ADVANCED DATA MINING AND MACHINE LEARNING FOR SMART REFRACTORY CONTROL*

Gregor Lammer[1]
Manuel Forrer[2]
Markus Feuerstein[3]
Franz Pernkopf[4]

Abstract
The full digitization of the industry promises significant efficiency gains. This development begins to have an impact on the operation in steel plants, when decisions are made based on traceable data.This paper presents an approach to discover patterns in big data sets and applying methods of artificial intelligence for interpretation. As example, the identification of the main refractory wear mechanism in the hot spots and improvements applying this approach will be presented. Further, we applied this intelligent system for process optimization to calculated to optimal campaign length considering production, maintenance and refractory parameters.The paper also examines and discusses the operational impact and future applications.
Keywords: Machine Learning, Big data, Refractory systems, AI, Smart scheduling, condition monitoring

1 GregorLammer, Dipl.-Ing., RHI Magnesita, Vienna, Austria
2 Manuel Forrer, Dipl.-Ing.,ferrisol, Graz, Austria
3 MarkusFeuersteinmFeuSolutions,Graz, Austria
4 Franz Pernkopf, Assoc.Prof. Dipl.-Ing. Dr.mont, Graz University of Technology, Graz, Austria

** Contribuiçãotécnicaao22° Seminário de Automação e TI, parte integrante da ABM Week, realizada de 02 a 04 de outubro de 2018, São Paulo, SP, Brasil.*

1Introduction

Industry 4.0 is a popular term to describe the imminent changes of the industrial landscape, particularly in the production and manufacturing industry of the developed world. Yet the term is still used in different contexts and lacks an explicit definition. In this paper we define Industry 4.0 as the fourth industrial revolution focusing on the establishment of intelligent production processes and products.

Within the steel plant of the future, also considered a smart factory, CPS (cyber physical systems) will enable the communication between humans, machines, and products alike[i,ii]. Especially for companies in the steel industry it will be important to offer customized products that are superior in quality and competitive in price. This can be achieved by intelligent automation and reorganization of labor within the production system[iii]. The resolution of the automation pyramid towards self-controlling systems leads to an extreme amount of data, which can be extracted, analyzed, and visualized[iv]. Visualization, usability and knowledge accessibility are key drivers for user acceptance and require special attention in the development of Industry 4.0 projects.

Condition monitoring techniques are normally used on rotating equipment and other machinery (e.g., pumps, electric motors, internal combustion engines and presses), while periodic inspection using nondestructive testing techniques and fit-for-service evaluation is used for stationary plant equipment such as steam boilers, piping and heat exchangers. Currently, decisions of process adaptions are predominately made by humans based on experience. This is also valid for refractories[v]. In the future, the decision process will be increasingly assisted by self-optimizing and knowledgeable manufacturing systems. In this paper, we discuss condition monitoring and advanced data analysis for metallurgical vessels where we apply machine learning algorithms and advanced data visualization methods to support the decision base for refractory maintenance scheduling.

2Discussion

Refractories are designed to withstand harsh environments. Temperatures of more than 1,600 °C and corrosion-triggered wear limit the lifetime of refractories. The unpredictability of refractory behavior makes it hard for operation to match refractory lifecycles with plant lifecycles. One relief is to maintain the vessel lining (e.g. by gunning with an automated manipulator as shown in Figure 1) to extend refractory lifetime to operational needs. Nowadays, the maintenance cycles

Figure 1 - EAF gunning operation

are defined by experience and do not always foresee future pre-wear. Based on human evaluation – sometimes supported by laser measurement – the maintenance cycles are defined, but the predictability of the lining lifetime remains often unsatisfactory.

2.1 Technical Approach

In data science, where the amount of available data has increased dramatically over the recent past, intelligent systems modeling complex dependencies are developed to support the supervision of hazardous production processes such as steel making. We are at the beginning of a decade's long trend toward data intensive, evidence based decision making across many aspects of science and commerce. Steadily increasing data volume impose new demands such as computationally tractable algorithms, sensitive data raise the need for protecting privacy issues, and large amounts of unlabeled data require machine learning methods to be fully utilised.

