https://dx.doi.org/10.24001/ijaems.3.7.3 ISSN: 2454-1311

Predicting Mechanical Properties of Galvanized Steels: Data Mining Approach

Zahra Ghasemivinche, Ali Zeinal Hamadani

Department of Industrial & System Engineering, Isfahan University of Technology, Isfahan, Iran

Abstract— *The purpose of this paper is to predict the mechanical properties of galvanized steel, using appropriate data mining techniques such as neural network, support vector machine, regression analysis and regression tree methods. It is found that by using the neural network technique one can get the best result for predicting the mechanical properties of galvanized steel according to the values of input parameters and also considering the effects of annealing temperature and line speed as the controlling parameters.*

Keywords— knowledge discovery, data mining, steel industry, prediction, galvanized line, neural network

I. INTRODUCTION

Knowledge is the most valuable asset of a manufacturing enterprise, as it enables a business to differentiate itself from competitors and to compete efficiently and effectively to the best of its ability [1]–[3]. In modern manufacturing environments, vast amounts of data are collected in database management systems and data warehouses from all involved areas, including product and process design, assembly, material planning, quality control, scheduling, maintenance, fault detection etc. and data mining has emerged as an important tool for knowledge acquisition from the manufacturing databases [4]–[6].

The constant search to improve product quality and to reduce the costs of production has a primary importance in any industrial plant [7]. One way of achieving these objectives is based on efficient methods and tools for data mining; an artificial intelligence that, by means of analyzing historical data, helps in understanding the industrial process more completely and developing strategies that lowers costs, improves product quality and increase production. In large steel and iron companies, one product of great interest is steel, coated in an immersion process using a bath of liquid zinc. This product, is known as "galvanized steel". Galvanized products have a long life and excellent corrosion resistance. Zinc produces twofold protection for the steel base, adding the galvanic action specific to this element to the physical barrier of the coating itself. Thus, the use of galvanized products is increasingly popular for a large number of applications. These applications can be used both indoors and outdoors. Construction, agriculture and domestic appliances are some of the most common applications. The use of galvanized products in the automotive industry has increased over the years as a response to the ever increasing requirement for improved corrosion resistance, paint adherence, surface finish, weldability and drawability [8].

In galvanizing line, the mechanical properties of sheet are of the most important properties and are very effective in determining the final quality of galvanized sheet. In almost all of galvanizing lines, after sheet production, the amounts of these properties are determined. In this paper, with using data mining techniques, the data are analyzed and a predictive model is presented that can predict the amounts of these properties before sheet is produced.

Furthermore, the effect of annealing temperature and the speed of galvanizing line as the most important controlling parameters in determining these properties will be studied.

Until now, in several researches, data mining techniques are applied in galvanizing line [8]–[10]. In all of them, one or more special techniques selected and applied. In this paper, for the first time all of applicable techniques for predicting a continuous variable are applied and the best technique is used for further analyses [11]. First, a description of the continuous galvanizing process and control system of mechanical properties is given. This process is currently used at the most important steel making company in Iran. Then, the steps of knowledge discovery in databases are performed. Finally, the best model is determined and is used for further analyses.

II. MATERIALS AND METHODS

In this section we explain the problem in galvanizing steel line and describe the data that we have used to implement our proposed model.

2.1 Problem Statement

The analyzed continuous galvanizing line produces galvanized sheets and coils using various grades of cold rolled steel strip base suited to the final use of the product required [12]. First, in order to form a continuous strip, coils are uncoiled and a shear cuts off the end of each coil so that they can be welded together. Then, the oil, dirt and oxides on the surface of the cold rolled coils are removed

https://dx.doi.org/10.24001/ijaems.3.7.3 ISSN: 2454-1311

before the strip enters the annealing section of the line. A good adherence, necessary to obtain an excellent coating quality, is achieved by perfect strip cleaning [13], [14]. The clean strip passes through the annealing furnace to give steel the desired properties by heating it to particular temperatures and profiles that determine the grain structure within the metal and prepare it for the galvanizing process. The entire process is carried out in a protective atmosphere that also reduces the surface of the strip used in the coating preparation step. The annealing cycle has the following phases:

(i) the cold strip is recrystallized by heating it to the highest temperature of the annealing profile
(ii) the strip temperature is maintained and grain growth takes place
(iii) An initial slow cooling period is used to control the metal texture
(iv) A fast cooling period prepares the steel for the strain aging treatment. The strip is cooled to a temperature appropriate for the coating step
(v) The over-aging step results in the precipitation of carbon to an extent that reduces the solute carbon. Thus, the strain aging tendency of the strip is reduced.

