A Forecasting Model-Based Discovery of Causal Links of Key Influencing Performance Quality Indicators for Sinter Production Improvement

Authors

Matej Vukovic
Pro2Future GmbH, Linz, Austria
matej.vukovic@pro2future.at

Vaishali Dhanoa
Pro2Future GmbH, Linz, Austria
vaishali.dhanoa@pro2future.at

Markus Jäger
Pro2Future GmbH, Linz, Austria; University of Applied Sciences Upper Austria, Wels, Austria
markus.jaeger@pro2future.at
markus.jaeger@fh-wels.at

Conny Walchshofer
Johannes Kepler University Linz, Linz, Austria
conny.walchshofer@jku.at

Josef Küng
Johannes Kepler University Linz, Linz, Austria
josef.kueng@jku.at

Petra Krahwinkler
Primetals Technologies Austria GmbH, Linz, Austria
petra.krahwinkler@primetals.com

Belgin Mutlu
Pro2Future GmbH, Linz, Austria;
belgin.mutlu@pro2future.at

Stefan Thalmann
Center for Business Analytics and Data Science, University of Graz, Graz, Austria
stefan.thalmann@uni-graz.at

Sintering is a complex production process where the process stability and product quality depend on various parameters. Building a forecasting model improves this process. Artificial intelligence (AI) approaches show promising results in comparison to current physical models. They are mostly considered black-box models because of their hidden layers. Due to their complexity and limited traceability, it is difficult to draw conclusions for real sinter processes and improving the physical models in a running plant. This challenge is addressed by focusing on detecting causal links from AI-based forecasting models in order to improve the understanding of sintering and optimizing existing physical models.

The recent advances in industry have driven an enormous increase in the amount of data generated due to tightly connected machines and services. This data is increasingly valuable for the industry as it contains information about correlations between production parameters, faults and/or disturbances in the production, causes of these problems, etc. Yet, this information is hidden in the data and its extraction is a tedious and time-consuming task.[1] Machine learning (ML) attempts to address these issues by combining the strength of human perception and intelligence with the processing power of computers.

As a result, it becomes easier to find statistical associations between the variables, identify previously undetected correlations and even answer critical questions such as "What caused a certain problem within production?" It stands to reason that when better understanding the data, the understanding of the key elements about the quality of a product, the reason and detail about what and why happened in the production is paramount.[2] Motivated by that, many industry sectors are increasingly applying ML methods to improve decision support for employees by better understanding the production data, extracting production insights, and defining methods that support product quality fine-tuning and production process optimization.[3]

Sintering, like other manufacturing processes, involves several stages and components, from initializing the physical model to the final product, each with its own effect on the product quality. Furthermore, modern sinter plants are endowed with sensors that have the ability to collect, produce and exchange data (machine-to-machine, machine-to-human) from the entire production process.[4] These data contain important information that can be utilized for estimating the production parameters with the biggest influence on quality and their causal relations. Based on these insights, it seems likely that the final product quality can be improved and the production efficiency increased. However, so far little is known about how ML applications can make use of such data in sinter production processes.

This paper presents an ML approach for estimating the potential influencing and process parameters in a sinter production. The pursued results and targeted impact contain findings about the process and relationships in the sinter plant and the sinter process and should pave the way to define methods to forecast and predict the quality of the produced sinter.

A modern sinter plant operation is often supported by a rule-based system.[4] This paper further investigates the rule-based system to understand the current system-changing events and the resulting rules. These findings should aid in adapting and optimizing the existing rules and inventing new rules to gain a higher production and quality increase on the sinter production machine. Therefore, an ML approach that improves the process of forecasting is used. ML approaches show promising results in comparison to current physical models. Due to their complexity and limited traceability, it is difficult to draw conclusions for real sinter processes and improve the physical models in a running plant. Hence, they are mostly considered as black-box models because of their hidden layers.[5] This challenge is addressed by focusing on detecting causal links from ML-based forecasting models in order to improve the understanding of sintering and optimizing existing physical models. Furthermore, an interactive visual analytics tool is proposed that should support the experts in exploring the collected data as well as the results of ML methods.

