Applied Mathematics & Information Sciences Letters
An International Journal

http://dx.doi.org/10.18576/amisl/060202

Quality Control and Classification of Steel Plates Faults Using Data Mining

Mohammad Ali Afshar Kazemi, Sima Hajian and Neda Kiani*

Department of Industrial Management, Central Tehran Branch, Islamic Azad University, Tehran, Iran

Received: 22 Jun. 2017, Revised: 2 Nov. 2017, Accepted: 8 Nov. 2017
Published online: 1 May 2018

Abstract: Evaluating the quality of steel Plate is vital for factories and doing this manually is associated with life-threatening risks and low efficiency. Given that the steel Plate faults often create problems in the production process and sometimes may lose market, need to check the quality of the Plate surface and find the failures are inevitable. This paper tends to first identify and recognize the data based on the CRISP-DM cycle in data mining and thus compares and evaluates four data mining models to classify seven commonly occurring faults of the steel plate. For this purpose, the models of C5.0 decision tree, Multi Perception Neural Network (MLPNN), Bayesian network (BN) and Ensemble model are used. A faults dataset of steel plates is taken from the University of California at Irvine (UCI) machine learning repository. The diagnostic accuracy of C5.0 decision tree obtained remarkable performance with an accuracy of 95.56 percent for the training data and the accuracy of 95.66 percent for testing data. The results of the model fitting of training and testing data indicated that the model C5.0 is superior to Multi Perception Neural Network (MLPNN), Bayesian network (BN) and Ensemble model.

Keywords: Fault diagnosis, Steel plate faults, Data Mining, classification, decision tree

1 Introduction

Quality control is one of the most important processes in steel plate manufacturing industry. Diagnosis of surface faults accounts for a high percentage of the quality control process[1]. The purpose of surface detection system is to identify and classify the faults of surfaces that influence the quality of surfaces according to the standards and user needs[2].

According to Himmelbau [3], the term fault is generally dened as a departure from an acceptable range of an observed variable or a calculated parameter associated with a process. A fault is dened as abnormal process behavior, whether associated with equipment failure, sensor degradation, and set point change or process disturbances. Fault diagnosis is aimed at discovering the time, location, and size of certain faults in the modern industrial process [4]. Fault diagnosis is the problem of detecting the potential faults hidden in the observed instances that relate to specic application domains [5]. It is usually based on available data on the spot and systematic fault classication records. An effective method that can determine fault types and causes will not only lower maintenance cost and unexpected waste, but also improve production efciency and quality level of products [6]. Besides, further treatments such as recycling, are also based on accurate fault diagnosis [7]. Traditionally, even experts with fault diagnosis manuals have to carefully analyze operational environments to infer potential causes of a particular fault. However, more intelligent means, derived from study of machine learning, have developed a lot to address this problem quickly and correctly[8]. Typically, they include articial neural networks (ANNs)[9], logistic regression (LR) [10], decision tree (DT)[11], principal component analysis (PCA) [12], and support vector machines (SVMs) [13]. The manual fault diagnosis system is the traditional way where an expert with electronic meter tries to obtain some information about relevant operational equipment, check the maintenance manual and then diagnosed the probable causes of a particular fault. However, intelligent fault diagnosis techniques can provide quick and correct systems that help to prevent product quality problems and facilitates precautionary maintenance. These intelligent systems have used different artificial intelligence and data mining models and they should be simple and efficient [14]. Particularly in fault detection, data mining can be

* Corresponding author e-mail: Sima.Hajian.877@gmail.com

Table 1: Types of faults and sample sizes.

Class	Type of faults	Number of samples
1	Pastry	158
2	Z_Scratch	190
3	K_Scratch	391
4	Stains	72
5	Dirtiness	55
6	Bumps	402
7	Other_Faults	673

utilized to identify patterns of defects, factors influencing the failure of processes, types of defects, and error rates in manufacturing [9]. The aim of this study is to evaluate the performances of the popular and effective data mining models to diagnose seven commonly occurring faults of the steel plate, namely; Pastry, Z_Scratch, K_Scatch, Stains, Dirtiness, Bumps and Other_Faults. The models include C5.0 decision tree (C5.0 DT), Multi Perceptron Neural Network (MLPNN) with pruning, Bayesian Network (BN) and Ensemble.

2 Description of data and method

2.1 Steel plate faults dataset

Steel plate is an important raw material in hundreds of manufacturing industries. The Steel Plates Faults Data Set used in the study was obtained from the University of California at Irvine (UCI) Machine Learning Repository [15]. In this dataset, the faults of steel plates are classied into 7 types. Since it has been donated on October 26, 2010, this dataset has been widely used in machine learning for automatic pattern recognition. Types of fault and corresponding numbers of sample are shown in Table 1 . As shown in Table 1, the numbers of sample vary a lot from one category to another. Meanwhile, fault 7 is a special class of fault because it contains all other faults except the rst six kinds of fault. In other words, samples in class 7 may have no obvious common characteristics. For every sample, 27 features are recorded, providing evidences for its fault class. All attributes were expressed by integers or real numbers. Detailed information about these 27 independent variables is listed in Table 2.

