Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 54 (2021) 178–183

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

10th CIRP Sponsored Conference on Digital Enterprise Technologies (DET 2021) – Digital Technologies as Enablers of Industrial Competitiveness and Sustainability

# A slag prediction model in an electric arc furnace process for special steel production

Maialen Murua [a,*] , Fernando Boto [a] , Eva Anglada [b] , Jose Mari Cabero [c] , Leixuri Fernandez [d]

[a] TECNALIA, Basque Research and Technology Alliance (BRTA), Mikeletegi 7, 20009 Donostia-San Sebastián (Spain).
[b] TECNALIA, Basque Research and Technology Alliance (BRTA), Mikeletegi 2, 20009 Donostia-San Sebastián (Spain).
[c] TECNALIA, Basque Research and Technology Alliance (BRTA), Astondo Bidea, Edificio 700, 48160 Derio (Spain).
[d] Sidenor I+D, Barrio Ugarte s/n, 48970 Basauri (Spain).

* Corresponding author. Tel.:+34-946430850 ; fax: 946073349. *E-mail address:* maialen.murua@tecnalia.com

**Abstract**

In the steel industry, there are some parameters that are difficult to measure online due to technical difficulties. In these scenarios, soft-sensors, which are online tools that aim forecasting of certain variables, play an indispensable role for quality control. In this investigation, different soft sensors are developed to address the problem of predicting the slag quantity and composition in an electric arc furnace process. The results provide evidence that the models perform better for simulated data than for real data. They also reveal higher accuracy in predicting the composition of the slag than the measured quantity of the slag.



*Keywords:* slag prediction model; soft-sensors; electric arc furnace process; Machine Learning

## 1. Introduction

In the industrial sector and in traditional industrial processes, such as the steel industry, more and more elements of the chain are digitized. The availability of data is increasing, but this data is difficult to structure, process and in some cases to collect, and it is also difficult to extract valuable information from it. In this context, efficiency in industrial processes is of great importance, especially when it is intended to reduce the ecological footprint, while maintaining the production availability.

The optimization of steel production tends to apply new technologies, combining automation, connectivity, digitization and artificial intelligence, which will make these processes more efficient. In addition, the intelligent combination of different tools, such as physical modeling and data-driven modeling, will play an important role in the digitization of the steel sector. In [2], the importance of digitization in this sector is exposed where it is argued that one of the most important lines is adaptive online control.

A soft-sensor (SS), is an online control tool which aims the online forecasting of certain variables that play an indispensable role in the quality control of an industrial process. In some industrial processes, there are variables which are difficult to measure online due to technical or economic limitations [18, 7]. A solution for this type of problem comes on the hand of the SS paradigm, which provides a reliable and stable estimation. Applications of this technology can be found in a multitude fields of industrial processes. The most common ones are: online forecasting, process monitoring and failure detection, as well as quality control.

The steel production based on electric arc furnace (EAF), consumes a large amount of energy, hence making the more efficient becomes a priority in any steel company. However, given the complicated characteristics of the process, online measurements with physical equipment are very complicated,

2351-9789 © 2021 The Authors. Published by Elsevier B.V.

10.1016/j.promfg.2021.07.027

and therefore SSs are appropriate in these harsh environments. For this reason, there are studies that try to obtain measurements online through inference, based on other variables that are easier to monitor. Electric energy consumption is one of the most important factor [10, 14], as well as the slag foaming [19, 13], which is a key factor in terms of quality and productivity in a EAF steelmaking process. The slag is a by-product of steelmaking process which is produced during the separation of the molten steel from impurities in steel making furnaces.

The objective of this work is to control by means of an SS the quantity and composition of the generated slag. The amount of slag is also a very important factor, which is an indicator of the efficiency of the process. Therefore, an estimation prior to the melting process is considered of great importance [8, 15] from an economic point of view. The online estimation of the composition of the slag allows to have a complete analysis of the process, as well as to increase the measurement frequency. This is especially relevant to provide a second life to the waste of a process, the composition is a characteristic to take into account.

The actual weighting of the slag is a complicated task. Due to the dynamic nature of the process the quantity removed from the furnace is not reliable enough. It is only reliable when all the steel is removed from the furnace, and this occurs once a week in this specific usecase. Thus, two strategies have been used when estimating the produced slag. First, a study of a model to estimate the real slag generated in a week has been made and second, a value calculated from the real composition has been taken as the variable to be estimated for the model. Finally, two elements of the composition of the slag are also predicted.