The novelty value and scientific relevance embrace the detection of causal links by combining data analytics and information visualization approaches in the area of a sinter plant in the steel industry; new scientific findings and contributions in the field of visual interactive prediction in the industrial sector; requirements and solution models for introducing data analytics in an industrial context; findings about the possible uses or the connection between data analytics and visual analytics; and insights into the interplay and connection of data and rule-based decision support in the industrial environment.

This paper describes the background and provides an overview of applied knowledge processing, rule-based expert systems, and artificial intelligence in the sinter production, followed by a description of the conducted and executed case study.

Background and Related Work

Being a process with a significant energy consumption in the steel and iron production process, sintering attracts an increasing business and research interest.[6] The nature of the main drivers of research efforts in this specific process is mostly economical but are often environmental and regulatory. Due to the increasing pressure to reduce conversion costs, the iron- and steelmaking industry is continuing its efforts to optimize production and processes.[7]

Research on computer-aided and data-driven methods for predicting process parameters like quality and throughput in the sintering process started with mathematical models and simulation and the efforts of, e.g., Kawaguchi et al.[8] Since the mid-1980s, researchers were mostly focused on the application of expert systems for plant control, automation[9] and capturing general operational knowledge.[7]

Current Rule-Based Expert Systems in Use — For monitoring the entire sintering process, several technical control systems are in use. These systems manage to control and operate the level 1 (L1) and level 2 (L2) layers of sinter machines and are named as "control systems," "expert systems," "rule-based systems" or "rule-based expert systems." A starting point and overview about knowledge processing systems can be found in References 10 and 11.

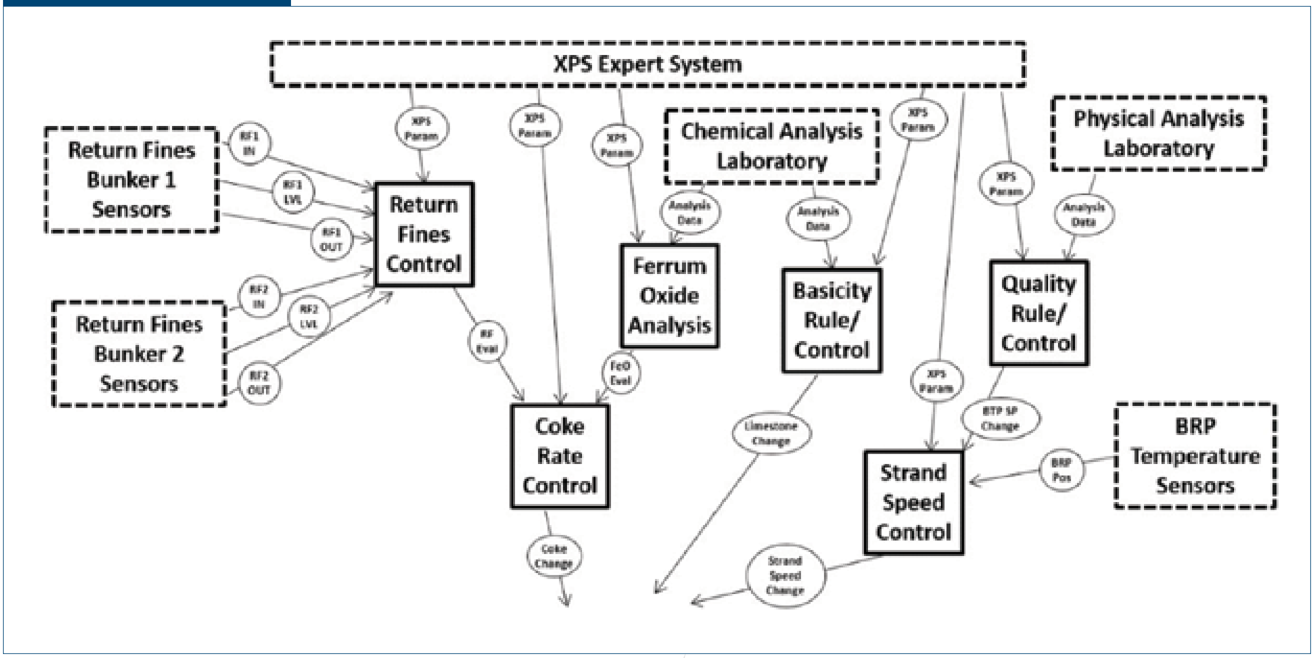
Factors that influence the quality of the final sinter product, which are quite similar in every sinter plant despite where it is located or from whom it is operated, prevail. These factors compound the dimensioning of the return fines which are added to the current recipe/production, the physical (e.g., coarseness or cohesiveness) or chemical (e.g., basicity) composition of the current sinter, and the burn rising point (BRP), as well as the burn-through point (BTP). These factors always influence the quality and quantity of the final product. An overview of the influencing factors in sinter plants as the conclusion from the analysis of running facilities and coherences of these influencing factors is shown in Fig. 1.

