Performance Analysis of Classification Learning Methods on Large Dataset using two Data Mining Tools

Mazhar Ali[1], Dr. Husnain Mansoor Ali[2]

[1]*Benazir Bhutto Shaheed University, Lyari Karachi, Pakistan*
[2]*Shaheed Zulfikar Ali Bhutto Institute of Science and Technology (SZABIST) Karachi Pakistan*

[1]mazharaliabro@gmail.com
[2]husnain.mansoor@szabist.edu.pk

Abstract - Data is increasing day to day thus, processing this data and selection of right method and tool is really a big problem. Computer scientists are processing and analysing data on different machine learning methods using various Data Mining tools to get the high accuracy of results and minimum time for building of Model. There are several data analysis and processing tools like WEKA, RapidMiner, Keel, and etc. available for the purpose of processing, analysis, modelling and etc. Still no single tool is perfect or nominated for data processing and analysis. In this concern, the authors present here a comparative and analytical research study on the performance of different classification machine learning algorithms like Naïve Bayes, KNN, IBK, Random Forest, C4.5, J48 and Data Mining tools which are WEKA and RapidMiner on a large datasets to evaluate their performance and analytical results with low cost of error. The data set Adult Income is taken from UCI Data repository for this research study. The significance and aim of this study is to evaluate and assess the range of performance of different machine learning methods and two diverse data mining tools on dissimilar datasets. The result of each classification method and Data mining tool is analysed and presented in the end.

Keywords - Data Mining, Classification, Machine Learning Methods, WEKA, RapidMiner.

I. INTRODUCTION

The data is increasing fast every day because of several reasons and sources like databases, flat files, data warehouses, and etc. To get the accurate information and knowledge from this data, there is need of suitable tools; otherwise, it is not possible to justify the results and therefore, right decisions cannot be taken. Data mining is technique that processes the data which is stored in different formats in data warehouses and databases etc. [1]. Different data mining techniques are used to analyse and process the large number of data by Data Mining tools such as WEKA, RapidMiner, Orange, Keel etc. [2].

The idea of this research is to compare and evaluate the performance of WEKA and RapidMiner on diverse Classification Algorithms as well as performance of dissimilar classifiers using large dataset. In this concern, accuracy, precision, recall and classification error of each classification method are investigated by using these two environments of Data Mining and the difference of performance of each tool and classification method is assessed.

II. BACKGROUND OF STUDY

Classification is one of the popular and essential procedures of Data mining that performs most important supervised tasks. Several analytical research studies are carried out on performance of a number of classifiers on diverse types of datasets using several types of Data Mining tools and various results have been obtained. This independent research study is different than others. Here, it is tried to assess the performance of different classifiers and Data Mining tools on large data set.

There are two dimensions of this research study; one is to estimate the performance of diverse classifiers using two types of open source data mining tools and other dimension is to analyse the performance of two data mining tools on classification algorithms using large dataset. In this regard, model building time and the performance of both tools are counted. The training set of Adult Income data is evaluated and analysed in two environments of Data Mining by applying classification methods.

In data mining, a number of researches have conducted to perform multiple tasks. Therefore, different tools and methods have been used to evaluate and process large scale data. Mikut et al [3], describe that Data mining techniques are forecasting the prospect trends, behaviours to take the positive, practical and knowledge based decisions. The use of proper and powerful Data Mining tool for the application of data mining algorithms is very important. Christa et al [4] portray in their research study that the primary challenge for any organization is to process large scale data; therefore, there is need of proper evaluation and assessment of a range of Data mining tools and methods accessible to mining specialists. In this concern, there is need of expert users to conduct the experiments by choosing proper techniques and tools to develop a knowledge model. Graczyk et al [5] discussed the results of six machine learning methods applied on dataset derived from the Cadastral Systems using WEKA, KEEL and Rapid Minor and describe that KEEL and WEKA give the same results but there was difference between RapidMiner and KEEL. Sharma et al [6] performed a study on spam email data analysis using WEKA and find out J48 as highest accuracy oriented classifier while CART, AD Tree and ID3 remained low than the J48. Rodriguez-Galiano et al [7] talk about the Random Forest classification method which achieved above the ground performance with highest accuracy on land-cover classification. In that study, it is shown that Random Forest method is acknowledged as one of the significant classification method for categorizing the land cover. A number of research studies are performed on the classifiers to evaluate their performance on different type of data. Chai et al [8] gave details of the accuracy of Naïve Bayes and Decision trees classifiers and mentioned the massive accomplishment of both classifiers in building a classification models with little cost of the error. . The importance of classification technique on Off-line signature recognition was shown in [9]. These researchers mentioned the K-NN on the highest rank by getting 96% accurate results [9].

