A New Approach Using Data Envelopment Analysis for Ranking Classification Algorithms

[1]A. Bazleh, [1]P. Gholami and [2]F. Soleymani
[1]Department of Industrial Engineering, Arak Branch, Islamic Azad University, Arak, Iran
[2]Department of Mathematics, Zahedan Branch, Islamic Azad University, Zahedan, Iran

Abstract: Problem statement: A variety of methods and algorithms for classification problems have been developed recently. But the main question is that how to select an appropriate and effective classification algorithm. This has always been an important and difficult issue. **Approach:** Since the classification algorithm selection task needs to examine more than one criterion such as accuracy and computational time, it can be modeled and also ranked by Data Envelopment Analysis (DEA) technique. **Results:** In this study, 44 standard databases were modeled by 7 famous classification algorithms and we have examined them by accreditation method. **Conclusion/Recommendation:** The results indicate that Data Envelopment Analysis (DEA) is an appropriate tool for evaluating classification algorithms.

Key words: Operations research, classification algorithms, ranking, Data Envelopment Analysis (DEA), accuracy parameter, relative efficiency, accreditation method

INTRODUCTION

Classification is an extensive and also important issue in various fields, including statistics, artificial intelligence, operations research, data mining and knowledge discovery. Depends on the number of classes, classification is divided into two groups, binary and multiclass. Due to the increasing use of classification in real systems such as network intrusion detection, credit analysis, classifying websites, diagnosis of disease, a lot of algorithms and methods have been presented to classify databases with binary and specially multi classes, (Peng *et al.*, 2008)

Allwein *et al.* (2001) used binary learning algorithm based on margin and presented a framework for multiclass classification. Crammer and Singer (2001) explained algorithmic implementation of multiclass kernel based on vector machine and it had been compared with prior works. A complex numbers programming approach was presented by Loucopoulos (2001) for minimizing misclassification costs. Rennie and Rifkin (2001) compared Naïve Bayes and support vector machine with each other for text classification. Har-Peled *et al.* (2002) introduced a constraint classification method and also a Meta algorithm for multiclass classifying. Kou *et al.* (2009) presented a multiple criteria mathematical programming for data

classification in large scale. A least squares support vector machine classifier was presented by Yu *et al.* (2009) for risk analysis.

To date a variety of classification algorithms, especially for multiclass data classification, have been presented. Note that choosing an effective classifier is an important and difficult issue. The algorithm selection problem is actually a central issue in many fields including artificial intelligence, operations research and learning machine is indeed active from the viewpoint of research (Smith-Miles, 2008). Classification algorithms evaluation usually have more than one criterion, accuracy, misclassification rate and computational time (Dietterich, 1998), therefore algorithm selection can be modeled by developed Data Envelopment Analysis (DEA).

DEA decision making technique has been proposed by Charnes *et al.* (1978), which is founded based on the relative efficiency of each decision making unit in comparison with the priors and has a lot of implications in evaluating and ranking of congenial units (independent units that have equal inputs and outputs), which (Cook and Bala, 2007; Avkiran, 2006; Lin *et al.*, 2009) can be noted among them.

One of the advantages of using DEA is that, when inputs and outputs do not have the same scales (in this study accuracy and time), DEA can be used for computing efficiency and ranking of units easily. The

Corresponding Author: Fazlollah Soleymani, Department of Mathematics, Zahedan Branch, Islamic Azad University, Zahedan, Iran

other advantage is to ponder the desirability of newly presented algorithms.

The goal of this study is to use DEA in order to rank classification algorithms and choose the best through this technique.

The results of experiments in this study indicate that the presented approach is capable of ranking the classifications in various fields.

Background: Here we will have a general review on the selected classification algorithms and developed Data Envelopment Analysis (DEA).

Classification algorithms: In this study, seven classification algorithms (Logistic, SMO, Naïve bayes, cart, random forest, bagging and C4.5) have been chosen.

Logestic linear regression (Le Cessie and Van Houwelingen, 1992) is modeled the probability of occurrence of an event by a linear function of predictor variables set. SMO (Premachandra *et al*., 2011) is an algorithm for solving optimization problems, which is raised from learning machines support vector. Simple Bayes (Domingos and Pazzani, 1997) is modeled probable relation between predictor variables and class variable. Regression tree and classification (Breiman *et al*., 1984) is a greedy algorithm for multiple variables learning decision trees, which can model crisp and continuous variables. Random tree (Breiman, 2001) is a grouping classifier, which is made of some decision trees and its output is a class includes single trees classes' mode.