A good approach of improving the current state of the art of these systems is to apply knowledge processing methods like machine learning, deep learning or other ML techniques that would lead to a higher complexity and inspection of non-linear relationships between the influencing parameters, which is discussed in the following sections.

Current Machine-Learning Approaches in Sinter Production — Research on application of data-driven approaches in the sinter production is attracting more interest.[6] Due to the limitations in traditional model-driven approaches and the recent advances in the domain of ML, there have been significant efforts in the application of these approaches in the iron- and steelmaking domain. With the increase in their complexity, ML models are becoming extremely difficult to explain and hence are often referred to as black-box models.[5] However, unlike the traditional model- or domain-driven approaches, ML approaches can address the three important aspects of the (industrial) data.[12] First, ML approaches can learn and model non-linear and complex relationships. Second, these approaches address the problem of generalization, i.e., once the model is trained it can capture possible hidden relationships, which enables better predictions on unseen data in the future. Third, these approaches do not impose any restrictions on the input variables and their distribution.

One of the first uses of these approaches was proposed by Shigaki and Narazaki.[7] In their work, an

Figure 1



Sketch of common implemented/triggered events in sinter plants.

approach for inducing the operational rules to obtain products that meet a given quality specification was presented. The approach was based on a multi-layered neural network and the black-box nature of the approach was addressed through a rule extraction algorithm. Aside from the inference of the operational rules, there are research efforts in the domain of optimization of energy consumption of the sinter process. Wang et al.[13] focused on developing an ML model for energy consumption using a combination of local outlier factor method for removing the outliers from the production data, RReliefF method for selecting the most important features, and bagging-enhanced extreme learning machine for final prediction. On the other hand, Pasha et al.[14] focused on the detection of air leaks in grate bars using deep learning for the classification of acoustic measurements with the goal of improving the energy efficiency by removing defected grate bars.

The area of sinter production that focuses on dosing, proportioning and optimization of sinter ingredients was recognized in the literature as another field with a significant research interest. Wu et al.[15] proposed and developed an intelligent integrated optimization system for proportioning of materials. The proposed system contains multiple steps and for each step a different ML method is used: least-squares support vector machine and the grey system theory model is used in the first step to predict the state parameters, following with the back propagation neural network for sinter quality prediction in the second step and concluding with the optimization algorithms for predicting the best proportioning scheme. On the other hand, the same problem was approached by Wu et al.[16] with the improved genetic algorithm method. Similarly, linear programming, genetic algorithm and particle-swarm optimization techniques for optimization of sinter ingredients were compared in the work of Sun et al.[17] Vannocci et al.[18] introduced the usage of fuzzy logic to optimize the control of the charging gates in a sinter production scenario.

In research focusing on material sciences for sinter production, ML methods such as genetic algorithms were used for parameter optimization in the mathematical modeling of sinter process[19] and the problem of sinter basicity analysis was approached with the kernel extreme learning machine and the Random Forest ensemble method.[20, 21]