Classifying Adult Income Dataset, [10] discussed that naïve Bayes gives good results of accuracy in the small datasets but it could not scale up the accuracy in large data sets of classification. This study suggested that Decision trees provide the higher accuracy in large datasets [10].

Viewing this suggestion, we used the Random forest and C4.5 / J48 decision trees along with other classification methods for the classification of Adult income in this study. Naïve Bayes algorithm is also used in this research to verify the mentioned paper's assessment. K-NN and IBK algorithms are also added in this research to see result differences.

III. DATA MINING TOOLS

In this age of computer and information technology, data processing and knowledge extraction is important due to overwhelming expansion of data.. Therefore, to process that large data and explore the knowledge from that rich data require sophisticated technology. Data Mining is a technology that mines the knowledge from abundance of data and provides a better environment with extracted knowledge to decision makers to make proper decision. There are numerous Data Mining and Knowledge discovering tools that process the data and summarizes the results to give concrete information to decision makers and researchers. This study is designed to assess the performance of different classification algorithms on large dataset of Adult Income using two environments of data processing and knowledge extraction namely WEKA and RapidMiner.

A) WEKA

"Waikato Environment for Knowledge Analysis" known as WEKA tool is open source Data mining tool developed by Java in the "University of Waikato" New Zealand. This tool is widely used by researchers and academicians for purpose of research and teaching. It is a workbench for Machine-Learning methods which plans to develop an application to solve the real world problems using various types of machine learning methods [11]. WEKA had been ranked high in comparison of Orange, KNIME, and Tanagra working on diverse type of classification methods [11].

B) RAPIDMINER

It is integrated and open source Data Mining environment which provides solutions, services and platform for Machine-Learning, analytics consisted of predictive analytics, data mining, text mining, and business analysis. This software was developed in year 2001 with name "YALE" at the "University of Dortmund" and then in year 2007 with name RapidMiner-1. RapidMiner is employed by businesses, researchers and academicians for research, data analysis and education. This tool keeps the data tables as the example set objects and store in memory [12-13].

There are diverse properties of WEKA and RapidMiner. Some are similar and others are different. Both tools are open source and have cross platform.

Table 1. The comparative analysis of RapidMiner and WEKA Data Mining tools [4].

Property	RapidMiner	WEKA
Dataset Partitioning For Training and Testing Set	Yes- Partial	Yes- Partial
Descriptor Scale	Yes	No
Selection of Descriptor	Yes	Yes- But it is not part of Knowledge-Flow
Parameter Optimization of Machine Learning/Statistical Methods	Yes	Yes-But not automatic
Model Validation using Cross Validation and / or Independent Variable set	Yes	Yes-But have to rebuild the model for every Future Dataset

IV. DATASET INFORMATION

The Adult Income data set is taken from UCI data repository for this research. It is a multivariate and categorical data set. It has 48842 instances and 14 attributes with one class attribute Income that has two categories one is Income <=50K and second is Income >= 50K. The class type is nominal. The training set of this data set is consisted of 32561 example set.

This set of data is processed in RapidMiner and WEKA tools for cleansing and classification process. All the values are processed and normalized. The class attribute is labelled as class to classify the adult income. The data set is converted into CSV format for both tools of data mining (Figure 1 and 2).

