© 2005 - 2010 JATIT & LLS. All rights reserved.

APPLYING DATA MINING METHODS TO PREDICT DEFECTS ON STEEL SURFACE

[1]SAYED MEHRAN SHARAFI, [2]HAMID REZA ESMAEILY

[1] Computer Engineering Department, Islamic Azad University, Najaf Abad Branch, Esfahan, Iran

[2] Computer Engineering Department, Islamic Azad University, Najaf Abad Branch, Esfahan, Iran

ABSTRACT

In the steel industry, especially alloy steel, creating different defected product can impose a high cost for steel producers. One common defect in producing low carbon steel grades is Pits & Blister defect. To eliminate this drawback, we need to grind the surface of the product. In some cases, the severity of defects may lead to scrap part of the product. Grinding cause waste of time and cost of production will be increased. Incidence of defects is related to several factors including material analysis and production processes. In this study we want to create a model to predict this fault with data mining methods including decision tree, neural network and association rules. We will compare the efficiency and accuracy of these models and select the appropriate model.

In this study, methodology used to perform data mining is CRISP (Cross Industry Standard Process for Data Mining). To create a decision tree and neural network respectively, entropy method and 24 hidden nodes are used. And to discover association rules, four members itemset is used. And applying data mining on data received from Iran Alloy Steel Company, the model created using the decision tree, has higher accuracy.

Keywords: *Data mining, neural network, Decision tree, Association rules, Alloy steel, Pits & Blister defect.*

1. INTRODUCTION

Data mining is an interdisciplinary science. It uses different methods including statistics, pattern recognition and machine learning to extract knowledge from mass of data [1]-[2]. In this study, predict Pits & Blister defect has been selected as a target. This type of surface defect is observed after rolling and it should be removed from the product with milling, or if not, it becomes scrap. This description is determined to reduce the incidence of this defect. It will cause reducing cost and time.

The origin of defects has been a topic of interest for many research projects, but it is still hard to find literature on the modeling of this defect. No physical model for this defect formation has been formulated so far. The problem for modeling is due to high dimensional variable group with their interactions.

In this study data mining techniques has been used to find the causes of this defect. Since data mining methods is based on data, selecting the correct data increase accuracy of models created by these methods [3].

Hence, we studied the production process and defected products with Pits & Blister in Iran Alloy Steel Company. Chemical analysis and casting parameters were identified as factors in the rise of this defect. And these data were selected as input to data mining algorithms.

Since we had a higher number of chemical analysis instances rather than casting parameters instances, first we applied only chemical analysis data to the algorithms and after that the composition of chemical analysis and casting parameters were applied to the algorithms. For predicting this defect we created models using decision tree, neural network and discovery association rules.

Several reasons for this defect are mentioned in the literature, including the effects of aluminum and nitrogen elements.

© 2005 - 2010 JATIT & LLS. All rights reserved.

The result of our research not only confirms the effects of aluminum in the defect but also identifies other elements as effective elements.

2. DATA GATHERING

At the beginning of the research, we held several meetings with experts to review and guess factors that cause this defect. The effective parameters in continuous casting and melting analysis known as factors in this defect. And since most pit & blister defects occur in low carbon steels, occurrence of this defect was evaluated in these steel grades.

Data used in this research belong to Iran Alloy Steel Company. This data was available in Progress database on the VAX machine. We had two different datasets for our research. First dataset includes only chemical analysis. It contains 2300 instances, and is related to the products from 2003 to 2009. Second dataset includes chemical analysis and effective parameters of casting. It contains 350 instances, and is related to the products from 2009. For data preparation, we either repaired instances with lost attributes or just deleted them. For repairing the instances we used most frequently values to fill the lost attribute. For example, to replace the amounts of silicon attribute, the most frequently value had replaced. Class attribute in this collection has two values. 523 indicates Pits & Blister defect and 0 indicates lack of defects.

Data mining operation was conducted in two stages. First phase was conducted only on melting analysis (dataset1) and second phase was done on the analysis of melting and effective parameters of casting (dataset2).

The set of melting analysis attributes is shown in Tables (1) and effective parameters in continuous casting attribute is shown in Tables (2).