However, most of the research work found in literature focused on forecasting various production parameters and sinter characteristics. Wang et al.[6] presented a prediction model for sintering characteristics such as solid fuel consumption, gas fuel consumption, burn-through point prediction and tumbler index. Their work shows promising results where an accuracy of 94–96% is achieved using a combination of methods such as AdaBoost.RT and extreme learning machine for the prediction process and RReliefF method for feature selection. Similarly, Laitnen et al.[22] use feedforward neural network for the prediction of sinter quality, sinter plant productivity, fuel consumption and the share of cold return fines. However, a similar task is approached with fuzzy logic in the work of Lei et al.[23] An interesting approach was presented in the work of Zhang et al.,[24] where the task of predicting sinter output, quality, energy consumption and the production cost was approached with the combination of several methods. At first, several back propagation neural network models were trained depending on the prediction task, following the usage of genetic algorithm for the selection of inputs to the previously trained models. Webster et al.[25] showed the usage of partial least-squares regression for the prediction of sinter strength using x-ray diffraction patterns with the goal of establishing a fast feedback loop to the plant operator on whether the process is operating within the acceptable limits. The previously mentioned task of predicting the sinter strength was addressed with artificial neural network approaches in several research papers[26–28] and with the Random Forest method.[29] The task of sinter basicity prediction was approached with least-squares support vector machine method in the work of Wang et al.[30] and Song et al.[31] A new combination of Bayesian evidence framework and support vector machine method was used by Qiang et al.[32] to predict a burn-through point where the Bayesian evidence framework was used to infer regularization and kernel parameters of the support vector machine.

From the previous work mentioned in this section, it can be seen that until now, research in the domain of data-driven sinter production was focusing on a few key areas: (1) optimization of energy consumption, (2) optimization of dosing, proportioning and sinter ingredients, (3) improvement of the physical models, and (4) prediction of sinter production parameters and sinter characteristics. It was determined that there is a lack of research work focused on the causality and approaches focusing on exchange with domain experts to further improve the production process, enable discovery of new insights and increase applicability of ML models in the real production scenarios.

General Options of Applying Modern Artificial Intelligence Approaches — Modern AI developments show a huge potential to improve many aspects of everyday life. However, AI is an umbrella term for all types of systems that are recognized as intelligent by humans. Different classes of approaches belong to AI; machine learning and reasoning are two of them. Whereas machine learning tries to find connections, relationships, dependencies within data autonomously, reasoning uses a symbolic knowledge representation and applies some inference mechanism to derive a solution. As described earlier, both research and application are found in the area of sinter production. Consequently, there are good chances to improve sinter production in terms of quality and quantity by applying methods from machine learning, but also from reasoning as well. The best results are expected in combining both approaches, first gaining insights into the data through machine learning and secondly using these results in reasoning solutions.

Case Study: Context and Procedure

For sinter plants, the detailed connections between potential influencing parameters and process parameters like return fines or production, and quality measures, e.g., grain size or strength, are not yet described precisely by a metallurgical model. A control for optimizing productivity while meeting quality requirements can help to make full use of the production capacity.

Data was collected in 2019 in the L2 system, which uses an open platform communications (OPC) interface to the L1 system. For this case study, a resolution of 5-minute average values was used, which allows for the consideration of time delays between the measuring points in the plant. The quality target variable was derived by physical analysis. Moreover, a 4-hour mixture probe, as well as several probes for the produced sinter during a 4-hour time span, were taken. These probes were mixed and this composition, which represents a 4-hour interval of the produced sinter quality, was then analyzed. This approach gives an overview of the performance in the 4-hour interval. However, for data analysis, it is difficult to relate the data from this 4-hour mixture probe to the process data, as process data can change or fluctuate during the 4-hour interval.

Measured data was stored as 5-minute average values and included information about waste gas (pressure, amount, fan power, etc.), material (amounts of input materials), ignition hood (temperatures, amount of air and gas), process (sinter band speed, feeding roll data, water addition, return fines, sinter temperature, etc.), burn-through point (position of burn-through point and burn rising point, flame front speed, etc.), windboxes (temperatures), cooler (temperatures, speed), production (amounts of mixed materials, produced sinter and return fines) and laboratory (chemical and physical analysis).

The sinter production was evaluated using a scale that weighs the material after screening. Harmonic diameter was used as the quality target variable. From the sample, taken as a 4-hour mixture probe, the grain size distribution was measured. The harmonic diameter d_h was then calculated as the harmonic mean of the grain size distribution.

