DECISION MAKING IN ENTERPRISE COMPUTING:

A DATA MINING APPROACH

D. Asir Antony Gnana Singh
Department of Computer Science and Engineering,
Bharathidasan Institute of Technology, Anna University
Tiruchirappalli, India.
asirantony@gmail.com

E. Jebamalar Leavline
Department of Electronics and Communication Engineering,
Bharathidasan Institute of Technology, Anna University,
Tiruchirappalli, India.
jebi.lee@gnail.com

Abstract
The quality of the life and economic growth of a country mainly depends on the successful operations of enterprises. The enterprise is an organization with business activity that renders the goods or services for improve the quality of life. Decision making plays an important role in the development of the enterprises. The enterprise computing utilizes the advantages of information and communication technology for computerizing the operation of the enterprises in management process, management level and the functional area of the enterprises. In general the decisions are made in various stages from receiving raw material to convert them into the finished goods or from initiating the services to until finish the service in an enterprise. Therefore, quality decisions are very essential for the growth of the enterprises. These decisions are made by the various data generated during the operations of the enterprises. The data mining is a powerful analytic process to obtain the knowledge from the generated data in order to make decision for improving the quality of operations of the enterprises. This paper presents the decision making approach in enterprise computing in data mining perspective. This paper also explores and compares the quality of decision making using various data mining algorithms on the enterprise data.
Index Terms— Decision making, Enterprise computing, Data mining, Management process, Data mining in management.

I.   Introduction

In the globe, the countries take the rigorous efforts to improve their economy level to improve the quality of life of the people. The gross domestic product (GDP) is an indicator of the economy level

of the country and its magnitude highly depends on the production and services that are offered by the country. The enterprises are the organizations that involve in the business activities to produce the goods and render the services for the humanity. The enterprises can be categorized based on their business activities such as retail, manufacturing, service, wholesale, government, educational, transportation and etc. The successful growth of the enterprises is very essential for improving the economy of the country in order to improve the well-being of the society and the individuals. The enterprise computing can computerize the enterprise operations using the information and communication technology [1] [2].

A. Operations of the enterprises

The outcome of the operations of the enterprises can be the finished goods from a raw material or the services. The operations of the enterprises are carried out in all the areas of enterprises including the management process, management level and functional areas of enterprises.

B. Management Process of the Enterprises

The management process can be the planning, organizing, leading, directing and controlling. Planning can be the sequence of activities that are carried out in order to achieve the desired of the organizational goal. In the organizing process, the relations can be established by providing the authority and the responsibilities to the human resources in order to play the role in the enterprise and the proper departmentalization can be made on the asserts or materials of the enterprise for easy identification , access and management. In the leading process, the directions are given to the workers for driving the enterprise in the right way for the growth of the enterprise. In directing process, the workers can be influenced, guided, supervised and motivated in order to extract the job with positive attitude. In the controlling process, the activity of the organization is monitored with the pre-determined standard and the preventive and corrective measures are carried out to sail the enterprise within the pre-determined standards.

C. Management levels in Enterprises

The management level can be classified as top level managers, chief executive officers (CEO), senior managers, middle managers and employees. The top level managers can develop the goals, strategic plans, company policies, and make decisions about the direction of the business and they can be designated as governing board or chairman or author or founder. The chief executive officers can execute the plans developed by the top level mangers and the managers can be responsible for the entire operations of the corporation and report directly to the chairman and board of directors. Senior manager can have the executive powers given to them by authority of the board of directors and they can be designated as heads of the enterprise or institution. The middle managers are the subordinates to the senior manager .Operational supervisors can be considered as the middle management and they can be designated as department heads. The employees can perform specific duties and provide services to the company on a regular basis in exchange for compensation.

D. The functional areas of enterprises

The functional areas of enterprise can be classified as production, sales, marketing, information and

communication technology department, human resources management, finance department, distribution department, customer service and relation department, administration department and research and development department. Each department carry out the operations based on the purpose of the department.