Breiman (1996) is a grouping Meta algorithm to improve accuracy in classification. C4.5 (Quinlan, 1993) is a decision tree algorithm which builds decision trees by division and dominance method in a reversing and up to down way.

DEA method: Developed Data Envelopment Analysis (CCR) is a methodology based upon an interesting application of linear programming, which measured the relative performance of Decision Making Units (DMUs) with different inputs and outputs. The basic efficiency measure used in DEA is the ratio of total outputs to total inputs (Charnes *et al*., 1978).

The ratio form of DEA for the mth DMU can be expressed as:

$$\text{Max} E_m \frac{\sum_{j=1}^{J} V_{jm} y_{jm}}{\sum_{j=1}^{I} u_{im} X_{im}}$$

Subject to Eq. 1:

$$0 \le \frac{\sum_{j=1}^{J} V_{jm} y_{jn}}{\sum_{j=1}^{I} u_{im} X_{in}}$$

$$\le 1 \ n = 1, 2, 3, ..., N; u_{im}, v_{jm} \ge 0 \qquad (1)$$

$$i = 1, 2, ..., I; j = 1, 2, ..., J;$$

Let us use x and y to represent inputs and outputs, respectively. For instance y_{jm} represents the amount of jth output and x_{im} represents the amount of ith input for the mth DMU. (n is the number of DMUs).

The point is to maximize the amount of Efficiency (E), so this problem can be solved by two different methods:

- Output-oriented: It means the same amount of inputs while producing more output
- Input-oriented: In this method for calculating efficiency, we should produce the same amount of output with fewer inputs

In this study, the first method is utilized and by placing denominator to 1 we reach to an input orientation CCR model as comes next:

$$\text{Max } Z = \sum_{j=1}^{J} V_{jm} y_{jm}$$

Subject to Eq. 2:

$$\sum_{i=1}^{J} u_{im} x_{im} = 1$$

$$\sum_{j=1}^{J} v_{jm} x_{jn} - \sum_{i=1}^{I} u_{im} x_{in} \le 0, n = 1, 2, ..., N; \qquad (2)$$

$$u_{im} x_{in} \ge \varepsilon \qquad i = 1, 2, ..., I \ j = 1, 2, ..., J;$$

where, Z the number of efficiency changes between 0-1 and if the efficiency of examined unit becomes 1, this unit is efficient unless it considers as an inefficient unit. In a situation that the efficiency of more than one DMU becomes 1, we use the Andersen-Peterson model for ranking DMUs. In this model, we let it have the efficiency more than 1, by omitting the constraint of that DMU, (Andersen and Petersen, 1993). Each decision making unit that can reach to a greater efficiency number, has a high level of performance among the efficient units, is something, which takes to account in this model.

Experiments: Experiments, which have been done on informational databases will be explained here.

Performance measures: There are an extensive number of performance measures for classification. Commonly used performance measures in software defect classification are accuracy, precision, recall, F-measure, the area under receiver operating characteristic (AUC) and mean absolute error UC Irvine Machine Learning Repository (Challagulla *et al*., 2005; Elish and Elish, 2008; Lessmann *et al*., 2008). Besides these popular measures, this study includes seven other classification measures. The following paragraphs briefly describe these measures.

Overall accuracy: Accuracy is the percentage of correctly classified modules (Mair *et al*., 2000). It is one the most widely used classification performance metrics Eq. 3:

$$\text{Overal accuracy} = \frac{TN + TP}{TP + FP + FN + TN} \tag{3}$$

True Positive (TP): TP is the number of correctly classified fault-prone modules. TP rate measures how well a classifier can recognize fault-prone modules. It is also called sensitivity measure Eq. 4:

$$\frac{\text{True positive rate}}{\text{sensitivity}} = \frac{TP}{TP + FN} \tag{4}$$

False Positive (FP): FP is the number of non-fault-prone modules that is misclassified as fault-prone class. FP rate measures the percentage of non-fault-prone modules that were incorrectly classified Eq. 5:

$$\text{False positive rate} = \frac{FP}{FP + TN} \tag{5}$$

True Negative (TN): TN is the number of correctly classified non-fault-prone modules. TN rate measures how well a classifier can recognize non-fault-prone modules. It is also called specificity measure Eq. 6:

$$\frac{\text{True negtive rate}}{\text{specificity}} = \frac{TN}{TN + FP} \tag{6}$$

