Asian Journal of Information Technology 13 (9): 530-537, 2014
ISSN: 1682-3915

Improving the Accuracy of the Supervised Learners using Unsupervised based Variable Selection

[1]Danasingh Asir Antony Gnana Singh,
[3]Subramanian Appavu Alias Balamurugan and [2]Epiphany Jebamalar Leavline
[1]Department of Computer Science and Engineering,
[2]Department of Electronics and Communication Engineering,
Bharathidasan Institute of Technology, Anna University,
Tiruchirappalli, 620 024 Tamil Nadu, India
[3]Department of Information Technology,
KLN College of Information Technology, Sivagangai, India

Abstract: Decision making is practiced in every moment of individual life. The correct decision leads the humanity in right path towards obtaining prosperity and secures the life from the losses. The good decision making mainly relay accurate prediction. The prediction is practiced in all the emerging fields to make decision. In medical field, prediction supports to diagnose the disease in order to prescribe the correct medicine. In finance, prediction assists to predict the future demand and supply in order to satisfy the customer needs. In management, prediction helps to predict the profit and losses to lead the organization with maximum profitable. In engineering, the prediction supports to conduct the research and development activities. In management the prediction facilitates to predict the natural calamities to save and secure the life form the calamity. This prediction is carried out by the supervised learners known as supervised learners. The accuracy of these learners is determined by the significant variables presents in training dataset to train the learners. This study propose a novel algorithm namely Clustering with Variable Ranking and Selection algorithm (CVRS) to select most significant variable from the training dataset and remove the redundant and irrelevant variables form the training dataset. The performance of the proposed algorithm is compared with the six existing algorithms by four supervised learners. This proposed algorithm produces higher accuracy compared to other algorithms compared for the supervised learners.

Key words: Variable selection, supervised learner, unsupervised learners, clustering, prediction, decision making

INTRODUCTION

In this world, people are very eager to guess their futures in order to make decisions and plans for securing their selves form the losses. Even the philosophy says "prevention is better than cure". This imparts the preventive measures. The prediction identifies the unknown data or event to guide the prevention by making decision and plans. This prediction is carried out by the supervised learner is known as supervised learner builds the predictive model by training dataset (Pan and Yang, 2010). This model employs to predict the unlabeled data or unknown parameter in many computer aided applications (Rahman and Hasan, 2011). The accuracy of the predictiveodel mainly depends on the dataset which used to train the model. In this digital era, various researches try to achieve higher accuracy in supervised learners through preprocessing the training dataset in terms of variable selection. Three types of variables present in the training dataset namely redundant, relevant and irrelevant. The process of selecting the relevant variable to build the predictive model is known as variable selection. Three methods are followed for variable selection namely Filter, Wrapper, Embedded and Hybrid Methods (Song *et al*., 2013). Filter method adopts any one of the statistical measure to select with the selection criteria to select the relevant variable for the training dataset. This filter approach is suitable for any one of the supervised learners since it posses more generality (Artur and Mario, 2012; Wu *et al*., 2012). The Wrapper Method adopts the supervised learner for performance evaluation to select the relevant variable for the training dataset hence its performance depends on the supervised learner adopted since it lags in more generality (Bermejo *et al*., 2012).

The Embedded Method uses a part of the training process of the supervised learner to perform the variable selection since not poses more generality (Hou *et al*.,

Corresponding Author: Danasingh Asir Antony Gnana Singh, Department of Computer Science and Engineering, Bharathidasan Institute of Technology, Anna University, Tiruchirappalli, 620 024 Tamil Nadu, India

2013). The Hybrid Method combines functionalities of the Wrapper and Filter Method (Song *et al*., 2013; Srivastava *et al*., 2013).

Many researches focus only on removing the irrelevant attribute form the training dataset and not consider about removing irrelevant variable. This decreases the accuracy of the supervised learners. This study proposes a filter based algorithm namely clustering with Variable Ranking and Selection algorithm (CVRS). This algorithm removes the redundant variables from the training dataset by the clustering technique and selects the relevant and removes the irrelevant variables by the statistical measure. This algorithm clusters the variables by k-mean clustering technique and applies the Chi-squared measure on the each clustered variables and ranks the variables based on the Chi-squared value with threshold function to select significant variable form the each clustered variables. Thus, the combination of this selected significant variables are considered as a selected most significant variable.

Literature review: This study discusses various types of variable selection, cluster and supervised learning algorithms as a part of the related works of this proposed algorithm.