E. Importance of Decision Making in Enterprise Computing

The economic growth of a country in terms of GDP highly depends on the productivity of goods and rendering services. These goods and services are produced and rendered by the enterprises. Therefore, the enterprises are directly participating in the economy growth of the country and improving the quality life of the citizens. The success of an enterprise relies on the good decision making at various levels in the organization operations right from the receipt of raw materials to the delivery of the finished goods and also from initiating the services to finish-up the services. Hence, the decision making is very essential in enterprises for the growth of the enterprises and country's economy [3] [4].

II. Theoretical view of decision making in Enterprise Computing

The development of the enterprises depends on its successful operations. The successful operations can be achieved only through good decision making at various level of the operations. In the enterprise, there are various stages in carrying out the tasks to produce the finished goods from a raw material or to render the services from the initial stage to completion stage. In each stage, making the correct decision is essential to fulfill the objectives or for further stages of operation. These decisions are made based on the enterprise data. These data are generated by recording the actions of the events carried out in every stage of production of the goods or providing the services. The data mining analytic process can be employed to extract the interested pattern from the enterprise data to explore the knowledge for making decision. The data mining process can be classified into two categories namely, classification and clustering. The classification algorithms can process the labeled data and the clustering algorithms can process the unlabeled data. The enterprise data must be preprocessing before given to the data mining process for decision making.

A. Process of Decision making in Enterprise Computing

The process of the decision making in the enterprises involves various activates such as data generation or collection, data storing, data selection and decision making using data mining algorithms. In the recent past, due to the growth of the information technology, data are being generated in high volume and dimension in the enterprises though a variety of data generating tools and techniques in the enterprises such as surveillances, sensors, video and motion capturing equipments, measurement devices, etc. Therefore these data are stored into the enterprise data warehousing for the further data mining process [5] [6] [7].

The enterprise data warehouse can collect the massive data from various sources such as functional areas, departments, and branches of the enterprise and store them in a unified scheme. In general there are two types of architectural schemas adopted in data warehouse. One is source driven scheme and another one is destination driven scheme. In the source driven architecture, the data are being collected from various sources to the warehouse periodically or in a predetermined

interval without expecting the request from the data warehouse. In the destination driven architecture, whenever the data warehouse needs the data, it will give a request and get the desired data from various or particular data sources. The data warehouse can be integrated with the data mart, online analytical processing (OLAP), online transaction processing (OLTP) and integrated big data environment. Data mart can be a less weight data warehouse that can have the data of a particular source such as a functional area of the enterprise. The online analytical processing (OLAP) can be employed to view, explore the data in the form of the multi-dimensional view such us slicing, roll-up, drill-down and pivoting the data from the data cube. The online transaction process can be employed to carryout shout query processing such as delete, update, and insert operations. This is quite faster in query processing.

The enterprise data can be classifies as transactional, analytical, and master data. The transactional data describes the transactional event that is associated with the time. The analytical data can be the outcome of the analysis of an activity or functional data of the enterprise such as quality, availability, efficiency, etc. The master data represents the key business entities of the enterprise such as customer, employee, product, supplier etc. In general, the real world data are noisy, redundant, irrelevant, having missing values and uncertainty in nature. Therefore the data has to be preprocessed before they are feed into the data mining algorithm for decision making. Preprocessing includes data cleaning, data integration, data transformation, data reduction, data discretization [8] [9].

Data mining is the analytic process and it can be employed for discovering the hidden patterns from the large volumes of the enterprise data to extract the knowledge in order to make decision. The data mining approach consists of various data mining algorithms such as associations, constructions and classification. The classification technique plays a vital role in the decision making. The classification algorithms are also known as supervised machine learning algorithms. The classification algorithm receives the enterprise data and develop a model called classification model. The classification model can predict the unknown data and help in decision making.

Classification algorithms can be classified according to the fashion of building the classification model such as tree-based model, probabilistic-based model, rule-based model, and etc. In the tree-based model construction, the attributes of the dataset are considered as nodes in order to form the decision tree. The information gain measure is used to measure the weight of each feature present in the dataset for constructing the decision tree. The information gain for any attribute is calculated as follows: Let p_i be the probability that an arbitrary tuple in D belongs to class C_i, estimated by |Ci, D|/|D| the expected information (entropy) needed to classify a tuple in D is calculated from the Equation 1.