APO builds on methods from machine learning and artificial intelligence to determine the condition of the lining based on several data sources without any human interference. Moreover, APO predicts the refractory wear and the lining lifetime. Furthermore, the influence of the production parameters on the refractory wear lining can be determined, and the most influential parameters are ranked. In addition to the visualization of the steel making process statistics, APO infers a maintenance proposal for optimal exploitation of the maintenance resources and refractory lining treatment.

Figure 2 shows the APO data processing pipeline. Currently, APO uses three main data sources, namely:

Laser measurements: During a production campaign, laser measurements are recorded to determine the remaining refractory lining thickness. These laser measurements are prone for optical insufficiencies such as dust which can lead to missing data values and insufficient measurement results. We introduced a pre-processing stage to remove outliers, fill measurement holesvi and to de-noise the laser measurements based on statistics of the local spatial neighborhood, to compensate these erroneous measurements.

Production parameters: During each heat several hundred production parameters are recorded such as temperatures, energies, durations, chemical ingredients, etc. A feature selection module, discussed further below, is introduced to determine a subset of production parameters which are useful for APO.

Maintenance data: Occasionally maintenance (gunning, fettling) is performed to repair the lining in zones of large wear rates (hot spots) to increase the lining lifetime. Here gunning data such as time, gunning mix, gunning consumption, maintenance areas, gunning mix per area are delivered to APO.

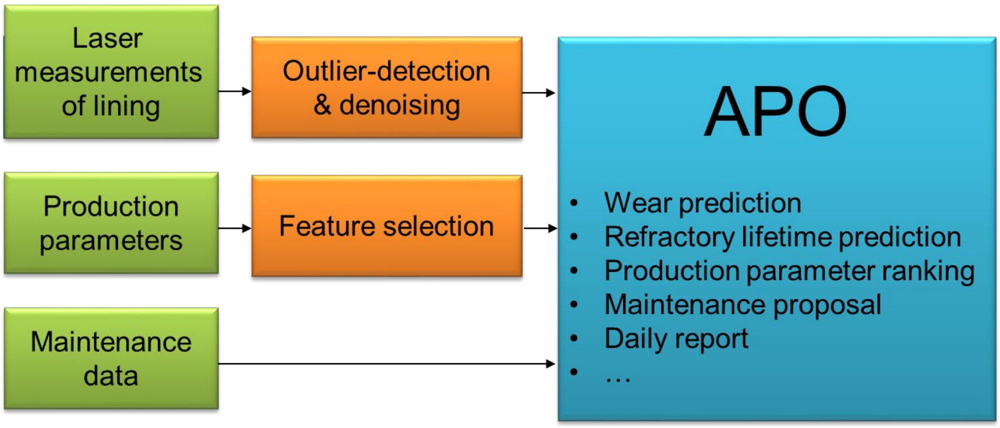

Figure 2 – APO data processing pipeline.

Feature selection: In real-world prediction problems the relevant features (i.e., production parameters) are often unknown a priori. Thus, the most useful features (with the highest informative content) for APO should be selected. Feature selection has become important for numerous pattern recognition and data analysis methods[vii,viii,ix].Many search heuristics have been proposed where an exhaustive search is usually computationally impractical. Even for a given cardinality of the final feature set d, the total number of different subsets

$$\mathrm{q} = \binom{\mathrm{D}}{\mathrm{d}} = \frac{\mathrm{D}!}{(\mathrm{D}-\mathrm{d})!\,\mathrm{d}!}$$

is too large for performing an exhaustive search, where D is the total number of production parameters. For this reason, many suboptimal deterministic and stochastic search heuristics have been proposed[x,xi]. Particularly interesting methods are based on genetic algorithms (GAs)[xii,xiii].GAs are optimization algorithms founded upon the principles of natural evolution discovered by Darwin. In nature, individuals have to adapt to their environment in order to survive in a process of further development. GAs turn out to be competitive for certain problems, e.g., large-scale search and optimization tasks.

APO Wear Prediction: To provide insights on APO we would like to focus on a simple wear approach. Laser measurements of the lining are not available for every heat due to the time needed to record a measurement. Between two consecutive laser measurements LM_{t} and $\mathrm{LM}_{\mathrm{t+1}}$ we do not know the current lining thickness. Let us introduce this time span as a slot.