Variable Selection algorithms: The Variable Selection algorithm selects the most significant variables from the dataset using the following three techniques: ranking based, subset based and unsupervised based. In the ranking based technique the individual variables 'f_i' are ranked by applying any one of the mathematical measures such as information gain, gain ratio, Chi-squared, etc., on Training Dataset (TD). The ranked variables are selected as significant variables for Learning algorithm by a threshold value 'T_V' calculated by the threshold function. In subset based technique, the variables of the training datasets are separated into maximum number of possible variable subsets 'S' and each subset is evaluated by an evaluation criteria to identify the significance of the variable subsets for selecting the variables. In the unsupervised based technique, the cluster analysis is carried out to identify the significant variables form the training dataset (Wei and Billings, 2007; Vinh and Bailey, 2013; Gheyas and Smith, 2010; Yu and Liu, 2004; Dash *et al*., 2000).

Variable selection based on correlation (CRR): In this variable subset selection, the entire variable set F = $\{f_1, f_2,...f_x\}$ of a Training Dataset (TD) is sub divided into variable subsets 'FS_i'. Then, two types of the correlation measures are calculated on each variable subset 'FS_i'. One is the variable-variable correlation that is the correlation measure among the variables present in a variable subset 'FS_i', another one is variable-class correlation that is the correlation measure between the individual variable and the class value of a the training dataset. These two correlation measures are computed for all the variable subsets of the training dataset. The significant variable subset is identified based on the comparison between the variable-variable correlation and variable-class correlation. If the variable-class correlation value is higher than the variable-variable correlation value, the corresponding variable subset is selected as a significant variable subset of the training dataset (Hall, 1999).

Variable selection based on Chi-squared (CQ): This is a ranking based variable selection technique. The Chi-squared statistical measure is applied on the Training Dataset (TD) to identify the significant level of each variable presents in the training dataset. The Chi-squared Value (CV) is computed the sum of ratio of the difference between the observed ($o_{i,j}$) and expected (e_{ij}) frequencies of the variables f_i ,f_j to the expected frequencies (e_{ij}) of the variables f_i ,f_j of the possible instance value combinations of the variables (Peng *et al*., 2005).

Variable selection based on Information Gain (IG): In this variable selection technique, the information gain measure is applied on the training dataset to identify the significant variables based on information gain value of the individual variables (Wei and Billings, 2007) in terms of entropy. The entropy value of each variable of the Training Dataset (TD) is calculated and ranked based on the information gain value (Uguz, 2011).

RELIEFF: This variable selection technique selects the significant variables from the training dataset 'TD' based on the weighted probabilistic function w(f) with nearest neighbor principle. If the nearest neighbors of a instance T belong to the same class, it is termed as 'nearest hit' and if the nearest neighbors of a instance T belong to different class, it is termed as 'nearest miss'. The probabilistic weight function value w(f) is calculated based on the distinct value of the variable 'f' that nearest neighbor instance T belongs to different class and nearest neighbor instance T belongs to the same class of given instance T (Robnik-Sikonja and Kononenko, 2003; Sun *et al*., 2010; Peng *et al*., 2013).

Variable selection based on Gain Ratio (GR): In this Variable Selection Method, the information gain ratio is calculated for each variable GR(f) of the training dataset 'TD' to identify the significant variable based on the information present in the variables of the 'TD' (Uguz, 2011).

Variable selection based on Symmetric Uncertainty (SU): This technique uses the correlation measure to select the significant variable form the Training Dataset (TD). In addition to that the Symmetric Uncertainty (SU) is calculated using the entropy measure to identify the similarity between the two variables f_i and f_j (Liu *et al*., 2009; Lee, 2009).

Variable selection based on Unsupervised Learning algorithm: Unsupervised learning is formally known as Clustering algorithm. This algorithm groups similar object with respect to the given criteria like density, distance, etc. The objects present in a group are highly similar than the outliers. This is technique is applied for selecting the significant variables from the training dataset. Each Unsupervised algorithm has its own advantages and disadvantages that determine the application of each algorithm. This study utilizes K-Means (KM) clustering technique to select the significant variables by identifying the independent variables in order to remove the redundant variables from a Training Dataset (TD) (Mitra *et al*., 2002; Handl and Knowles, 2006; Liu and Yu, 2005).

Supervised Learning algorithm: The Supervised Learning algorithm builds the Predictive Model (PM) by learning the Training Dataset (TD) to predict the unlabeled instance T. This model can be built by various supervised learning techniques such as tree based, probabilistic and rule based. This study uses the supervised learners namely Naive Bayes (NB), instance based IB1 and C4.5/J48 supervised learners to evaluate and compare the performance of the proposed Variable Selection algorithm in terms of prediction accuracy and time taken to build the prediction model with the existing algorithms (Garcia *et al*., 2013; Balamurugan and Rajaram, 2009; Mangai *et al*., 2012).