The remainder of the paper is organized as follows. Section 2 gives some background of the EAF process and special considerations for this work. Section 3 describes the proposed approach to build the soft-sensors of the slag. Section 4 is devoted to the discussion. Finally, Section 5 draws some conclusions.

## 2. The electric arc furnace process

The steel production can be divided in two different phases: primary and secondary metallurgy. During the primary metallurgy, the raw material (iron ore or scrap) is converted into steel (that is, an iron alloy with a C content lower than 2%). During the secondary metallurgy the alloy is refined and its chemical composition is adjusted [21].

The primary metallurgy is achieved across two main routes: Blast Furnace-Basic Oxygen Furnace (BF-BOF) route and EAF route. For the BF-BOF route the raw materials are predominantly iron ore, coal, and recycled steel (10-30%), while the EAF route produces steel using mainly recycled steel (about 100%) and electricity. Although nowadays only about the 28% of steel is produced by EAF [1], this route is a very relevant actor thanks to its major flexibility and sustainability [16]: steel can be produced from 100% of scrap instead of using iron ore contributing to the Circular Economy, and it uses electricity instead of coal contributing to the industry decarbonization [12].

An EAF usually consists of a bottom vessel lined with refractory and a cover with openings for the electrodes. The graphite electrodes (three in this case) are connected to alternate current and create the electric arc between them and the scrap. In order to do that, they are connected to a transformer which provides suitable current and voltage conditions to turn in the electric arc. The furnace has also different openings to extract the liquid steel and the slag, to add lime and to insert lances for the injection of oxygen, carbon or gas [6]. The operation typically takes place in several stages: First, part of the scrap is introduced in the furnace by means of one or several enormous baskets (in the studied cases, one basket loading between 50 and 90 Tn of scrap is used). The furnace is closed, and the scrap is partially melted by the electrodes to reduce its volume and release space in the furnace. Next the furnace is opened, the rest of scrap is introduced again by the basket in the furnace, which is closed again, and the scrap is completely melted. This process, ideally lasted around one hour, however sometimes there are slight variations in the process duration. Part of the additives, such as part of the lime, are added in each basket and the rest are added in this last stage together with the injection of oxygen, carbon and/or gas [20]. The scrap selection is done depending on the final composition desired for the produced steel, in order to control the residual elements (Cu, Sn, Ni, Cr,). In this investigation, the study of the steel production is limited to special steels as the employed data belongs to a special steel production. However, it would be possible to generalize to other steel production processes.

Table 1.Predictive variables summarized in groups of variables indicating the number of variables per group.

| Group of variable | Number of variables |
|---|---|
| Injection modules | 12 |
| Additions | 5 |
| Scrap (basket 1) | 12 |
| Scrap (basket 2) | 11 |
| Others | 4 |

During the melting process, the main chemical phenomenon that takes place is the Fe oxidation/reduction. It is achieved thanks to the lime (CaO) addition and the injection of oxygen. The iron oxidation $Fe+O_2 \rightarrow FeO$ is followed by a reduction process $FeO+C \rightarrow CO$. In addition, the oxygen oxidizes the excess of carbon ($C+O \rightarrow CO$), phosphorous ($2P+5/2O_2 \rightarrow P_2O_5$), silicon ($Si+O_2 \rightarrow SiO_2$), and manganese ($Mn+O_2 \rightarrow MnO$). These oxidation reactions are exothermic, so they also have an energetic contribution. This process generates a layer of oxides, which forms part of the slag, avoiding energy losses and protecting the melt steel. The CO bubbles floating up through the melt result in refining of the steel from non-metallic inclusions and hydrogen removal. Moreover, as the CO tend to exit to the atmosphere across the slag it produces a foamy effect. These CO partially burns in the atmosphere $CO+O_2 \rightarrow CO_2$ , and both, CO and $CO_2$ are removed by the exhausting system. The slag layer is later removed, and the steel is poured to a smaller furnace where the secondary metallurgy takes place [4].