After the annealing step, the strip enters the molten zinc bath in order to form a zinc coating that is metallurgically bonded to the steel surface. The coating thickness is controlled by air knives installed after zinc bath. The control of the coating thickness is one of the most critical areas of development for coated sheets.

Finally, the coated strip is subjected to a chromate conversion treatment by the application chromate solutions to the strip surface. This chromate treatment results in a surface resistant to corrosion during storage and transport until the steel can be used in other applications [8]. Nowadays, the mechanical properties of galvanized sheets and coils are measured after their fabrication. Owing to the offline control, a large dead time occurs which makes the control solution inefficient. That is, the continuous galvanizing line produces a large amount of sheets or coils with undesired properties until appropriate actions are taken. Such a delay results in the cost for each coil of an inappropriate quality. The most important mechanical properties are yield strength, tensile strength and elongation.

2.2 Data Gathering and Preparation

First, all of effective variables in determining the mechanical properties discovered. These variables were: chemical composition, dimensional properties, annealing temperature and speed of strip in the annealing furnace. We gathered a data bank including 5210 records from effective variables was gathered. Then, data was preprocessed. After that, 14 variables and 875 records remained. Table 1 shows some statistical indices for both input and output variables.

Table.1: Variable used in developing the model

Variable	Unit	Min	Max	Mean	STD
Input variables					
Phosphorous	wt,%	3	21	7.995	3.441
sulfur	wt,%	1	17	9.265	2.924
Aluminum	wt,%	18	63	47.153	5.514
Nitrogen	ppm	20	113	38.144	9.850
Vanadium	wt,%	1	70	2.591	4.105
Manganese	wt,%	177	644	220.358	30.455
Silicon	wt,%	1	157	10.365	6.343
Carbon	wt,%	29	162	46.714	10.226
Al/N2	-	0.425	2.75	1.323	0.395
Annealing Temperature	° C	710	748.8	733.38	6.826
Strip speed in furnace	m/min	35.1	89.5	82.172	12.434
Thickness	mm	0.4	2	0.574	0.232
Width	mm	777	1250	1136.8	128.7
Output variables					
Yield strength	N/m2	147	384	294.159	23.557
Tensile strength	N/m2	278	422	368.097	15.5
Elongation	-	20	45	36.048	3.163

As a previous step to the modeling, it is often useful to visualize the experimental data in order to observe their structure, possible outliers, different groups, etc. In this work, principal component analysis is used for this purpose. The principal component analysis transforms a set of correlated variables into a number of uncorrelated variables, called principal components, which are ordered by reducing variability. The uncorrelated variables are linear combination of the original variables. The first new variable contains the maximum amount of variation; the second one contains the maximum amount of variation unexplained by the first and orthogonal to the first, etc. PCA mapping in Fig.1 reveals the existence of one main cluster and some outliers.

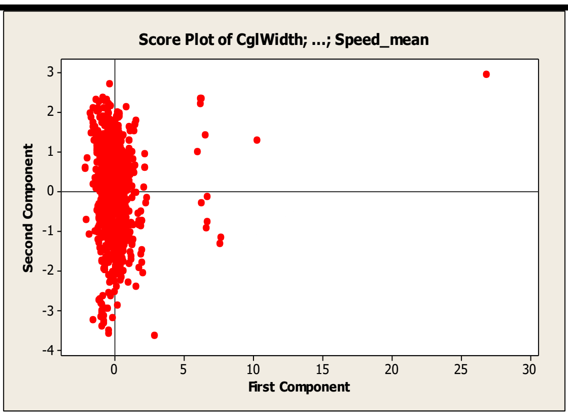

Fig. 1: PCA mapping

The outliers omitted from the model and PCA applied for second time.Fig.2, shows the result. After that, data is ready for further analysis.

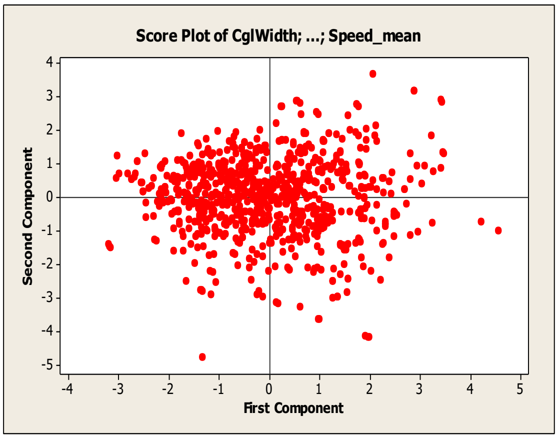

Fig. 2: PCA mapping after omitting outliers

2.3 Proposed Methods

All techniques that can be used for predicting the amounts of a continuous variable are:

(i) Support Vector Regression (SVR)
(ii) Neural Networks
(iii) Regression
(iv) Regression tree

For the case of SVR, two types of kernel function (polynomial and sigmoid) are more common and for the neural networks, a lot of networks can be used. For doing a proper comparison between different methods first, the best SVR and neural network will be defined then, they will be compared with regression and regression tree. Finally, the best model will be selected and used for further analysis. In order to improve accuracy of the results, a model was trained for each output instead of training a model with all of mechanical properties.