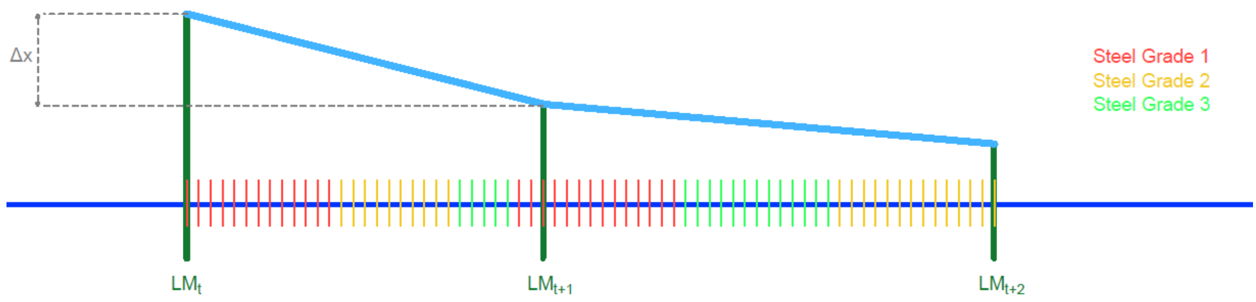

Figure 3 – Refractory wear (blue line) over heats with different steelgrades per slot

Figure 3 shows a sketch of the refractory wear over the heats including steel grades and laser measurements. On the one hand, the slot sizes may vary, on the other hand, within a slot, it is possible that several different steel grades are produced. Moreover, several maintenance actions could have occurred in each slot (not visualized in Figure 3).
The aim of this approach is to predict the refractory lining thickness based on produced steel grades, assuming, that each steel grade has its individual wear on the lining.

Least Squares Approach:
Having the definitions above in mind a simple first linear approach can be postulated. For this wear prediction approach, least squares methods are used where the weights $w_i$model the wear per heat for each steel grade. The least squares solution for $[w_1 \cdots w_{n+1}]^T$ of the following system of equations

$$\begin{bmatrix} SG_{1,1} & \cdots & SG_{1,n} & GC_1 \\ \vdots & \ddots & & \vdots \\ SG_{m,1} & \cdots & SG_{m,n} & GC_m \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_{n+1} \end{bmatrix} = \begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_m \end{bmatrix}$$

can be determined, where Δx_imodels the wear in a slot, GC_imodels the gunning frequency count per slot, $SG_{i,j}$ is the frequency count of produced steel grade $j$between two laser measurements, n and $m$denote the number of steel grades and data samples, respectively. Each line of the system of equations corresponds to the recorded data per slot. As a result, this simple approach performs well as long as the data noise is low. The prediction accuracy is insufficient in the case of noisy data.

For this reason, this approach is currently extended in various directions. In doing so, a subset of selected production parameter is included in the model. Furthermore, this model is extended by a Kalman-Filter to account for parameter adaptation over the campaign[xiv]. Additionally, we introduced a hybrid optimization objective which increases the prediction accuracy[xv]. Moreover, a Gaussian processes approach using selected production parameters instead of the steel grades is part of APO, too.

According to Bishop et al., clustering is the task of finding groups within a set of ofmulti- dimensional data points[xvi]. Hence, obviously, clustering is a representative of unsupervised learning, since the input data that is to be divided into groups, is unlabelled. Nevertheless, the partitioning into a certain number K of subsets, of course, should be optimal in some sense. The type of multi-dimensional input points (which we will also refer to as input vectors throughout this article) and therefore the meaning of each of the points' coordinates, depends on the data that is to be clustered and the clustering application itself. Obviously, finding suitable vectors and their respective coordinates is a rather crucial point. Hence, for the time being, simply assume that, in some mystical way, we managed to obtain our N multi-dimensional (or, more precisely D-dimensional) input vectors X = {x1, ..., xN } and that we use squared Euclidean distances as a measure of dissimilarity. Given this, we can intuitively restate our previous definition of a cluster more precisely: A cluster is a subset of X , where the distances between the vectors belonging to this cluster are small compared to the ones to other vectors not belonging to this cluster[xvii].