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (1)$$

Information needed (after using A to split D into *v* partitions) to classify D is calculated using the Equation 2.

$$Info_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} \times I(D_j) \qquad (2)$$

Information gained by branching on attribute A is calculated using the equation 3.

$$Gain(A)= Info(D) - Info_A(D) \qquad (3)$$

The attribute with higher information gain weight can be considered as the root node of the decision tree and the remaining nodes can be considered as the child nodes based on their information gain values. The leaf nodes can be the labels of the target attributes. Using this decision tree, the decisions are made to predict the unknown label of the known data [10-14].

III. Framework for Decision Making in Enterprise Computing

This section discusses the framework for decision making in enterprise computing. The framework is illustrated in the Figure1. Decision making is a continuous process in the organizational life cycle. The enterprise data are collected from the enterprise that is generated from the operation of the enterprise. Then the collected data are feed into the data mining algorithm for decision making. Based the decision taken, the corrective or preventive actions can be carried out to improve the operations of the enterprise. Flowchart representation of the data mining and decision making in enterprise computing is depicted in Figure 2. Initially, the enterprise data are loaded into the classification algorithm, and the classification algorithm develops the classification model. From the classification model, the decisions are made from the predicted unknown label of the known enterprise data.

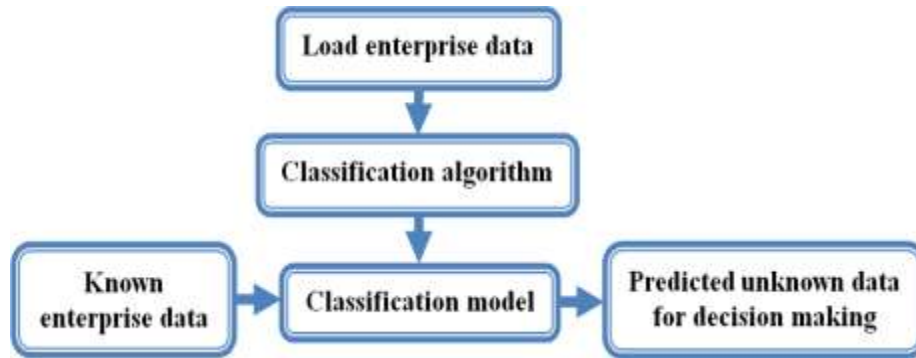Figure 1. Framework of decision making in enterprise computing

Figure 2. Flowchart representation of the data mining and decision making in enterprise computing

IV. Experimental Setup With Experimental Procedure

In order to conduct the experiment, the knowledge flow environment of WEKA data mining software is employed to analyze the performance of the various classification algorithms on different enterprise dataset collected from the UCI dataset repository and WEKA software repository. The performance evaluation metrics such as classification accuracy, root mean squared error (RMSE) and the receiver operating characteristics (ROC) are used to assess the performance of the classification algorithms. In order to evaluate the performance of the classification algorithm totally four classification algorithms are used namely probabilistic-based Naïve Bayes classifier (NB), instance-based classifier IB1, tree-based classifier J48, and the function-based classifier radial basis function network (RBFN).

A. Details of enterprise dataset

The enterprise datasets are collected from the UCI dataset repository and WEKA software dataset. These datasets are tabulated in Table 1. The Bank-marketing dataset was prepared at Portuguese banking institution. The features of clients of the bank such as age, education, marital status, job, etc were recorded. Also, the interest of a client to subscribe for the term deposit was investigated by calling them over phone. This attribute represented desired target attribute to describe whether the client is interested in subscribing for term deposit or not. This dataset can be used to make decision on the whether a person can subscribe for the term deposit or not. The dataset Credit-g can be used to make decision on a person whether his credit risk is good or bad for approval of the loan application in baking sector. In this dataset, the first twenty features are the attributes of the individuals and the 21st attribute is the desired target attribute for make decision.