Table 1.Attribiutes of melting analysis

Attribute	Ni	Cr	S	P	Mn	Si	C
Description	Nickel	Chromium	Sulfur	Phosphorus	Manganese	Silicon	Carbon

Continue Table 1.

Attribute	Sn	V	W	Cu	Co	Ti	Al
Description	Tin	Vanadium	Tungsten	Copper	Cobalt	Titanium	Aluminum

Continue Table 1.

Attribute	As	B	Ca	Ce	Nb	Sb	Mo
Description	Arsenic	Blond	Calcium	Cerium	Niobium	Antimony	Molybdenum

Table 2.Attribiutes of casting parameters

Attribute	**Description**
Param_Date	Sampling time
Cast_Speed	Casting speed
Cool_Flow1	Water flow rate(zone 1)
Cool_Flow2	Water flow rate(zone 2)
Cool_Flow3	Water flow rate(zone 3)
Bloom_Temp	Bloom temperature under straightening and withdrawal units
Air_press1	Air pressure(zone 1)
Air_press2	Air pressure(zone 2)
Air_press3	Air pressure(zone 3)
Water_press1	water pressure(zone 1)
Water_press2	water pressure(zone 2)
Water_press3	water pressure(zone 3)
Mould_Water_Temp	Mould water temperature
Secondary_Water_Temp	Secondary water temperature
Mould_Oscil_Freq	Mould frequency

3. APPLYING METHODS

© 2005 - 2010 JATIT & LLS. All rights reserved.

For applying decision tree, neural network and association rules, Microsoft Sql Server 2008 is used for modeling [4]. To perform data mining, data set was divided to three categories with different levels of carbon. This division is shown in Table (3).

Table 3.Records categories

Description	Name
Carbon<0.15%	product with very low carbon
0.15%<carbon<0.25%	product with low carbon
Consist Molybdenum	product with Molybdenum

To create a model, data randomly divided into two parts, namely training and testing datasets.

Training dataset includes 70 percent of our data and testing dataset includes 30 percent of our data. In training stage, we used decision tree with entropy settings to compute the weight of nodes for break points [5]. And we used 24 hidden nodes for neural network [6]. Also we used quaternary itemset for discovery association rules.

According to Figure (1), model of association rules for low carbon products (0.15% to 0.25% carbon) has 64 percent accuracy and is more accurate than the models created with decision tree and neural network.

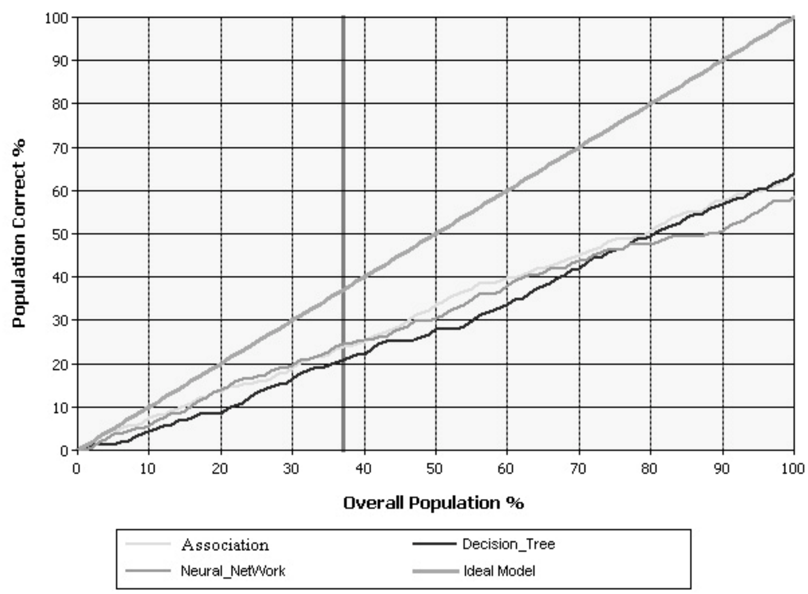

Figure 1. A comparison between algorithms for low carbon products (0.15% to 0.25% carbon)

According to Figure (2), the model created with decision tree for very low carbon products (0.0% to 0.15% carbon) has 83 percent accuracy and is more accurate than the models created with association rules and neural network.

© 2005 - 2010 JATIT & LLS. All rights reserved.