State of the art cluster algorithms like the K-Means clustering algorithm, which is a well investigated and frequently used approach, is usually one of the first off- the-shelf algorithms that one tries out when it comes to clustering data of any kind.
As described in, the K-Means clustering algorithm is based on the assumption that we have K D-dimensional vectors μk, the so-called prototypes. Each ofthese prototypes μk is associated with one of the K clusters. Given this, our goal is to

(1) assign each of the points to one (the closest or the optimal) of the K clusters and

(2) findoptimal values for each of the K prototype vectors μk.

Optimal in this context means that the resulting prototype assignment as well as the set of prototype vectors {μk} should minimize the sum of squared distances

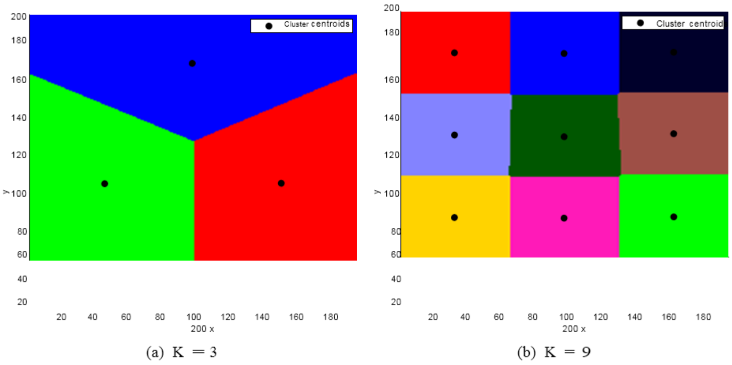

Figure 4 Example data after being clustered by K-Means algorithm with different numbers clusters of centroids K showing the typical Voronoi tessellation

between each point and the prototype vector that is assigned to it. By applying this optimization rule and given that we are using squared Euclidean distances as a measure of dissimilarity, the resulting partitioning is called the D-dimensional Voronoi tessellation[xviii] having the prototype vectors μk as the K centroids. See Figure 4 for an example of this tessellation, where we used two-dimensional vectors comprising the x and y coordinates of the respective points in the 2D-plane as input vectors.

3 Results

3.1 Lining Clustering

The refractory wear of different areas of the EAF's inner surface is influenced by the chosen refractory concept (mainly quality) and different production parameters. By applying a clustering framework we seek to identify areas where refractory wear characteristics show similar temporal evolution. This can be done by applying various clustering algorithms like k-means[xix] or affinity propagation[5] to the laser measurement data. We also introduced different transformations, the so-called input generation schedules, and applied them to the laser data in order to construct multidimensional vectors that serve as input for the clustering algorithm.

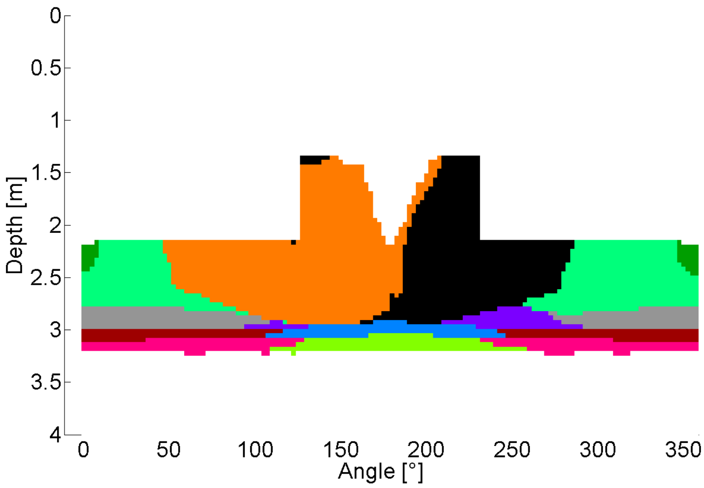

Figure 5 – Example of refractory lining clustering

For the analysis shown in Figure 5, we used more than 240 scans. One laser scan has more than 500.000 points. Areas in one color show similar wear behavior.

