

## A novel hybrid support vector machine with decision tree for data classification



Mehdi Zekriyapanah Gashti \*

Department of Computer Engineering, Payame Noor University, Tehran, Iran

### ARTICLE INFO

#### Article history:

Received 18 March 2017

Received in revised form

20 May 2017

Accepted 5 July 2017

#### Keywords:

Classification

Support vector machine

Decision tree

### ABSTRACT

The purpose of this paper is to increase the accuracy of a proposed support vector machine model using hybrid model of SVM and ID3. Then the hybrid approach based on SVM and ID3 tree will be evaluated focusing on analyzing the impact of ID3 on SVM performance. The evaluation process was carried out on the global dataset and Adult reference extracted from KEEL dataset repository. The obtained results demonstrate higher classification accuracy (0.9125) of the proposed model compared to SVM and ID3.

© 2017 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### 1. Introduction

Machine learning focuses on how to write a program that is capable of improving its performance through training and learning. The major objective in machine learning deals with choosing a model for current data that is favorably capable of estimating data features and deducing by using statistical principles and rules considering previously-occurred samples or information. The process of structuring a machine that is capable of deducing may be accomplished by under-supervision learning or without-supervision learning (Deniz et al., 2017)

Supervision learning is a process that is overseen. Such systems involve labeled data that are required to be recovered during testing phase by the system that learns by making use of training data that the positive or negative samples within them are determined for the supervisor when there are two classes of data. In the case of learning without supervision, the labels are not clear and there is no supervisor, so the system has to identify categories or classes among data based on the features of various samples that exist (Begum et al., 2015).

Classification can be carried out by various approaches including Fisher's Linear Discriminant Analysis (FLDA) (Giron-Sierra, 2017), Weighted Voting (Sun et al., 2005), Native Bayes (NB) (Giron-Sierra, 2017), Neural Networks (Giron-Sierra, 2017), Decision Tree (Giron-Sierra, 2017), Nearest

Neighbor (NN) (Giron-Sierra, 2017), SVM, Boosting, etc. Each approach has both advantages and drawbacks presenting various accuracy and precision regarding to attributes of data sets. The approach of decision tree covers a wide range of applications including identification of patterns, classification of patterns, classification, etc. (Song et al., 2013). The major advantage of decision tree method lies in identifying solutions (Mehenni, 2015). In certain situations when we confront a large sample space, this approach can make data preparation much easier and more understandable for users without technical knowledge compared with remaining methods (Mehenni, 2015).

SVM has been proved to be successful in several real-life applications such as handwritten figures, particle identification, face recognition, and bioinformatics (Vijayan et al., 2016). Applying SVM to these cases has yielded higher levels of generalization in comparison with other techniques. SVM model in testing phase is considerably slower than other methods (Cortes and Vapnik, 1995). The reason is that the computational complexity changes decision-making function of SVM depending on the number of support vectors. Thus, as the number of support vectors increases, SVM will spend much longer time sorting the new data points. Approaches such as feature selection and learning algorithms are used to improve SVM performance during testing phase (Cortes and Vapnik, 1995). The major purpose of reducing the number of support vectors is shortening the testing duration. The current study proposes approximation of SVM decision boundary by using decision tree in order to accelerate SVM in testing phase. The suggested approach is a combination of SVM and DT with a focus on

\* Corresponding Author.

Email Address: [gashti@pnu.ac.ir](mailto:gashti@pnu.ac.ir)

<https://doi.org/10.21833/ijaas.2017.09.019>

2313-626X/© 2017 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

achieving a quick classification without sacrificing its precision or accuracy.

Classification of Adult data was accomplished (Deepajothi and Selvarajan, 2012) by some data analysis techniques such as Naive Bayesian, K Star, Random Forest, and Zero R. Preprocessing and size reducing operations were also utilized in order to maximize classification accuracy. Precision of Naive Bayesian model was estimated 84.37 based on its evaluation by WEKA program. PSO and K-Anonymity models (Mohana et al., 2016) were proposed for classification of Adult data. K-Anonymity model is usually used for discovering hidden patterns within an enormous mass of data. The results indicate that K-Anonymity model has a higher identifying precision than PSO.

