



Analysis of the factors affecting the iron ore pellet FEO index using data mining technique



Mahdi Nakhaeinejad^{1*}, Farzaneh Zarei²

¹ Department of Industrial Engineering, Yazd University, Safaeih, Yazd, 8915818411, Iran

² Department of Industrial Engineering, Science and Arts University, Daneshjoo Blvd. Yazd, 891671333, Iran

ARTICLE INFORMATION

Article history:

Received: September 22, 2020

Revised: November 23, 2020

Accepted: December 05, 2020

Keywords:

Data mining
Decision tree
CRISP-DM
Iron pellet
Regression

A B S T R A C T

One of the most critical factors affecting iron pellet quality is the reduction in FeO (Iron Oxide) index in the final product. This study aims to predict factors affecting the FeO (Iron Oxide) of iron pellets and find out the contribution of each factor to reduce the pellets FeO (the ideal amount is between 0.4 to 0.6) using data mining techniques. When the FeO index's value is in the optimal amount, the quality and price of pellets are higher. The data used in this study was collected from the pelletizing plant of Gol-E-Gohar in Sirjan, Iran, and the decision tree and regression algorithms are used in this analysis. Forty-five factors that can affect the FeO (Iron Oxide) index of the final product were studied, showing that the Magnesium Oxide and Airflow of the inlet fan of the indurating machine had the greatest impact on the FeO (Iron Oxide) of iron pellets.

*Corresponding Author

Mahdi Nakhaeinejad
E-mail: m.nakhaeinejad@yazd.ac.ir



This is an open-access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



© 2020 Some rights reserved

1. INTRODUCTION

Today, iron and steel industries are essential foundations of economic and industrial activities in many countries. One of the most important plants in the chain of steel production is the iron pelletizing plant, so any improvement in this plant's final production, i.e., iron pellets, is desired and can increase the steel's productivity production process. The iron pellet is used as the raw material for sponge iron and steel production, so its mechanical, chemical and thermal properties are important. Efficiency improvement in the steel production cycle is one of the goals of the industries engaged in this field, and many studies have been conducted in this field to improve the factors affecting the steel production cycle and

increase the quality or decrease the costs. To control and understand different factors for better production, plants' raw and interpreted data should be used to analyze this data and get meaningful and useful results. In this research, the increase in pellet production productivity has been investigated by decreasing FeO (Iron Oxide) values in pellets, and the ideal amount is between 0.4 to 0.6 [1]. In fact, in iron pellets, several criteria determine their quality; one of the most important of them is the amount of FeO. To analyze the raw data, data mining techniques can be used. Today the development of database systems and the high volume of stored data require a tool to process stored data and provide meaningful results for the users. To achieve this goal in the past, pure

statistical methods were used, but more recently the data mining techniques are also used [2] which are an interdisciplinary science providing various methods including statistics, pattern recognition, machine learning, and database to extract meaningful results out of the enormous raw data available in the database [3].

Numerous studies have used data mining techniques to improve efficiency in steel production industries. Due to the complexity of structure and process, assembling metal sheet products such as the body and dimensional control are challenging tasks to improve quality. Lian et al. [4] used a data mining technique to develop a diagnostic method. They used a multivariate statistical analysis method to study major changes in patterns and decision tree based on production and assembly processes. The cooling process of iron ore pellets in a circular cooler affects the pellets' quality and energy use. Yang et al. [5] optimized the cooling process of Iron ores pellets using mathematical modeling and data mining techniques. They studied the effects of important process parameters on pellet temperature profiles in the circular cooler to understand the whole cooling process better and find an optimal control strategy. Ronowicz et al. [6] used the decision tree algorithm in an industrial application to study the effects of formulation and sphericity on the pellet formation since the design of a new product always requires full awareness of the influence of alloying elements on the microstructure. Wilk-Kolodziejczyk et al. [7] improved the design of modern casting materials using the data mining technique. Since fatigue resistance is an important mechanical property and testing steel fatigue in laboratories is costly and time-consuming, Agrawal et al. [8] provided a fast prediction model of fatigue resistance with 98 % accuracy using data mining techniques. Akinci et al. [9] use the K-Means algorithm as data mining techniques to find the production defects in a system. They do this as a part of clustering analysis. Ghasemvinche & Hamadani [10] use the neural network technique to determine galvanized steel's mechanical properties and determine the most important affecting parameters in the production process. The steel production process requires a monitoring strategy such as fault detection to reduce the number of errors leading to significant losses. Correct fault diagnosis can help in making accurate decisions. Sharafi & Esmaeily [11] predicted the superficial steel defect using the