The generated slag quantity per casting is one of the most critical process parameter, and should be under control. The slag is extracted from the furnace using a slag pot to be measured. Every casting, the generated slag is removed from the furnace using the slag pot, nevertheless there is always some quantity that remains in the furnace.

Table 1 shows the predictive variables considered in this investigation, summarized in groups indicating the number of variables in each of the groups. The variables of group *Others* are the number of uses of the furnace bottom, the furnace shell id, the presence/absence of the direct purging plug (DPP) and the day of the week. The furnace's bottom is repaired every two weeks and once a month, the furnace bottom is changed. Every time that the furnace bottom is changed, the use of the furnace bottom is initialized to zero. Regarding to the furnace shells, two different furnace shells are used to perform the EAF process. Sometimes, DPPs are used to insert argon to facilitate chemical reactions. The EAF process is organized in cycles. One cycle corresponds to a week where at the end of a week the furnace is cleared out. There is a time gap between the last casting of a cycle and the first casting of the next cycle.

Table 2. Minimum and maximum permitted ranges used to filter the castings.

| Variable | Minimum | Maximum |
|---|---|---|
| Scrap basket 1 + Scrap basket 2 | 135000 | 150000 |
| Temperature | 1600 | 1700 |
| Uses of the furnace bottom | 0 | 600 |
| Electrical consumption | 50000 | 80000 |

## 3. Predictive model generation

The presented problem of predicting the slag quantity was addressed as a regression task, where the regression was based on a number of predictive variables that describe the EAF process. The proposed Machine Learning-based approach consisted of first applying data filtering techniques and selecting the predictive variables. The slag prediction model was addressed using three different focuses: 1) a weekly model of the generated slag; 2) a calculated slag model per casting; 3) different slag composition models.

### *3.1. Data pre-processing*

Machine Learning algorithms consist of constructing associations between several predictive variables and a target variable. However, before starting building the models, some data pre-processing is commonly needed [11].

In this case, the data pre-processing was focused on filtering some castings that were not suitable for the analysis. Note, that one data sample corresponds to one casting. On the one hand, this step of filtering the castings was carried out with the help of an expert person with experience and knowledge in the EAF process. Table 2 shows the feasible ranges of values that are permitted to take the variables. The values that are out of the defined ranges are considered anomalies of the process. Note that here we include the variables of electrical consumption and temperature that were not mentioned in Section 2 as predictive variables. These variables give us valuable information, although they cannot be used as predictive variables, as they are outcome of the process, hence they are unknown before performing a casting.

On the other hand, the castings were filtered also using some statistical analysis. The feasible ranges for the variables belonging to the group injection modules and additions were calculated as shown in Equation (1) (with the exception of the minimum for the additions, that is 0).

$$min = \mu - d \times \sigma \qquad max = \mu + d \times \sigma \qquad (1)$$

where $\mu$ is the mean value of the variable, $\sigma$ the standard deviation and $d = 2$.

For illustrative purposes, Fig. 1 shows, the permitted range for the variable primary oxygen consumption oxipri4. The castings that are out of the band represented by a green line are removed from the analysis.

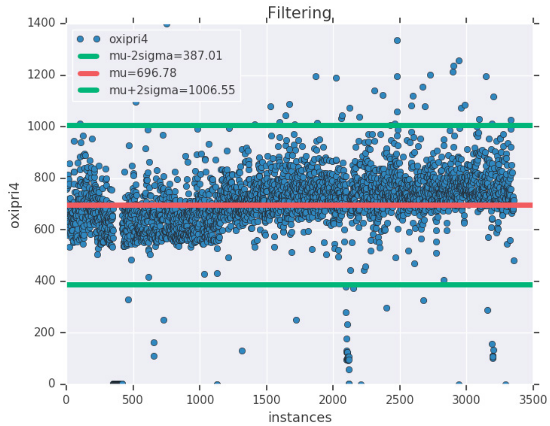


Fig. 1: Considered range for the variable primary oxygen consumption oxipri4. The permitted minimum and maximum are computed using statistical analysis.

### *3.2. Model Validation*

To define training and testing sets different techniques are used depending on the scenario. In k-fold cross-validation, the dataset D is randomly split into k subsets (the folds) $D_1, D_2, ..., D_k$ of approximately equal size. The sets are used for training and testing k times; at each time $t \in \{1, 2, ..., k\}$ the regressor is trained on $D \backslash D_t$ and tested on $D_t$ [9]. Leave-one-out cross-validation is a specific case of *k*-fold cross-validation when $k = N$, where *N* is the number of observations in the dataset.