Based on such information, the zoning for different refractory material can be determined. High wear resistance bricks in the hot spots (green) and less performing bricks in the other areas. For each zone a refractory model can be trained and used for further analysis.

3.2 Daily Report

In data processing operational reporting is reporting about operational details that

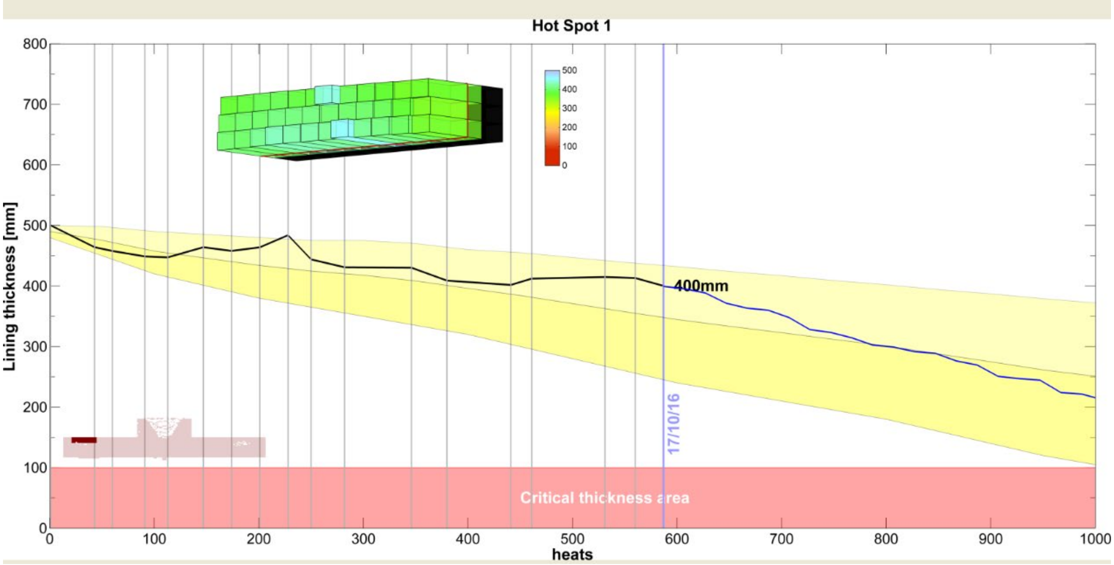

Figure 6 – Example for refractory wear prediction (blue line)

reflects current activity. Operational reporting is intended to support the day-to-day activities of the organization. "Examples of operational reporting include bank teller end-of-day window balancing reports, daily account audits and adjustments, daily production records, flight-by-flight traveler logs and transaction logs"[xx]. In APO we generate a Daily Report (DR) giving the operation manager an easy accessible overview of the current state of the furnace lining. The DR shows the refractory thickness over time or heats for specific areas (black line in Figure 6). The red zone labeled 'Critical thickness area' indicates the minimum required lining thickness for that area. The light blue vertical lines represent laser measuring. The 3D model shown is based on the last laser measuring. Next to the 3D model is a scale indicating the condition of each brick. This makes the assessment very easy. The blue line starting right of the last laser scan is the refractory wear prediction and allows to evaluate if the targeted lifetime of the lining can be reached.

APO can also include maintenance data in the DR. In the example given in Figure 7 the gunning data are represented with green vertical bars. The height of the bars indicates the amount of gunning material applied in this area. Each bar stands for one maintenance intervention, in this case gunning with the TERMINATOR XL. The doted blue line shows the cumulated gunning mix consumption from campaign start.