decision tree technique and neural networks, as any steel production defect can lead to high steel industries' high costs. The data mining technique is used in several papers as a prediction tool in the steel industries. Cuzzola & Aurora [12] used the data mining technique to predict the rolling process properties as a tool in the steel industry. Bustillo et al. [13] also use data mining techniques as a prediction tool for steel threading wear. Nkonyana et al. [14] evaluated the predictive methods based on data mining compare to classical predictive methods. The data mining technique is also used for effects of different parameters in a different process, Wilk-Kolodziejczyk et al. [15] used the data mining technique to investigate the process effect on the austempered ductile iron, and Regulski [16] used the data mining technique to explore the formation of ausferrite in compacted graphite iron. Effective parameters in the bond-valence model are also studied by Zheng et al. [17] With the help of the data mining technique. The data mining technique was also used to validate the recorded data of drilled section properties that are important for decisions taken on the mining process as in the study done by Wedge et al. [18].

The present research investigated the various factors that could affect the iron pellets quality and determined the most influential ones to increase iron ore pelletizing plants' final product quality. In none of the studies done before, the FeO (Iron Oxide) of pellets was studied as an affecting contributing factor in the index of iron pellets quality separately using the data mining techniques. This study's results can be used as a strategy for better control of FeO (Iron Oxide) and thus improve the quality of the produced iron pellets.

2. RESEARCH METHODS

One of the thriving markets that are constantly looking for better productivity is the base metals and steel industries. The steel industry is constantly demanding high-quality raw materials. This study investigated the increase in iron pellet quality by decreasing the FeO (Iron Oxide) of the pellets between 0.4 and 0.6. The iron pellet is one of the stages in the production cycle of iron ore to steel. In this study, the iron pellets' FeO is studied as an influential index of pellets quality.

The data mining techniques discover meaningful and new relationships, patterns, and trends by exploring large amounts of data stored in the raw data using pattern recognition techniques with mathematical and statistical methods

[19]. CRISP-DM model was used in this research, which presents a process model for data mining, which provides an overview of each data mining project's life cycle and includes the steps associated with a project, its related tasks, and the relationship between these tasks. A data mining project's life cycle consists of six stages: Understanding the business, understanding the data, preparation, modeling, evaluation, and development [20]. The chart of the CRISP-DM model is shown in Fig. 1.

The regression and decision tree algorithms are used to find the best method to analyze the raw data. The importance of various factors affecting the FeO (Iron Oxide) amount of iron pellets is determined. The industrial data used in this study were obtained from the Gol-E-Gohar Pelletizing Plant (Iran-Sirjan). Forty-five factors were studied using Clementine and Minitab software, and the accuracy of these models was determined.

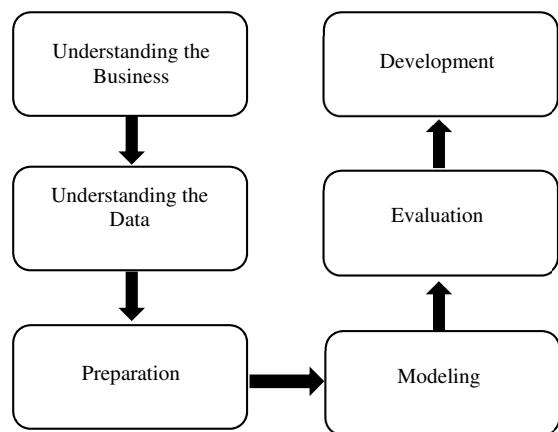


Fig. 1. CRISP-DM model

3. RESULTS AND DISCUSSION

The six stages that form every data mining project were listed in the section above, and the data related to this study for each stage is discussed here.

3.1. Understanding the Business

Successful deployment of data mining requires a good understanding of the area in which data mining is intended to be used. This study was conducted in the Gol-E-Gohar Mining & Industrial Company in Sirjan, Iran. Gol-E-Gohar Mining & Industrial Co., with its rich iron ore mines and plants, has many potentials and could be one of the pioneer industrial mining poles in the middle east and worldwide. Gol-E-Gohar Mining and Industrial Company emphasize the value and

importance of research, innovation, creation as a key element to development. This company is one of the main producers of iron ore pellets in Iran, and this study was conducted based on the data collected in the iron ore pellets plants of this company.

3.2. Understanding the Data

This step relates to the process of collecting data and storing them in a database. The data required for this study was gathered from the Gol-E-Gohar Palletization Plant. At this stage, all the variables and factors that may affect the process were identified through several meetings with the presence of professionals and process experts. In these brainstorming meetings, 45 factors that could affect the pellets FeO (Iron Oxide) were identified. These data are divided into two categories, the process variables and laboratory variables. The process variables data were recorded automatically, and laboratory variables data was recorded manually (Table 1 and Table 2).