The mean absolute error (MAE) is used to calculate the errors. It measures the average of the absolute errors [17]. If $\hat{y}=(\hat{y}_1,...,\hat{y}_N)$ is a vector of *N* predictions and $y=(y_1,...,y_N)$ is the vector of observed values, then MAE is calculated as shown in Equation (2).

$$MAE = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N} \qquad (2)$$

For better interpretability of the MAE, the normalized MAE (NMAE) is also used. The NMAE is calculated as shown in Equation (3) where $\boldsymbol{y}_{max}$ and $\boldsymbol{y}_{min}$ are the maximum and minimum values of $\boldsymbol{y}$, respectively.

$$NMAE = \frac{MAE}{(\boldsymbol{y}_{max} - \boldsymbol{y}_{min})} \qquad (3)$$

Moreover, the coefficient of determination, $R^2$ , is used which is the proportion of the variance in the target variable that is predictable from the input variables.

### *3.3 Model generation*

Various Machine Learning algorithms were considered in this research. The methods belong to the class of Ensemble Learning techniques. Random Forest regressor [3] is an Ensemble Learning method that combines Decision Trees with the notion of an ensemble. It belongs to the bagging family, which combines the results of multiple regressors modelled on different subsamples in order to reduce the variance of the predictions. Gradient-boosting regressor [5], is an Ensemble Learning algorithm and belongs to the family of boosting. It starts by giving the same weight to all observations and uses a weak regressor. The population distribution is updated at each stage.

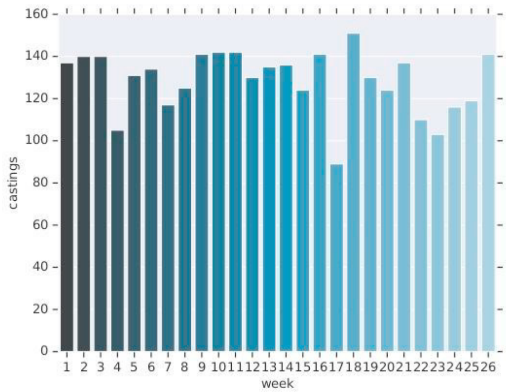


Fig. 2. Bar plot indicating the number of castings that belong to each week.

### 3.3.1. Weekly model of the generated slag

In this section, an SS for the weekly slag quantity is generated. In this scenario, the weekly measurements of the slag are close to the generated real slag as at the end of the week the EAF is clear out. The data contains the castings performed in 26 weeks and 42 predictive variables are considered in the analysis.

- Predictive variables: Weekly scrap quantity belonging to basket 1 and 2, weekly quantity of additions, weekly quantity of injection modules, furnace shell id, furnace bottom uses and number of castings per week.
- Target variables: Slag quantity per week.

In the case of most of the variables, the casting values are added per week. All the castings belonging to the same week are performed in the same furnace shell, thus variable furnace shell id is the same for all castings. In the case of the variable furnace bottom uses, the maximum value is considered. The performed castings per week are shown in Fig. 2, where X axis indicates the week number (from 1 to 26) and Y axis the performed number of castings. As it can be observed there is a slight variation in the number of castings. The average number of castings per week is $\mu = 128.46$ with a standard deviation of $\sigma = 14.44$.

The Random Forest method was used to model the data with leave-one-out cross-validation, and the obtained error was of MAE = 138250. Note that this measure is the average of absolute errors per week. If we divide it by the average number of castings per week, the MAE is 1076.21. In this case, the $R^2$ is not computed as it is not possible to calculate this metric at each iteration (we only have one casting at each iteration).

A more rigorous selection of predictive variables was employed in order to improve the accuracy of predictions. For this purpose, feature importance was computed using the Random Forest regressor. The selection of variables was carried out using a threshold value $\lambda$ (variables that present a value lower that $\lambda$ were removed).

The regression method was applied using a reduced set of selected variables with a threshold value of $\lambda = 0.001$ resulting in 23 variables instead of 42. With the reduced set of variables the obtained error was smaller *(MAE = 118803*, resulting on *MAE = 924.82* per casting). With this reduced set the NMAE = 0.11. Fig. 3 shows the scatter plot of real and predicted values of the slag quantity per week. The values are standardized to a [0, 1] range.