The main question is that how have to control steel plates faults quality and do classify those using data mining?

Table 2: Independent attributes of steel plates.

Number	Attribute
1	X_Minimum
2	X_Maximum
3	Y_Minimum
4	Y_Maximum
5	Pixels_Areas
6	X_Perimeter
7	Y_Perimeter
8	Sum_of_Luminosity
9	Minimum_of_Luminosity
10	Maximum_of_Luminosity
11	Length_of_Conveyer
12	TypeOfSteel_A300
13	TypeOfSteel_A400
14	Steel_Plate Thickness
15	Edges_Index
16	Empty_Index
17	Square_Index
18	UmbriOutside_X_Indexa
19	Edges_X_Index
20	Edges_Y_Index
21	Outside_Global_Indexxx
22	LogofAreas
23	Log_X_Index
24	Log_Y_Index
25	Orientation_Index
26	Luminosity_Index
27	SigmoidOfAreas

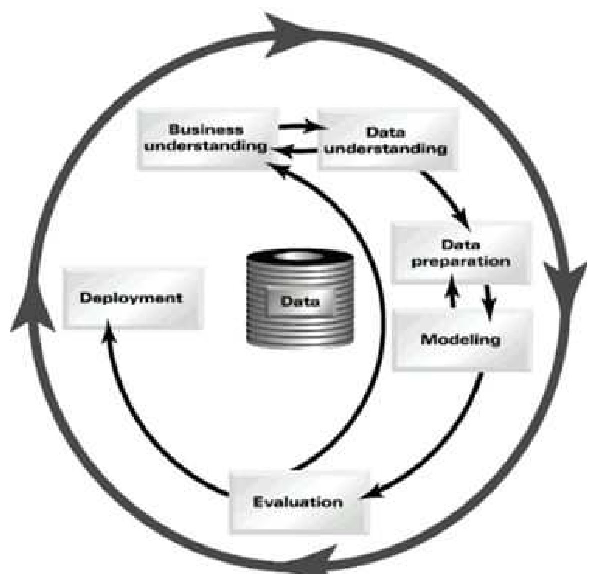


Fig. 1: CRISP-DM Cycle in data mining

2.2 Data mining

The researcher, the analysis of the data is based on CRISP-DM cycle. The general CRISP-DM process model includes six phases that address the main issues in data mining. The common steps of this model in data mining are shown in Figure 1[18].

What is important is that there should be a proper understanding of the CRISP-DM cycle steps, different modeling techniques should be selected and used and the parameters should be evaluated according to the optimum values. There are several modeling techniques in every data mining issue and some data require a special format. Due to the close relationship between the data preparation and modeling and mostly an initial understanding of the data on an issue and acquiring an initial idea for the upcoming limitations, this cycle contributes a lot in choosing the kind of the model and doing the modeling[16].

After the preparation, identifying the model based on the main goal, which is classification the data based on the quality of steel plates faults and determinate the best classification method, thus types of classification method was used. The last phase is CRISP-DM cycle that provides the strategy for developing the model [17,18].

2.3 Classification models

Classication builds up and utilizes a model to predict the categorical labels of unknown objects to distinguish between objects of different classes. These categorical labels are predened, discrete and unordered [19]. Zhang and Zhou [20] stated that classication and prediction is the process of identifying a set of common features and models that describe and distinguish data classes or concepts. Common classication techniques include neural networks, the Nave Bayes technique, decision trees and support vector machines [21].

Decision tree: The decision tree criteria separate important from unimportant branches so that only strong relationships between inputs and the target variable are retained [22]. One of the main advantages of decision trees is the ability to generate understandable knowledge structures, i.e., hierarchical trees or sets of rules, a low computational cost when the model is being applied to predict or classify new cases, the ability to handle symbolic and numeric input variables, provision of a clear indication of which attributes are most important for prediction or classication [23]. Decision trees are predictive decision support tools that create mapping from observations to possible consequences [19,24]. These trees can be planted via machine-learning-based algorithms such as the $ID3$, CART and C5.0. Predictions are represented by leaves, and the conjunctions of features by branches. Decision trees are commonly used in credit card, automobile insurance, and corporate fraud [25].

C5.0 algorithm top-down decision tree base was proposed by Quinlan. The algorithm is a successor of $ID3$, which determines at each step the most predictive attribute, and splits a node based on this attribute. Every node represents a decision point over the value of some attribute. The split criterion can be calculated as follows [26]:

- Calculate the expected information needed to classify a tuple in D

$$Info(D) = \sum_{j=1}^{m} p_i \log_2(p_i) \tag{1}$$

where, P_i is the probability that a tuple in D belong to class C_i.

Calculate the expected information required to classify a tuple from D based partitioning by A

$$Info_A(D) = \sum_{j=1}^{m} \frac{|D_f|}{|D|} * Info(D_j) \tag{2}$$

The term $\frac{|D_f|}{|D|}$ acts as the weight of the j^{th} partition.