Case Study: Results

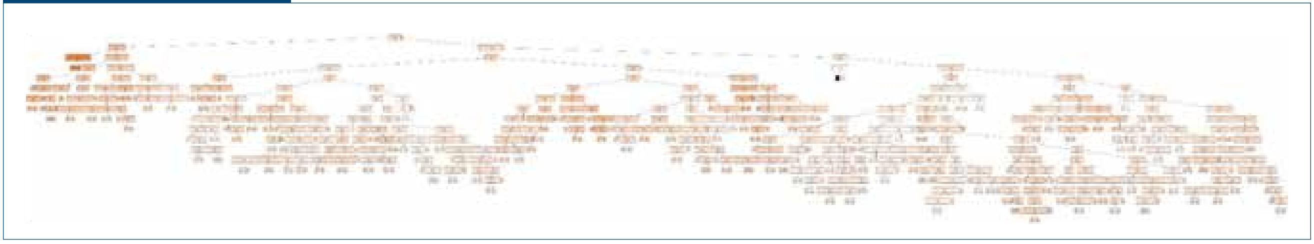
Forecasting Model — Formerly, for analyzing the key production and quality influencing indicators, model-driven (domain-driven) approaches were applied to evaluate the previously described use-case scenario. Model-driven approaches are powerful as they (1) require profound understanding of the production and production processes, (2) base their assumptions on the previously proven physical relationships, and (3) are easy to interpret by the production engineers. However, relying on the linear relationships restricts these approaches and implies other limitations[33] that can be complemented with ML methods that cover interdependencies that have been neglected by the first principles models.

To address these challenges, a forecasting model was developed based on the production data collected in a real sintering plant. Data was processed in three automated steps: initial data cleaning, uptime filtering and time modeling. In the first step, data is denoised by removing the possible outliers and unreliable data is removed. Understanding the mapping of different attributes of the production process, filtering, combining different attributes, and the imputation of missing data required a substantial domain knowledge and intensive exchange between analysts and domain experts. In the second step, data collected during the production downtime is filtered to prevent possible production outliers and finally, due to the changes in the production speed, data is adjusted time-wise. Due to the continuous nature of the production and the frequent change in speed of the production process, every data point required time adjustment, making time modeling the most challenging task in data pre-processing. During this process, a reference point (end of the sinter strand) was defined and every data point was adjusted time-wise to match this point in the production process. This pre-processing was crucial to correctly assign the parts of the production data that were relevant for the target value. Afterwards, features were extracted from cleaned and adjusted time-wise data using Time Series Feature extraction based on scalable hypothesis tests approach.[34]

The forecasting model is a ML regression model based on the Random Forest ensemble method.[35] During the development of the model, several other approaches such as support vector machine regressor, multi-layer perceptron and K nearest neighbors regressor were tested, with Random Forest showing the best results. The predicted value was a harmonic diameter of the sinter. The Random Forest has shown promising results with root mean square error (RMSE) of 0.209 (Normalized RMSE of 8.9%) on the prediction of the target value. However, the model with the best performance was too complex to be interpreted and explained (see Fig. 2, which shows just one of 225 estimators). The interpretation of such model would be extremely time-consuming and despite not being a black-box model per se, a control strategy or a new production insight is difficult to infer. To improve the interpretability of the model and increase its applicability in a real sinter plant, it was necessary to use an additional approach to understand the parameters that influence the final sinter quality.

Causal Approach — To discover parameters of the production that have influence on the sinter quality, a new requirement was added, namely the requirement of explainability. It can be associated with the general requirements of algorithmic decision-making processes: fairness, accountability and transparency.[36] The fairness requirement means that algorithmic decisions should not create discriminatory or unjust consequences.[36] It is not applicable in the presented context as decisions do not influence humans. However, accountability is addressed through the review and auditing process of the forecasting model by domain experts. The requirement of transparency is not addressed with the current (black box) forecasting model. The fulfillment of this requirement would allow for better understanding of the model and the interplay of different process variables, and assist the process engineers in developing new control strategies with new insights. Addressing the requirement of transparency, an approach was developed to increase the explainability of the forecasting model enabling an easier discovery of new insights. This approach is depicted in Fig. 3 and it consists of five steps: forecasting model development, discovery interview with the domain experts, verification through visual analytics, verification through forecasting model analysis, and

Figure 2



One out of 255 estimators (decision trees) from the Random Forest model.