### 3.3.2. Simulated slag model per casting

In this section, an SS for the simulated slag model per casting is generated. The simulated slag is used instead of the measured one, as it is not assured that all the slag is removed when finishing the process. The simulated slag is calculated using a mass-balance modeling focused in the lime (Equation (4)).

$$slag_{sim} = \frac{100}{\mathrm{CaO\%}} \times (cc \times \mathrm{CaO\%}_{cc} \times R_{cc} + cd \times \mathrm{CaO\%}_{cd} \times R_{cd} + caa \times \mathrm{CaO\%}_{caa} \times R_{caa}) \quad (4)$$

where *cc*, *cd* and *caa* are different kinds of limes injected in the process. $\mathrm{CaO\%}_{cc}$ refers to the CaO% quantity in *cc* and $R_{cc}$ is the efficiency of *cc*. This is extrapolable to the variables *cd* and *caa.*

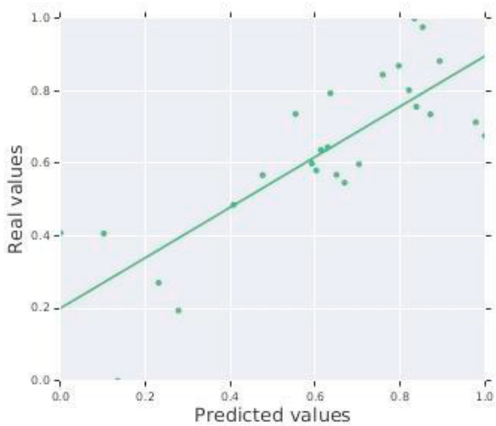


Fig. 3. Scatter plot of the real and predicted values of the slag quantity per week using a reduced set of predictive variables.

The following variables are considered in the model.

- Predictive variables: Variables related to injection modules, additions, scrap from basket 1 and 2 and others.
- Target variables: Simulated slag quantity per casting.

The model has been trained with 387 castings and 45 variables. The reduced set of castings is due to the fact that not in all castings an analysis of the composition of the slag is performed. For that reason, there is no information of the CaO% for all the castings and it is not possible to know the slag sim for those castings. The Gradient Boosting regressor was used to model the data using 5-fold cross-validation. The obtained error was of *MAE = 1302.67,* $R^2$ *= 0.7 and NMAE=0.05*. Fig. 4 shows the scatter plot of real and predicted values of the simulated slag per casting. The values are standardized to a [0, 1] range.

*3.3.3. Models of different compositions of the slag*

The composition of the slag is measured for some castings and predictive models of the lime percentage (CaO%) and silicon oxide percentage (SiO 2 %) are carried out.

As occurred with the simulated casting model, these two models were trained with 387 castings and 45 variables. The reduced set of castings is because of the composition of the slag is only measured for a reduced set of castings. The predictive variables involved in the analysis are the ones in Section 3.3.2.

- Target variables: CaO% and SiO 2 %.

The Random Forest regressor was used to model the data using 5-fold cross-validation. The obtained errors for the CaO% were *MAE = 2.22,* $R^2$ *= 0.35 and NMAE = 0.07;* whereas for the $SiO_2$ % model, *MAE = 1.31,* $R^2$ *= 0.41 and NMAE = 0.1*. Fig. 5 and Fig. 6 show the real and predicted values obtained from the generated two models of the slag composition.

The values are standardized to a [0, 1] range. The $R^2$ can be interpreted as the proportion of the variance in the target variable that is predictable from the input variables. In both cases the obtained $R^2$ s are quite low.

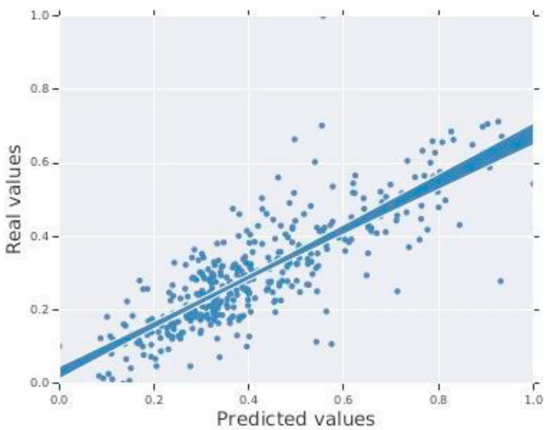


Fig. 4: Scatter plot of the real and predicted values of the simulated slag quantity per casting.