The present paper has been organized as follows: SVM and decision tree models, two data analysis methods, will be discussed in the Section 2. Section 3 deals with describing the proposed model in more details. Section 4, explains evaluation of the model and its results. Finally, Section 5 deals with conclusions and potential studies for the future.

**2. Data mining**

Data mining is used for analyzing large-size data sets to find exact relationships among data (Prakash et al., 2015). Databases contain lots of relationships among existing data that are impossible to be discovered without resorting to data mining techniques. Classification algorithms are among the most important data mining techniques from which SVM and DT are explained in this section.

**2.1. Support vector machine**

SVM (Cortes and Vapnik, 1995) is actually a binary classifier that separates two classes by a boundary line. SVM is capable of reducing the error of experimental categorizing and increasing class reparability using various transformations simultaneously. This ability has endowed SVM with a superior performance in dealing with large-size data or classes with multi-dimensional distribution space. Linear dividing of data is mainly aimed at acquiring a function that determines a hyper plane with the widest margin. As the margin reaches the maximum range, separation between classes will be maximized. Assume that  $S = \{x_i, y_i\}$  is a testing sample containing two  $y_i = \pm 1$  classes, and each class is composed of  $x_i, i = 1, \dots, m$  attribute. Fig. 1 illustrates a dividing hyper plane within SVM.

As can be seen from Fig. 1, the line  $w \cdot x_i + b = 0$  classifies the existing data into two  $\pm 1$  classes. This line is called dividing hyper plane. Two lines, namely  $w \cdot x_i + b = +1$  and  $w \cdot x_i + b = -1$  represent  $y=+1$  and  $y=-1$  sets respectively. Distance of each data corresponding to  $X_i$  from hyper plane is calculated by Eq. 1.

$$d(x_i) = \frac{y_i(w \cdot x_i + b)}{\|w\|} \tag{1}$$

Eq. 1 is used when data are separable in non-linear way. In Eq. 2,  $y$  and  $y_i$  represent equation output and value of testing sample  $X_i$  class respectively. The vector  $x = (x_1, x_2, \dots, x_n)$  indicates an input data, and  $x_i, i = 1, 2, \dots, N$  are support vectors.

$$y = \text{sign}(\sum_{i=1}^N y_i \alpha_i K(X, X_i) + b) \tag{2}$$

$K(X, X_i)$  is the kernel function that generates internal pulses to produce various types of machine from non-linear classes within data space. Radial kernel function usually shows a better performance for prediction. Radial kernel function is defined by Eq. 3 in which  $\|X-X_i\|$  is a set of training data and  $\gamma$  represents a parameter that is defined by user for the kernel width.

$$K(X, X_i) = \exp(-\gamma \|X - X_i\|^2) \text{ for } \gamma > 0 \tag{3}$$

**2.2. Decision tree**

Data classification using decision tree is one the well-recognized approaches to this purpose that obviates the needs for having parameters in advance and previous knowledge about data. This approach is categorized as a supervision method that is able to label and identify the individual attributes of testing data based on the training data that has been provided to form a tree called decision tree. Decision tree can be used to design rules for deduction system and label the unlabeled data (Nasridinov et al., 2013). Decision tree is a well-known approach to classification the results of which are provided on a flow chart similar to a tree structure. On this tree, each node represents a test on feature value, each branch indicates results of each test, and leaves represent classes. Complexity of a decision tree increases by higher number of features. Although only a few features have been observed in certain situations to be capable of determining the belonging class of an object, remaining features have weak or no effect (Nowak et al., 2013).