Laboratory data were collected by daily testing of produced iron ore pellets and are listed in the following Table 1. Process data are divided into different characteristics of the raw pellets' process to form the final product. These data in the factories with similar technology for producing the iron ore pellets are nearly the same. So, for the 30 process variables used in this study, just the overall data of each variable was provided compared with the data collected in similar factories. The detailed data of the process variable can be confusing to the readers of this paper. These data are listed in Table 2. The laboratory data, which are 15 variables, are measured daily; the total number in this paper is 5375 records for one year. The control data, which are 30 variables, are measured three times a day, and for one year, it is a total of 32850 records. Therefore, the total number of records used in this paper is 39225.

A brief description of the Gol-E-Gohar pelletizing plant is provided here for a better understanding of the process data. In Fig. 2, the overall process of this factory is shown. Most of the data that can affect the FeO factor is related to the indurating machine that is shown in Fig. 2. A detailed and overall scheme of the indurating machine is shown in Fig. 3. All the process data are listed in Table 2, and their location is shown in Fig. 3. These process data are the main factors to control producing final iron ore pellets from the raw pellets.

Table 1. Laboratory variables

Parameter	Description
FeO	Iron Oxide
Al ₂ O ₃	Aluminum Oxide
MgO	Magnesium Oxide
CaO	Calcium Oxide
PORO	Porosity of Iron Ore Pellet
S	Sulfur
P	Phosphor
SiO ₂	Silicon Oxide
Fetot	Iron (Total)
+ 16	Pellets Size Greater Than 16mm
+ 14 -16	Pellets Size Between 16mm, 14mm
+ 10 - 12.5	Pellets Size Between 12.5mm, 10mm Pellets
+ 8 – 10	Size Between 10mm, 8mm
+ 6.3 – 8	Pellets Size Between 8mm, 6.3mm
- 6.3	Pellets Size Smaller Than 6.3mm

Table 2. Process variables

Parameter	Description
G3_00902_P01_LST	Air pressure at the exit of drying I section
G3_00903_T01_LST	Air temperature at the exit of the drying II section
G3_00904_P01_LST	Air pressure at the exit of drying II section
G3_00905_P01_LST	Air pressure at the exit of preheating section
G3_00905_T01_LST	Air temperature at the exit of preheating section
G3_00908_P01_LST	Air pressure at the exit of firing I section
G3_00910_P01_LST	Air pressure at the exit of firing II section
G3_00912A_T01_LST	Air temperature at the exit of firing II section
G3_00913_P01_LST	Air pressure at the exit of after firing I section
G3_00915_P01_LST	Air pressure at the exit of after firing II section
G3_00915_T01_LST	Air temperature at the exit of after firing I section
G3_00916_T01_LST	Air temperature at the exit of after firing II section
G3_00917_T01_LST	Air temperature at the exit of after firing III section
G4_000_T03_LST	Air temperature at hood section 7
G4_000_P09_LST	Air pressure at hood section 7
G4_000_P08_LST	Air pressure at hood section 6
G4_000_P07_LST	Air pressure at hood section 4
G4_000_P03_LST	Air pressure at hood section 7
G4_000_P02_LST	Air pressure at hood section 1
G4_000_P01_LST	Air pressure at hood section 2
G5_002M02_Z01_LST	Airflow of fan 02
G5_012M02_Z01_LST	Airflow of fan 12
G5_013_T01_LST	Air temperature after fan 12
G5_021AM01_Z01_LST	Airflow before fan 22
G5_021BM01_Z01_LST	Airflow before fan 12
G5_022M02_Z01_LST	Airflow of fan 22
G5_023_T01_LST	Air temperature after fan 22
G5_038_T01_LST	Air temperature after fan 37
G5_039_T01_LST	Air temperature after fan 38
G3_00702M01_S01_LST	Pellet car speed