ExampleSet (32561 examples, 1 special attribute, 14 regular attributes)

Role	Name	Type	Statistics	Range	Missings
label	income	binominal	mode = <=50K (24720), least = >	<=50K (24720), >50K (7841)	0
regular	age	integer	avg = 38.582 +/- 13.640	[17.000 ; 90.000]	0
regular	workclass	polynominal	mode = Private (22696), least = I	State-gov (1298), Self-emp-not-ir	0
regular	fnlwgt	integer	avg = 189778.367 +/- 105549.97	[12285.000 ; 1484705.000]	0
regular	education	polynominal	mode = HS-grad (10501), least =	Bachelors (5355), HS-grad (105(	0
regular	education-num	integer	avg = 10.081 +/- 2.573	[1.000 ; 16.000]	0
regular	marital-status	polynominal	mode = Married-civ-spouse (149	Never-married (10683), Married-(	0
regular	occupation	polynominal	mode = Prof-specialty (4140), lea	Adm-clerical (3770), Exec-manag	0
regular	relationship	polynominal	mode = Husband (13193), least	Not-in-family (8305), Husband (1	0
regular	race	polynominal	mode = White (27816), least = O	White (27816), Black (3124), Asia	0
regular	sex	binominal	mode = Male (21790), least = Fe	Male (21790), Female (10771)	0
regular	capital-gain	integer	avg = 1077.649 +/- 7385.292	[0.000 ; 99999.000]	0
regular	capital-loss	integer	avg = 87.304 +/- 402.960	[0.000 ; 4356.000]	0
regular	hours-per-week	integer	avg = 40.437 +/- 12.347	[1.000 ; 99.000]	0
regular	native-country	polynominal	mode = United-States (29170), le	United-States (29170), Cuba (95	0

Fig. (1). Shows missing values of dataset of Adult Income processed in RapidMiner.

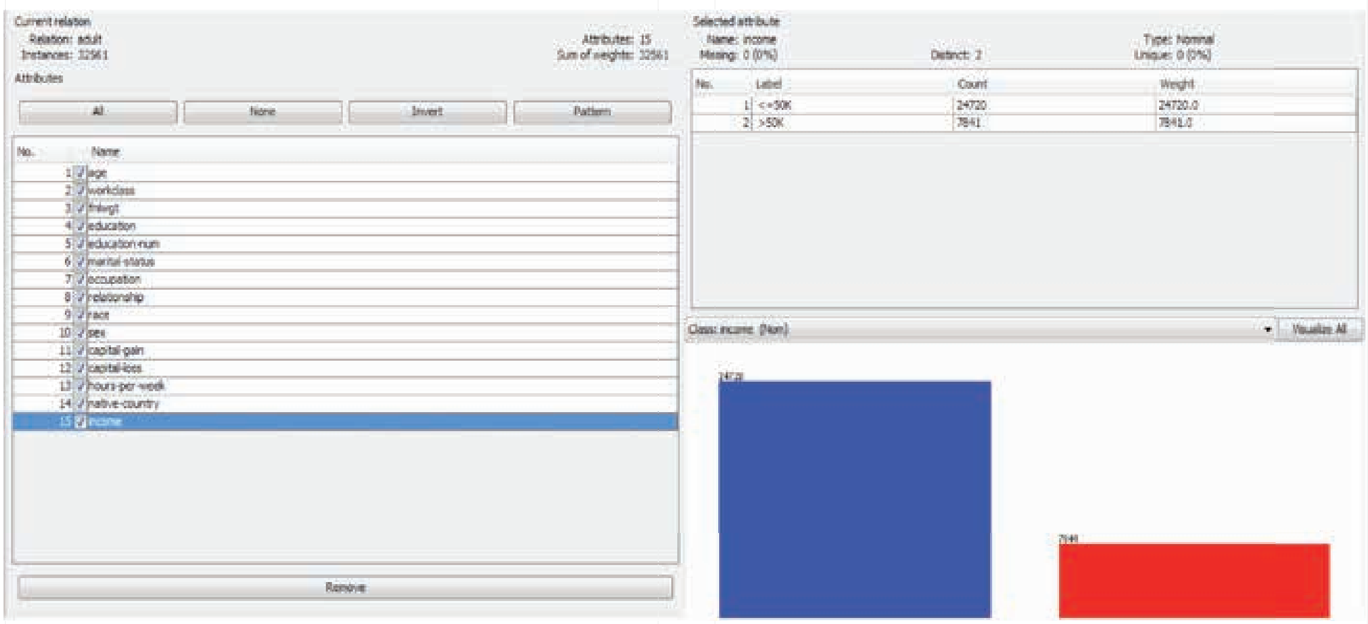

Fig. (2). Shows status of data of Adult Income dataset processed in WEKA

V. MATERIALS AND METHOD

Data understanding is the fundamental step to find out the proper and useful data set for research; especially, when it comes to application of data mining techniques. Data preparation measures are required to make the data set clean and usable for the classification.