- calculate information gain of attribute A.

$$GAIN(A) = Info(D) - Info_A(D) \tag{3}$$

-Calculate split information of attribute A.

$$SplitInfo_A(D) = \sum_{j=1}^{v} \frac{|D_j|}{|D|} * \log_2\left[\frac{|D_j|}{|D|}\right] \tag{4}$$

- calculate gain ratio:

$$GainRation(A) = \frac{Gain(A)}{SplitInfo_A(D)} \tag{5}$$

The C5.0 Decision Tree Classifier uses the information gain ratio; the metric which tests each node and selects the subdivisions of the data in order to maximize the Entropy Decrease of the connected node. Thus, the best attributes (features) obtained are used to describe each case belonging to only one separate class. These attributes are tested in the nodes of each split, and partitioning processes are continued until reaching to a terminal (leaf) node downwards the tree [27].

Neural networks and their applications: An Articial Neural Network (ANN) model is a structure that can be adjusted to produce a mapping from a given set of data and it features or relationships among the data. One of the well-known structures of ANN is Multilayer Feed-Forward Neural Network (FFNN) or Multilayer Perceptron Neural Network (MLPNN) [28]. Structure of the MLPNN includes one input layer, one output layer, and one or more hidden layers, which are connected to the next layers through some pre-specied weights. All nodes, except those of the input layer, are composed of several neurons. The output of each layer is connected to the input of the next layer. In the MLPNN structure, the nodes are connected in one direction and it has no feedback loop. The number of nodes in each layer varies depending on the problem. Large number of the hidden layers and nodes of the MLPNN architecture lead to more complexity. Hidden layers help to extract higher-order properties from the input by the MLPNN. Each neuron in MLPNN contains a continuously differentiable activation function. Training of a MLPNN is a process by which one can nd a set of weights that would give desired output values of the ANN. Fig.2 illustrates the general structure of a MLPNN [29].

There are a set of types of classical neural networks such as radial basis function networks, neural

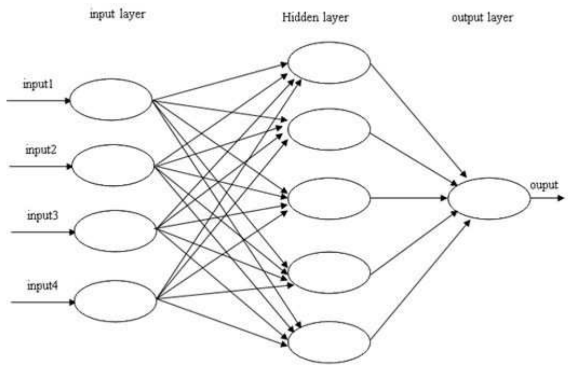


Fig. 2: Feed forward MLP, with a single hidden layer and a single output neuron

classification, competitive neural networks, , gene regulatory networks, fuzzy recurrent neural networks, neural nets, back-propagation artificial neural networks,... etc [30,34]. M. Zidan et al. proposed a novel two autonomous neural network algorithms to make the learning process of neural notworks done after only iteration [35,36]. Another technique to enhance the efficiency of MLP is called ELM algorithm [37].

Bayesian networks for classification: Bayesian networks (BNs) are a popular medium for graphically representing and manipulating attribute interdependencies, and represent a joint probability distribution over a set of discrete, stochastic variables. Bayesian networks are directed acyclic graphs (DAG) that allow for efficient and effective representation of joint probability distributions over a set of random variables.

The nave Bayesian classifier: The nave Bayesian (NB) classifier is one of the most computationally efficient algorithms for machine learning and data mining. It has been shown in many domains to be surprisingly accurate compared to alternatives including decision tree learning, rule learning, neural networks and instance based learning Fig.3.

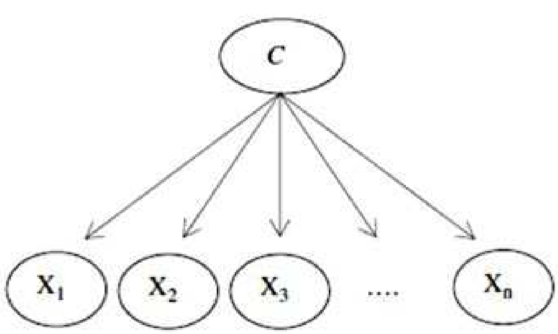


Fig. 3: The nave Bayes classifiers

This classifier learns the class-conditional probabilities $P(X_i = x_i|C = c_1)$ of each variable X_i given the class label c_1. A new test case $X_i = x_i, \ldots, X_n = x_n$ is then classified by using Bayes rule to compute the posterior probability of each class c_1 given the vector of observed variable values:

$$P = (C = c_1|X_1 = x_1, \ldots, X_n = x_n) = \frac{P(C = c_1)P(X_1, \ldots, X_n = x_n|C = c_1)}{P(X_1 = x_1, \ldots, X_n = x_n)} \quad (6)$$

The simplifying assumption behind the nave Bayes classier then assumes that the variables are conditionally independent given the class label. Hence,

$$P = (X_1 = x_1, \ldots, X_n = x_n|C = c_1) = \prod_{i=1}^{n} P(X_i = x_i|C = c_i) \quad (7)$$

Fig 3 shows the network structure of a Bayesian network that represents a nave Bayes classifier. Because the structure is static, there is no need to perform structural learning, only probability tables $P(c)$ and $P(X_i|c)$ need to be assessed [31].

