rated on several additional characteristics. After selecting the 10 software packages, authors have used expert evaluators and re-rated each tool's characteristics, and the top five tools were selected for extensive hands-on evaluation. The selected tools and software were Clementine, Darwin, Enterprise Miner, Intelligent Miner, and PRW. The tools and software properties evaluated included the areas of client-server compliance, automation capabilities, breadth of algorithms implemented, ease of use, and overall accuracy on fraud-detection test data. Results have showed that the evaluated five products by authors would all display excellent properties; however, each may be best suited for a different environment. Authors have concluded that Intelligent Miner has the advantage of being the current market leader. Clementine excels in support provided and in ease of use. Enterprise Miner would especially enhance a statistical environment. Darwin is best when network bandwidth is at a premium. Finally, PRW is a strong choice when it's not obvious what algorithm will be most appropriate, or when analysts are more familiar with spreadsheets than UNIX.

Hen and Lee [1] have compared and analyzed the performance of five known data mining tools namely, IBM intelligent miner, SPSS Clementine, SAS enterprise miner, Oracle data miner, and Microsoft business intelligence development studio. 38 metrics were used to compare the performance of the selected tools. Test data was mined by various data mining methods ranging from different types of algorithms that are supported by the five tools, these includes classification algorithms, regression algorithms, segmentation algorithms, association algorithms, and sequential analysis algorithms. Results have provided a review of these tools and have proposed a data mining middleware adopting the strengths of these tools.

III. The Comparative Study

The methodology of the study constitute of collecting a set of free data mining and knowledge discovery tools to be tested, specifying the data sets to be used, and selecting a set of classification algorithm to test the tools' performance. Fig. 1 demonstrates the overall methodology followed for fulfilling the goal of this research.

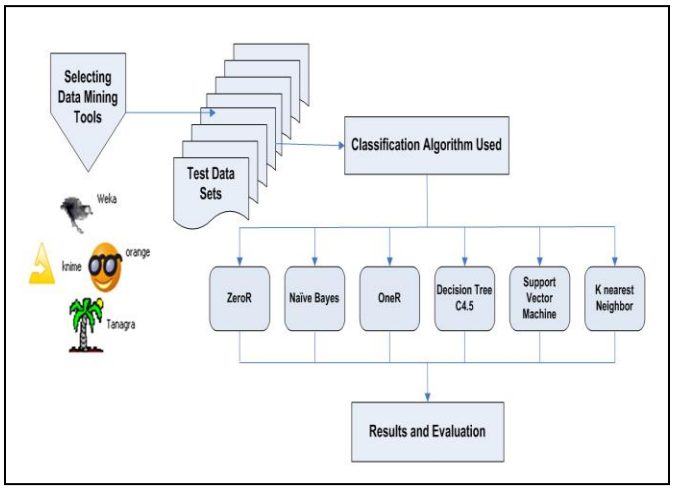

Figure 1. Methodology of Study.

A. Tools Description

The first step in the methodology consists of selecting a number of available open source data mining tools to be tested. Many open data mining tools are available for free on the Web. After surfing the Internet, a number of tools were chosen; including the Waikato Environment for Knowledge Analysis (WEKA), Tanagra, the Konstanz Information Miner (KNIME), and Orange Canvas.

- WEKA toolkit [12] is a widely used toolkit for machine learning and data mining that was originally developed at the University of Waikato in New Zealand. It contains a

large collection of state-of-the-art machine learning and data mining algorithms written in Java. WEKA contains tools for regression, classification, clustering, association rules, visualization, and data pre-processing. WEKA has become very popular with the academic and industrial researchers, and is also widely used for teaching purposes.