## 4. Discussion

The problem of predicting the slag quantity has been a challenging task due to the non-reliability of the data and the complexity of the EAF process itself. Thus, three different procedures were explored to address the problem.

Table 3 shows the NMAE of the four models considered in this investigation. The lowest value was achieved for the simulated slag model continued by the CaO%, whereas the worst ones were weekly and SiO 2 % models.

The relatively modest results obtained for the weekly model are due to the low quantity of weeks that are available. As it can be observed in Fig. 3, there are some predicted values that are far away from the regression line. A similar phenomenon occurs to the predicted values of the SiO 2 % model, shown in Fig. 6.

Usually, the data-driven models with simulated data perform better than the models fitted with real data. This explains that the lowest error was achieved for the simulated slag model.

Although the models generated for predicting CaO% and $SiO_2$% are not direct approaches to deal with the slag prediction problem, they gave information of interest and the predicted CaO% can be used to calculated the generated slag with Equation (4). The relatively modest results obtained in the predictive models are because of the uncertainties related to the quality of the data and complexity of the EAF process itself.

Table 3. Comparison of the generated models where achieved NMAE values are shown.

| Models | NMAE |
|---|---|
| Weekly model | 0.1 |
| Simulated slag | 0.05 |
| CaO% | 0.07 |
| $SiO_2$% | 0.1 |

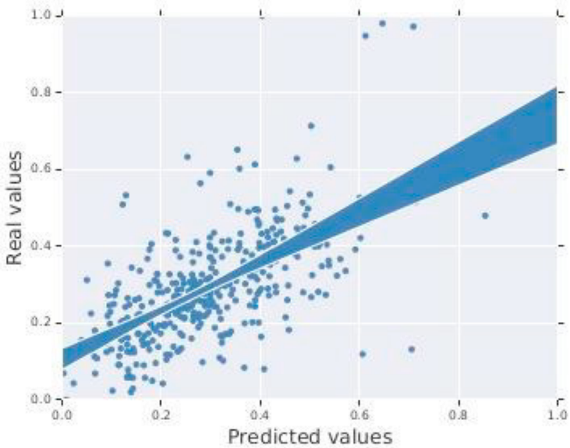


Fig. 5. Scatter plot between real and predicted values per casting for CaO % model.

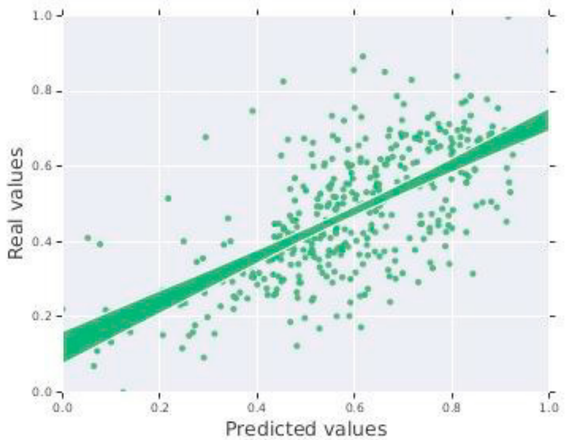


Fig. 6. Scatter plot between real and predicted values per casting for $SiO_2$ % model.

## 5. Conclusions and future work

This investigation introduces different approaches to deal with the slag prediction problem. This problem has been challenging because of non-reliability of the data and the complexity of the EAF process.

The best results were obtained for the slag simulated model, achieving highest NMAE. However, this model is built with simulated data, in contrast to the rest of the models. The only way to use the measured slag was by considering weekly, as it is not assured that all the generated slag is extracted at the end of a casting. It would be interesting to address the slag prediction model per casting, with measured slag quantity, assuring that the measured and generated quantities agree.

This study reveals the need to go deeper into the slag prediction problem with more reliable data. The analysis might be improved by adding the following information that currently is not available. The characterization of the scrap (e.g. composition) used in the castings, knowing the specific supplier of the additions of lime and coal per casting, and which of the injection points for coal is used for a casting.