Table1. Details of enterprise dataset

S.No.	Dataset	Features	Instances
1	Bank-marketing	16	4521
2	Credit-g	20	1000

B. Experimental Setup

In order to carry out the experiment, the WEKA knowledge flow environment is used as illustrated in Figure 3. For this experimental setup, the knowledge flow layout is constituted by various data mining modules such as arff loader, class assigner, cross validation fold maker, classifiers, classifier performance evaluator, text viewer, and the model performance chart. The arff loader loads the enterprise data in the form of dataset with attribute relation file format (arff). The class assigner assigns the desired target attribute for the classification algorithm. The cross validation fold maker is used to split the training and testing dataset for evaluating the performance of the classifier. In this experiment, 10 fold cross validation method is adopted for testing. The classification modules observe the training dataset and build the classification model. The classifier performance evaluator performs the evaluation process for determining the performance of the classification model built by the classifier module in terms of various performance evaluation metrics. In this experiment, the classification accuracy measure, root mean squared error (RMSE) and the receiver operating characteristics (ROC) are used as performance evaluation metrics. The text viewer helps to display the performance metric values. The performance chart model is used to draw the ROC chart based on the performance of the classification mode with respect to the dataset and performance evaluation metrics. In this experiment, the false positive rate and true positive rate are used in order to plot the ROC chart.
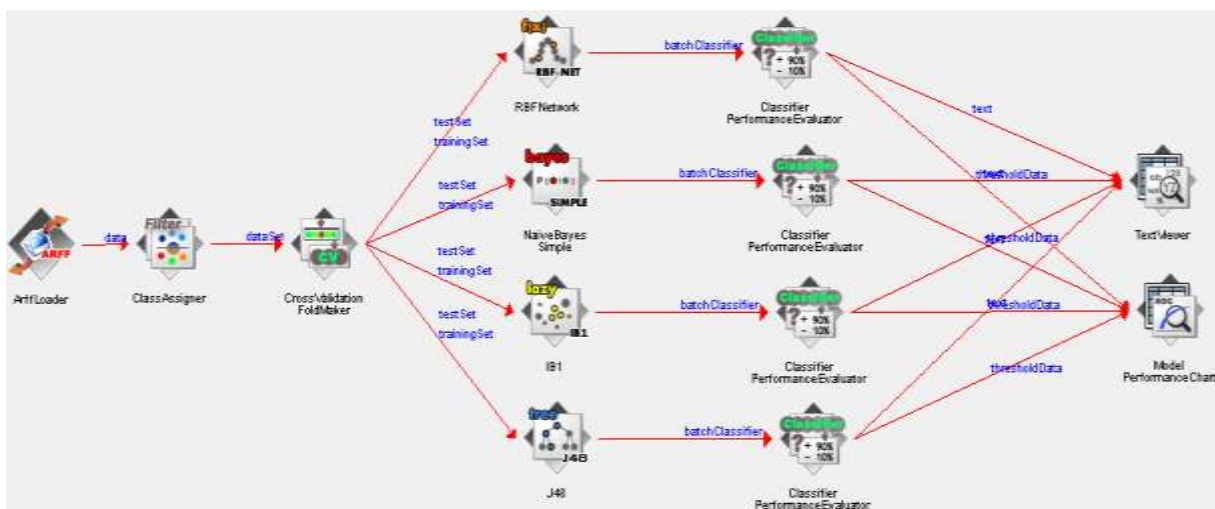

Figure 3. WEKA knowledge flow layout for the conduced experiment

C. Experimental Procedure

As shown in Figure 3, all the data mining modules are connected together using the rubber band connecters and the enterprise dataset is loaded using the arff loader. The desired class attributes are set in class assigner module. The cross validation fold maker is set with the 10 fold cross validation. The results are obtained from the test viewer and model performance chart and recorded in Table 1, Table 2 and Figure 1 to Figure 3.

V.   Results and Discussion

The obtained results are tabulated in Table 2 and Table 3 and illustrated in Figure 4 to Figure 7. From Table 2 and Figure 4, it is observed that for the Bank-marketing dataset J48 performs better than other classifiers in terms of classification accuracy. For Credit-g dataset NB classifier produces higher accuracy than other classification methods. From Table 3 and Figure 5, it is observed that for the Bank-marketing dataset, RBFS performs better than the other classifiers in terms of root mean squared error (RMSE). For the Credit-g dataset NB classifiers produces lesser RMSE than the other classification methods. From Figure 6 and Figure 7, it is observed that for the Bank-marketing and Credit-g dataset the J48 and the NB classifier perform better in terms of receiver operating characteristics (ROC) compared to other classification methods.