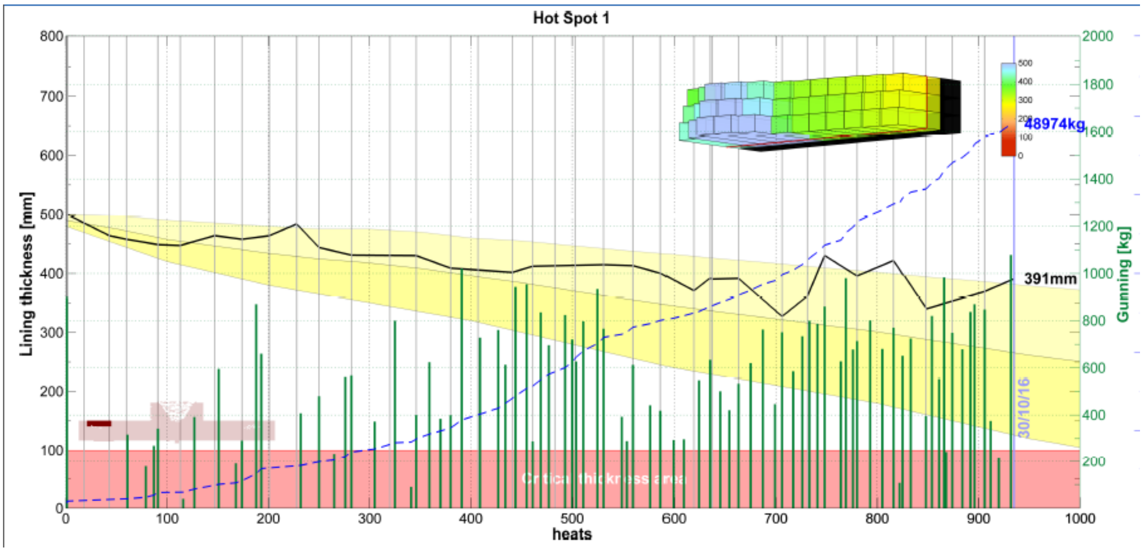

Figure 7 – Example for Daily Report (excerpt)

3.4 Data mining – Parameter Ranking

Data mining is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. It is an interdisciplinary subfield of computer science. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. In this case to identify and rank wear influencing parameters. Figure 8 shows an example of the ranking for one specific campaign.

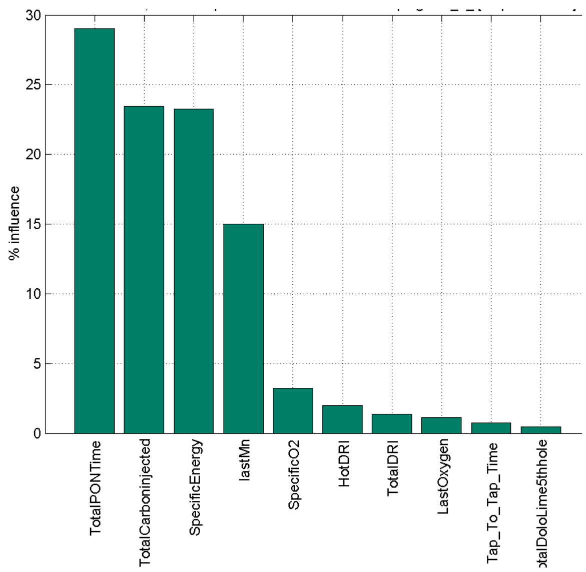

Figure 8 – Example: wear influencing parameters

When thinking about optimization of the furnace productivity the knowledge of the wear influencing parameters become obvious.

Depending on process and product mix the ranking of the parameter can vary in a large scale.

4Conclusion and Outlook

Customer front end

Command and execution, thought and action: the balance between HMI (human machine interface) devices and software in Industry 4.0 is becoming more and more important. Nowadays, mobile devices are widely used for communication and data sharing via the internet. Bringing interactive refractory reports onto such devices is a challenging task, as comprehensible representations on small screens are not easy to make. Managers are used to having reports on one page in a larger format (e.g. paper format A4 in Europe). Condensing information details and highlighting specific importance is a topic of current research. The domain of HMI science is exploring different interfaces to interact with smart systems where VR (virtual reality) or AR (augmented reality) are two of multiple research directions. RHI Magnesita assumes that mobile devices will continue to play an important role as communication devices in steel plants for the coming years. Figure9 shows a wireframe for refractory control and analysis developed by RHI Magnesita.