Overall Scheme of Iron Ore Pelletizing Plant



Fig. 2. Overall scheme of iron ore pelletizing

Indurating Machine

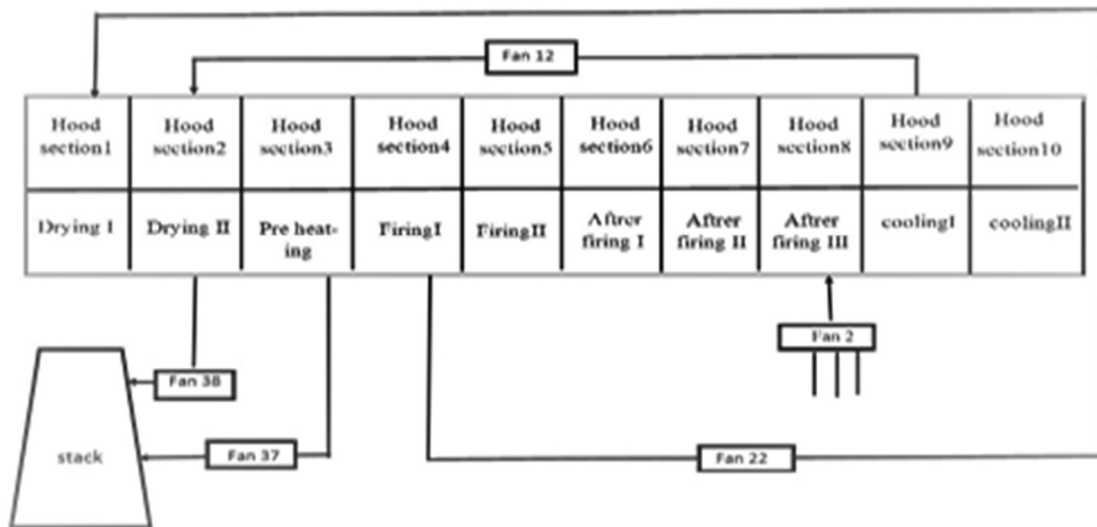


Fig. 3. Indurating machine

3.3. Preparation

After data collection, any possible error should be eliminated. At the data preparation stage, the collected data for one year was provided and the outliers were identified using Minitab software. To identify outliers, statistical histograms and outlier tests were used. As shown in Fig. 4, in FeO (Iron Oxide) histograms, most of the data are in the range of 0 to 1.5, and a few of the data are in the range of 1.8 and above, which are considered as outliers and incompatible with other data, shows the record errors caused by the operator. As can be seen in Fig. 4, there are several data in the zero range. This was investigated, showing that these data were recorded during the plant shutdown. The Minitab software was used to identify and clear the outlier data, and the red dots represent outliers, as shown in Fig. 5. After identifying the outliers, they were deleted. As shown in Fig. 6, all the remaining data are in the range of 0 to 1.4, and they are ready for analysis.

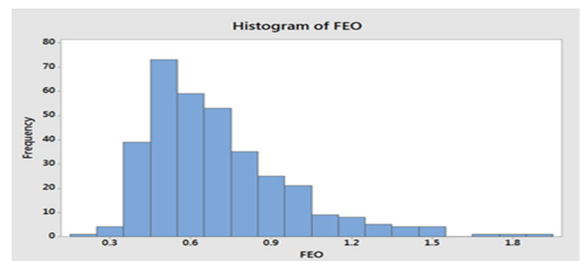


Fig. 4. FeO data histogram before eliminating the outliers

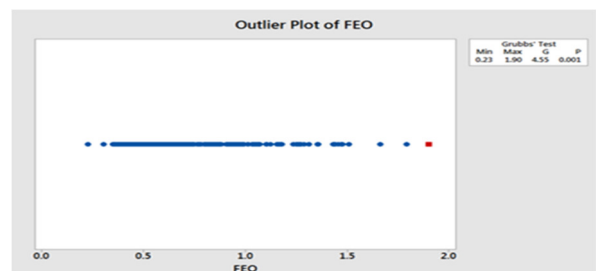


Fig. 5. FeO outlier data

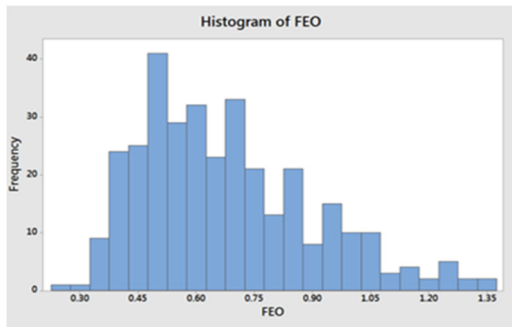


Fig. 6. Clear FeO data

3.4. Modeling

For the modeling stage, Clementine software was used to conduct the analysis and data modeling. This was done using a set of nodes called Stream. These nodes represent the set of activities that must be performed on the data, and the relationship between them indicates the direction of information transfer [2].

As you can see in Fig. 7, the input data was extracted from an Excel file, and data types were specified using the Type node. The Partition node was used to create a segmented field separating the data into subdirectories for training, testing, and validating different models. The decision tree and regression node were applied to the data using the table node to bring the data into a table format and use the Analysis node to analyze and evaluate the predictive models' ability to make accurate predictions. To model the data, operations were performed on each algorithm and the algorithm with the highest accuracy was used for the modeling.