Naive Bayes (NB) supervised learner: This supervised learning works based on the Bayesian theory. The probabilistic function is applied on Training Dataset (TD) that contains the variables $F = \{f_1, f_2,..., f_x\}$, instance $T = \{t_1, t_2,..., t_y\}$ and classes $C = \{c_1, c_2,..., c_z\}$. The unlabeled instance 'T' is predicted to identify its class C_i with the probabilistic condition $P(C_k|T)>P(C_w|T)$ where $k\neq w$ (Hung and Hsu, 2002).

Decision tree based C4.5/J48 supervised learner: This tree based supervised learning constructs the Predictive Model using decision tree. Basically the statistical tool information gain is used to learn the dataset and construct the decision tree. Information gain is computed for each attribute present in the Training Dataset (TD). The variable with higher information value is considered as the route node and the data set is divided into further levels. The information value is computed for all the nodes and this process is iterated until a single class value is obtained in all the nodes (Ruggieri, 2002; Polat and Gunes, 2009).

IB1: This supervised learner utilizes the nearest neighbor principle to measure the similarity among the objects to predict the unlabeled instance T. The Euclidean distance measure is used to compute the distance between various instances present in the training dataset (Cheng and Hullermeier, 2009).

Unsupervised learning with Ranking algorithm (CVRS): This algorithm follows a sequence of steps to select the significant variables from the training dataset. Initially, the training dataset is transposed, the variables are clustered and the variables in each cluster are ranked based on the Chi-squared value. Then, the threshold value is computed to select highly significant variables. All the selected variables from different clusters are combined together as candidate variables from the training dataset.

Algorithm (CVRS):

Require: Dataset 'D' contains variables $V = \{v_1, v_2,..., v_l\}$, instances $I = \{i_1, i_2,..., i_m\}$ and classes $C = \{c_1, c_2,.., c_n\}$

Ensure:

Significant Variable Subset (SVS) = $\{v_1, v_2,..., v_i\} \in V$

Steps:

(1) Initiate;
(2) Reverse (D)
(3) Return←Reversed dataset $RD = D^T$ }
// transpose the instances (I) into variables (V)
(4) K_Mean(RD)
(5) {Return←Clustered variables $CV = \{CV_1, CV_2.., CV_n\}$}//$CV \in V$
(6) for (X = 1, X≤n, X++)
(7) {C_squared(CV_i,C)
(8) {Return←Ranked variables for the cluster $RV_{K=1..n}$ with Chi-squared weight W_C} } // calculating the Chi-square weight for all the variable V_i present in the each clusters
(9) Cut_off (Max_W_C , Min_W_C) // Compute Cut-off Chi-squared weight α
//Max_W_C-Maximum weight of Chi-squared value
// Min_W_C –Minimum weight of Chi-squared value
(10) {Return←α} // α–Cut-off Chi-squared weight
(11) for (Y = 1,Y≤n, Y++)
(12) {Variable_Seleation (RV_K, $W_{C\text{-}K}$, α) // $W_{C\text{-}K}$- Corresponding Chi-squared weight of ranked variable RV_K
(13) {Return←$CRSV_M$}}//$CRSV_M$-Cut-off ranked selected variable CRSV by the Cut-off weight-α
(14) for (Z = 1, Z≤n, Z++)
(15) {Return SVS ={Union($CRSV_M$)}//Combining $CRSV_M$ as a Significant Variable Subset SVS
(16) Finish;

Phase 1: In this phase, the CVRS algorithm receive the training dataset 'D' as input with the variable $V = \{v_1,$

$v_2,..., v_l\}$, I = $\{i_1, i_2,..., i_m\}$ and classes C = $\{c_1, c_2,..., c_n\}$. Then, the class C is removed. The function Reverse (.) transpose the variables 'V' into instances 'I' and returns transposed dataset 'RD' for grouping the variables v_i.