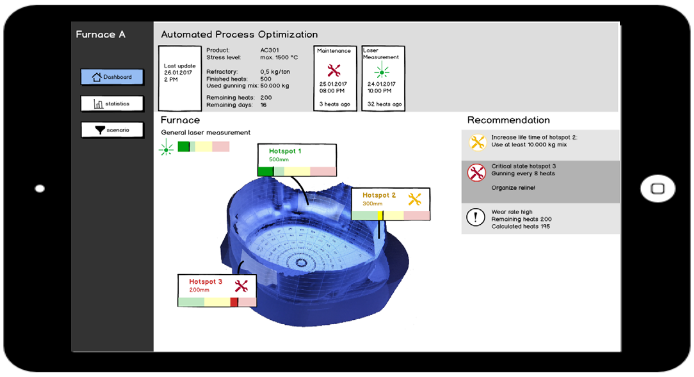

Figure 9 – HMI example for EAF

5 References

[i]. C. Brecher et al., *Integrative Produktionstechnik für Hochlohnländer*, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 17–81.

[ii]. J. P. Womack, D. T. Jones, and D. Roos, *The Machine That Changed the World: The Story of Lean Production*, 1990.

[iii].MalteBrettel, NiklasFriederichsen, Michael Keller, Marius Rosenberg , *How Virtualization, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 PerspectiveWorld Academy of Science*, Engineering and Technology International Journal of Mechanical, Aerospace, Industrial, Mechatronic and Manufacturing Engineering Vol:8, No:1, 2014 , pp 43

[iv]. D. Spath et al., *Produktionsarbeit der Zukunft – Industrie 4.0*, Stuttgart, 2013.

[v]R. Steiner, G. Lammer, C. Spiel, C. Jandl, "Refractories 4.0," BHM, 2017, Vol. 162, No. 11, pp. 514–520

[vi]. M. Forrer, *Prediction of refractory wear with Machine Learning methods*, Master Thesis, Graz University of Technology, 2012, pp 126.

[vii]. A.K. Jain and D. Zongker, *Feature selection: Evaluation, application,and small sample performance*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(2), pp 153–158.

[viii]. R. Kohavi and G.H. John, *Wrappers for feature subset selection*, Artificial Intelligence, 1997, pp 273–324.

[ix]. I. Guyon and A. Elisseeff, *An introduction to variable and feature selection,* Journal of Machine Learning Research, 2003, pp 1157–1182.

[x]. F. Pernkopf and P. O`Leary, *Visual inspection of machined metallic high-precision surfaces*, Special Issue on Applied Visual Inspection, Eurasip Journal on Applied Signal Processing, 2002, pp 667–678.

[xi]M. Forrer, *Prediction of refractory wear with Machine Learning methods*, Master Thesis, Graz University of Technology, 2012, pp 126

[xii]. W. Siedlecki and J. Sklansky, *A note on genetic algorithms for large-scale feature selection*, Pattern Recognition Letters, 1989, pp 335–347.

[xiii]. F. Pernkopf and P. O`Leary, *Feature Selection for Classification Using Genetic Algorithms with a novel Encoding*, International Conference on Computer Analysis of Images and Patterns, 2001, pp 161–168.

[xiv]. M. Feuerstein, *Refractory Wear Modeling using Statistical Methods*, Master Thesis, Graz University of Technology, 2016, pp 61.

[xv]. F. Pernkopf, M. Forrer, M. Feuerstein, N. Mutsam, *Wear Prediction @ SMP2 in Abu Dhabi*, Presentation at RHI, 2016.

xvi C. M. Bishop. Pattern recognition and machine learning. Springer, New York, 10. (corrected at 8th printing) edition, 2009. 6, 7, 9, 19, 26

xvii C. M. Bishop. Pattern recognition and machine learning. Springer, New York, 10. (corrected at 8th printing) edition, 2009. 6, 7, 9, 19, 26 xviii T. Hastie, R. Tibshirani, and J. H. Friedman. The elements of statistical learning: data mining, inference, and prediction: with 200 full-color illustrations. Springer, New York, 2001. 7, 9, 10, 11, 12

[xix]. C.M. Bishop, *Pattern recognition and machine learning*, Springer, New York, 10. Edition 2009, pp 6,7,9,18.

[xx.] Inmon, Bill (Jul 1, 2000), *Operational and Informational Reporting,* Information Management, 2013, Retrieved June 22, 2016 .