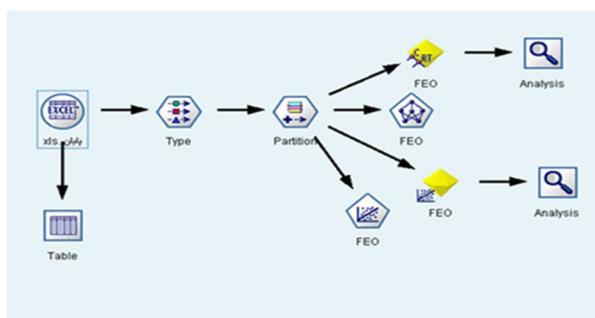


Fig. 7. Modeling plan

3.4.1. Based on Decision Tree Algorithm

Among the decision tree algorithms, the C&R algorithm was selected to fit the data, which is one of the most commonly used decision tree algorithms. The data type used in this paper is numerical. Dealing with numerical data in a decision tree algorithm is considered too, that the

algorithm can use the numerical data without any problem. The C&R algorithm is used in this paper, which is one of the most widely used decision tree algorithms. This algorithm can process predictor variables and numerical or categorical targets. The decision trees produced by C&R are binary, and there are exactly two branches for each decision node. The C&R algorithm divides the training data into subcategories based on the target values. The C&R algorithm performs a thorough search in all available and all possible divisions and develops the decision tree from the optimal point. The optimal distribution point is the one that has the highest value among all the dividing points in the parent node. After constructing the tree, the maximum pruning operation is performed by one of the pruning methods in the decision tree algorithm, starting from the leaves to the roots. A good tree is identified by evaluating its effectiveness on independent experimental data.

This process includes comparing each of the values in each test record's characteristics with the rules in the model that can be performed by the Clementine software so that this time-consuming operation be done automatically and quickly. Therefore, pruning the tree will make the results more general and reduces the errors. Fig. 8 illustrates the factors affecting FeO (Iron Oxide) pellets according to the decision tree model. MgO (Magnesium Oxide) and G4_000_P03_LST (Air pressure at hood section 7) are the most important variables affecting FeO (Iron Oxide) of the iron pellets (these variables are shown in figure 3). The constructed tree has 4 depths. Part of the constructed tree has been shown in Fig. 9. According to the decision tree shown in Fig. 9, the first branching is formed on the MgO (Magnesium Oxide) variable. If the MgO (Magnesium Oxide) is less than 1.797, which comprises 76% of the total data, and if other variables remain constant, then FeO (Iron Oxide) is predicted to be 0.6. The next branch is formed, the G4_000_P03_LST (Hood Air pressure). Suppose the G4_000_P03_LST (Hood Air pressure) is less than -6.6 and the MgO (Magnesium Oxide) is less than 1.797 out of 56 record, which comprises 24% of the total data, and other variables remain constant. In that case, FeO (Iron Oxide) is predicted to be 0.5.

3.4.2. Based on Regression

In the regression model, the Enter method was used for fitting. Enter method is an approach used to select variables in which all variables are

entered in a single step in the regression determination. The regression algorithm prioritizes the importance of the variables according to Fig. 10. As seen in Fig. 10, MgO (Magnesium Oxide) and G5_002M02_Z01_LST (Airflow of fan 02) are the most important variables affecting FeO (Iron Oxide) of iron pellets (these variables are shown in Fig. 3).

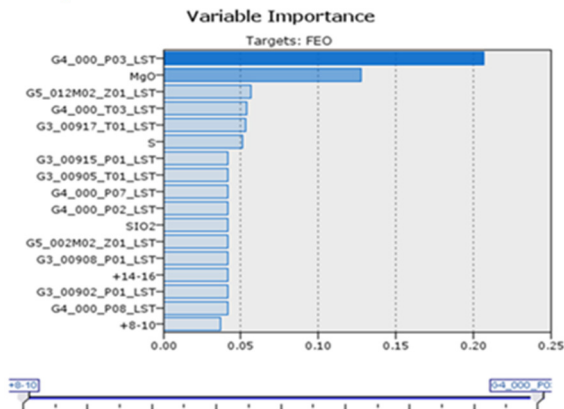


Fig. 8. The importance of indicators have been shown, based on the decision tree algorithm