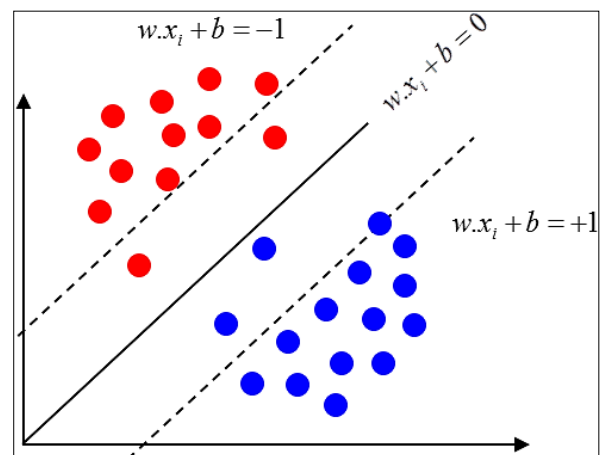


Fig. 1: Dividing hyper plane within SVM

This approach is one of the nonparametric classification methods that may be classified into two groups with regard to type of dependent

variable: tree sorting for discrete variable and variable batch, and tree regression for continuous variable. Tree classification is in line with some methods such as auditing analysis of discriminant function and logistic regression. This method involves a set of logic conditions in the form of an algorithm with tree-like structure used for classification or prediction (Salleh, 2014). Generally, test data that are used to create a tree differ from data that are used for tree evaluation, and the number of errors in identifying the data class is a criterion for assessing appropriateness of the algorithm. Most learning algorithms of decision tree operate on the basis of an up-down voracious searching action within the space of existing trees. ID3 algorithm is an algorithm that is used for building decision tree (Quinlan, 1986). Within the ID3 tree, the lack of order for all features is calculated through entropy as the first step, the rate of information is estimated using the obtained value for all features. Entropy shows the randomness in the form of a mathematical criterion. If the set S includes both positive and negative samples of one dataset, entropy of S in relation to Boolean classification is defined by Eq. 4 (Quinlan, 1986).

$$E(S) = -P_{\oplus} \log_2 P_{\oplus} \oplus -P_{\ominus} \log_2 P_{\ominus} \quad (4)$$

In Eq. 4,  $P_{\oplus}$  represents ratio of positive samples to all samples, and  $P_{\ominus}$  is the ratio of negative samples to all samples. Making decision about choosing a feature that should be at the root of tree depends on information rate of each feature. Eq. 5 is used to calculate the information rate (Quinlan, 1986).

$$IG(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \times E(S_v) \quad (5)$$

In Eq. 5, the parameter Values (A) is a value set of A features, and  $S_v$  is a subset of S for which A has the value of V. This algorithm is only capable of data classification with limited and discrete range of features, thus it cannot be applied to noisy and distorted data (Quinlan, 1986). The algorithm C4.5 is a completed form of algorithm ID3. This algorithm is capable of classification continuous and noisy data as well. To do this, data are initially ordered then the useful amounts for all possible modes of segregating the ordered data are acquired. Finally, a separation corresponding to the largest useful amount is chosen as a separator.

### 3. The proposed model

When dataset is large, complexity of training phase and required storage memory for saving these data will increase accordingly. Therefore, it's necessary to have a model that is able to reduce the complexity. The proposed model is a combination of SVM and ID3 in order to achieve an efficient integrated procedure for classification Adult data from KEEL dataset repository (sci2s.ugr.es/keel/

datasets.php) that contain 48842 samples and 14 features. Fig. 2 shows the flow chart of the suggested model.