2.4 Experiment Evaluation

2.4.1 Confusion matrix

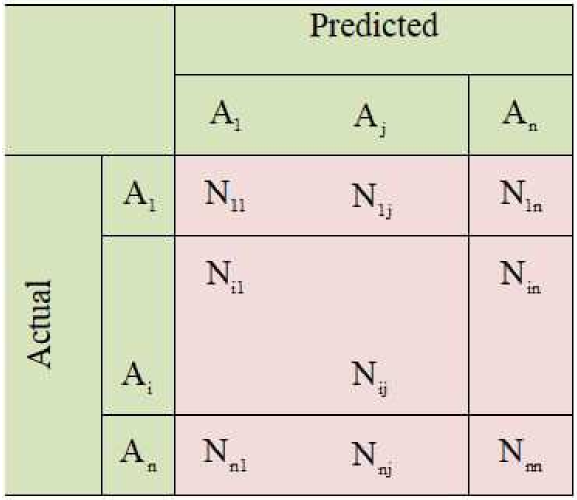
Confusion matrix [32] is a concept from machine learning, which contains information about actual and predicted classications done by a classication system. *A* confusion matrix has two dimensions, the actual class of an object indexes one dimension, the other is indexed by the class that the classier predicts. Fig 4 presents the basic form of confusion matrix for a multi-class classication task, with the classes A_1, A_2, and A_n. In the confusion matrix, N_{ij} represents the number of samples actually belonging to class A_i but classied as class A_j.

A number of measures of classication performance can be dened based on the confusion matrix. Some common measures are given as follows. Precision is a measure of the accuracy if a specic class has been predicted. It is dened by:

$$Precision_i = \frac{N_{ii}}{\sum_{k=1}^{n} N_{ki}} \quad (8)$$

Recall is a measure of the ability of a prediction model to select instances of a certain class from a data set; it is dened by the formula: [33]

$$Recall_i = \frac{N_{ii}}{\sum_{k=1}^{n} N_{ik}} \quad (9)$$



		Predicted		
		A_1	A_j	A_n
Actual	A_1	N_{11}	N_{1j}	N_{1n}
		N_{i1}		N_{in}
	A_i		N_{ij}	
	A_n	N_{n1}	N_{nj}	N_{nn}

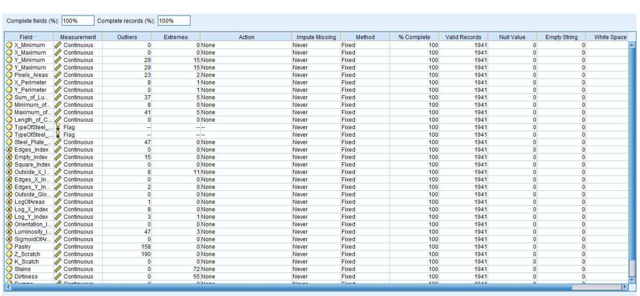
Fig. 4: Confusion matrix

3 Implementation of data

In this study, the implementation of the data mining process is done in two steps of preparation and modeling which are to be discussed:

- Preparation process:

This step includes the initial investigation of the data, correcting the current mistakes and increasing the quality of the data set. The data is purged and at this step one starts to investigate the lost values. The qualitative reporting of Data Audit is used for the investigation of lost values. Are shown in Figure 5.