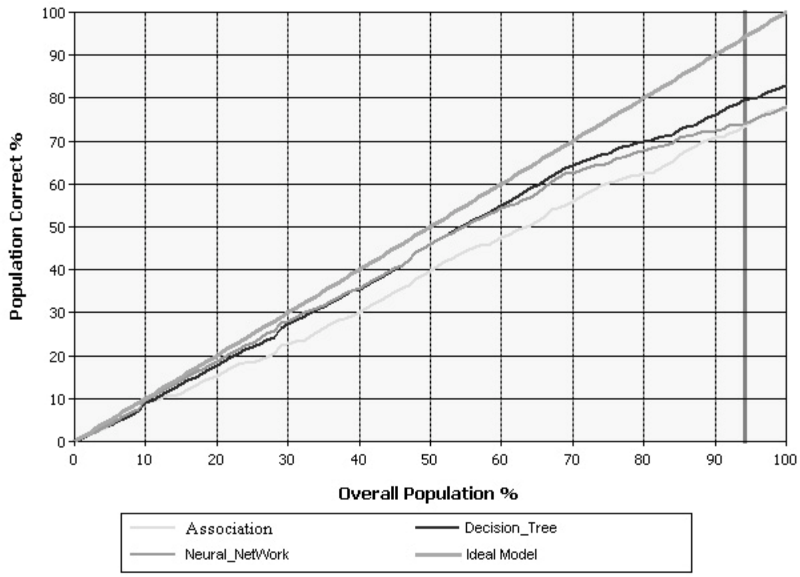

Figure 2. A comparison between algorithms for very low carbon products (0% to 0.15% carbon)

According to Figure (3), model of neural network for product with Molybdenum element has 76 percent accuracy and is more accurate than the models created with association rules and decision tree.

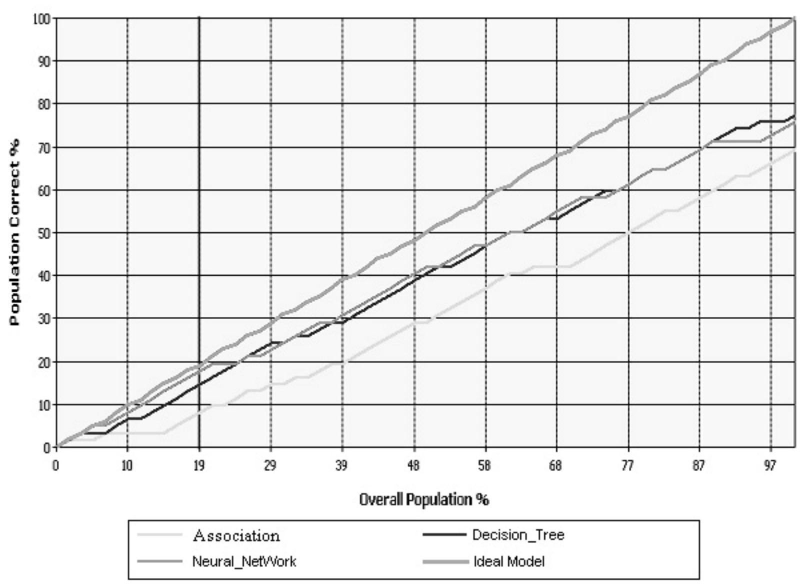

Figure 3. A comparison between algorithms for products with Molybdenum

Comparing figures 1, 2 and 3 shows that the decision tree model, resulting from data mining on very low carbon has the highest accuracy and it is appropriate.

Figure (4) shows the resulted tree after applying decision tree method on very low carbon data.

© 2005 - 2010 JATIT & LLS. All rights reserved.