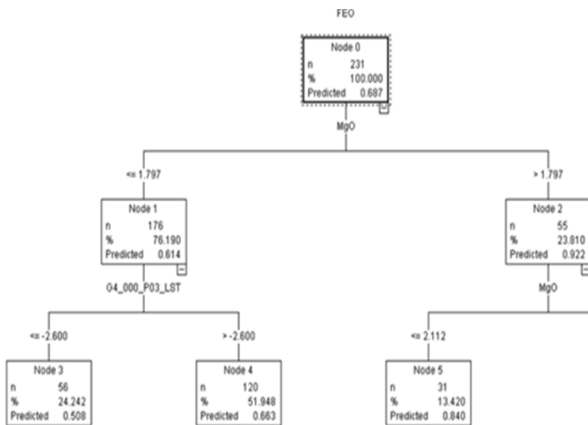


Fig. 9. Decision tree algorithm

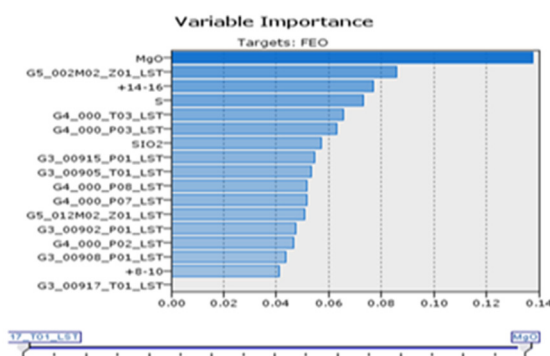


Fig. 10. The importance of indicators have been shown, based on the decision tree algorithm

3.5. Evaluation

At the evaluation stage, a model with higher quality is selected. Since the proposed method's validity has to be evaluated, the sample records have been divided into two training and testing datasets to evaluate the model's validity. Validity was tested with the comparison of the model results and the testing dataset. In this study, 70% of records were selected as the training set data, and the remaining 30% were used as the testing data. The validity of the model depends on the accuracy of the classification that was done with the testing data. This study used cross-validation with 4 iterations; each time 70% of the training data and 30% of the testing data were selected, and classification accuracy was measured. After 4 iterations, the final accuracy of the model could be determined. Table 3 and Table 4 show the decision tree's evaluation results and regression, and the final accuracy is summarized in Table 5; the regression with a value of 0.7 is more accurate than the decision tree.

Table 3. The results of the regression evaluation

Partition	1-Training	2-Testing
Minimum Error	-0.287	-0.422
Maximum Error	0.621	0.464
Mean Error	-0.0	0.006
Mean Absolute Error	0.115	0.138
Standard Deviation	0.149	0.184
Linear Correlation	0.742	0.707
Occurrences	248	117

Table 4. The results of the decision Tree evaluation

Partition	1-Training	2-Testing
Minimum Error	-0.347	-0.669
Maximum Error	0.574	0.814
Mean Error	-0.0	0.014
Mean Absolute Error	0.109	0.184
Standard Deviation	0.142	0.252
Linear Correlation	0.791	0.403
Occurrences	248	117

Table 5. The accuracy of algorithms

Accuracy	Algorithm
0.4	Decision tree
0.7	Regression

Linear correction is used to determine the type and degree of relation between a numerical

variable with another numerical variable. The correlation between two variables shows how the decrease or increase of one variable affects the decrease and increase of another variable. Therefore, the higher the correlations between variables, the more accurate the predictions. Therefore, the knowledge extracted from the regression algorithm is the most reliable knowledge derived from the studied data and can be the basis for extracting data mining rules.

3.6. Development

The most important variables in the regression algorithm are MgO (Magnesium Oxide) and Airflow of fan 02 (variable G5_002M02_Z01_LST) that are illustrated in Fig.11 by drawing the graph of FeO (Iron Oxide) intervals relative to MgO (Magnesium Oxide) and G5_002M02_Z01_LST (Airflow of fan 02), which is less than 1.4 for MgO (Magnesium Oxide) and greater than 50 for G5_002M02_Z01_LST (Airflow of fan 02), FeO (Iron Oxide) is in the range (0.6, 0.4).