- Tanagra is free data mining software for academic and research purposes. It offers several data mining methods like exploratory data analysis, statistical learning and machine learning. The first purpose of the Tanagra project is to give researchers and students easy-to-use data mining software. The second purpose of TANAGRA is to propose to researchers an architecture allowing them to easily add their own data mining methods, to compare their performances. The third and last purpose is that novice developers should take advantage of the free access to source code, to look how this sort of software was built, the problems to avoid, the main steps of the project, and which tools and code libraries to use for. In this way, Tanagra can be considered as a pedagogical tool for learning programming techniques as well [13].
- KNIME (Konstanz Information Miner) is a user-friendly and comprehensive open-source data integration, processing, analysis, and exploration platform. From day one, KNIME has been developed using rigorous software engineering practices and is currently being used actively by over 6,000 professionals all over the world, in both industry and academia. KNIME is a modular data exploration platform that enables the user to visually create data flows (often referred to as pipelines), selectively execute some or all analysis steps, and later investigate the results through interactive views on data and models [14].
- Orange is a library of C++ core objects and routines that includes a large variety of standard and not-so-standard machine learning and data mining algorithms, plus routines for data input and manipulation. This includes a variety of tasks such as pretty-print of decision trees, attribute subset, bagging and boosting, and alike. Orange also includes a set of graphical widgets that use methods from core library and Orange modules. Through visual programming, widgets can be assembled together into an application by a visual programming tool called Orange Canvas. All these together make the Orange tool, a comprehensive, component-based framework for machine learning and data mining, intended for both experienced users and researchers in machine learning who want to develop and test their own algorithms while reusing as much of the code as possible, and for those just entering who can enjoy in powerful while easy-to-use visual programming environment [15].

B. Data Set Description

Once the tools have been chosen, a number of data sets are selected for running the test. For bias issues, several data sets have been downloaded from the UCI repository [16]. Table 1 shows the selected and downloaded data sets for testing purposes as shown in the table, each dataset is described by the data type being used, the types of attributes; whether they are categorical, real, or integer, the number of instances stored within the data set, the number of attributes that describe each dataset, and the year the dataset was created. Also, the table demonstrates that all the selected data sets are used for the classification task which is the main concentration of this paper.

These data sets were chosen because they have different characteristics and have addressed different areas, such as the number of instances which range from 100 to 20,000. Also, the number of attributes; which range from 5 to 70, and the attribute types; where some data sets contain one type while others contain two types. Such characteristics reflect different dataset shapes where some data sets contain a small number of instances but large number of attributes and vice versa.

TABLE1: UCI DATA SET DESCRIPTION

Data Set Name	Data Type	Default Task	Attribute Type	# Instances	# Attributes
Audiology (Standardized)	Multivariate	Classification	Categorical	226	69
Breast Cancer Wisconsin (Original)	Multivariate	Classification	Integer	699	10
Car Evaluation	Multivariate	Classification	Categorical	1728	6
Flags	Multivariate	Classification	Categorical, Integer	194	30
Letter Recognition	Multivariate	Classification	Integer	20000	16
Nursery	Multivariate	Classification	Categorical	12960	8
Soybean (Large)	Multivariate	Classification	Categorical	638	36
Spambase	Multivariate	Classification	Integer, Real	4601	57
Zoo	Multivariate	Classification	Categorical, Integer	101	17

C. Data Classification

Data classification is a two-step process: in the first step; a classifier is built describing a predetermined set of data classes or concepts. This is the learning step (or training phase), where a classification algorithm builds the classifier by analyzing or "learning from" a training set made up of database tuples and their associated class labels. In the second step, the model is used for classification; the predictive accuracy of the classifier is estimated using the training set to measure the accuracy of the classifier. The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier. The associated class label of each test tuple is compared with the learned classifier's class prediction for that tuple. If the accuracy of the classifier is considered acceptable, the classifier can be used to classify future data tuples for which the class label is not known [2].

1. Classification Algorithm Description

After selecting the data sets, a number of classification algorithm are chosen for conducting the test. Many classification algorithms mentioned in literature are available for users such as Naïve Bayes (NB) algorithm [17] [18], K Nearest Neighbor (KNN) algorithm [18] [19] [20], Support Vector Machine (SVM) algorithm [21], and C4.5 algorithm [22. For testing purposes, we selected well known classifiers that are almost available in every open source tool, namely; Naïve Bayes (NB) classifier, One Rule (OneR) classifier, Zero Rule (ZeroR) classifier, Decision Tree Classifier; which is represented by the C4.5 Classifier, Support Vector Machine (SVM) classifier, and the K Nearest Neighbor (KNN) classifier.