Phase 2: In this phase, the k-Means clustering technique is used (Hall *et al*., 2009) to group the variables using the function K_Mean(.). It receives the variables V = $\{v_1, v_2,..., v_l\}$ with its corresponding instance vector I = $\{i_1, i_2,..., i_m\}$ to cluster the variable in to 'k' number of clustered variables CV = $\{CV_1, CV_2,..., CV_n\}$ with minimizing the cluster sum of squares is computed as shown in Eq. 1 where μ is the mean points in CV:

$$\arg\min_{CV}\sum_{i=1}^{k}\sum_{v_i\in cv_i}\left\|v_j-\mu_i\right\|^2 \quad (1)$$

Phase 3: In this phase the most significant variables are extracted from the clustered variable subset by the C_squared(.) function. It receives the clustered variable set CV_N with the class variable C and computes the Chi-squared value as the weight W_C for all the variables with the Eq. 2:

$$W_c=\sum_{i=1}^{m}\sum_{j=1}^{n}\frac{o_{ij}-e_{ij}}{e_{ij}} \quad (2)$$

The variable v_i contains distinct instance values tv_1, $tv_2,..., tv_m$ and $f_j$contains distinct instance values iv_1, $iv_2,..., iv_n$. Then, the Chi-square value 'W_C' is computed for the variables V_i, V_j. o_{ij} is the observed frequency and e_{ij} is the expected frequency of possible instance values. To identify the significant variables, the variables are ranked by their Chi-squared value W_C and the ranked grouped variables set RV_K is obtained. The Cut_off(.) function receives the Max_W_C, Min_W_C values and returns the break off threshold value α as shown in Eq. 3. The function Variable_Seleation (.) recognizes the ranked grouped variable set RV_K, W_{C-K} with α and cut off the top ranked RV up to the value α and returns them as break off variable subset CRSV:

$$\alpha=H\left(\frac{\lambda-\beta}{2}+\beta\right) \quad (3)$$

Where:

α = Break off-threshold value
H = Threshold function
λ = Maximum Chi-squared value (Max_W_C) of the variable
β = Minimum Chi-squared value (Min_W_C) of the variable

Table 1: Datasets

Dataset	Variables	Instances	Classes
Breast cancer	9	286	2
Contact lenses	24	5	5
Credit (g)	20	1000	2
Diabetes	8	768	2
Glass	9	214	7
Iris 2D	2	150	3
Labor	16	57	2
Segment challenge	19	1500	7
Vote	435	17	3
Weather numeric	14	5	2
Dermatology	34	366	3
E. coli	7	336	8
Cylinder bands	39	540	2
Anneal	39	898	6
Car	6	1728	4

Phase 5: This phase combines the significant variables thresholded from the clustered variable by the function Union(.) and produces the selected significant variable subset SVS as the candidate selected variable.

MATERIALS AND METHODS

Experimental setup: Totally 15 datasets were used to conduct the experiment with number of variables ranging from 2-435, number of instances from 5-1728 and number of class labels from 2-8 as listed in the Table 1. These datasets are taken from the weka tool (Hall *et al*., 2009) and UCI repository (Bache and Lichman, 2013). The performance of the proposed algorithm FSCVRS is analyzed and compared with the other Variable Selection algorithms CRR, CQ, GR, RELIEFF, SU and IG with NB, J48 and IB1 supervised learners (Appendix I).

RESULTS AND DISCUSSION

Experimental procedure: This experiment is conducted with the fifteen well known publically available datasets as seen in the Table 1 by six variable selection algorithms namely CRR, CQ, GR, RELIEFF, SU and IG. Initially, the datasets are fed as an input to the Variable Selection algorithms and the selected variables are obtained as output from the Variable Selection algorithms. These selected variables are then given to the NB, J48 and IB1 supervised learners and the predictive accuracy and the time taken to build model are calculated with the 10 fold cross validation test mode (Appendix II).

The performance of the proposed CVRS algorithm is analyzed in terms of predictive accuracy, time to build the Predictive Model and number of variables reduced. Figure 1 expresses that the overall performance of CVRS in producing the predictive accuracy is better than all other Variable Selection algorithms compared. The second and third position retained by the CRR and CQ, respectively.

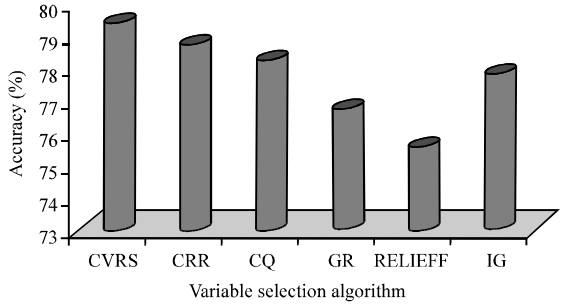

Fig. 1: Comparison of overall prediction accuracy with respective Variable Selection algorithm