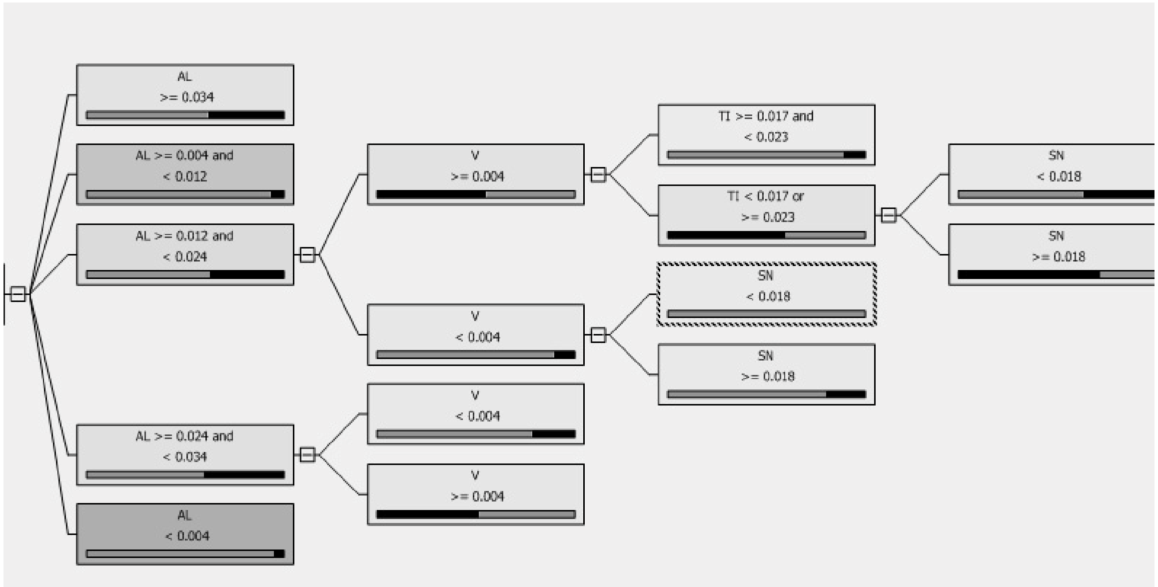

Figure 4. Tree resulting from the decision tree method for very low carbon products

No defect product with gray color and defect product with black color were shown. And as it is clear, the aluminum element in the creation of this defect is very impressive and products with aluminum less than 0.004 do not have defect. Also vanadium, tin and titanium elements are effective in creation or prevention of this defect. In the next step data mining algorithms applied on data sets including analysis of melting and casting parameters. Accuracy of created models in this step was about 50 percent, so unacceptable.

4. CONCLUSIONS

An effort to produce high quality steels in the steel industry continues. One way to achieve quality products is using the knowledge and experience from the past. And since usually high level of automation are used in steel production, after some time mass data is collected. Data mining algorithms can discover very useful knowledge and good patterns from the data.
The non-uniform data and the complexity of the underlying process made the data synchronization a challenging task, which was successfully accomplished before the data analysis. Pre-processing continued with the selection of the most meaningful variables, and the number of variables was reduced from 186 to 36 at the modeling phase.
The study also tried using these algorithms to reduce product defects with Pits & Blister defect. Applying different inputs on data mining algorithms can realize the fact that only chemical analysis has better results and especially very low carbon steel model gave us a better answer. And also confirmed the influence of Aluminum element as that is mentioned in various articles [7]. In addition, the effect of titanium in the prevention of this defect was revealed. Using this model, the expected benefits are the following:

- Reduce the process time due to omitting grinding step.
- Reduce energy costs.
- There is an optimal tool for predicting defect and also considering that for producing products with no defect.

© 2005 - 2010 JATIT & LLS. All rights reserved.

Acknowledgements

The authors would like to express their gratitude to Mr Jafarzadeh, Mr Safarzadeh and Mrs Jafari at Iran Alloy Steel Company for their expert knowledge support.

REFRENCES:

[1] D.Olson;D.Delen; *Advanced Data Mining Techniques*,Springer,2008.
[2] J.Han;M.Kamber*;"Data Mining: concept and Techniques"*, Second Edition, Elsevier Inc, 2006.
[3] D.Hand;H.Mannila;P.Smyth; *Principles of data mining*, The MIT Press . 2001.
[4] J.MacLennan; B.Crivat; "*Data Mining With Microsoft SqlServer 2008*", Wiley Publishing, 2009.
[5] P.Tan;M.Steinbach;V.Kumar;" *Introduction to Data Mining*", Addison Wesley,2005
[6] M.Bramer; *Principle of data mining*, Springer, 2007.
[7]. Z.Ma,P.Valentine,A.Heinen, "*Improvement of castability and quality of continuously cast stee*l", ISIJ International,vol.40,No.1,pp.31-39,2000.