2. Evaluation of Classificatiuon Algorithms

For evaluation purpose, two test modes were used; the k-fold Cross Validation (k-fold CV) mode and the Percentage Split (also called Holdout method) mode. The k-fold CV refers to a widely used experimental testing procedure where the

database is randomly divided into k disjoint blocks of objects, then the data mining algorithm is trained using k-1 blocks and the remaining block is used to test the performance of the algorithm; this process is repeated k times. At the end, the recorded measures are averaged. It is common to choose k = 10 or any other size depending mainly on the size of the original dataset.

In percentage split (Holdout) method, the database is randomly split into two disjoint datasets. The first set; which the data mining system tries to extract knowledge from, is called the training set. The extracted knowledge may be tested against the second set which is called the test set. In machine learning, to have the training and test sets, it is common to randomly split a dataset under the mining task into two parts. It is common to have 66% of the objects of the original database as a training set and the rest of objects as a test set [23].

The accuracy measure refers to the percentage of the correctly classified instances from the test data. The goal of testing, using the two modes, is to check whether there is an improvement in the accuracy measure when moving from the first test mode to the second test mode for all tools. Once the tests are carried out using the selected data sets, then using the available classifiers and test modes, results are collected and an overall comparison is conducted in order to determine the best tool for the classification purposes.

IV. Experiments and Evaluations

To evaluate the selected tools using the given datasets, several experiments were conducted. This section presents the results obtained after running the four data mining tools using the selected data sets described in Table 1.

A. Experiments Setup and Preliminaries

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

As for experiments' setup, all tests were accomplished as follows: the holdout method has used 66% of each data set as training data and the remaining 34% as test data while the Cross Validation method used k = 10. The accuracy measure is used as a performance measure to compare the selected tools.

After running the four tools, we have obtained some results regarding the ability to run the selected algorithms on the selected tools. All algorithms ran successfully on WEKA; the six selected classifiers used the nine selected data sets.

As for the Orange tool, all classification techniques run successfully, except the OneR classifier; which is not implemented in Orange. For KNIME and Tanagra, Table 2 showed that some of the algorithms are unable to run some of the selected data sets. We noticed that this is due to one of the following three reasons; the first one is that the classifier is unable to run against the dataset because it is a multi-class data set and the classifier is only able to deal with binary classes; which are referenced in the tables with entry (MC). The second reason is that the classifier is unable to run the selected dataset because it contains discrete values and the algorithm is unable to deal with such kind of values; referenced in tables with (D) entry. The third reason is that the tool itself does not have an implementation for some classifiers; this reason is referenced in tables with not applicable (NA) entry.

We can notice that the One Rule algorithm (OneR) has no implementation in KNIME, Tanagra and Orange. Also, the ZeroR has no implementation in KNIME and Tanagra tools, and hence, it is referenced in tables as NA. On the other hand, tables shows that the K Nearest Neighbor (KNN) algorithm does not run against part of the data sets such as Audiology, Car, Nursery, and the SoyBean data sets because they contain some discrete values where the KNN algorithm cannot deal with. Finally, the Support Vector Machine does not run against any data sets; except the Breast-W and SpamBase data sets. This is because the other data sets are either containing a multi class data set and/or containing discrete values.

TABLE 2: ABILITY TO RUN SELECTED ALGORITHMS ON KNIME AND TANAGRA

	Audiology	Breast-W	Car	Flags	Letters	Nursery	SoyBean	SpamBase	Zoo
NB	OK	OK	OK	OK	OK	OK	OK	OK	OK
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	OK	OK	OK	OK	OK	OK	OK	OK	OK
SVM	MC/D	OK	MC/D	MC	MC	MC/D	MC/D	OK	MC
KNN	D	OK	D	OK	OK	D	D	OK	OK
ZeroR	NA	NA	NA	NA	NA	NA	NA	NA	NA

* OK: Algorithm Run Successfully. NA: Algorithm has no Implementation. D: Discrete Value. MC: Multi Class

B. Evaluating the Performance of the Algorithms

For performance issues, Table 3 shows the results after running algorithms using WEKA toolkit. For the NB classifier, the accuracy measure has ranged between 44%-97%, while the OneR classifier accuracy has ranged between 5%-92%. For the C4.5 and SVM classifiers, results were almost the same for all data sets; where it ranged between 49%-96%. The KNN classifier has achieved accuracy measure values between 59%-98%. Finally, the ZeroR classifier has achieved the lowest accuracy measure for all of the data sets with accuracy measures ranging between 4%-70%.