False Negative (FN): FN is the number of fault-prone modules that is misclassified as non-fault-prone class. FN rate measures the percentage of fault-prone modules that were incorrectly classified Eq. 7:

$$\text{False negative rate} = \frac{FN}{FN + TP} \tag{7}$$

Precision: This is the number of classified fault-prone modules that actually are fault-prone modules Eq. 8:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

Recall: This is the percentage of fault-prone modules that are correctly classified:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{9}$$

F-measure: It is the harmonic mean of precision and recall. F-measure has been widely used in information retrieval (Han and Kamber, 2000) Eq. 10:

$$F - \text{measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{10}$$

AUC: ROC stands for receiver operating characteristic, which shows the tradeoff between TP rate and FP rate. AUC represents the accuracy of a classifier. The larger the area, the better the classifier.

Kappa Statistic (KapS): This is a classifier performance measure that estimates the similarity between the members of an ensemble in multi-classifiers systems (Baeza-Yates and Ribeiro-Neto, 2011) Eq. 11:

$$\text{KapS} = \frac{P(A) - P(E)}{1 - P(E)} \tag{11}$$

P(A) is the accuracy of the classifier and P(E) is the probability that agreement among classifiers is due to chance Eq. 12:

$$P(E) = \frac{\sum_{k=1}^{c} \left(\left[\sum_{j=1}^{c} \sum_{i=1}^{m} f(i,j) \right] \right) \left(\left[\sum_{j=1}^{c} \sum_{i=1}^{m} f(i,j) C(i,k) \right] \right)}{m^2} \tag{12}$$

m is the number of modules and c is the number of classes. f (i, j) is the actual probability of i module to be of class $\sum_{i=1}^{m} f(i,j)$ is the number of modules of class j. Given threshold θ, C_e (i, j) is 1 if j is the predicted class for i obtained from P(i, j); otherwise it is [0, 1].