Complete fields (%): 100% Complete records (%): 100%

Field	Measurement	Outliers	Extremes	Action	Impute Missing	Method	% Complete	Valid Records	Null Value	Empty String	White Space
X_Minimum	Continuous	0	0	None	Never	Fixed	100	1941	0	0	
X_Maximum	Continuous	0	0	None	Never	Fixed	100	1941	0	0	
Y_Minimum	Continuous	29	15	None	Never	Fixed	100	1941	0	0	
Y_Maximum	Continuous	29	15	None	Never	Fixed	100	1941	0	0	
Pixels_Areas	Continuous	23	2	None	Never	Fixed	100	1941	0	0	
X_Perimeter	Continuous	8	1	None	Never	Fixed	100	1941	0	0	
Y_Perimeter	Continuous	0	1	None	Never	Fixed	100	1941	0	0	
Sum_of_Lu...	Continuous	37	5	None	Never	Fixed	100	1941	0	0	
Minimum_of...	Continuous	8	0	None	Never	Fixed	100	1941	0	0	
Maximum_of...	Continuous	41	5	None	Never	Fixed	100	1941	0	0	
Length_of_C...	Continuous	0	0	None	Never	Fixed	100	1941	0	0	
TypeOfSteel_...	Flag	--	--	--	Never	Fixed	100	1941	0	0	
TypeOfSteel_...	Flag	--	--	--	Never	Fixed	100	1941	0	0	
Steel_Plate_...	Continuous	47	0	None	Never	Fixed	100	1941	0	0	
Edges_Index	Continuous	0	0	None	Never	Fixed	100	1941	0	0	
Empty_Index	Continuous	15	0	None	Never	Fixed	100	1941	0	0	
Square_Index	Continuous	0	0	None	Never	Fixed	100	1941	0	0	
Outside_X_I...	Continuous	8	11	None	Never	Fixed	100	1941	0	0	
Edges_X_In...	Continuous	0	0	None	Never	Fixed	100	1941	0	0	
Edges_Y_In...	Continuous	2	0	None	Never	Fixed	100	1941	0	0	
Outside_Glo...	Continuous	0	0	None	Never	Fixed	100	1941	0	0	
LogOfAreas	Continuous	1	0	None	Never	Fixed	100	1941	0	0	
Log_X_Index	Continuous	8	0	None	Never	Fixed	100	1941	0	0	
Log_Y_Index	Continuous	3	1	None	Never	Fixed	100	1941	0	0	
Orientation_I...	Continuous	0	0	None	Never	Fixed	100	1941	0	0	
Luminosity_I...	Continuous	47	3	None	Never	Fixed	100	1941	0	0	
SigmoidOfAr...	Continuous	0	0	None	Never	Fixed	100	1941	0	0	
Pastry	Continuous	158	0	None	Never	Fixed	100	1941	0	0	
Z_Scratch	Continuous	190	0	None	Never	Fixed	100	1941	0	0	
K_Scatch	Continuous	0	0	None	Never	Fixed	100	1941	0	0	
Stains	Continuous	0	72	None	Never	Fixed	100	1941	0	0	
Dirtiness	Continuous	0	55	None	Never	Fixed	100	1941	0	0	

OK

Fig. 5: Data Audit

As seen in the figure above, there are no missing values in the data set. However, you can see that the field contains values are outliers. Since the number of fields is high and the number of outliers in some fields are too large to examine the existence of outlier values, using single-variable methods based on box diagrams and IQR index, can remove or modify a significant part of the data. In order to the Clustering Technique of Anomaly Detection is used as a Multivariate Method. The Anomaly Detection clustering technique was used to determine the different records of the original data body, are shown in Figure 6.

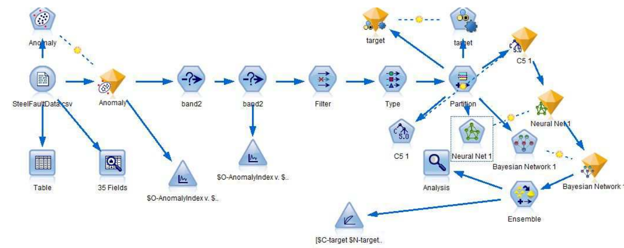


Fig. 6: Body model

In this method, first the data are clustered. Then through determining the norm in every cluster and the distance of every record from the cluster norm, the straggled values are analyzed. Figure 7 shows the distance of every record from the cluster norm individually:

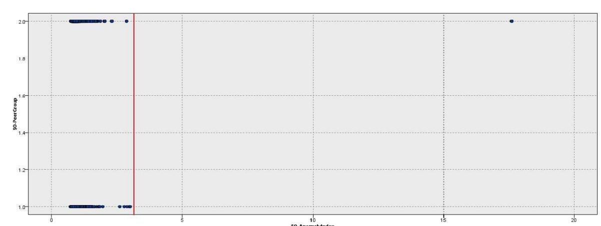


Fig. 7: considering the distribution of the Anomaly Index values

As shown in the figure 7, considering the distribution of the Anomaly Index values, the 3.1 threshold value is determined for the identification of the straggled records. As regards existence of this outlier data in chart is caused that the position of other data had unclear, this chart is been again draw by removing its.

As shown in the figure 8, considering the distribution of the Anomaly Index values, the 2 threshold value is determined for the identification of the straggled records. Based on this, 12 observations are considered as the straggled records of the data set and removed.

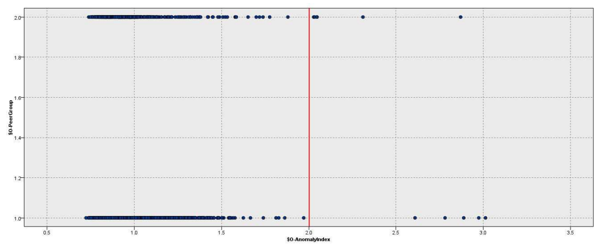


Fig. 8: considering the distribution of the Anomaly Index values

4 Results

- Modeling step

In order to classify the number of defects based on qualitative indicators of steel Plate, three models of C5.0 decision tree, Multi Perception Neural Network (MLPNN), Bayesian network, as well as a model of Ensemble of these models have been used. In the collective model of confidence indicator, each of the models is considered as the weight of each model and Ensemble model vote will be based on weighted voting of classification of three models. In each of these models, field of the number of steel Plates faults has been considered as the target field, and the fields relating to qualitative indicators have been considered as input fields. In order to enable assessment of models, initially 70 percent and 30 percent of record are considered as training and testing records, respectively; and then the models have been fitted on the training data. Since the purpose in this paper is to classify steel sheets, as we see in Figure 6, the main body, by choosing the classification models and aggregation of these models, to determine the best model according to the results presented in Table 3. According to the results presented in Table 3, the prediction accuracy of C5.0 decision tree model in the training and testing data is 95.56 percent and 95.66 percent, respectively, which has the highest accuracy in the training and testing data. Then the Ensemble model is the most accurate. Two models of MLPNN and BN in terms of accuracy in the classification of the number of faults are not much different from each other. Closeness of accuracy in both training and testing data sets on each model indicates that the model has high stability.