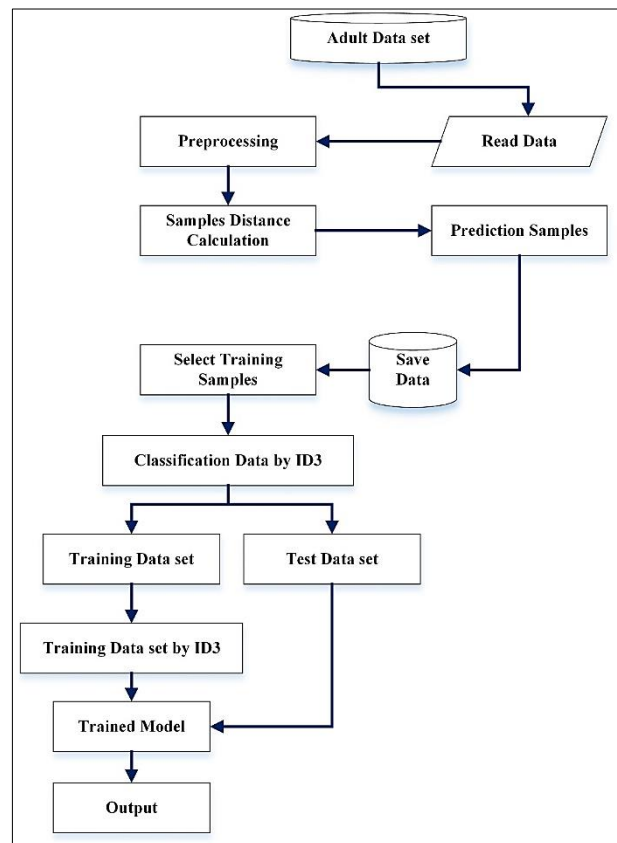


Fig. 2: Proposed model flowchart

The proposed model for data classification includes following phases:

- Phase 1: Data reading
- Phase 2: Preprocessing of data (including normalization and removing deviated data)
- Phase 3: Applying support vector machine on dataset
- Phase 4: Calculating the distance of sample data for support vectors relevant to each class
- Phase 5: Predicted label and obtained distance as data membership in one class is calculated together with actual label of each sample class.
- Phase 6: The obtained results from phase 5 are stored into a new dataset.
- Phase 7: Classification of training data is carried out using the new dataset by decision tree.
- Phase 8: Data testing
- Phase 9: Evaluating outcomes of the proposed model

The proposed method involves dividing all data into two groups of experimental and test data in a random way following a proportion of 70 to 30. Then the experimental data are fed into standard SVM, and output is estimated. Data are classified again with the aid of SVM using the obtained coefficients. The estimated class was called new target. As the next step, distance between individual data is obtained by support vectors corresponding to the estimated class then their average value is calculated.

The estimated class and the value of calculated distance for each experimental data as the feature vector together with actual class of data are fed into the decision tree classifier to get the results recalculated.

The above steps are repeated during test phase in such a way that individual test data are initially obtained with the aid of SVM model. Then they are classified, and the estimated class for each test data is considered as the new target. As the next step, distance between test data from support vectors corresponding to new target is averaged, and the resultant value is considered as the second feature. Two obtained features (estimated class and distance) are fed into the already acquired decision tree classifier in order to verify the class of test data.

#### 4. Evaluation and results

The proposed model implementation was accomplished by MATLAB 2016b program, and the required evaluation has been carried out on Adult dataset containing 48842 samples and 14 features. Various criteria have been utilized for the purposes of evaluation (Bratko et al., 1999) in order to demonstrate the improvements made in the suggested model compared with SVM and ID3.

$$Precision = \frac{TP}{TP+FP} \tag{6}$$

$$Recall = TPR = \frac{TP}{TP+FN} \tag{7}$$

$$TNR = \frac{TN}{TN+FP} \tag{8}$$

$$FPR = \frac{FP}{TN+FP} \tag{9}$$

$$FNR = \frac{FN}{TP+FN} \tag{10}$$

$$F - Measure = \frac{2*Precision*Recall}{(Precision+Recall)} \tag{11}$$

$$AUC = \frac{TPR+TNR}{2} \tag{12}$$

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \tag{13}$$

$$Error Rate = \frac{(FP+FN)}{(TP+TN+FP+FN)} = 1 - Accuracy \tag{14}$$

$$BCR = 0.5(TPR + TNR) \tag{15}$$

$$BER = 1 - BCR \tag{16}$$

In the defined equations, the parameter TN represents those records that actually belong to negative classification, and classifying algorithm has accurately detected them as negative. TP represents those records that actually belong to positive classification, and classifying algorithm has accurately detected them as positive. FP represents those records that actually belong to negative classification, but classifying algorithm has mistakenly detected them as positive. FN represents those records that actually belong to positive classification, but classifying algorithm has mistakenly detected them as negative.