Table 1: Accuracy and time training of algorithms using Weka

	J48		RF		Bagging		SMO		Logistic		NB		SC	
Name	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy
Abalone	1.36	21.20	3.500	22.36	4.02	24.61	24.98	25.26	35.02	27.040	0.05	23.84	8.61	26.07
Adult + Strech	0.00	100.00	0.030	100.00	0.00	100.00	0.02	100.00	0.05	100.000	0.00	100.00	0.03	100.00
Anneal	0.09	92.90	0.860	92.86	0.36	92.86	1.05	92.86	10.67	92.860	0.02	92.86	0.86	92.86
Breast Cancer	0.02	95.30	0.090	96.28	0.20	95.28	0.20	96.71	0.60	95.570	0.00	95.85	0.13	94.42
Car	0.02	92.40	0.120	92.84	0.08	92.07	1.06	93.75	1.06	93.110	0.00	85.53	0.75	97.11
Cardiotocoraphy	0.31	92.40	0.620	94.45	0.70	93.88	0.44	88.66	0.81	89.130	0.03	81.56	1.09	92.66
CMC	0.16	53.20	0.420	50.10	0.36	55.19	0.70	50.98	0.16	51.590	0.02	49.29	0.80	55.26
Contact-Lenses	0.00	83.30	0.000	70.83	0.00	66.67	0.22	70.84	0.02	66.660	0.00	70.83	0.00	79.17
Ecoli	0.02	84.20	0.060	83.33	0.06	86.31	2.01	83.63	0.25	86.310	0.00	85.42	0.06	83.93
Glass	0.00	96.70	0.060	98.60	0.05	97.66	0.94	80.84	0.14	92.990	0.00	84.11	0.05	98.13
Haber man	0.00	71.90	0.060	69.28	0.03	73.20	0.02	73.53	0.02	74.840	0.02	74.84	0.08	72.55
Hayes-Roth	0.00	81.10	0.050	82.58	0.02	81.82	0.02	56.06	0.02	55.300	0.00	76.51	0.03	79.54
Iris	0.02	96.00	0.020	95.33	0.02	94.00	0.02	96.00	0.03	96.000	0.02	96.00	0.03	95.33
Labor	0.02	73.90	0.030	87.72	0.05	85.96	0.22	89.47	0.08	92.980	0.00	89.47	0.08	78.95
Lenses	0.00	79.20	0.020	75.00	00.00	70.83	0.34	79.17	0.02	66.670	0.00	75.00	0.02	79.17
Mammographic Masses		0.03	82.400	0.17	76.89	0.11	82.93	0.05	79.29	0.030	82.62	0.01	82.52	0.34
Pima-Indians	0.04	73.80	0.190	72.53	0.20	74.61	0.21	77.34	0.05	77.210	0.01	76.30	0.29	75.13
Sanor	0.08	71.20	0.015	80.77	0.30	77.40	0.05	95.96	0.08	73.080	0.00	67.79	0.52	71.15
Transfusion	0.01	77.80	0.120	72.86	0.08	77.67	0.03	76.20	0.02	77.140	0.00	75.40	0.14	77.00
Weather	0.00	64.30	0.010	57.14	0.00	50.00	0.01	57.14	0.01	57.140	0.00	64.28	0.00	50.00
Weather. Nominal	0.00	50.00	0.000	71.43	0.00	50.00	0.00	64.28	0.00	71.430	0.00	57.14	0.00	42.86
Adult- Strech	0.00	100.00	0.000	100.00	0.00	100.00	0.01	100.00	0.00	100.000	0.00	100.00	0.00	100.00
Breast Cancer Wisconsin		0.04	94.600	0.08	96.00	0.08	95.71	0.02	96.71	0.060	95.57	0.01	96.00	0.15
Bridges 1	0.40	67.90	0.010	62.36	0.07	41.51	2.31	67.92	0.39	56.600	0.01	68.17	0.02	59.66
Bridges 2	0.01	63.20	0.050	55.66	0.04	41.51	2.20	67.92	0.37	55.660	0.00	64.16	0.03	58.17
CRX	0.05	86.10	0.150	84.35	0.13	85.07	0.78	84.93	0.17	85.220	0.00	85.22	0.75	77.68
Dermatology	0.04	95.90	0.130	96.99	0.13	95.63	1.25	97.27	0.63	98.090	0.02	97.54	0.13	94.26
Flag	0.03	57.70	0.260	59.28	0.17	35.57	2032.00	58.76	5.53	39.040	0.00	40.72	1.20	35.57
Hepatitis	0.02	58.10	0.090	58.06	0.06	60.64	0.02	66.45	0.02	63.220	0.10	70.97	0.10	61.29
Hill-Vally with Noise	0.09	50.50	1.970	55.44	6.45	52.97	0.32	55.11	0.13	80.669	0.04	48.68	4.88	52.64
Hill-Vally without Noise	0.08	50.30	2.320	59.07	7.21	61.39	0.33	58.74	0.18	90.920	0.04	50.66	6.38	58.74
Horse-Vote-84	0.01	77.30	0.080	77.34	0.05	79.46	0.03	78.25	0.03	75.530	0.00	75.53	0.11	71.20
Ionosphere	0.09	93.50	0.180	93.45	0.27	90.89	0.07	88.60	0.09	88.890	0.01	82.62	0.29	89.74
Sensor Reading 2	0.04	100.00	0.200	100.00	0.18	100.00	0.90	77.20	1.21	99.670	0.03	90.58	0.39	100.00