Table 3: Performance of the four classifiers.

	Testing		Training	
Classifier	Correct percent	Wrong percent	Correct percent	Wrong percent
C5.0	95.66	4.34	95.56	4.44
MLPNN	73.29	26.71	78.03	21.97
BN	77.63	22.37	76.07	23.93
Ensemble	85.48	14.52	85.03	14.97

The predictions of all models are compared to the original classes to identify the values of true positives, true negatives, false positives and false negative. These values have been computed to construct the confusion matrix as shown in Table 4. Recall and Precision parameters obtained from testing data of each of the models are shown in Table 5 by classes of target field. Recall (proportion of a class that is correctly predicted) and Precision (proportion of a proper predicted class) indicators calculated in model C5.0 are high in all classes that show the proper classification of the model. In the other three models, Recall and Precision indicators are high in some classes and moderate in other classes.

Comparing accuracy, Recall and Precision, it is concluded that among the four models fitted, C5.0 decision tree has the best performance in the target field classification. The rules generated by this model with a minimum support of 50 records are presented in Fig.9. These rules by fishbone diagram are shown in Fig. 9. The fishbone diagram is a causeandeffect diagram that can be used to identify the potential (or actual) cause(s) for a performance problem. Fishbone diagrams provide a structure for a groups discussion around the potential causes of the problem. In a typical Fishbone diagram, the effect is usually a problem needs to be resolved, and is placed at the ”fish head”. The causes of the effect are then laid out along the ”bones”, and classified into different types along the branches. Further causes can be laid out alongside fur

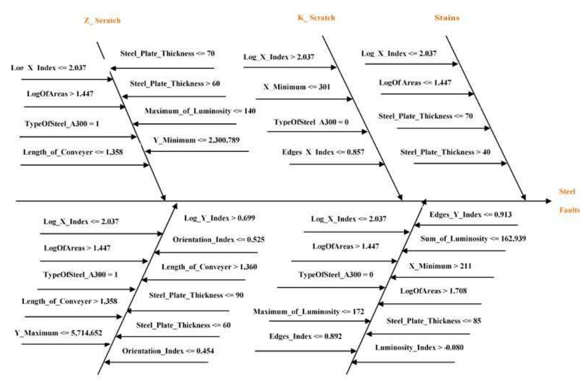


Fig. 9: The Rules generated by C5.0 model

The least amount of confidence in the above rules is 96 percent. The confidence amount in the rules listed above in data test set is calculated. Calculated accuracies are 100 percent, 98.7 percent, 100 percent, 95.65 percent and 100 percent, respectively; high level of confidence indicates the stability of the rules.

The gain chart provides a visual summary of the usefulness of the information provided by the classification models for predicting categorical dependent attributes. Gain in each class is defined as a proportion of

Table 4: Confusion matrices of individual models

		predicted outcome						
Model	Actual Outcome	1	2	3	4	5	6	7
c5.0	1	42	1	0	0	0	3	3
	2	0	59	0	0	0	0	0
	3	0	0	106	0	0	0	0
	4	0	0	0	27	0	0	0
	5	0	0	0	0	10	0	1
	6	1	0	1	0	0	118	7
	7	3	0	2	0	1	3	211
MLPNN	1	1	1	1	0	0	4	16
	2	0	50	1	0	1	1	6
	3	0	0	99	0	0	1	6
	4	0	0	0	24	0	0	3
	5	0	0	0	0	10	0	1
	6	1	4	2	0	1	76	43
	7	17	6	7	1	5	31	153
BN	1	33	4	0	0	1	5	6
	2	0	53	0	0	0	2	4
	3	0	0	99	0	0	1	6
	4	0	0	0	27	0	0	0
	5	0	0	0	0	10	0	1
	6	2	4	0	0	1	92	28
	7	17	6	4	1	3	38	151
Ensemble	1	39	2	0	0	0	4	4
	2	0	56	0	0	0	0	3
	3	0	0	103	0	0	1	2
	4	0	0	0	27	0	0	0
	5	0	0	0	0	10	0	1
	6	1	2	0	0	0	98	26
	7	13	1	2	0	3	22	179

Table 5: Indicators of Recall and Precision of each model in terms of classes of target field.