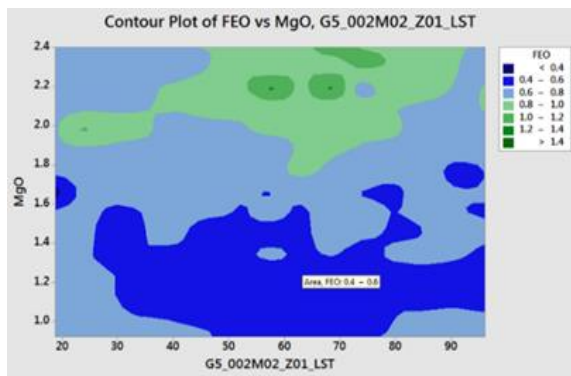


Fig. 11. Graph of the FeO relative to MgO and G5_002M02_Z01_LST

4. CONCLUSION

In this study, the most important factors affecting FeO (Iron Oxide) of pellets were identified using data mining knowledge based on the Gol-E- Gohar Sirjan Pelletizing Plant data. The present research investigated the various factors that could affect the iron pellets quality by reducing the FeO index (between 0.4 to 0.6) and determined the most influential ones that increase the final product quality of iron ore pelletizing plants. As described in the literature section in any of the previous study the data mining techniques did not use to analyze the FeO factor, FeO is one of the main factors to determine the quality of iron ore pellets and can even affect the price of the

produced pellets. The unique work done in this study is first of all to determine the main variables affecting the FeO factor with the help of experts of this industry with so many years of experience and factory data, and almost all the factors that can play a role in FeO of the iron ore pellets are considered. The results showed that matching of the regression algorithm was more than that of the decision tree, so the results extracted from the regression algorithm are the most reliable results from the data under study and can be the basis for data mining rules extraction. Then, data analysis revealed that from the 45 variables studied for affecting the pellet FeO (Iron Oxide) the two variable, the MgO (Magnesium Oxide) and G5_002M02_Z01_LST (Airflow of fan 02), had the strongest effect on pellet FeO index and thus more control over these two factors can help improve pellets production.

According to these results, magnesium oxide should be reduced to 1.4 to reduce the FeO content of iron ore (between 0.6 to 0.4) and a high-quality pellet can be achieved by combining the iron concentrate feed from different iron pits. The other factor to reduce the pellets FeO (between 0.6 to 0.4) is by adjusting the hood pressure (to zero), which increases the Airflow of fan 02 (G5_002M02_Z01_LST variable). An increase in the amount of air entering the high oxygen furnace improves oxidation in the preheating area and decreases FeO (Iron Oxide). So, based on the findings of this paper the operators of iron ore pellets production with the typical process as the one considered in this paper can produce pellets with higher quality by focusing more on these two factors (Magnesium Oxide and Airflow of the ambient fan), and knowing how important the other factors are by having this knowledge that these two factors have a greater impact on the FeO index.

REFERENCES

- [1] T. Umadevi, N. F. Lobo, S. Desai, P. C. Mahapatra, R. Sah, and M. Prabhu, "Optimization of Firing Temperature for Hematite Pellets," *ISIJ Int.*, vol. 53, no. 9, pp. 1673–1682, 2013, doi: [10.2355/isijinternational.53.1673](https://doi.org/10.2355/isijinternational.53.1673).
- [2] M. Ghazanfari, S. Alizadeh, and B. Teymourpour, *Data mining and knowledge discovery*, 2nd edition. Tehran: Publish center of Iran University of Science & Technology, 2011.