Sensor Reading 4	0.11	100.00	0.360	100.00	0.37	100.00	1.23	77.28	1.78	99.800	0.03	89.11	0.57	100.00
Soybean	0.01	97.90	0.010	100.00	0.02	100.00	0.37	100.00	0.00	100.000	0.00	97.87	0.03	97.87
Tae	0.02	59.60	0.030	66.89	0.03	56.39	0.18	54.30	0.02	54.300	0.00	52.98	0.05	52.32
Vowel-Context	0.21	83.10	0.580	94.04	0.92	85.86	5.07	69.60	2.33	69.800	0.02	67.07	0.95	79.70
Wine Quality-Red	0.26	62.00	0.630	68.23	0.82	64.54	1.36	58.35	2.42	59.790	0.01	55.03	1.08	60.35
Wine quality-White	1.27	58.00	2.570	67.64	3.55	62.78	19.37	52.10	8.31	53.710	0.06	44.26	5.37	59.35
WPBC	0.15	33.50	0.420	41.75	0.06	42.78	14.06	44.85	74.38	32.100	0.01	18.04	0.67	29.50
Yellow-Small + Adult	0.00	62.50	0.010	81.25	0.01	50.00	0.01	75.00	0.01	62.500	0.00	68.75	0.04	62.50
Yellow-Small	0.00	100.00	0.010	100.00	0.01	95.00	0.02	100.00	0.00	100.000	0.00	100.00	0.00	95.00
Zoo	0.00	92.10	0.030	91.08	0.03	42.57	1.18	74.20	0.16	86.140	0.01	93.07	0.20	40.59
Abalone	1.36	21.20	3.500	22.36	4.02	24.61	24.98	25.26	35.02	27.040	0.05	23.84	8.61	26.07
Adult + Strech	0.00	100.00	0.030	100.00	0.00	100.00	0.02	100.00	0.05	100.000	0.00	100.00	0.03	100.00
Anneal	0.09	92.90	0.860	92.86	0.36	92.86	1.05	92.86	10.67	92.860	0.02	92.86	0.86	92.86
Breast Cancer	0.02	95.30	0.090	96.28	0.20	95.28	0.20	96.71	0.60	95.570	0.00	95.85	0.13	94.42
Car	0.02	92.40	0.120	92.84	0.08	92.07	1.06	93.75	1.06	93.110	0.00	85.53	0.75	97.11
Cardiotocoraphy	0.31	92.40	0.620	94.45	0.70	93.88	0.44	88.66	0.81	89.130	0.03	81.56	1.09	92.66
CMC	0.16	53.20	0.420	50.10	0.36	55.19	0.70	50.98	0.16	51.590	0.02	49.29	0.80	55.26
Contact-Lenses	0.00	83.30	0.000	70.83	0.00	66.67	0.22	70.84	0.02	66.660	0.00	70.83	0.00	79.17
Ecoli	0.02	84.20	0.060	83.33	0.06	86.31	2.01	83.63	0.25	86.310	0.00	85.42	0.06	83.93
Glass	0.00	96.70	0.060	98.60	0.05	97.66	0.94	80.84	0.14	92.990	0.00	84.11	0.05	98.13
Haber man	0.00	71.90	0.060	69.28	0.03	73.20	0.02	73.53	0.02	74.840	0.02	74.84	0.08	72.55
Hayes-Roth	0.00	81.10	0.050	82.58	0.02	81.82	0.02	56.06	0.02	55.300	0.00	76.51	0.03	79.54
Iris	0.02	96.00	0.020	95.33	0.02	94.00	0.02	96.00	0.03	96.000	0.02	96.00	0.03	95.33
Labor	0.02	73.90	0.030	87.72	0.05	85.96	0.22	89.47	0.08	92.980	0.00	89.47	0.08	78.95
Lenses	0.00	79.20	0.020	75.00	0.00	70.83	0.34	79.17	0.02	66.670	0.00	75.00	0.02	79.17
Mammographic Masses		0.03	82.400	0.17	76.89	0.11	82.93	0.05	79.29	0.030	82.62	0.01	82.52	0.34
Pima-Indians	0.04	73.80	0.190	72.53	0.20	74.61	0.21	77.34	0.05	77.210	0.01	76.30	0.29	75.13
Sanor	0.08	71.20	0.015	80.77	0.30	77.40	0.05	95.96	0.08	73.080	0.00	67.79	0.52	71.15
Transfusion	0.01	77.80	0.120	72.86	0.08	77.67	0.03	76.20	0.02	77.140	0.00	75.40	0.14	77.00
Weather	0.00	64.30	0.010	57.14	0.00	50.00	0.01	57.14	0.01	57.140	0.00	64.28	0.00	50.00
Weather. Nominal	0.00	50.00	0.000	71.43	0.00	50.00	0.00	64.28	0.00	71.430	0.00	57.14	0.00	42.86
Adult- Strech	0.00	100.00	0.000	100.00	0.00	100.00	0.01	100.00	0.00	100.000	0.00	100.00	0.00	100.00
Breast Cancer Wisconsin		0.04	94.600	0.08	96.00	0.08	95.71	0.02	96.71	0.060	95.57	0.01	96.00	0.15
Bridges 1	0.40	67.90	0.010	62.36	0.07	41.51	2.31	67.92	0.39	56.600	0.01	68.17	0.02	59.66
Bridges 2	0.01	63.20	0.050	55.66	0.04	41.51	2.20	67.92	0.37	55.660	0.00	64.16	0.03	58.17
CRX	0.05	86.10	0.150	84.35	0.13	85.07	0.78	84.93	0.17	85.220	0.00	85.22	0.75	77.68
Dermatology	0.04	95.90	0.130	96.99	0.13	95.63	1.25	97.27	0.63	98.090	0.02	97.54	0.13	94.26