TABLE 3: THE ACCURACY MEASURES GIVEN BY WEKA TOOL USING PERCENTAGE SPLIT.

	Audiology	Breast-W	Car	Flags	LETTERS	Nursery	SoyBean	SpamBase	Zoo
NB	71.43%	94.96%	87.59%	43.94%	64.47%	90.63%	90.56%	78.02%	97.14%
OneR	42.86%	92.02%	69.56%	4.55%	16.82%	70.41%	39.06%	77.83%	37.14%
C4.5	83.12%	95.38%	90.99%	48.48%	85.47%	96.48%	90.56%	92.20%	94.29%
SVM	84.42%	95.38%	93.37%	59.09%	81.13%	92.83%	93.99%	90.54%	94.29%
KNN	58.44%	95.38%	90.65%	51.52%	93.57%	97.53%	89.70%	89.27%	77.14%
ZeroR	27.27%	63.87%	69.56%	34.85%	3.90%	32.90%	13.30%	60.58%	37.14%

For the Orange toolkit, results are shown in Table 4. The NB classifier has achieved accuracy measures ranging between 52%-96%. The OneR classifier has no results as it has no implementation. For the C4.5 classifier, results have ranged between 51%-96%, while the SVM and KNN classifiers have achieved measures between 55%-97% and 56%-96% respectively. Finally, the ZeroR classifier has achieved the lowest measures for almost all data sets with values ranging between 4%-70%.

TABLE 4: THE ACCURACY MEASURES GIVEN BY ORANGE TOOL USING PERCENTAGE SPLIT.

	Audiology	Breast-W	Car	Flags	LETTERS	Nursery	SoyBean	SpamBase	Zoo
NB	70.13%	96.22%	86.90%	51.52%	61.44%	90.04%	92.24%	89.51%	88.24%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	72.73%	95.80%	91.50%	51.52%	85.50%	95.87%	79.74%	89.96%	91.18%
SVM	54.55%	95.38%	94.39%	56.06%	74.72%	97.07%	89.22%	92.26%	88.24%
KNN	76.62%	94.54%	88.78%	56.06%	95.84%	92.76%	92.67%	85.29%	85.29%
ZeroR	24.68%	65.55%	70.07%	34.85%	4.06%	33.33%	13.36%	60.61%	41.18%

Table 5 shows results achieved using the KNIME toolkit; for the NB classifiers results have ranged between 42%-95%. On the other hand, the ZeroR and OneR classifiers have no results because they have no implementation. The C4.5 classifier has achieved accuracy ranging between 43%-97%. The SVM and KNN classifiers did not run using some of the data sets because of the presence of one of the three reasons mentioned before; however, these classifiers have achieved measures between 67%-98% and 26%-97% respectively.

Finally, for the Tanagra tool, results are shown in Table 6 where the NB classifier has achieved an accuracy ranging between 60% and 96%. ZeroR and OneR classifiers have no results; because they have no implementation. C4.5 classifier has achieved results between 39% and 96%. On the other hand, SVM and KNN did not run using all the data sets as happened with KNIME; however, they both have achieved results ranging between 91%-97 and 29%-99% respectively.

TABLE 5: THE ACCURACY MEASURES GIVEN BY KNIME TOOL USING PERCENTAGE SPLIT

	Audiology	Breast-W	Car	Flags	Letters	Nursery	SoyBean	SpamBase	Zoo
NB	53.20%	95.00%	86.10%	42.40%	62.90%	90.50%	85.40%	89.80%	82.90%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	68.80%	95.00%	93.50%	43.10%	85.30%	96.70%	66.10%	91.10%	94.30%
SVM	MC/D	97.90%	MC/D	MC	MC	MC/D	MC/D	67.00%	MC
KNN	D	96.60%	D	25.80%	95.00%	D	D	80.90%	45.70%
ZeroR	NA	NA	NA	NA	NA	NA	NA	NA	NA