Details about precision of classification models on Adult data with 100 iteration cycles within the program are provided in Table 1. As can be seen, precision of the proposed model is higher compared with SVM and ID3. Moreover, precision of ID3 is higher than SVM.

**Table 1:** Comparing accuracy of classification models on adult data with 100 iteration cycles

Criteria's	Models		
	SVM	ID3	Proposed Model
Precision	0.6384	0.6479	0.7214
Recall=TPR	0.6523	0.6632	0.7532
TNR	0.6632	0.6712	0.7164
FPR	0.6487	0.6234	0.6847
FNR	0.6341	0.6584	0.6924
F-Measure	0.6453	0.6555	0.7370
AUC	0.6710	0.6932	0.8034
Accuracy	0.6914	0.7165	0.8103
Error Rate	0.3086	0.2835	0.1897
BCR	0.6577	0.6672	0.7348
BER	0.3423	0.3328	0.2652

Fig. 3 depicts a diagram comparing models on Adult dataset with 100 iteration cycles based on the criteria of Precision, Recall, F-Measure, and Accuracy. Table 2 summarizes classification precision of the models on Adult dataset with 200 iteration cycles. As can be seen from Table 2, precision of the proposed model is higher compared with 100 iteration cycles. The reason can be attributed to this fact that the suggested model is able to examine the searching space more and generate more precise rules by making use of the ID3.

**Table 2:** Comparing precision of classification models on Adult data with 200 iteration cycles

Criteria's	Models		
	SVM	ID3	Proposed Model
Precision	0.7012	0.7361	0.8334
Recall=TPR	0.7219	0.7594	0.8617
TNR	0.7495	0.7690	0.8436
FPR	0.7137	0.7304	0.8297
FNR	0.7024	0.7110	0.8208
F-Measure	0.7114	0.7476	0.7924
AUC	0.7598	0.7932	0.9012
Accuracy	0.7620	0.8094	0.9125
Error Rate	0.2380	0.1906	0.0875
BCR	0.7357	0.7642	0.8526
BER	0.2643	0.2358	0.1474

Fig. 4 depicts a diagram comparing models on Adult dataset with 200 iteration cycles based on the criteria of Precision, Recall, F-Measure, and Accuracy. Fig. 5 depicts a diagram comparing the models with regard to implementation of the iterations and Error Rate criterion on Adult dataset.

#### 5. Conclusion and future works

Classification is an operation for data analysis and modeling with the major focus on describing significant classes of data. For the purposes of classification, similarity of each classification is estimated on the basis of existing differences among predefined data. Classification is actually a supervised learning in which classes are already specified. The main objective of classification is evaluating features of a dataset and assigning them to a set of classes. The current paper suggests a combined model integrating SVM and ID3 tree in order to sort Adult data. The ID3 tree was utilized to increase precision of classification in SVM model. The obtained results indicate that the suggested model has better detection accuracy compared with

SVM and ID3 because this model has verified the acquired classifications from SVM using the ID3 tree.

Furthermore, ID3 provides better accuracy and lower error by comparison with SVM.