There are different indices for determining the accuracy of a model. Two most applicable indices are mean of absolute deviation MAD) and mean of squared error (MSE).8

$$MSE = \frac{\sum_{i=1}^{d}(y_i - \hat{y}_i)^2}{d}, \quad (1)$$

$$MAD = \frac{\sum_{i=1}^{d}| y_i - \hat{y}_i |}{d}, \quad (2)$$

III. RESULTS

In this section result of implementing different methods on predicting galvanized steel line presented and discussed.

3.1 Support Vector Regression (SVR)

Support Vector regression (SVR) is a discriminative classifier formally defined by a separating hyper-plane. In other words, given labeled training data (supervised learning), the algorithm outputs an optimal hyper-plane which categorizes new examples. To determine the best kernel function, data is divided in to two groups; the first group for training or modeling which contains 80% of the data and the second group for testing the results or validation of the method with the reaming 20% of the data set. With training data set, the model is produced and its accuracy is checked by the second data set. In all models, sigmoid function gives a smaller deviation for the second data set and for the training data set, polynomial kernel showed better results. Since the accuracy of predicting test data set is more important and shows the degree of model generalization, therefore the kernel function is the best approach and it will be used in final comparisons.

3.2 Neural Network

Neural Network is a computer system modeled on the human brain and nervous system [15]. For this method, the first step is determining the structure of the network. Hence, some appropriate parameters should be specified for the type of network, number of hidden layers, number of neurons in hidden layers, activation function of hidden and output layers and training method. To obtain the best generalization, the data set was randomly split into three parts:

(i) Training set for training the neural network (60%)
(ii) Validation set for determining the performance of the neural network on unseen patterns during learning (30%). Learning process would be stopped at the minimum amount of validation set error
(iii) Testing set for checking the general performance of the neural network (10%). After determining the structure of a network, the training process will be started. Every network will be trained 1000 epoch and for the prevention of over-fitting, if in 100 consecutive epochs, the error of predicting validation set does not improve, the training process will be stopped.

Table 2: Comparison between different models in predicting mechanical properties

Output of network							
		Yield strength		**Tensile strength**		**Elongation**	
		MAD	MSE	MAD	MSE	MAD	MSE
Methods	SVR	16.9	475	11.1	134.6	2.3	10.5
	Regression tree	17.2	503.4	11.3	211.9	2.3	8.3
	Neural network	15.6	436.7	9.6	152.5	2.2	7.4

Finally, the best model with lower amount of validation error will be compared with other models. According to the neural network's properties, proper types and parameters for predicting mechanical properties, as continuous variables, are selected.

3.1 Result of Applying Different Methods

For the final comparison, 80% of the data is selected randomly for training and the remaining 20% for testing. To have a good validation result, the training and the test data were the same for all methods.

The regression method is used and the results show that the p-values for models are more than 0.05 for nearly all important input variables such as annealing temperature and strip speed. It means that these important variables are not effective in determining the mechanical properties. Furthermore, R2 values were less than 12% in all models. Accordingly, we can conclude that the regression method does not give an acceptable result. The results of comparison between SVR with sigmoid kernel function, the best structure of neural networks and regression trees for each of mechanical properties are summarized in Table 2.

As one can see, neural network is the best method for predicting the mechanical properties of the galvanized steel and therefore it is used for further analysis.

IV. DISCUSSION

In order to predict the mechanical properties of galvanized sheets, one can use the collected information about the chemical compositions, width and thickness of sheets, annealing temperatures and strip speeds through the predictive models, such that when a sheet is going to be produced in a galvanizing line, all of the uncontrollable variables and temperature and speed as controllable variables will be inputted to the model and obtain the proper results. Finally, by changing the controllable variables, the best value for them will be defined before starting the production.

It is important to point out, in predicting galvanized sheet is to predict the effect of annealing temperature and strip speed on mechanical properties. For this purpose, all of the uncontrollable variables will be fixed as the most occurred values (mode). It is interesting to use the obtained models for prediction of line speed effect on mechanical properties. For this purpose the line speed is changed from 35 to 90(m/min.) and the results are summarized in Fig. 3. As one can see, with increasing speed, yield strength and tensile strength will increase and elongation will decrease.