TABLE 6: THE ACCURACY MEASURES GIVEN BY TANAGRA TOOL USING PERCENTAGE SPLIT

	Audiology	Breast-W	Car	Flags	Letters	Nursery	SoyBean	SpamBase	Zoo
NB	66.23%	95.80%	87.24%	63.64%	59.59%	90.74%	89.70%	87.54%	88.57%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	81.82%	92.44%	89.97%	39.39%	86.34%	96.32%	90.56%	90.73%	88.57%
SVM	MC/D	96.64%	MC/D	MC	MC	MC/D	MC/D	90.73%	MC
KNN	D	98.74%	D	28.79%	94.75%	D	D	79.17%	82.86%
ZeroR	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 7 shows the results obtained after running the algorithms against test data sets using the second test mode with 10-folds-CV. As shown in the table, the NB classifier has achieved accuracy measures ranging between 56%-96% while the OneR classifier has achieved measures ranging between 17%-93%. Both C4.5 and SVM classifiers have achieved accuracy ranging between 59%-97% and 61%-97% respectively. Accuracy measures ranging between 57%-98% were achieved using the KNN classifier. ZeroR rule classifier has achieved the lowest accuracy measures ranging between 4%-70%.

TABLE 7: THE ACCURACY MEASURES GIVEN BY WEKA USING 10-FOLDS-CV.

	Audiology	Breast-W	Car	Flags	LETTERS	Nursery	SoyBean	SpamBase	Zoo
NB	73.45%	95.99%	85.53%	55.15%	64.12%	90.32%	92.97%	79.29%	95.05%
OneR	46.46%	92.70%	70.02%	4.64%	17.24%	70.97%	39.97%	78.40%	57.43%
C4.5	77.87%	94.56%	92.36%	59.28%	87.98%	97.05%	91.51%	92.98%	92.08%
SVM	81.85%	97.00%	93.75%	60.82%	82.34%	93.08%	93.85%	90.42%	96.04%
KNN	62.83%	96.71%	93.52%	57.22%	95.52%	98.38%	90.19%	90.42%	95.05%
ZeroR	25.22%	65.52%	70.02%	35.57%	4.07%	33.33%	13.47%	60.60%	40.59%

Table 8 shows the accuracy measures using the Orange toolkit with 10-folds CV. As the table demonstrates, the NB classifier has achieved an accuracy measure ranging between 58%-97%. On the other hand, OneR has no accuracy measures because it has no implementation. For the C4.5 and SVM classifiers, the accuracy measures have ranged between 54%-96% and 64%-98% respectively. The KNN classifier has achieved accuracy measures ranging between 58%-96%. However, ZeroR has achieved the lowest measures ranging between 4%-70%.

TABLE 8: THE ACCURACY MEASURES GIVEN BY ORANGE USING 10-FOLDS-CV.

	Audiology	Breast-W	Car	Flags	LETTERS	Nursery	SoyBean	SpamBase	Zoo
NB	73.10%	97.14%	85.70%	58.24%	60.01%	90.29%	93.86%	89.31%	91.18%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	76.13%	95.85%	93.58%	54.03%	87.96%	96.57%	89.61%	90.64%	94.18%
SVM	64.23%	96.57%	95.54%	66.47%	76.58%	97.78%	93.27%	85.79%	92.09%
KNN	79.21%	95.71%	88.42%	57.61%	96.48%	92.63%	81.55%	88.71%	96.09%
ZeroR	25.24%	65.52%	70.02%	35.58%	4.06%	33.33%	13.18%	50.02%	40.46%

For the KNIME toolkit, Table 9 shows the results obtained using 10-folds CV as the test mode. The results of Table 9 shows that the NB classifier has achieved accuracy measures ranging between 52%-95%, the OneR and ZeroR classifiers have no accuracy measures because they have no implementation. The C4.5 classifier has achieved accuracy measures ranging between 55%-97%. Finally, both SVM and KNN classifiers have problems running some of the data sets, however, they have achieved accuracy measures ranging between 67%-96% and 33%-98% respectively.

TABLE 9: THE ACCURACY MEASURES GIVEN BY KNIME USING 10-FOLDS-CV.