Class	Recall				Precision			
	C5.0	MLPNN	BN	Enemble	C5.0	MLPNN	BN	Enemble
Class1	0.86	0.55	0.67	0.80	0.91	0.6	0.63	0.74
Class2	1	0.85	0.9	0.95	0.98	0.82	0.79	0.92
Class3	1	0.93	0.93	0.97	0.97	0.9	0.96	0.98
Class4	1	0.89	1	1	1	0.96	0.96	1
Class5	0.91	0.91	0.91	0.91	0.91	0.59	0.67	0.77
Class6	0.93	0.6	0.72	0.77	0.95	0.67	0.67	0.78
Class7	0.96	0.7	0.69	0.81	0.95	0.67	0.77	0.83

the whole class in each percentile, and is obtained from the following formula:

$$\frac{\textit{The number of records of the class in each percentile}}{\textit{The total number of records in the class}} * 100$$

The gain charts for C5.0 Decision tree, neural network, Bayesian network and Ensemble models trained in SPSS Clementine for training and test subsets are shown in Fig. 10. The higher lines in the gain charts indicate better models, especially on the left side of the chart. These charts depict that the performances of the decision tree with C5.0 learning algorithm is the best model for training and test subsets. This suggests that the achievement of any class compared to the random detection, C5.0 decision tree model acts better than the other three models. In contrast, the neural network model is much poorer than the other three models.

5 Conclusion

The aim of this study was to compare classifier models in classification of steel Plates in terms of the number of faults. For this purpose, the models of C5.0 decision tree, Multi Perception Neural Network (MLPNN) and

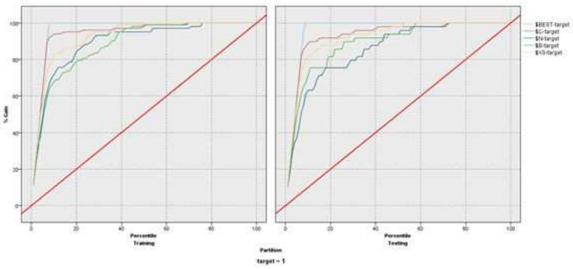


Fig. 10: The cumulative gain charts of the four models for training and test subsets

Bayesian network (BN) that have different classification methods as well as the Ensemble model in which the weighted voting is based on the confidence of these three models have been used. Qualitative indicators of steel sheets and the number of faults are considered as input and target fields, respectively. Results of the models fitting on training and testing data showed that the C5.0 model is superior to other models and Ensemble model in terms of accuracy, the indexes of Recall, Precision and Gain; and neural network model has weaker performance compared to other models. In addition to high indexes of models quality measurement, C5.0 decision tree model has more advantages than other models for providing rules. The results of the model fitting of training and testing data indicated that the model C5.0 is superior to Multi Perception Neural Network (MLPNN), Bayesian network (BN) and Ensemble model.

References

[1] Yazdchi M, Yazdi M, Golibagh A, Steel Surface Defect Detection Using Texture Segmentation Based on Multifractal Dimension, The 1nd International Conference on Digital Image Processing (ICDIP 2009), pp. 346-350, 2009, IEEE.

[2] Newman T.S , Jain A.K , A survey of automated visual inspection, Comput. Vision Image Understanding 61 (1995) 231-262.

[3] Debon A, Garcia-Daz J.C, Fault diagnosis and comparing risk for the steel coil manufacturing process using statistical models for binary data, J. Reliability Engineering and System Safety 2012; 100: 102114.

[4] Yang T, Mengyu F, Fang W, Steel plates fault diagnosis on the basis of support vector machines, J. Neurocomputing 2015;151: 296303.

[5] Yu LH, Ran W, Sam K, Xi ZW, Bayesian classiers based on probability density estimation and their applications to simultaneous fault diagnosis, J. Information Sciences 2014; 259: 252268.

[6] Dong H, Wang Z, Gao H, Fault detection for markovian jump systems with sensor saturations and randomly varying nonlinearities, Circuits and Systems I: Regular Papers, IEEE Transactions on 59 (10) (2012) 23542362.

[7] Yin S, Ding S.X, Haghani A, Hao H, Zhang P, A comparison study of basic data-driven fault diagnosis and process monitoring methods on the bench- mark Tennessee Eastman process, J. Process Control 22 (9) (2012) 15671581.

[8] Widodo A, Yang B.S , Support vector machine in machine condition monitoring and fault diagnosis, Mech. Syst. Signal Process. 21 (6) (2007) 25602574.

[9] Chongwatpol J, Prognostic analysis of defects in manufacturing, J. Industrial Management & Data Systems, Vol. 115, 2015. pp. 64 87.

[10] Yin S, Ding S , Xie X, Luo H, A review on basic data-driven approaches for industrial process monitoring, IEEE Trans. Ind. Electron. 61 (11) (2014) 64186428.

[11] Du W, Zhan Z, Building decision tree classier on private data, in: Proceedings of the IEEE International Conference on Privacy, Security and Data Mining, vol. 14, Australian Computer Society, Inc., 2002, pp. 18.

[12] Zou H, Hastie T, Tibshirani R, Sparse principal component analysis, J. Comput. Graphical Stat. 15 (2) (2006) 265286.

[13] Yin S, Zhu X, Kaynak O, Improved pls focused on key performance indictor related fault diagnosis, IEEE Trans. Ind. Electron. (2014).