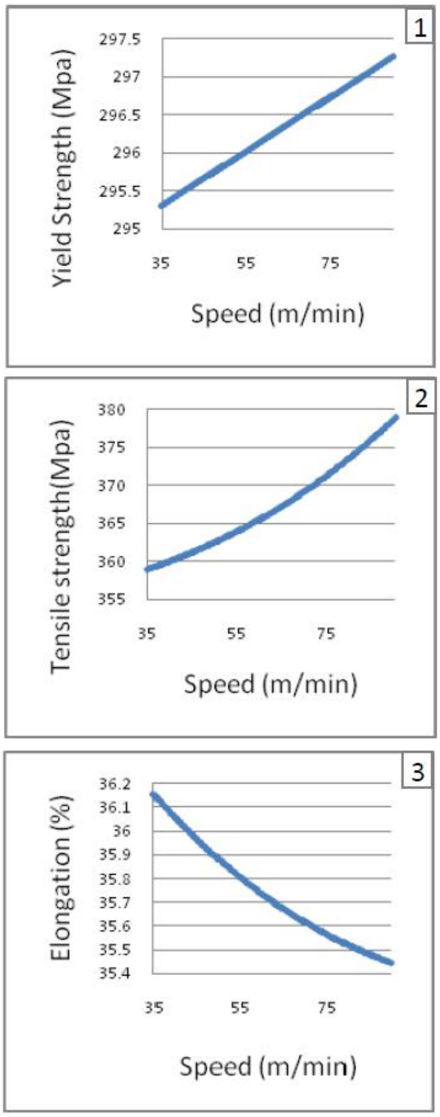

Fig. 3: Effect of speed on mechanical properties.

Also to see the effect of temperature on mechanical properties, the temperature is changed from 710 to 750 (C) by steps of 0.5 (C). The results are presented in Fig. 4. It is clear that with increasing temperature, yield strength will decrease; tensile strength will decrease until about

730 (C) and after that, it will increase and elongation will increase.

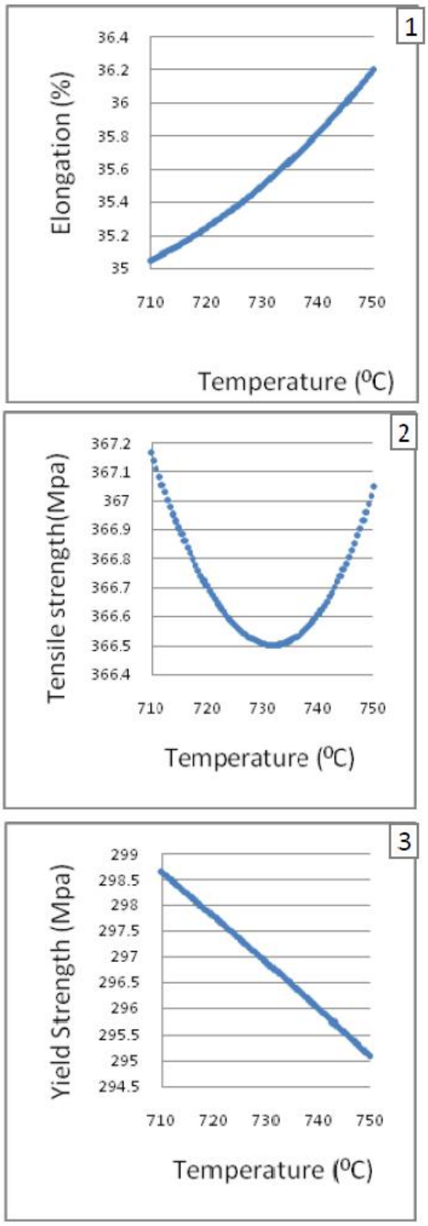

Fig. 4: Effect of temperature on mechanical properties.

V. CONCLUSION

In this paper, data mining techniques were applied for predicting the mechanical properties of galvanized sheets. For this purpose, after preprocessing, special techniques for predicting continuous response variables such as SVR, neural networks, regression and regression trees were used. It is concluded that the neural networks gives the best results and it is used for further analysis. By the selected model one can predict the mechanical properties of the galvanized sheets using the input variables before it is produced. Also it is possible to find out the optimum values for temperature and line speed to obtain the best results for mechanical properties.