Table 1: Continue

Flag	0.03	57.7	0.26	59.28	0.17	35.57	2032.00	58.76	5.53	39.040	0.00	40.72	1.20	35.57
Hepatitis	0.02	58.1	0.09	58.06	0.06	60.64	0.02	66.45	0.02	63.220	0.10	70.97	0.10	61.29
Hill-Vally with Noise	0.09	50.5	1.97	55.44	6.45	52.97	0.32	55.11	0.13	80.669	0.04	48.68	4.88	52.64
Hill-Vally without Noise	0.08	50.3	2.32	59.07	7.21	61.39	0.33	58.74	0.18	90.920	0.04	50.66	6.38	58.74
Horse-Vote-84	0.01	77.3	0.08	77.34	0.05	79.46	0.03	78.25	0.03	75.530	0.00	75.53	0.11	71.20
Ionosphere	0.09	93.5	0.18	93.45	0.27	90.89	0.07	88.60	0.09	88.890	0.01	82.62	0.29	89.74
Sensor Reading 2	0.04	100.0	0.20	100.00	0.18	100.00	0.90	77.20	1.21	99.670	0.03	90.58	0.39	100.00
Sensor Reading 4	0.11	100.0	0.36	100.00	0.37	100.00	1.23	77.28	1.78	99.800	0.03	89.11	0.57	100.00
Soybean	0.01	97.9	0.01	100.00	0.02	100.00	0.37	100.00	0.00	100.000	0.00	97.87	0.03	97.87
Tae	0.02	59.6	0.03	66.89	0.03	56.39	0.18	54.30	0.02	54.300	0.02	52.98	0.05	52.32
Vowel-Context	0.21	83.1	0.58	94.04	0.92	85.86	5.07	69.60	2.33	69.800	0.02	67.07	0.95	79.70
Wine Quality-Red	0.26	62.0	0.63	68.23	0.82	64.54	1.36	58.35	2.42	59.790	0.01	55.03	1.08	60.35
Wine quality-White	1.27	58.0	2.57	67.64	3.55	62.78	19.37	52.10	8.31	53.710	0.06	44.26	5.37	59.35
WPBC	0.15	33.5	0.42	41.75	0.06	42.78	14.06	44.85	74.38	32.100	0.01	18.04	0.67	29.50
Yellow-Small +Adult	0.00	62.5	0.01	81.25	0.01	50.00	0.01	75.00	0.01	62.500	0.00	68.75	0.04	62.50
Yellow-Small	0.00	100.0	0.01	100.00	0.01	95.00	0.02	100.00	0.00	100.000	0.00	100.00	0.00	95.00
Zoo	0.00	92.1	0.03	91.08	0.03	42.57	1.18	74.20	0.16	86.140	0.01	93.07	0.20	40.59

Mean Absolute Error (MAE): This measures how much the predictions deviate from the true probability. $P(i, j)$ is the estimated probability of i module to be of class j taking values in [0,1] (Kuncheva, 2004; Ferri *et al*., 2009) Eq. 13:

$$MAE = \frac{\sum_{j=1}^{C}\sum_{j=1}^{m}|f(i,j)|) - p(i,j)|}{m.c} \qquad (13)$$

- Training time: the time needed to train a classification algorithm or ensemble method
- Test time: the time needed to test a classification algorithm or ensemble method

MATERIALS AND METHODS

Data sources: We have gathered approximately 300 binary and multi class databases and 44 among them in various domains and fields have been chosen. Choosing this amounts of data bases has accomplished based on the Morgan table which says if the statistical society was 300, it would be enough to choose 44 samples for statistical analysis.

Proposed approach: Experiments are done based on the following procedure: Input: 44 chosen data bases, Output: ranking of classifiers.

Step 1: After preparing data bases using Weka, we proceed with the data analysis and calculate the accuracy and constructing time of model in each algorithm by 7 famous algorithms in classification (Table 1)

Step 2: Since we are planning to gain the efficiency of every classification algorithms, using DEA CCR output oriented, each of these algorithms is considered as a Decision Making Unit (DMU), we assume the amount of accuracy of every algorithm is pondered as its output, in analysis of every data bases with method. Moreover in some analysis the accuracy of

classification is just the most effectual factor in the best algorithm selection and in some other analysis algorithm performance time in addition to accuracy is effective too, to determine the efficiency, two methods are presented which result to rank the algorithms.