V.I CLASSIFICATION METHOD

Classification is a patent and important technique that is applicable to numerous fields and domains [14]. There are several classification algorithms but authors have concentrated on those learning methods of classification which are available in both WEKA and RapidMiner DM tools and are capable of processing chosen dataset. Therefore, Naïve Bayes, Random Forest, K-NN, IBK, C4.5 and J48 algorithms are preferred to process and analyse the selected dataset of Adult Income. These algorithms are categorized and assessed on basis of their accuracy.

A) NAIVE BAYES

Though Naïve Bayes is good for spam filtering and text classification but it cannot be ignored in the other areas of classification. In this concern, a number of modifications have been revised to formulate and build this technique more flexible by the diverse statistical, machine learning, pattern recognition and data mining groups [15]. On basis of its ease, solidity, significance, worth and simple to apply on big set of data, Naïve Bayes is considered as prominent and useful algorithm of data mining and machine learning [16]. Naïve Bayes is supervised method of classification. This algorithm resolves the problems of diagnostic and prediction [17].

B) RANDOM FORESTS

Random Forest also called the RF algorithm and is one of the machine learning methods that is commonly used for the developing continuous variables as well as image classification. This classifier is an ensemble model that gathers the outcomes of diverse models to work out and then gives an estimated output. In this way, a number of decision trees or classifiers are established and result of them is calculated to present an outcome [18]. To build a RF framework, two constraints are required; the amount of Decision Tree together or in the ensemble and an amount of input predictors randomly chosen from each node. Usually it is one third of the whole contributed variables of a classification process [19]. Random Forest algorithm has several applications including the cloud, shadow detection and Land cover mapping [18]

C) K-NN / IBK

K-Nearest Neighbour is classification algorithm which is used and widely accepted. This is easily understandable and instance-based algorithm. It collects the nearest objects of data of similar kind from training data set; therefore, it is appropriate for multi-function or model class decision. It locates the nearby elements called K elements from training set whereas the value of K is defined [20].

The RapidMiner tool use K-NN classifier while WEKA use IBK. The IBK is a java implementation of the K-nearest Neighbour in WEKA. These types of classifiers classify the instance by major votes of close elements called K elements. Here K is positive integer and it may be equal to 1, 3 or 5 etc. [21]. In this research, the value of K is set to 3.

D) C4.5 / J48

Decision trees or DTs perform an important role in classification technique of Data Mining. The C4.5 algorithm is one of the Decision Tree algorithms used for the classification problems of Data Mining as well as Machine Learning. This algorithm is successor of ID3 and developed by J. Ross Quinlan. C4.5 algorithm divides the data into small sub sets through testing specific attribute of node till the entire sub sets are purified resulting in instances of sub sets falling in similar class. The test is performed thorough Gain Ratio or Gain [16]. Meera [22] described C4.5 decision tree as good classifier for prediction than the instance based and Naïve Bayes classification methods. RapidMiner tool uses the C4.5 decision tree algorithm for classification while WEKA tool is using J48 in place of C4.5 for classification process which is java implementation of C4.5 method [6].

VI. MODELING AND TESTING

Classification is supervised method that classifies the data on basis of target class. The model of Adult Income data set defined in figure 3 to 7 is constructed separately by cross validation with the intention of estimation of the statistical performance of learning methods and in this research, the number of cross-validation is set to 10-fold cross-validation and sampling type is stratified sampling.