Table 2 Classification accuracy of various classifiers in decision making against the enterprise dataset

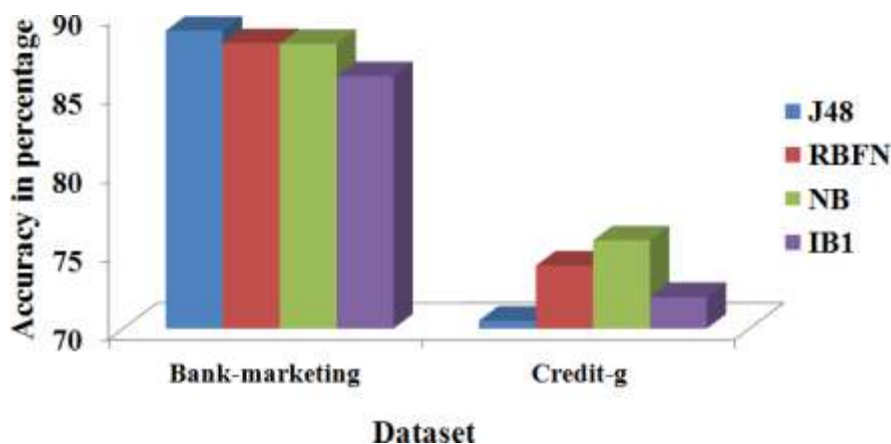| Dataset	Supervised machine learning algorithms			
	J48	RBFN	NB	IB1
Bank-marketing	88.914	88.1221	88.0806	86.0208
Credit-g	70.500	74.0000	75.6000	72.0000

Figure 4. Classification accuracy of various classifiers in decision making against the enterprise dataset

Table 3 Root mean squared error (RMSE) of various classifiers in decision making against the enterprise dataset

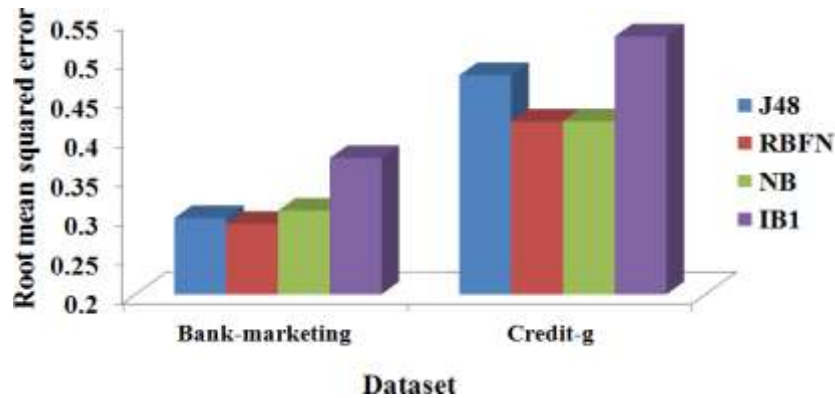| Dataset	Supervised machine learning algorithms			
	J48	RBFN	NB	IB1
Bank-marketing	0.2974	0.2898	0.3069	0.3739
Credit-g	0.4796	0.4204	0.4209	0.5292