	Audiology	Breast-W	Car	Flags	Letters	Nursery	SoyBean	SpamBase	Zoo
NB	59.30%	94.80%	85.80%	51.50%	61.60%	90.30%	91.20%	89.90%	88.10%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	70.70%	94.30%	93.50%	54.50%	87.50%	97.30%	72.00%	91.30%	93.10%
SVM	MC/D	96.30%	MC/D	MC	MC	MC/D	MC/D	67.30%	MC
KNN	D	97.50%	D	33.00%	95.40%	D	D	80.90%	71.30%
ZeroR	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 10 shows the accuracy measures achieved using Tanagra with 10-folds CV test mode. As this table demonstrates, the NB classifier has achieved accuracy measures that have ranged between 63% and 96%. For the OneR and ZeroR classifiers, Tanagra has no implementation for such classifiers. On the other hand, the SVM and KNN classifiers have not achieved accuracy measures for all data sets because of some reasons; however, they have achieved accuracy measures that ranged between 90%-97% and 25%-97% respectively. Finally, the C4.5 classifier has achieved accuracy measures ranging between 57%-96%.

TABLE 10: THE ACCURACY MEASURES GIVEN BY TANAGAR USING 10-FOLDS-CV.

	Audiology	Breast-W	Car	Flags	Letters	Nursery	SoyBean	SpamBase	Zoo
NB	70.00%	95.80%	84.30%	62.63%	59.98%	89.91%	89.85%	88.28%	93.00%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	71.36%	93.33%	86.45%	56.84%	85.84%	95.83%	90.24%	91.54%	88.00%
SVM	MC/D	96.96%	MC/D	MC	MC	MC/D	MC/D	89.98%	MC
KNN	D	96.81%	D	25.26%	95.76%	D	D	79.00%	92.00%
ZeroR	NA	NA	NA	NA	NA	NA	NA	NA	NA

C. Performance Improvement

In this section, it is worth to measure the effect of using different evaluation methods for the tools under study. Fig. 2 shows the performance improvements in accuracy when moving from the percentage split test mode to the 10-folds CV mode. This figure demonstrates that WEKA toolkit has achieved the highest improvements in accuracy with a 32 accuracy measures increase, when moving from the percentage split test to the CV test. Orange toolkit on the other hand, has achieved the second highest improvement with a 29 accuracy measures increase, when moving from the percentage split test to CV test. Finally, both KNIME and Tanagra toolkits have achieved the lowest improvements with 12 and 8 accuracy measures increase respectively.

In addition, Fig. 2 shows that the KNIME toolkit has achieved the best rate in terms of the number of accuracy measures decreased; only 4 accuracy measures are decreased when moving from the percentage split test to the CV test in KNIME. For the Orange and WEKA toolkits, the number of accuracy measures decreased where 6 and 7 respectively Finally, the Tanagra toolkit has achieved the least rate with the number of 9 accuracy measures decrease when moving from the percentage split test to the CV test.

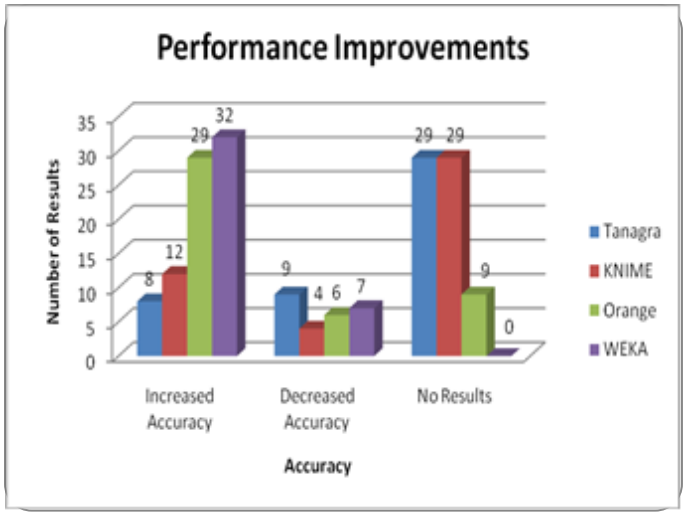

Figure 2. *Performance Improvements*.