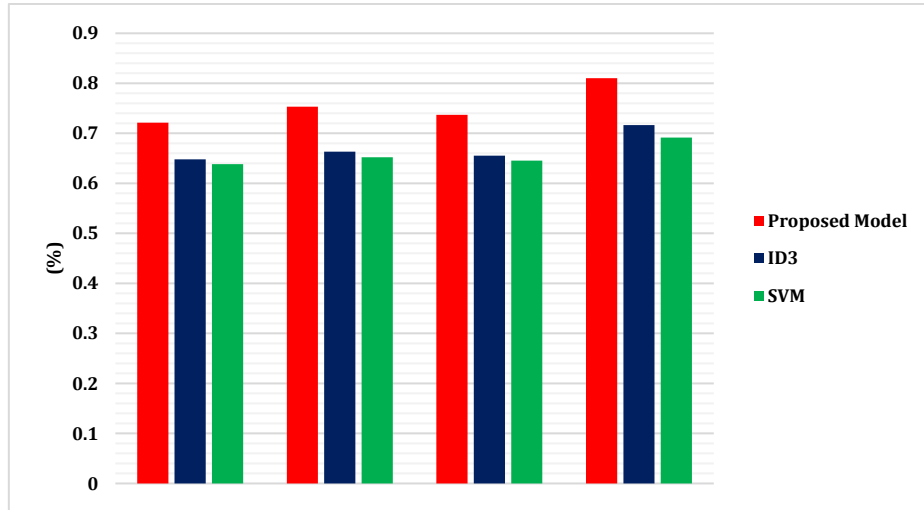


Fig. 3: A diagram for comparing precision among classification models on adult dataset with 100 iteration cycles

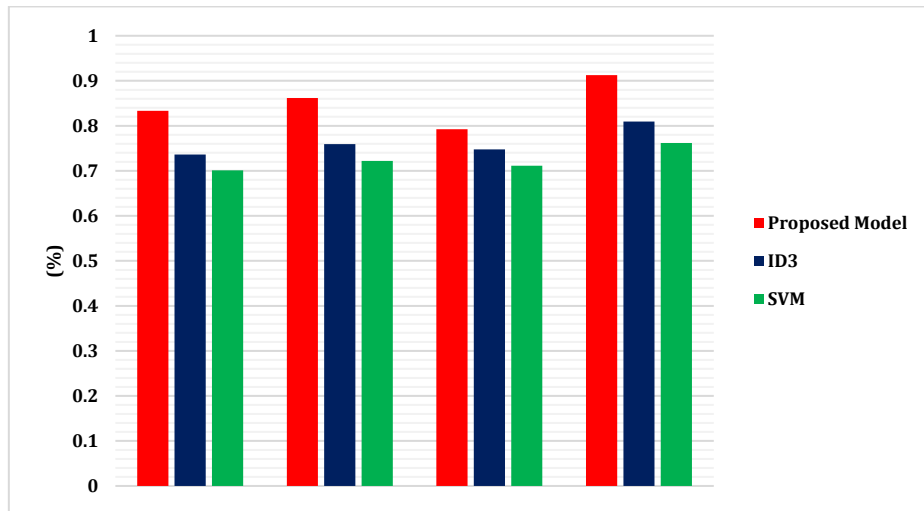


Fig. 4: A Diagram for comparing precision among classification models on adult dataset with 100 iteration cycles

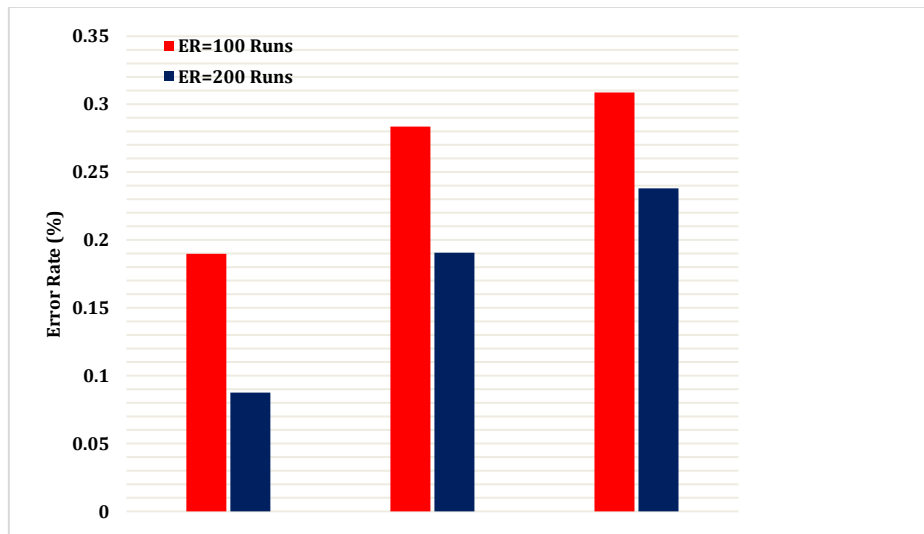


Fig. 5: A diagram for comparing the models considering error rate criteria on adult dataset with various iterations

**References**

Begum S, Chakraborty D, and Sarkar R (2015). Data classification using feature selection and kNN machine learning approach. In the International Conference on Computational Intelligence and Communication Networks, IEEE, Jabalpur, India: 811-814. <https://doi.org/10.1109/CICN.2015.165>

Bratko I, Michalski RS, and Kubat M (1999). Machine learning and data mining: methods and applications. John Wiley and Sons, Hoboken, USA.