Figure 5. Root means squared error (RMSE) of various classifiers in decision making against the enterprise dataset
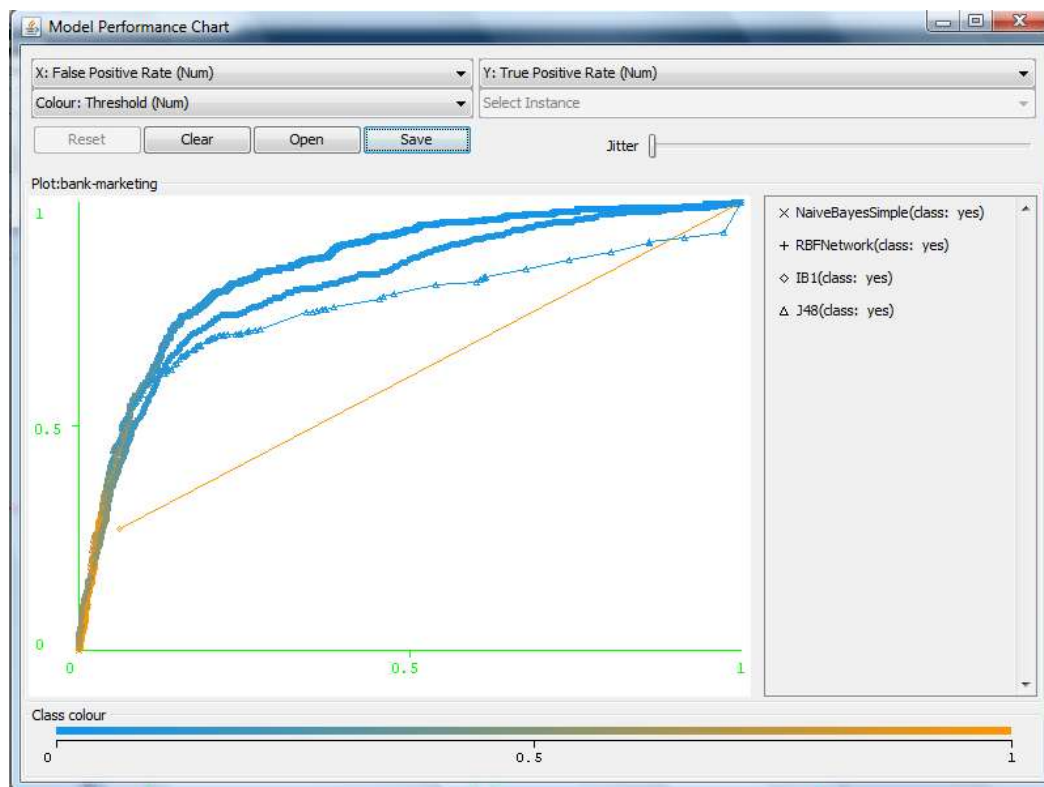

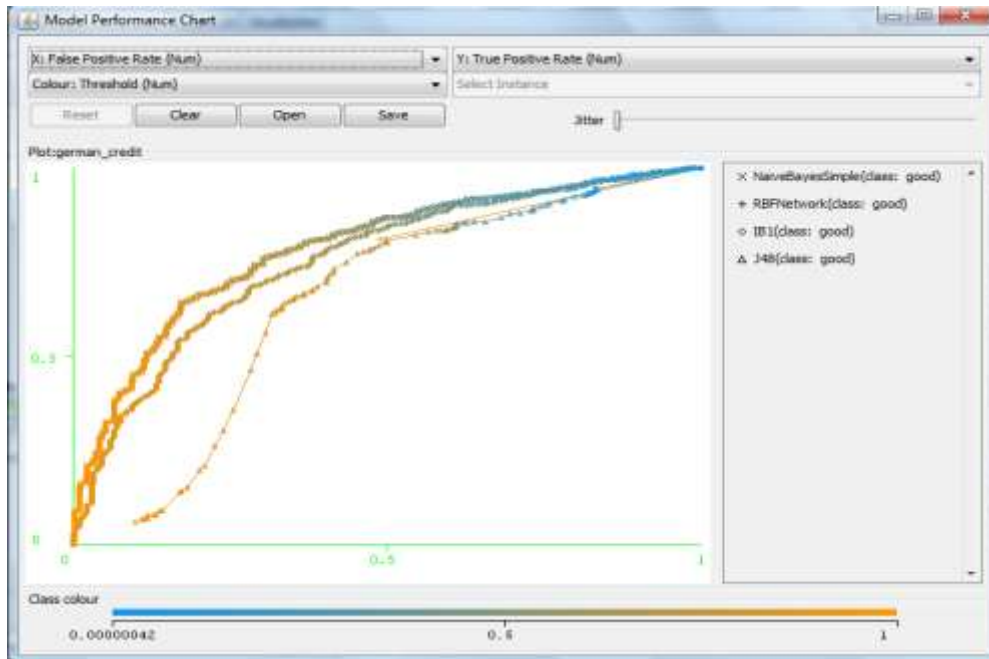Figure 6. Receiver operating characteristics (ROC) graph for the classifiers on the dataset Bank-marketing