[14] Seng-Yi, S.D.L. and W.L. Chang, Application of multiclass support vector machines for fault diagnosis of field air defense gun. Expert Syst. Appli.38: 6007-6013. DOI: 10.1016/j.eswa.2010.11.020, 2011.

[15] Frank A, Asuncion A, UCI machine learning repository Irvine. University of California, 2010.

[16] Berry, M., Linoff, G., Data Mining Techniques for Marketing, Sales, and Customet Relationship Management. John Wiley & Sons, New York, 2004.

[17] Dheeraj R, Xiaogang S, Patricia A.P, Lori A.L, Mary S.M, Exploring factors associated with pressure ulcers: A data mining approach, J. Nursing Studies 2015;52: 102111.

[18] Rudiger, W., Jochen, H, 2000, CRISP-DM: Towards a Standard Process Model for Data Mining, 6:11.

[19] Han J, Kamber M, Data Mining: Concepts and Techniques, Second ed, Morgan Kaufmann Publishers, 2006, pp. 285464.

[20] Zhang D, Zhou L, Discovering golden nuggets: data mining in nancial application, IEEE Transactions on Systems, Man and Cybernetics 34 (4) (2004) Nov.

[21] Ngai E.W.T, Hu Y, Wong Y.H, Chen Y, Sun X, The application of data mining techniques in nancial fraud detection: A classication framework and an academic review of literature, J. Decision Support Systems 2011;50: 559569.

[22] Ronowicz J, Thommes M, Kleinebudde P, Krysiski J, A data mining approach to optimize pellets manufacturing process based on a decision tree algorithm, European Journal of Pharmaceutical Sciences, 2015.

[23] He Y, Cheung Hui S, Exploring ant-based algorithms for gene expression data analysis. Art Intell Med 2009;47:23541.

[24] Li J, Huang K, Jin J, Shi J, A survey on statistical methods for health care fraud detection, Health Care Management Science 11 (3) (2008) 275287.

[25] Al Snousy MB, EI-Deeb HM, Badran K, Al Khlil IA, Suite of decision tree-based classification algorithms on cancer gene expression data, J. Egyptian Informatics 2011;12: 7382.

[26] Andrej K, Item response theory modeling for microarray gene expression data. Met Zvezki 2009;6(1):5167.

[27] Akkas E, Akin L, ubuku H.E, Artuner H, Application of Decision Tree Algorithm for Classification and Identification

of Natural Minerals Using SEM-EDS, J. Computers and Geosciences, 2015.

[28] Josh L.P, Chris P, Rachel L.G, Comparison of response surface methodology (RSM) and artificialneural networks (ANN) towards efficient extraction of artemisininfrom Artemisia annua, J. Industrial Crops and Products 2014;58: 1524.

[29] Samet H, Hashemi F, Ghanbari T, Minimum non detection zone for islanding detection using an optimal Articial Neural Network algorithm based on PSO, J. Renewable and Sustainable Energy Reviews 2015;52: 118.

[30] Shu HL, Pei HC, Pei YH, Data mining techniques and applications A decade review from 2000 to 2011, J. Expert Systems with Applications 2012;39: 11303-11311.

[31] Abdelaziz O, Amar RC, Marie OK, Data mining based Bayesian networks for best classification, J. Computational Statistics & Data Analysis 2006;51: 1278 1292.

[32] Chang SS, Yoon NK, Hyung SK, Hyoung SP, Min SK, Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches, J. Biomedical Informatics 2012 ;45: 9991008.

[33] Deng X, Liu Q, Deng Y, Mahadevan S, An improved method to construct basic probability assignment based on the confusion matrix for classication problem, J. Information Sciences 2016;340-341,pp.250-261.

[34] M. Hagan, H. Demuth and M. Beale, Neural Network Design (PWS publishing Company, USA, 1996).

[35] M. Zidan, A. Sagheer and N. Metwally, An Autonomous Competitive Learning Algorithm using Quantum Hamming Neural Networks, Proceedings of the International Joint Conference on Neural Networks (IJCNN, Killarney, Ireland, 2015), IEEE, 1-7 (2015).

[36] A. Sagheer and M. Zidan, Autonomous quantum perceptron neural network, arXiv: 1312.4149, 1-11 (2013).

[37] Huang, Guang-Bin; Zhu, Qin-Yu; Siew, Chee-Kheong (2006). "Extreme learning machine: theory andapplications". Neurocomputing. 70 (1): 489501.

Mohammad Ali Afshar Kazemi has received his Ph.D in Operations Research from Science and Research Branch, Islamic Azad University in Iran, in 2003. He has published more conference papers and journal papers. He is a faculty member of Islamic Azad University, Central Tehran Branch and has an associate degree.His main research interests are: dynamical systems, optimization theory, intelligent network, neural network, Data Mining, simulation, Queue theory.

Sima Hajian is Ph.D. student in Industrial Management Department from Islamic Azad University Tehran Central Branch in Iran.Her main research interests are: optimization theory, Data Mining.

Neda Kiani is Ph.D. student in Industrial Management Department from Central Tehran Branch, Islamic Azad University in Iran.Her main research interests are: dynamical systems, optimization theory, intelligent network, neural network, Data Mining.