For further research in this area, it is possible to assign some weights to the input variables according to the expert's opinion or using statistical analysis. Also, it's possible to develop models for predicting proper amounts of controlling variables with training models by some data which resulted good response values [16]. In addition, it is suggested to combine these predictive models with optimizing models for finding the best values of input variables [17].

REFERENCES

[1] S. Ameri, M. J. Fard, R. B. Chinnam, and C. K. Reddy, "Survival Analysis based Framework for Early Prediction of Student Dropouts," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management - CIKM '16*, 2016, pp. 903–912.

[2] J. A. Harding, M. Shahbaz, Srinivas, and A. Kusiak, "Data Mining in Manufacturing: A Review," *J. Manuf. Sci. Eng.*, vol. 128, no. 4, p. 969, Nov. 2006.

[3] M. J. Fard, S. Ameri, R. B. Chinnam, A. K. Pandya, M. D. Klein, and R. D. Ellis, "Machine Learning Approach for Skill Evaluation in Robotic-Assisted Surgery," in *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2016*, 2016.

[4] A. G. Marcos, A. V. Pernía Espinoza, F. A. Elías, and A. G. Forcada, "A neural network-based approach for optimising rubber extrusion lines," *Int. J. Comput. Integr. Manuf.*, vol. 20, no. 8, pp. 828–837, Dec. 2007.

[5] M. J. Fard, S. Chawla, and C. K. Reddy, "Early-Stage Event Prediction for Longitudinal Data," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2016, pp. 139–151.

[6] A. K. Choudhary, J. A. Harding, and M. K. Tiwari, "Data mining in manufacturing: a review based on the kind of knowledge," *J. Intell. Manuf.*, vol. 20, no. 5, pp. 501–521, Oct. 2009.

[7] M. J. Fard, S. Ameri, S. R. Hejazi, and A. Zeinal Hamadani, "One-unit repairable systems with active and standby redundancy and fuzzy parameters," *Int. J. Qual. Reliab. Manag.*, vol. 34, no. 3, pp. 446–458, Mar. 2017.

[8] J. B. Ordieres Mer', A. Gonz'lez Marcos, J. A. Gonz'lez, and V. Lobato Rubio, "Estimation of mechanical properties of steel strip in hot dip galvanising lines," *Ironmak. Steelmak.*, vol. 31, no. 1, pp. 43–50, Feb. 2004.

[9] M. J. Fard, P. Wang, S. Chawla, and C. K. Reddy, "A Bayesian Perspective on Early Stage Event Prediction in Longitudinal Data," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3126–3139,

Dec. 2016.

[10] F. J. Mart'nez-de-Pis?n, F. Alba-El'as, M. Castej'n-Limas, and J. A. Gonz'lez-Rodr'guez, "Improvement and optimisation of hot dip galvanising line using neural networks and genetic algorithms," *Ironmak. Steelmak.*, vol. 33, no. 4, pp. 344–352, Aug. 2006.

[11] A. Pern'a-Espinoza, M. Castej'n-Limas, A. Gonz'lez-Marcos, and V. Lobato-Rubio, "Steel annealing furnace robust neural network model," *Ironmak. Steelmak.*, vol. 32, no. 5, pp. 418–426, Oct. 2005.

[12] F. J. Martinez de Pison, J. Ordieres, A. Pernia, F. Alba, and V. Torre, "Reduce of adherence problems in galvanised processes through data mining techniques," *Rev. Metal.*, vol. 43, no. 5, pp. 325–336, 2007.

[13] M. J. Fard, S. Ameri, R. B. Chinnam, and R. D. Ellis, "Soft Boundary Approach for Unsupervised Gesture Segmentation in Robotic-Assisted Surgery," *IEEE Robot. Autom. Lett.*, vol. 2, no. 1, pp. 171–178, Jan. 2017.

[14] A. Bustillo, L. N. López de Lacalle, A. Fernández-Valdivielso, and P. Santos, "Data-mining modeling for the prediction of wear on forming-taps in the threading of steel components," *J. Comput. Des. Eng.*, vol. 3, no. 4, pp. 337–348, Oct. 2016.

[15] M. W. Craven and J. W. Shavlik, "Using neural networks for data mining," *Futur. Gener. Comput. Syst.*, vol. 13, no. 2–3, pp. 211–229, Nov. 1997.

[16] M. J. Fard, S. Ameri, and A. Z. Hamadani, "Bayesian Approach for Early Stage Reliability Prediction of Evolutionary Products," in *International Conference on Operations Excellence and Service Engineering (IEOM)*, 2015, pp. 1–11.

[17] E. Ge, R. Nayak, Y. Xu, and Y. Li, "Data Mining for Lifetime Prediction of Metallic Components."