Finally, when comparing the four tools in terms of the number of tests that produced no accuracy; the Tanagra and KNIME toolkits have achieved the highest number of tests with no accuracy measures; with a 29 measures each. For the Orange toolkit, only 9 tests have no accuracy measures. On the other hand, WEKA toolkit has 0 tests with no accuracy measures.These results showed that no tool is better than the other to be used for a classification task, this is may be due to the kind of data sets used, or maybe there are some differences in the way the algorithms were implemented within the tools themselves (for example the SVM classifier implemented in WEKA and Orange can handle the problem of multiclass data sets; which is not the case in Tanagra and KNIME that were designed to handle only two class problems).

In terms of applicability (the ability to run a specific algorithm on a selected tool), the WEKA toolkit has achieved the highest applicability, since it is able to run the six selected classifiers using all data sets. Orange Canvas toolkit has scored the second place in terms of applicability, since it run five classifiers out of the six selected classifiers with no ability to run the OneR Classifier. Finally; the KNIME and Tanagra toolkits have both achieved the lowest applicability with the ability to run two classifiers namely; NB and C4.5 on all data sets completely, and partially using another two classifiers namely; SVM and KNN classifiers, while it has no ability to run the last two classifiers namely; OneR and ZeroR classifiers.

In terms of performance improvements, we can judge that WEKA and Orange toolkits have achieved the highest improvements with a 32 and 29 values increased respectively and only 7 and 6 values decreased respectively. On the other hand, the KNIME and Tanagra toolkits have achieved the lowest improvements with 12 and 8 values increased in accuracy respectively and 4 and 9 values decreased respectively.

V. Conclusion and Future Work

This research has conducted a comparison between four data mining toolkits for classification purposes, nine different data sets were used to judge the four toolkits tested using six classification algorithms namely; Naïve Bayes (NB), Decision Tree (C4.5), Support Vector Machine (SVM), K Nearest Neighbor (KNN), One Rule (OneR), and Zero Rule (ZeroR). This study has concluded that no tool is better than the other if used for a classification task, since the classification task itself is affected by the type of dataset and the way the classifier was implemented within the toolkit. However; in terms of classifiers' applicability, we concluded that the WEKA toolkit was the best tool in terms of the ability to run the selected classifier followed by Orange, Tanagra, and finally KNIME respectively.

Finally; WEKA toolkit has achieved the highest performance improvements when moving from the Percentage Split test mode to the Cross Validation test mode followed by Orange, KNIME, and then Tanagra Respectively. As a future research, we are planning to test the selected data mining tools for other machine learning tasks; such as clustering, using test data sets designed for such tasks and the known algorithms for clustering and association.

References

[1] Goebel, M., Gruenwald, L., A survey of data mining and knowledge discovery software tools, ACM SIGKDD Explorations Newsletter, v.1 n.1, p.20-33, June 1999 [doi>10.1145/846170.846172].

[2] Han, J., Kamber, M., Jian P., Data Mining Concepts and Techniques. San Francisco, CA: Morgan Kaufmann Publishers, 2011.

[3] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann P., Witten, I., H., The WEKA data mining software: an update, ACM SIGKDD Explorations Newsletter, v.11 n.1, June 2009 [doi>10.1145/1656274.1656278].

[4] Hornik, K., Buchta, C., Zeileis, A., Open-Source Machine Learning: R Meets Weka, Journal of Computational Statistics - Proceedings of DSC 2007, Volume 24 Issue 2, May 2009 [doi>10.1007/s00180-008-0119-7]. Hunyadi, D., Rapid Miner E-Commerce, Proceedings of the 12th WSEAS International Conference on Automatic Control, Modelling & Simulation, 2010.

[5] Al-Harbi, S., Almuhareb, A., Al-Thubaity, A., Khorsheed, S., and Al-Rajeh, A., "Automatic Arabic Text Classification", 9th International journal of statistical analysis of textual data, pp. 77-83, 2008.

[6] King, M., A., and Elder, J., F., Evaluation of Fourteen Desktop Data Mining Tools, in Proceedings of the 1998 IEEE International Conference on Systems, Man and Cybernetics, 1998.

[7] Giraud-Carrier, C., and Povel, O., Characterising Data Mining software, Intelligent Data Analysis, v.7 n.3, p.181-192, August 2003

[8] Carey, B., Marjaniemi, C., Sautter, D., Marjaniemi, C., A Methodology for Evaluating and Selecting Data Mining Software, Proceedings of the Thirty-second Annual Hawaii International Conference on System Sciences-Volume 6, January 05-08, 1999.