Figure 7. Receiver operating characteristics (ROC) graph for the classifiers on the dataset Credit-g

VI. Conclusion

This paper presented decision making in enterprise computing with the data mining approach. This paper insists on decision making in enterprise computing by exploring the concepts of operations of the enterprises, management process of the enterprises, management levels in enterprises, and the functional areas of enterprises. This paper also provides the theoretical view of decision making in enterprise computing. Also, a framework for decision making in enterprise Computing is designed. Further, the experiments are conducted using the knowledge flow environment of WEKA data mining software for analyzing the performance of various classification algorithms namely probabilistic-based Naïve Bayes classifier (NB), instance-based classifier IB1, tree-based classifier J48, and the function-based classifier radial basis function network (RBFN) in terms of classification accuracy, root mean squared error (RMSE) and the receiver operating characteristics (ROC) on the different enterprise dataset collected from the UCI dataset repository and WEKA software repository. The obtained results are illustrated and discussed.

REFERENCES

[1] Grossman, Gene M., and Elhanan Helpman. Innovation and growth in the global economy. MIT press, 1993.

[2] Ahlstrom, David. "Innovation and growth: How business contributes to society."The Academy of Management Perspectives, vol. 24, pp. 11-24, 2010.

[3] Umble, Elisabeth J., Ronald R. Haft, and M. Michael Umble. "Enterprise resource planning:

Implementation procedures and critical success factors."European journal of operational research, vol.146, pp. 241-257, 2003.

[4] Scheer, August-Wilhelm. Business process engineering: reference models for industrial enterprises. Springer Science & Business Media, 2012.

[5] Carley, Kathleen M. "Computational and mathematical organization theory: Perspective and directions." Computational & mathematical organization theory, vol.1, pp. 39-56, 1995.

[6] Rouse, William B. "A theory of enterprise transformation." Systems, Man and Cybernetics, 2005 IEEE International Conference on. Vol. 1. IEEE, 2005.

[7] Courtney, James F. "Decision making and knowledge management in inquiring organizations: toward a new decision-making paradigm for DSS." Decision Support Systems vol. 31, pp.17-381, 2001.

[8] Moody, Daniel L., and Mark AR Kortink. "From enterprise models to dimensional models: a methodology for data warehouse and data mart design."DMDW. 2000.

[9] Mattison, Rob, Rick Alask, and Brigitte Illustrator-Mattison. Data warehousing: Strategies, technologies, and techniques. McGraw-Hill, Inc., 1996.

[10] D. Asir Antony Gnana Singh, S. Balamurugan, E. Jebamalar Leavline, 'An Empirical Study on Dimensionality Reduction and Improvement of Classification Accuracy Using Feature Subset Selection and Ranking', Proceedings of the IEEE International Conference on Emerging Trends in Science, Engineering and Technology, pp. 102–108, 2012.

[11] D. Asir Antony Gnana Singh, S. Balamurugan, E. Jebamalar Leavline , 'Towards Higher Accuracy in Supervised Learning and Dimensionality Reduction by Attribute Subset Selection - A Pragmatic Analysis', Proceedings of the IEEE International Conference on Advanced Communication Control and Computing Technologies, pp. 125–130, 2012.

[12] D. Asir Antony Gnana Singh, S. Balamurugan, E. Jebamalar Leavline, 'Improving performance of supervised learners using unsupervised variable selection algorithm: A novel approach', International Journal of Soft Computing , vol. 9, pp.303–307, 2014.

[13] D. Asir Antony Gnana Singh, S. Balamurugan, E. Jebamalar Leavline, 'Improving the Accuracy of the Supervised Learners using Unsupervised based Variable Selection', Asian Journal of Information Technology, vol. 13, pp.530–537, 2014

[14] D. Asir Antony Gnana Singh, E. Jebamalar Leavline, Data Mining In Network Security-Techniques & Tools: A Research Perspective. Journal of Theoretical & Applied Information Technology, vol.57, 2013.