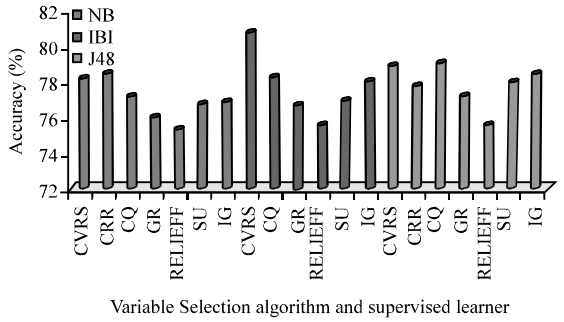

Fig. 2: Comparison of prediction accuracy with respect to the Variable Selection algorithm

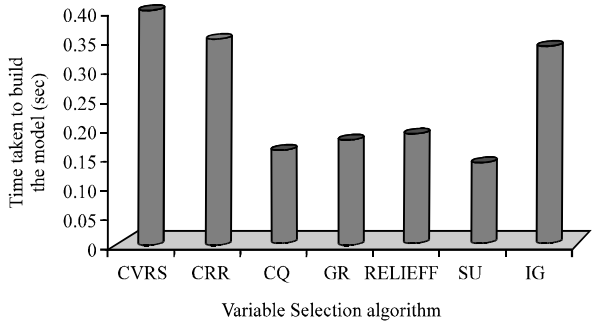

Fig. 3: Comparison of overall time taken to build the predictive model in seconds with respect to the Variable Selection algorithm

Figure 2 shows that the proposed CVRS achieves better accuracy than other algorithms compared for IB1 supervised learners. For NB and J48 supervised learner the CRR and CQ achieves better results, respectively than all other Feature Selection algorithm compared.

From Fig. 3, it is evident that shows that CVRS takes much time to build the predictive model compared to all other algorithms compared and it is observed that the GR and SU require less time to build the model compared to other algorithms compared. From Fig. 4, the GR takes lesser time to build model for NB supervised learner. The

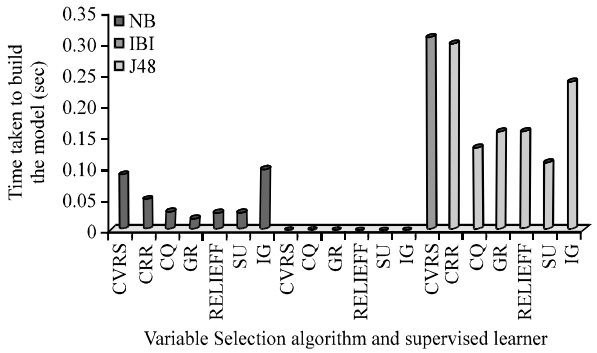

Fig. 4: Comparison of time taken to build the Predictive Model in seconds with respect to the Variable Selection algorithms

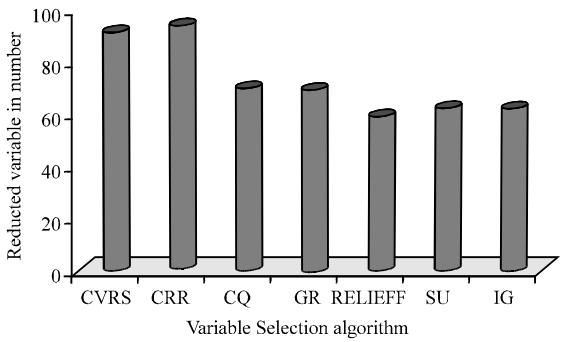

Fig. 5: Comparison of number of variable reduction with respect to the Variable Selection algorithm

SU consume lesser time to build the Predictive Model for IB1 supervised learner. Figure 5 exhibits that RELIEFF reduces the number of variables significantly than other algorithms compared and IG reduces the least number of variables (Appendix 3).

CONCLUSION

This study proposed a Variable Selection algorithm namely Clustering with Variable Ranking and Selection algorithm (CVRS). Performance of this algorithm is analyzed in terms of accuracy produced for supervised learner, time to build Predictive Model and variable reduction. The performance of this algorithm is compared with other Variable Selection algorithms namely correlation based CRR, Chi-squared based CQ, Gain ratio based GR, RELIEFF, symmetric uncertainty based SU and information gain based IG algorithms with NB, J48, IB1 supervised learners. The CVRS achieves better prediction accuracy than other Variables Selection algorithms and achieves higher accuracy for IB1 supervised learners compared to other Variable Selection algorithms. The

CVRS is considerably good in reducing the time to build model for NB compared to IG. CVRS is considerably good in reducing the time to build model for IB1 supervised learner. In future, this research can be extended with other statistical measures for ranking and other clustering techniques.