- [3] H. Baghchehband and A. A. Niknafs, "Predicting the amount of molybdenum in mines using data mining tools," *1th Iranian student Conference on Information Technology*. Iran, 2010. Available: <https://civilica.com/doc/88107>.
- [4] J. Lian, X. . Lai, Z. . Lin, and F. . Yao, "Application of data mining and process knowledge discovery in sheet metal assembly dimensional variation diagnosis," *J. Mater. Process. Technol.*, vol. 129, no. 1–3, pp. 315–320, Oct. 2002, doi: [10.1016/S0924-0136\(02\)00691-X](https://doi.org/10.1016/S0924-0136(02)00691-X).
- [5] G. Yang, X. Fan, X. Chen, X. Huang, and X. Li, "Optimization of Cooling Process of Iron Ore Pellets Based on Mathematical Model and Data Mining," *J. Iron Steel Res. Int.*, vol. 22, no. 11, pp. 1002–1008, Nov. 2015, doi: [10.1016/S1006-706X\(15\)30103-5](https://doi.org/10.1016/S1006-706X(15)30103-5).
- [6] J. Ronowicz, M. Thommes, P. Kleinebudde, and J. Krysiński, "A data mining approach to optimize pellets manufacturing process based on a decision tree algorithm," *Eur. J. Pharm. Sci.*, vol. 73, pp. 44–48, 2015, doi: [10.1016/j.ejps.2015.03.013](https://doi.org/10.1016/j.ejps.2015.03.013).
- [7] D. Wilk-Kołodziejczyk, K. Regulski, G. Gumienny, B. Kacprzyk, S. Kluska-Nawarecka, and K. Jaskowiec, "Data mining tools in identifying the components of the microstructure of compacted graphite iron based on the content of alloying elements," *Int. J. Adv. Manuf. Technol.*, vol. 95, no. 9, pp. 3127–3139, 2018, doi: [10.1007/s00170-017-1430-7](https://doi.org/10.1007/s00170-017-1430-7).
- [8] A. Agrawal and A. Choudhary, "A Fatigue Strength Predictor for Steels Using Ensemble Data Mining: Steel Fatigue Strength Predictor," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016, pp. 2497–2500, doi: [10.1145/2983323.2983343](https://doi.org/10.1145/2983323.2983343).
- [9] İ. B. Akinci and F. Ersoz, "Determination of Production Defects in Iron and Steel Sector by Data Mining," in *3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 2019, pp. 1–10, doi: [10.1109/ISMSIT.2019.8932741](https://doi.org/10.1109/ISMSIT.2019.8932741).
- [10] Z. Ghasemivinche and A. Z. Hamadani, "Predicting Mechanical Properties of Galvanized Steels: Data Mining Approach," *Int. J. Adv. Eng. Manag. Sci.*, vol. 3, no. 7, pp. 724–729, 2017, doi: [10.24001/ijaems.3.7.3](https://doi.org/10.24001/ijaems.3.7.3).
- [11] S. M. Sharafi and H. R. Esmaeily, "Application of data mining in predicting steel surface defect," *Artif. Intell. Appl.*, vol. 20, no. 2, pp. 87–92, 2010. Available: <http://www.jatit.org/volumes/research-papers/Vol20No2/4Vol20No2.pdf>.
- [12] F. A. Cuzzola and C. Aurora, "A Data-Mining Technology for Tuning of Rolling Prediction Models: Theory and Application," in *International Conference on Computational Science and Its Applications*, Springer, 2017, pp. 633–647. doi: [10.1007/978-3-319-62392-4_46](https://doi.org/10.1007/978-3-319-62392-4_46).
- [13] 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, doi: [10.1016/j.jcde.2016.06.002](https://doi.org/10.1016/j.jcde.2016.06.002).
- [14] T. Nkonyana, Y. Sun, B. Twala, and E. Dogo, "Performance Evaluation of Data Mining Techniques in Steel Manufacturing Industry," *Procedia Manuf.*, vol. 35, pp. 623–628, 2019, doi: [10.1016/j.promfg.2019.06.004](https://doi.org/10.1016/j.promfg.2019.06.004).
- [15] D. Wilk-Kołodziejczyk, B. Mrzygłód, K. Regulski, I. Olejarczyk-Woźeńska, and K. Jaśkowiec, "Influence of process parameters on the properties of austempered ductile iron (adi) examined with the use of data mining methods," *Metalurgija*, vol. 55, no. 4, pp. 849–851, 2016. Available: <https://hrcak.srce.hr/157481>.
- [16] K. Regulski, G. Rojek, D. Wilk-Kołodziejczyk, G. Gumienny, B. Kacprzyk, and B. Mrzygłód, "Approximation of Ausferrite Content in the Compacted Graphite Iron with the Use of Combined Techniques of Data Mining," *Arch. Foundry Eng.*, vol. 17, no. 3, pp. 117–122, 2017. Available: <http://journals.pan.pl/dlibra/show-content?id=102059&>.
- [17] H. Zheng *et al.*, "Data mining of iron(II) and iron(III) bond-valence parameters, and their relevance for macromolecular crystallography," *Acta Crystallogr. Sect. D*

- Struct. Biol.*, vol. 73, no. 4, pp. 316–325, Apr. 2017, doi: [10.1107/S2059798317000584](https://doi.org/10.1107/S2059798317000584).
- [18] D. Wedge, A. Lewan, M. Paine, E.-J. Holden, and T. Green, "A Data Mining Approach to Validating Drill Hole Logging Data in Pilbara Iron Ore Exploration," *Econ. Geol.*, vol. 113, no. 4, pp. 961–972, Jun. 2018, doi: [10.5382/econgeo.2018.4578](https://doi.org/10.5382/econgeo.2018.4578).
- [19] A. Berson, S. Smith, and K. Thearling, *Building data mining applications for CRM*. McGraw-Hill Professional, 1999. Available: <https://dl.acm.org/doi/book/10.5555/555454>.
- [20] H. Wiemer, L. Drowatzky, and S. Ihlenfeldt, "Data Mining Methodology for Engineering Applications (DMME)—A Holistic Extension to the CRISP-DM Model," *Appl. Sci.*, vol. 9, no. 12, pp. 1–18, Jun. 2019, doi: [10.3390/app9122407](https://doi.org/10.3390/app9122407).