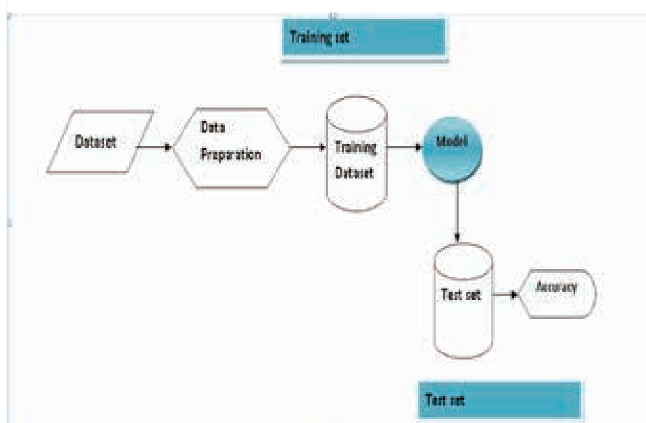

Fig. (3). Describes the dataset processing model of this study

Fig. (4). Describes Process Model of Naïve Bayes

Fig. (5). Describes Process Model of K-NN

Fig. (6). Describes Process Model C4.5 Decision Tree

VI.I. EVALUATION AND ANALYSIS

Both tools are analysed on their performance including the time taking for the construction of models. The classification learning methods are also evaluated and their results are

Fig. (7). Describes Process Model Random Forest

analysed on the subject of the accuracy, precision, recall, classification error and Kappa values. In this way, confusion matrices are derived and analysed after developing the models.

VII. RESULTS

This research is planned to assess the performance of data mining tools WEKA and RapidMiner in terms of how much time these tools take in building the models of classifiers like Naïve Bayes, Random forest, K-NN/IBK and C4.5/J48. Along with this, performance of mentioned classification methods is also measured on data set of Adult income.

In this way, accuracy, classification error, precision, recall and kappa is evaluated and compared. The parameters are set same in both tools. The Figure 8 shows the performance of all mentioned classification methods which processed in both tools.

The Figure 8 shows the performance of classification methods executed in WEKA and RapidMiner. WEKA and RapidMiner have different performance on classification methods. Performance is measured in form of accuracy, classification error and kappa values. Training data set is processed in Naïve Bayes, Random Forest, K-NN vs. IBK and C4.5 vs. J48 classifiers. The accuracy of J48, IBK and

Random Forest was measured higher than the C4.5, K-NN and Random Forest while accuracy of Naïve Bayes remained same in both tools. The figure 8 shows the good performance of WEKA on large data set. The cost of error of classifiers is less in WEKA in contrast of RapidMiner. The parameter settings are kept same for all mentioned classification methods in both tools.

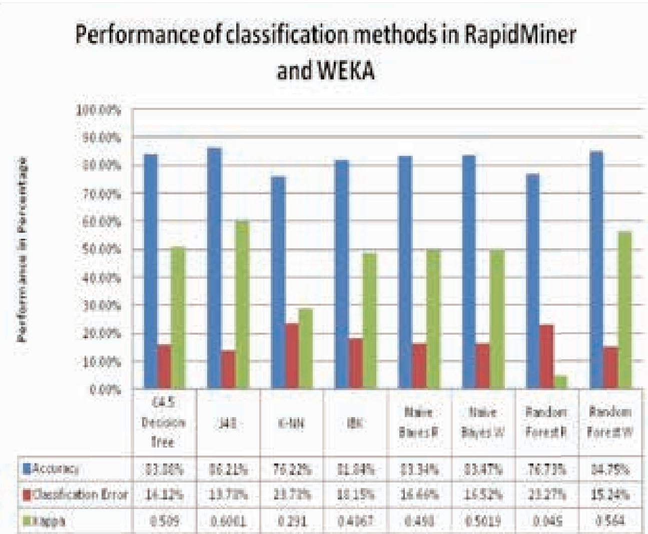

Fig. (8). Describes the performance of classification methods processed in RapidMiner and WEKA

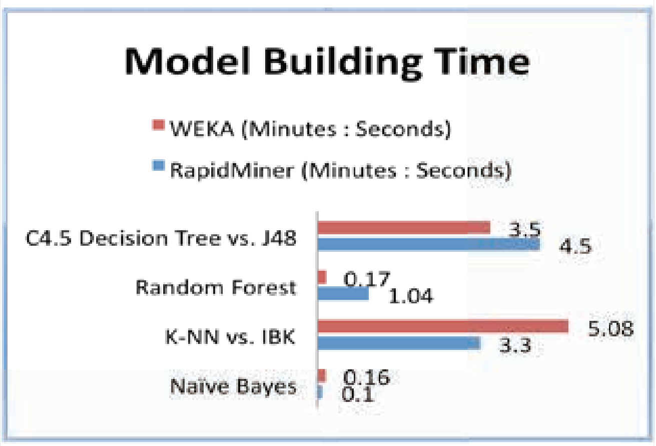

Fig. (9). Shows the performance difference of WEKA and RapidMiner in building the models.