APPENDICES

Appendix I: Number of features reduced by the Variable Selection algorithm respect to the dataset

Datasets	CVRS	CRR	CQ	GR	RELIEFF	SU	IG
Breast cancer	3	3	4	3	3	3	4
Contact lenses	3	1	2	3	1	2	2
Credit-g	6	3	1	1	1	1	1
Diabetes	5	4	1	3	2	3	1
Glass	5	8	6	4	1	6	7
Iris 2D	2	2	2	1	1	2	1
Labor	4	7	3	3	8	3	3
Segment challenge	11	6	11	11	11	11	11
Vote	6	17	4	3	1	3	3
Weather numeric	2	2	1	1	1	1	1
Dermatology	20	19	20	17	12	16	15
E. coli	5	6	5	5	3	4	3
Cylinder bands	8	6	2	6	5	2	2
Anneal	7	9	6	6	5	3	6
Car	4	1	2	2	4	2	2

Appendix II: Accuracy produced by the corresponding Variable Selection algorithm for respective unsupervised learners

	CVRS			CRR			CQ			GR		
Datasets	NB	J48	IB1	NB	J48	IB1	NB	J48	IB1	NB	J48	IB1
Breast cancer	73.77	73.07	67.83	73.07	75.52	67.48	72.72	73.07	66.78	73.77	73.07	67.83
Contact lenses	75.00	87.50	75.00	70.83	70.83	66.66	87.50	87.50	75.00	83.33	83.33	83.33
Credit-g	73.70	73.30	66.90	74.40	70.50	64.60	67.60	70.00	65.90	67.60	70.00	65.90
Diabetes	75.13	74.08	69.27	77.47	74.86	68.35	75.00	73.04	65.23	73.43	72.52	69.40
Glass	55.14	65.88	77.10	47.66	67.75	71.02	50.00	68.69	69.62	45.32	66.35	62.14
Iris 2D	96.00	96.00	96.66	96.00	96.00	96.66	96.00	96.00	96.66	95.33	94.66	90.00
Labor	85.96	82.45	84.21	91.22	77.19	84.21	84.21	80.70	80.70	84.21	80.70	80.70
Segment challenge	78.93	94.86	96.13	81.73	95.33	95.06	78.93	94.86	96.13	78.93	94.86	96.13
Vote	91.03	95.63	94.71	100.00	93.75	93.75	92.87	95.63	93.79	94.71	95.63	93.10
Weather numeric	57.14	42.85	78.57	57.14	42.85	78.57	50.00	50.00	64.28	50.00	50.00	64.28
Dermatology	87.15	82.24	84.97	97.81	94.80	95.90	87.15	82.24	84.97	85.79	81.69	86.61
E. coli	86.01	83.03	79.16	85.41	84.22	80.05	85.41	82.73	80.35	78.86	79.76	76.19
Cylinder bands	73.14	57.77	70.92	68.14	56.66	71.85	65.37	57.77	64.44	68.14	56.66	71.85
Anneal	89.08	98.21	98.44	87.19	96.88	97.55	89.08	98.21	98.44	84.63	84.63	71.38
Car	76.90	78.29	72.97	70.02	70.02	66.84	76.85	76.56	73.49	76.85	76.56	73.49

	RELIEFF			SU			IG		
Datasets	NB	J48	IB1	NB	J48	IB1	NB	J48	IB1
Breast cancer	71.67	72.72	62.93	73.77	73.07	67.83	72.72	73.07	66.78
Contact lenses	70.83	70.83	66.66	87.50	87.50	75.00	87.50	87.50	75.00
Credit -g	67.60	70.00	65.90	67.60	70.00	65.90	67.60	70.00	65.90
Diabetes	75.00	73.04	65.23	76.43	74.60	70.18	75.00	73.04	65.23
Glass	35.51	45.79	34.57	50.00	68.69	69.62	49.06	69.62	77.57
Iris 2D	96.00	96.00	96.66	95.33	94.66	90.00	96.66	94.00	91.33
Labor	87.71	78.94	85.96	84.21	80.70	80.70	84.21	80.70	80.70
Segment challenge	78.93	94.86	96.13	78.93	94.86	96.13	78.93	94.86	96.13
Vote	95.63	95.63	91.72	94.71	95.63	93.10	94.71	95.63	93.10
Weather numeric	50.00	50.00	64.28	50.00	50.00	64.28	50.00	50.00	64.28
Dermatology	80.60	73.77	79.23	86.88	81.14	84.97	86.88	81.42	84.69
E. coli	79.16	76.48	73.80	78.57	77.38	73.51	79.16	76.48	73.80
Cylinder bands	73.70	57.77	76.48	65.37	57.77	64.44	65.37	57.77	64.44
Anneal	85.96	92.31	90.20	86.52	88.19	86.97	90.08	96.99	98.10
Car	83.10	86.34	85.59	76.85	76.56	73.49	76.85	76.56	73.49