This work also can be extended with an optimization scenario where the generated slag can be minimized and by predicting the foaminess of the slag. The foaminess of the slag is an interesting parameter to consider, as indicates how much slag can be extracted from the furnace (higher the foaminess easier to extract slag).

## Acknowledgements

The project leading to this research work has received funding from the European Unions Horizon 2020 research and innovation programme under grant agreement No 820670.

## References

[1] W. Association. World Steel in Figures, 2020. Worldsteel Association, 2021.

[2] T. A. Branca, B. Fornai, V. Colla, M. M. Murri, E. Streppa, and A. J. Schröder. The challenge of digitalization in the steel sector. Metals, 10(2):288, 2020.

[3] L. Breiman. Random forests. Machine Learning, 45:5–32, 2001.

[4] U. Camdali and M. Tunc. Calculation of chemical reaction energy in an electric arc furnace and ladle furnace system. Metallurgist, 60:669–675, 2016.

[5] J. Friedman. Stochastic gradient boosting. Computational Statistics &Data Analysis, 38:367–378, 2002.

[6] J. Jones, B. Bowman, and P. Lefrank. Electric furnace steelmaking. In R. Fruehan, editor, The Making, Shaping and Treating of Steel, chapter 10,pages 525–660. The AISE Steel Foundation, Pittsburgh, 1998.

[7] P. Kadlec, B. Gabrys, and S. Strandt. Data-driven soft sensors in the process industry. Computers & chemical engineering, 33(4):795–814, 2009.

[8] A. Kamaraj, G. Mandal, and G. Roy. Control of slag carryover from the BOF vessel during tapping: BOF cold model studies. Metallurgical and Materials Transactions B, 50:438–458, 2018.

[9] R. Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence, Montreal, Québec, Canada, pages 1137–1143, 20-25 August, 1995.

[10] M. Kovačič, K. Stopar, R. Vertnik, and B. Šarler. Comprehensive electric arc furnace electric energy consumption modeling: A pilot study. Energies, 12(11):2142, 2019.

[11] A. Kusiak. Feature transformation methods in data mining. IEEE Transactions on Electronics Packaging Manufacturing, 24:214–221, 2001.

[12] S. Lechtenbhmerab, M. Nilsson, Lars J.and hmanb, and C. Schneidera. Decarbonising the energy intensive basic materials industry through electrification - implications for future eu electricty demand. Energy, 115(3):1623–1631, 2016.

[13] J. Lee, J. Kim, H. Hwang, K. Son, W. Jeon, Y. Kim, H. Um, and H. Yang. Modeling of slag foaming height of electric arc furnace using stepwise regression analysis. Metallurgical Research & Technology, 117(1):114, 2020.

[14] V. Logar. Modelling and simulation of the electric arc furnace processes. In Proceedings of The 9th EUROSIM Congress on Modelling and Simulation, EUROSIM 2016, The 57th SIMS Conference on Simulation and Modelling SIMS 2016, number 142, pages 1–7. Linköping University Electronic Press, 2018.

[15] T. Prithiv, G. Thirumurugan, M. Madan, and A. Kamaraj. Thermodynamic assessment of steelmaking practices for the production of re-sulfur steels. Transactions of the Indian Institute of Metals, 73:1595–1603, 2020.

[16] L. Reh. Process engineering in circular economy. Particuology, 11(1):119–133, 2013.

[17] G. Seber and A. Lee. Linear Regression Analysis. Wiley, 2003.

[18] C. Shang, F. Yang, D. Huang, and W. Lyu. Data-driven soft sensor development based on deep learning technique. Journal of Process Control, 24(3):223–233, 2014.

[19] K. Son, J. Lee, H. Hwang, W. Jeon, H. Yang, I. Sohn, Y. Kim, and H. Um. Slag foaming estimation in the electric arc furnace using machine learning based long short-term memory networks. Journal of Materials Research and Technology, 2021.

[20] F. M. Studios. Electric arc furnace operations. https://www.youtube.com/watch?v=HKQ2GaXFI3w.

[21] L. Verdeja, J. Sancho, and J. Verdeja. Iron and steelmaking in the hird millenium. CIM Bulletin, 95(1063):88–95, 2002.