Step 3: Ranking algorithms by considering accuracy parameter: to compute the efficiency of every DMU, let us consider 1 as input of DMU and the accuracy of every algorithm as output. This analysis is done by DEAP software and if the efficiency of several algorithms becomes 1, using DEA Andersen-Peterson model, their efficiency is found by lingo software.

Step 4: Ranking algorithms by considering accuracy and performance time parameters: to compute the efficiency of every DMU in this condition, it is enough to consider learning time as input (time is cost attribute) and the accuracy of classification as output (accuracy is benefit attribute) of that algorithm and do the same as previous step.

Step 5: Considering obtained efficiency, we rank algorithms.

RESULTS AND DISCUSION

According to the analysis, two tables due to the steps 3 and 4 are obtained: Table 2 represents the efficiency of every algorithm just by considering the accuracy parameter.

Since the CCR model efficiency of all algorithms equals 1, therefore all algorithms are desirable algorithms and Andersen-Peterson model is used to rank them. According to the efficiency number which is obtained by Andersen-Peterson model, final ranking of algorithms is as follows:

Logistic> Random Forest> SMO >
C4.5 > Naïve Bayse> Cart> Bagging

Har-Peled, S., D. Roth and D. Zimak, 2002. Constraint classification: A new approach to multiclass classification. Lecture Notes Comput. Sci., 2533: 267-280. DOI: 10.1007/3-540-36169-3_29

Kou, G., Y. Peng, Z. Chen and Y. Shi, 2009. Multiple criteria mathematical programming for multi-class classification and application in network intrusion detection. Inform. Sci., 179: 371-381. DOI: 10.1016/j.ins.2008.10.025

Kuncheva, L.I., 2004. Combining Pattern Classifiers: Methods and Algorithms. 1st Edn., Wiley, Hoboken, NJ., ISBN: 0471210781, pp: 350.

Le Cessie, S. and J.C. Van Houwelingen, 1992. Ridge estimators in logistic regression. Applied Stat., 41: 191-201.

Lessmann, S., B. Baesens, C. Mues and S. Pietsch, 2008. Benchmarking classification models for software defect prediction: A proposed framework and novel findings. IEEE Trans. Software Eng., 34: 485-496. DOI: 10.1109/TSE.2008.35

Lin, T.T., C.C. Lee and T.F Chiu, 2009. Application of DEA in analyzing a bank's operating performance. Expert Syst. Appli., 36: 8883-8891. DOI: 10.1016/j.eswa.2008.11.018

Loucopoulos, C., 2001. Three-group classification with unequal misclassification costs: A mathematical programming approach. Omega, 29: 291-297. DOI: 10.1016/S0305-0483(01)00023-8

Mair, C., G. Kadoda, M. Leflel, K. Phapl and C. Schofield *et al.*, 2000. An investigation of machine learning based prediction systems. J. Syst. Software, 53: 23-29. DOI: 10.1016/S0164-1212(00)00005-4

Peng, Y., .G. Kou, Y. Shi and Z. Chen, 2008. A descriptive framework for the field of data mining and knowledge discovery. Int. J. Inform. Technol. Decision Mak., 7: 639-689.

Premachandra, I.M., Y. Chen and J. Watson, 2011. DEA as a tool for predicting corporate failure and success: A case of bankruptcy assessment. Omega, 39: 620-626. DOI: 10.1016/j.omega.2011.01.002

Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. 1st Edn., Morgan Kaufmann, San Mateo, California, ISBN: 1558602380, pp: 302.

Rennie, J.D.M. and R. Refkin, 2001. Improving multiclass text classification with the support vector machine. Massachusetts Institute of Technology as AI Memo, pp: 1-14.

Smith-Miles, K.A., 2008. Cross-disciplinary perspectives on meta-learning for algorithm selection. ACM Comput. Surv., 41: 1-25. DOI: 10.1145/1456650.1456656

Yu, L., S. Wang and J. Cao, 2009. A modified least squares support vector machine classifier with application to credit risk analysis. Int. J. Inform. Technol. Decision Mak., 8: 697-710.