Figure 9 shows the performance difference of WEKA and RapidMiner in form of building the models and executing the classifiers. Both tools take dissimilar time in building models. Even the performance accuracy of classification methods is different. WEKA takes more time than RapidMiner in building models of IBK and Naïve Bayes while it takes less time in contrast of RapidMiner in building models of J48 and Random Forest.

Figure 10 shows the precision disparity of classification methods executed in WEKA and RapidMiner and Fig 11 shows the recall disparity of classification methods executed in WEKA and RapidMiner. In this research study, classification methods are analysed on basis of precision and recall using WEKA and RapidMiner data mining tools. The class precision and recall are found almost same in Naïve Bayes and little disparity is seen in C4.5 decision tree vs. J48 but a difference is observed in precision and recall of K-NN vs. IBK and Random Forest.

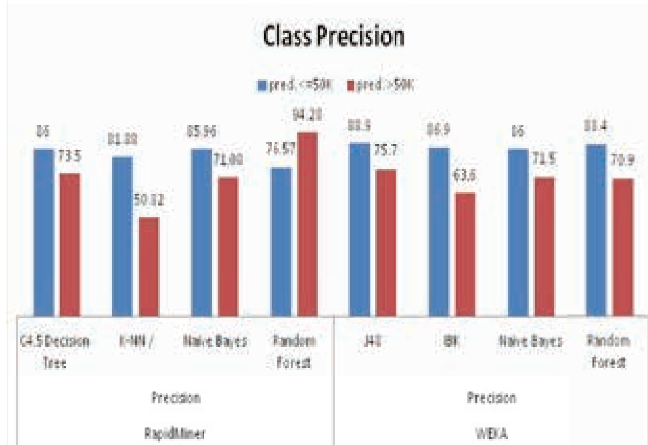

Fig. (10). Shows the performance of classification methods in form of class precision which are executed in WEKA and RapidMiner.

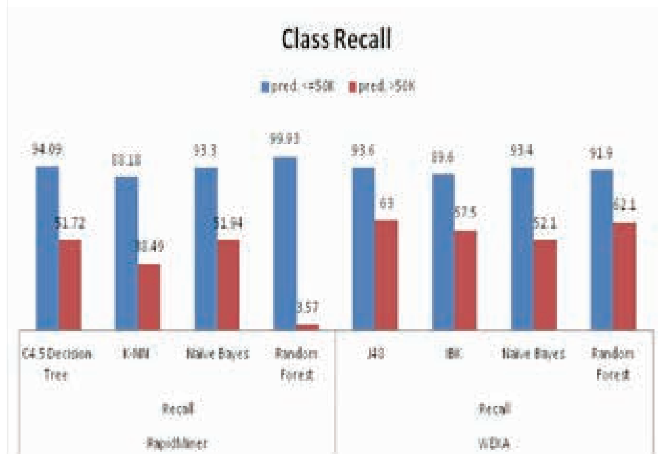

Fig. (11). Shows the Recall disparity of classification methods executed in WEKA and RapidMiner.

VIII. DISCUSSION

This comparative research is performed on analytical assessment and evaluation of performance of two data mining tools RapidMiner and WEKA as well as performance of diverse classifiers on large dataset of Adult income. The WEKA and RapidMiner tools are emerging and significant tools in the field of Data Mining. The performance of both tools is evaluated on basis of comparison of time taken for the model building by each tool including accuracy and precision measurement of classifiers. The training sets and models are built and tested by cross validation. The cross-validation is set to 10-fold cross-validation. The efficiency of every tool in terms of time is discussed in figure 9. This study ranks WEKA better for J48. Random Forest and ranks RapidMiner tool measured good for K-NN and Naïve Bayes on basis of time efficiency in building models.