[9] Abbot, D. W., Matkovsky, I. P., Elder IV, J. F., An Evaluation of High-end Data Mining Tools for Fraud Detection, IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA, October, pp. 12--14, 1998.

[10] Hen, L., E., and Lee, S., P., Performance analysis of data mining tools cumulating with a proposed data mining middleware, Journal of Computer Science: 2008.

[11] WEKA, the University of Waikato, Available at: http://www.cs.waikato.ac.nz/ml/weka/, (Accessed 20 April 2011).

[12] Tanagra – a Free Data Mining Software for Teaching and Research, Available at: http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html, (Accessed 20 April 2011).

[13] KNIME (Konstanz Information Miner), Available at: http://www.knime.org/, (Accessed 20 April 2011).

[14] Orange – Data Mining Fruitful and Fun, Available at: http:// http://orange.biolab.si/, (Accessed 20 April 2011).

[15] UCI Machine Learning Repository, Available at: http:// http://archive.ics.uci.edu/ml/, (Accessed 22 April 2011).

[16] Flach, P., A., Lachiche, N., Naive Bayesian Classification of Structured Data, Machine Learning, v.57 n.3, p.233-269, December 2004.

[17] Heb, A., Dopichaj, P., Maab, C., Multi-value Classification of Very Short Texts, KI '08 Proceedings of the 31st annual German conference on Advances in Artificial Intelligence, pp. 70-77, 2008,

[18] Zhou, S., Ling, T., W., Guan, J., Hu, J., Zhou, A., Fast Text Classification: A Training-Corpus Pruning Based Approach. Proceedings of the Eighth International Conference on Database Systems for Advanced Applications, pp.127, 2003.

[19] Pathak, A., N., Sehgal M., Christopher, D., A Study on Selective Data Mining Algorithms, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011.

[20] Li, Y., Bontcheva, K., dapting Support Vector Machines for F-term-based Classification of Patents, Journal ACM Transactions on Asian Language Information Processing, Volume 7 Issue 2, June 2008.

[21] Al-Radaideh, Q., Al-Shawakfa, E., Al-Najjar, M., I., Mining Student Data Using Decision Trees, The 2006 International Arab Conference on Information Technology (ACIT2006), December 19-21, 2006.

[22] Al-Radaideh, Q., The Impact of Classification Evaluation Methods on Rough Sets Based Classifiers, Proceedings of the 2008 International

Arab Conference on Information Technology (ACIT2008). University of Sfax, Tunisia. December 2008.

AUTHORS PROFILE

Abdullah H. Wahbeh is a lecturer in the department of Computer Information Systems at Yarmouk University in Jordan. He obtained his Master and bachelor degrees in Computer Information Systems (CIS) from Yarmouk University, Irbid-Jordan. His research interests include: data mining, web mining and information retrieval.

Qasem A. Al-Radaideh is an Assistant Professor of Computer Information Systems at Yarmouk University. He got his Ph.D. in Data Mining field from the University Putra Malaysia in 2005. His research interest includes: Data Mining and Knowledge Discovery in Database, Natural Language Processing, Arabic Language Computation, Information Retrieval, and Websites evaluation. He has several publications in the areas of Data Mining and Arabic Language Computation. He is curently the national advisor of Microsoft students partners program (MSPs) and MS-Dot Net Clubs in Jordan.

Mohammed N. Al-KABI is an Assistant Professor in the Department of Computer Information Systems at Yarmouk University. He obtained his Ph.D. degree in Mathematics from the University of Lodz (Poland). Prior to joining Yarmouk University, he spent six years at the Nahrain University and Mustanserya University (Iraq). His research interests include Information Retrieval and Search Engines, Data Mining, and Natural Language Processing.

Emad M. Al-Shawakfa is an Assistant Professor at the Computer Information Systems Department at Yarmouk University since September 2000. He holds a PhD degree in Computer Science from Illinois Institute of Technology (IIT) – Chicago, USA in the year 2000. His research interests are in Computer Networks, Data Mining, and Natural Language Processing. He has several publications in these fields and currently working on others.