- Cortes C and Vapnik V (1995). Support-vector networks. *Machine Learning*, 20(3): 273-297.
- Deepajothi S and Selvarajan S (2012). A comparative study of classification techniques on adult data set. *International Journal of Engineering Research and Technology*, 1(8): 1-8.
- Deniz A, Kiziloz HE, Dokeroglu T, and Cosar A (2017). Robust multiobjective evolutionary feature subset selection algorithm for binary classification using machine learning techniques. *Neurocomputing*, 241: 128-146.
- Giron-Sierra JM (2017). Data analysis and classification. In: Giron-Sierra JM (Ed.), *Digital signal processing with matlab examples*: 647-835, Springer Singapore, India.
- Mehenni T (2015). Integration of useful links in distributed databases using decision tree classification. In the 6<sup>th</sup> International Conference on Information Systems and Economic Intelligence, IEEE, Hammamet, Tunisia: 5-9. <https://doi.org/10.1109/ISEI.2015.7358717>
- Mohana S, Sahaaya SA, and Mary A (2016). A comparative framework for feature selection in privacy preserving data mining techniques using pso and k-anonumization. *Iioab Journal*, 7(9): 804-811.
- Nasridinov A, Ihm SY, and Park YH (2013). A decision tree-based classification model for crime prediction. In *Information Technology Convergence*, Springer, Dordrecht, Netherlands: 531-538. [https://doi.org/10.1007/978-94-007-6996-0\\_56](https://doi.org/10.1007/978-94-007-6996-0_56)
- Nowak BA, Nowicki RK, and Mleczo WK (2013). A new method of improving classification accuracy of decision tree in case of incomplete samples. In the International Conference on Artificial Intelligence and Soft Computing, Springer Berlin Heidelberg, Heidelberg, Germany: 448-458. [https://doi.org/10.1007/978-3-642-38658-9\\_40](https://doi.org/10.1007/978-3-642-38658-9_40)
- Prakash VA, Ashoka DV, and Aradya VM (2015). Application of data mining techniques for defect detection and classification. In the 3<sup>rd</sup> International Conference on Frontiers of Intelligent Computing: Theory and Applications, Springer International Publishing: 387-395. [https://doi.org/10.1007/978-3-319-11933-5\\_42](https://doi.org/10.1007/978-3-319-11933-5_42)
- Quinlan JR (1986). Induction of decision trees. *Machine Learning*, 1(1): 81-106.
- Salleh MNM (2014). Improving weighted fuzzy decision tree for uncertain data classification. In: Herawan T, Ghazali R, and Deris MM (Eds.), *Recent advances on soft computing and data mining*: 249-259. Springer International Publishing, Berlin, Germany.
- Song J, Zhu Z, Scully P, and Price C (2013). Selecting features for anomaly intrusion detection: A novel method using fuzzy C means and decision tree classification. In: Wang G, Ray I, Feng D, and Rajarajan M (Eds.), *Cyberspace Safety and Security*: 299-307. Springer International Publishing, Berlin, Germany.
- Sun Y, Kamel M, and Wong A (2005). Empirical study on weighted voting multiple classifiers. In: Singh S, Singh M, Apte C, and Perner P (Eds.), *Pattern recognition and data mining*: 335-344. Springer, Berlin, Germany.
- Vijayan A, Kareem S, and Kizhakkethottam JJ (2016). Face recognition across gender transformation using SVM classifier. *Procedia Technology*, 24: 1366-1373.