The performance of classifiers is remained different in both tools in this study. The J48 classification method performed with high accuracy of 86.21%, with lower cost of classification error 13.78% and Kappa value 0.600 in WEKA in contrast of C4.5 and other methods of classification in RapidMiner. C4.5 was 83.88% accurate with high classification error 16.12% and Kappa value 0.50 on same dataset with same parameter settings. Random Forest performed well in WEKA with 84.75% accuracy, 15.24% classification error and Kappa value 0.56 but same method provides less accuracy in RapidMiner with 76.73% accuracy, 23.27% cost of classification error and Kappa value 0.04. Naïve Bayes learning method's performance is almost same in both tools. It gives 83.34% accuracy in RapidMiner and 83.47% accuracy in WEKA. The Kappa value is less in RapidMiner with 0.049 and high in WEKA with 0.501. The IBK learning method's performance in form of accuracy remained high in WEKA than performance of K-NN in RapidMiner. The accuracy of IBK is 81.84% with classification error 18.15% while performance of K-NN in RapidMiner is 76.22% with classification error of 23.78%. The Kappa values remained less in RapidMiner with 0.29 than in WEKA with 0.48

As a result, J48 classifier is proper learning method for prediction on large dataset like Adult Income in contrast of C4.5 decision tree. Random Forest classifier is a better option for prediction on large dataset using WEKA tool. Therefore, WEKA performed well on large dataset using different classifiers than RapidMiner. In scenario of modelling, WEKA took less time in building models of J48 and Random Forest. Both classifiers gave good result in this study than the other classifiers but took more time in building of models of Naïve Bayes and IBK. RapidMiner took less time in building models of Naïve Bayes and K-NN but more time in building models of Random Forest and C4.5 Decision Trees.

IX. CONSLUSION

Data mining is best source for knowledge extraction. It provides platform to decision makers to take proper decisions. This research is carried out on large data set of adult income which is obtained through census. Data scientists are worried in selecting the tool for proper prediction. Viewing this problem, present research study was planned and two data mining tools were selected which are WEKA and RapidMiner. Four classifications methods Naïve, Bayes, K-NN vs. IBK, Random forest and C4.5 vs. J48 were chosen for processing. The parameter settings were kept same in both tools. In this study, WEKA performed well on large dataset using dissimilar classifiers than RapidMiner. In development of models, WEKA took less time in building models of J48 and Random Forest but took more time in building models of Naïve Bayes and IBK. The RapidMiner took less time in building models of Naïve Bayes and K-NN but took more time in building models of C4.5 Decision tree and Random Forest. The performance of diverse classifiers remained different. As a result, J48 classifier performed better on mention dataset and proved itself as proper learning method for prediction on large dataset of little element of class in contrast of C4.5 decision tree. Random Forest classifier is also good for prediction on large dataset using WEKA tool. The K-NN and Naïve Bayes performance remained less in RapidMiner in contrast of IBK and Naïve Bayes in WEKA.

There is a need of further analytical and comparative study using these tools on large and small dataset so that more difference can be found.

REFERENCES

[1] N. Padhy and R. Panigrahi "Multi Relational Data Mining Approaches: A Data Mining Technique," *International Journal of Computer Applications,* vol. 57, no. 17, pp: 23-32, 2012.

[2] R. S. Priyaa, and S. S. Kumar. "Efficient Aircraft Traffic Scheduling using Population Heuristics with CART Algorithm," *International Journal of Science and Research,* vol. 3, no. 1, pp: 248-252, 2014.

[3] R. Mikut, and M. Reischl. "Data mining tools." Wiley *Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* vol. 1, no. 5, pp: 431-443, 2011.

[4] S. Christa, K. L. Madhuri and V. Suma. "A Comparative Analysis of Data Mining Tools in Agent Based Systems." unpublished.

[5] M. Graczyk, T. Lasota and B. Trawiński. "Comparative analysis of premises valuation models using KEEL, RapidMiner, and WEKA." In *Proceedings of the 1st International Conference on Computational Collective Intelligence. Semantic Web, Social Networks and Multi-agent Systems (ICCCI '09),* Berlin, 2009, pp: 800-812.