Appendix III: Time taken to build the predictive model by the corresponding Variable Selection algorithm for respective unsupervised learners

	CVRS			CRR			CQ			GR		
Datasets	NB	J48	IB1	NB	J48	IB1	NB	J48	IB1	NB	J48	IB1
Breast cancer	0	0	0	0	0.03	0	0	0	0	0	0	0
Contact lenses	0	0	0	0	0	0	0	0	0	0	0	0
Credit-g	0.01	0.05	0	0.02	0	0	0	0	0	0	0	0

Appendix III: Continue

	CVRS			CRR			CQ			GR		
Datasets	NB	J48	IB1	NB	J48	IB1	NB	J48	IB1	NB	J48	IB1
Diabetes	0.01	0.04	0	0	0.03	0	0	0	0	0	0.03	0
Glass	0	0.03	0	0	0.03	0	0	0	0	0	0	0
Iris 2D	0	0	0	0	0	0	0	0	0	0	0	0
Labor	0	0	0	0	0.03	0	0	0	0	0	0	0
Segment challenge	0.06	0.08	0	0.02	0.08	0	0.02	0.08	0	0.02	0.08	0
Vote	0	0.01	0	0	0	0	0	0	0	0	0	0
Weather numeric	0	0	0	0	0	0	0	0	0	0	0	0
Dermatology	0	0.01	0	0	0	0	0	0	0	0	0	0
E. coli	0	0.03	0	0	0.03	0	0	0.02	0	0	0.01	0
Cylinder bands	0	0.01	0	0	0.02	0	0	0	0	0	0.01	0
Anneal	0.01	0.03	0	0.01	0.05	0	0.01	0.03	0	0	0.02	0
Car	0	0.01	0	0	0	0	0	0	0	0	0.01	0

	RELIEFF			SU			IG		
Dataset	NB	J48	IB1	NB	J48	IB1	NB	J48	IB1
Breast cancer	0	0	0	0	0	0	0	0.02	0
Contact lenses	0	0	0	0	0	0	0	0	0
Credit-g	0	0	0	0	0	0	0	0	0
Diabete	0	0	0	0	0	0	0	0	0
Glass	0	0	0	0	0	0	0	0.02	0
Iris 2D	0	0.03	0	0	0	0	0	0	0
Labor	0	0	0	0	0	0	0	0.02	0
Segment challenge	0.03	0.09	0	0.02	0.08	0	0.03	0.11	0
Vote	0	0	0	0	0	0	0	0	0
Weather numeric	0	0	0	0	0	0	0	0	0
Dermatology	0	0	0	0	0.02	0	0.02	0.03	0
E. coli	0	0.02	0	0	0	0	0.02	0.02	0
Cylinder bands	0	0	0	0	0	0	0.02	0	0
Anneal	0	0.01	0	0.01	0.01	0	0.01	0.02	0
Car	0	0.01	0	0	0	0	0	0	0

REFERENCES

Artur, J.F. and A.T. Mario, 2012. Efficient variable selection filters for high-dimensional data. Pattern Recognit. Lett., 33: 1794-1804.

Bache, K. and M. Lichman, 2013. UCI machine learning repository. University of California, School of Information and Computer Science, Irvine, CA., USA.

Balamurugan, S.A.A. and R. Rajaram, 2009. Effective and efficient feature selection for large-scale data using Bayes' theorem. Int. J. Automation Comput., 6: 62-71.

Bermejo, P., L. de la Ossa, J.A. Gamez and J.M. Puerta, 2012. Fast wrapper feature subset selection in high-dimensional datasets by means of filter re-ranking. Knowledge-Based Syst., 25: 35-44.

Cheng, W. and E. Hullermeier, 2009. Combining instance-based learning and logistic regression for multilabel classification. Mach. Learn., 76: 211-225.

Dash, M., H. Liu and H. Motoda, 2000. Consistency based variable selection. Proceedings of the 4th Pacific Asia Conference on Knowledge Discovery and Data Mining, April 18-20, 2000, Kyoto, Japan, pp: 98-109.

Garcia, S., J. Luengo, J.A. Saez, V. Lopez and F. Herrera, 2013. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. IEEE Trans. Knowledge Data Eng., 25: 734-750.

Gheyas, I.A. and L.S. Smith, 2010. Variable subset selection in large dimensionality domains. Pattern Recognit., 43: 5-13.

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann and I.H. Witten, 2009. The WEKA data mining software: An update. SIGKDD Explorations Newslett., 11: 10-18.

Hall, M.A., 1999. Correlation-based variable selection for machine learning. Ph.D. Thesis, The University of Waikato, New Zealand.

Handl, J. and J. Knowles, 2006. Feature subset selection in unsupervised learning via multiobjective optimization. Int. J. Comput. Intell. Res., 2: 217-238.

Hou, C., F. Nie, X. Li, D. Yi and Y. Wu, 2013. Joint embedding learning and sparse regression: A framework for unsupervised feature selection. IEEE Trans. Cybernetics, Vol. 10. 10.1109/TCYB. 2013.2272642.

Hung, H.J. and C.N. Hsu, 2002. Bayesian classification for data from the same unknown class. IEEE Trans. Syst. Man Cybernetics B: Cybernetics, 32: 137-145.

Lee, M.C., 2009. Using support vector machine with a hybrid feature selection method to the stock trend prediction. Expert Syst. Appl., 36: 10896-10904.

Liu, H. and L. Yu, 2005. Toward integrating feature selection algorithms for classification and clustering. IEEE Trans. Knowledge Data Eng., 17: 491-502.

Liu, H., J. Sun, L. Liu and H. Zhang, 2009. Feature selection with dynamic mutual information. Pattern Recogn., 42: 1330-1339.

Mangai, J.A., V.S. Kumar and S.A.A. Balamurugan, 2012. A novel feature selection framework for automatic web page classification. Int. J. Automation Comput., 9: 442-448.

Mitra, P., C.A. Murthy and S.K. Pal, 2002. Unsupervised feature selection using feature similarity. IEEE Trans. Pattern. Anal. Mach. Intell., 24: 301-312.

Pan, S.J. and Q. Yang, 2010. A survey on transfer learning. IEEE Trans. Knowledge Data Eng., 22: 1345-1359.

Peng, H., L. Fulmi and C. Ding, 2005. Variable selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Machine Intell., 27: 1226-1238.

Peng, W., S. Cesar and S. Edward, 2013. Prediction based on integration of decisional DNA and a variable selection algorithm RELIEF-F. Cybernetics Syst., 44: 173-183.

Polat, K. and S. Gunes, 2009. A novel hybrid intelligent method based on C4.5 decision tree classifier and one-against-all approach for multi-class classification problems. Expert Syst. Appl., 36: 1587-1592.

Rahman, R.M. and F.R.M. Hasan, 2011. Using and comparing different decision tree classification techniques for mining ICDDR, B hospital surveillance data. Expert Syst. Appl., 38: 11421-11436.

Robnik-Sikonja, M. and I. Kononenko, 2003. Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learn., 53: 23-69.

Ruggieri, S., 2002. Efficient C4. 5 [classification algorithm]. IEEE Trans. Knowledge Data Eng., 14: 438-444.

Song, Q., J. Ni and G. Wang, 2013. A fast clustering-based feature subset selection algorithm for high-dimensional data. IEEE Trans. Knowledge Data Eng., 25: 1-14.

Srivastava, A., S. Ghosh, N. Anantharaman and V.K. Jayaraman, 2013. Hybrid biogeography based simultaneous feature selection and MHC class I peptide binding prediction using support vector machines and random forests. J. Immunol. Methods, 387: 284-292.

Sun, Y., S. Todorovic and S. Goodison, 2010. Local-learning-based feature selection for high-dimensional data analysis. IEEE Trans. Pattern Anal. Machine Intell., 32: 1610-1626.

Uguz, H., 2011. A two-stage variable selection method for text categorization by using information gain, principal component analysis and genetic algorithm. Knowledge-Based Syst., 24: 1024-1032.

Vinh, N.X. and J. Bailey, 2013. Comments on supervised feature selection by clustering using conditional mutual information-based distances. Pattern Recognit., 46: 1220-1225.

Wei, H.L. and S.A. Billings, 2007. Feature subset selection and ranking for data dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell., 29: 162-166.

Wu, J., Y. Feng, Z. Zheng, M.C. Zhou and Z. Wu, 2012. Predicting quality of service for selection by neighborhood-based collaborative filtering. IEEE Trans. Syst. Man Cybernetics: Syst., 43: 428-439.

Yu, L. and H. Liu, 2004. Efficient variable selection via analysis of relevance and redundancy. J. Mach. Learn. Res., 5: 1205-1224.