[6] A. K. Sharma and S. Sahni. "A comparative study of classification algorithms for spam email data analysis," *International Journal on Computer Science and Engineering,* vol. 3, no. 5, pp: 1890-1895, 2011.

[7] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo and J. P. Rigol-Sanchez. "An assessment of the effectiveness of a random forest classifier for land-cover classification." *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 67, pp: 93-104, 2012.

[8] X. Chai, L. Deng, Q. Yang, and C. X. Ling. "Test-cost sensitive naive Bayes classification," In *Proceedings of Fourth IEEE International Conference on Data Mining (ICDM'04),* 2004, pp: 51-58.

[9] K. Ubul, A. Adler, G. Abliz, M. Yasheng, and A. Hamdulla. "Off-line Uyghur signature recognition based on modified grid information features." In *Proceedings of 11th International Conference on Information Science, Signal Processing and their Applications (ISSPA),* 2012, pp: 1056-1061.

[10] R. Kohavi. "Scaling Up the Accuracy of Naive-Bayes Classifiers: A Decision-Tree Hybrid." In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining,* 1996, pp: 202-207.

[11] A. H. Wahbeh, Q. A. Al-Radaideh, M. N. Al-Kabi, and E. M. Al-Shawakfa. "A comparison study between data mining tools over some classification methods," *International Journal of Advanced Computer Science and Applications, pp: 18-26, 2011.*

[12] *RapidMiner* [Online]. Available: http://en.wikipedia.org/wiki/RapidMiner.

[13] Z. Prekopcsak, G. Makrai, T. Henk and C. Gaspar-Papanek, "Radoop: Analyzing big data with RapidMiner and Hadoop," In *Proceedings of the 2nd RapidMiner Community Meeting and Conference (RCOMM 2011),* 2011, pp: 865-874.

[14] A. Srivastava, E. H. Han, V. Kumar, and V. Singh. "Parallel formulations of decision-tree classification algorithms," *Data Mining and Knowledge Discovery,* vol. 3, no. 3, pp: 237-261, 1999.

[15] J. Cheng and R. Greiner, "Learning Bayesian belief network classifiers: Algorithms and system," in *Advances in Artificial Intelligence* (Lecture Notes in Computer Science), vol. 2056, pp: 141-151, 2001

[16] X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z. Zhou, M. Steinbach, D. J. Hand and D. Steinberg, "Top 10 algorithms in data mining," *Knowledge and Information Systems,* vol. 14, pp: 1-37, 2008.

[17] *Naive-Bayes Classification Algorithm* [Online]. Available:http://software.ucv.ro/~cmihaescu/ro/teaching/AIR/docs/Lab4-NaiveBayes.pdf.

[18] N. Horning, "Random Forests: An algorithm for image classification and generation of continuous fields data sets," In *Proceeding of International Conference on Geo-informatics for Spatial Infrastructure Development in Earth and Allied Sciences,* 2010.

[19] A. Ghosh, R. Sharma and P. K. Joshi. "Random forest classification of urban landscape using Landsat archive and ancillary data: Combining seasonal maps with decision level fusion." *Applied Geography,* vol. 48, pp: 31-41, 2014.

[20] M. Kuramochi and G. Karypis, "Gene classification using expression profiles: A feasibility study," *International Journal on Artificial Intelligence Tools,* vol. 14, no. 4, pp: 641-660, 2005.

[21] G. M. Soares, A. G. B. Almeida, R. M. Mendes, E. C. Teixeira, H. A. C. Braga, M. N. Machado, R. S. Broetto, M. V. H. B. Castro, H. O. G. Filho, F. M. Varejao, J. G. P. Filho, A. B. Candeia, and R. A. A. Sousa, "Detection of street lighting bulbs information to minimize commercial losses." In *Proceedings of 2013 IEEE Seventh International Conference on Sensing Technology (ICST),* 2013, pp: 895-900.

[22] G. Meera Gandhi, "Machine learning approach for attack prediction and classification using supervised learning algorithms." *International Journal of Computer Science & Communication,* vol. 1, no. 2, pp: 247-250, 2010.