Figure 3

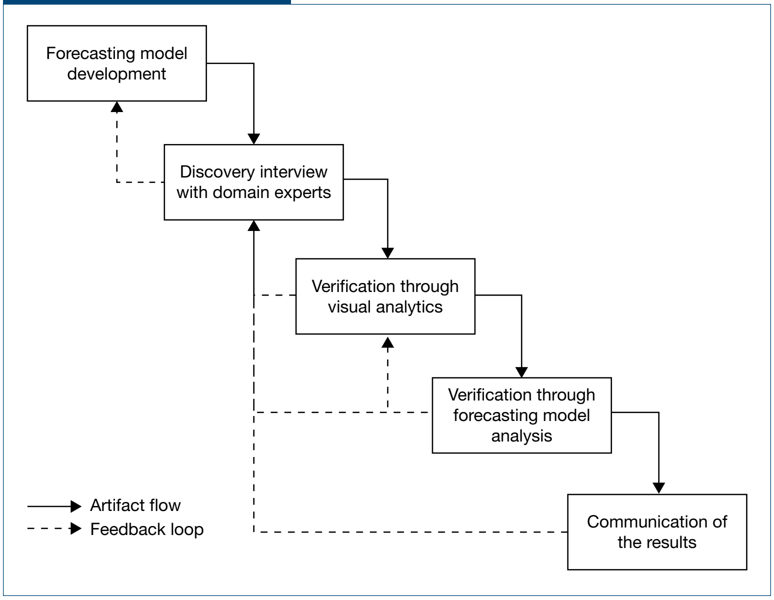


Diagram of the proposed approach.

communication of the results. The steps are interconnected and each one produces an output artifact that is used in another step.

First, a forecasting model was developed and used as an input artifact for other steps. This step represents a data-driven approach that is applied to a real-world use case. The developed forecasting model satisfied the accountability requirement as it underwent an audit that verified its performance and applicability. In the next step, a discovery interview was conducted with the domain experts to gather domain knowledge about the use case as well as materialize implicit knowledge. More specifically, a rough approximation of the diagram of influences was developed. This diagram contained the main entities of the production process, i.e., production and process parameters, target values, measured and non-measured variables, controllable and non-controllable production variables, known causal links, and an approximation of correlation between different process variables. An artifact developed in this step is a diagram of influencing parameters that contains the nature of connections between them. It can be seen as a set of hypotheses that are tested in the following steps to ensure that model-driven intuitions of the domain experts are present in the data, and that the developed forecasting model covers possible control strategies implemented in the production process.

In the third step of the approach (verification of the diagram through visual analytics), the previously developed diagram of influences underwent a verification process. Connections between process variables were tested to verify their existence and their nature through different types of correlations. During this process, feedback was provided to the domain experts to fine tune the diagram of influences from the second step. The fourth step in the approach was the verification of the diagram through forecasting model analysis. At this point, the diagram of influences was in a matured state that targeted the detailed description of the connections between different process parameters. Thus, connections that had undefined effects but were observable could be explored. The main goal of this step was to define the nature and the details of the relationship between different process parameters and to explore possible non-linear relationships. Aside from the analysis of the relationship nature, it was possible to quantify the impact of different variables on the end target through feature importance. The final step in the approach dealt with the communication of the results. It is of utmost importance to ensure the applicability of the forecasting model and discovery of the new insights and control strategies and communicate the satisfied accountability requirements to the end users, domain experts, and the applicants of the model.

In this case study, the proposed approach was applied. In the first step, a discovery interview was conducted. Three process engineers were interviewed to map their implicit knowledge about the relationships in the sinter production and the types of correlation between different process variables. In the second step, each of these relationships was viewed as an assumption. The validity of these assumptions was checked to verify if they corresponded with the real-world data and the results of the forecasting model. This process was done in several iterations with the feedback loop to the domain experts. After a discovery interview with the production engineers, a diagram of influencing parameters was developed. This diagram is depicted in Fig. 4. It contains the most important production parameters and defines connections between them. Connections are defined as different correlation types between the parameters and for this use case, correlations are defined as: (1) positive, (2) negative or (3) varying correlation.

A positive correlation (1) exists when values of one production parameter increase with respect to another production parameter. For instance, the amount of added sinter return fines (kg/metric ton) and the specific coke consumption (kg/metric ton) shows a characteristic pattern for a positive correlation. Similarly, if the revolutions per minute of the exhaust fan increase, the amount of created pressure

Figure 4

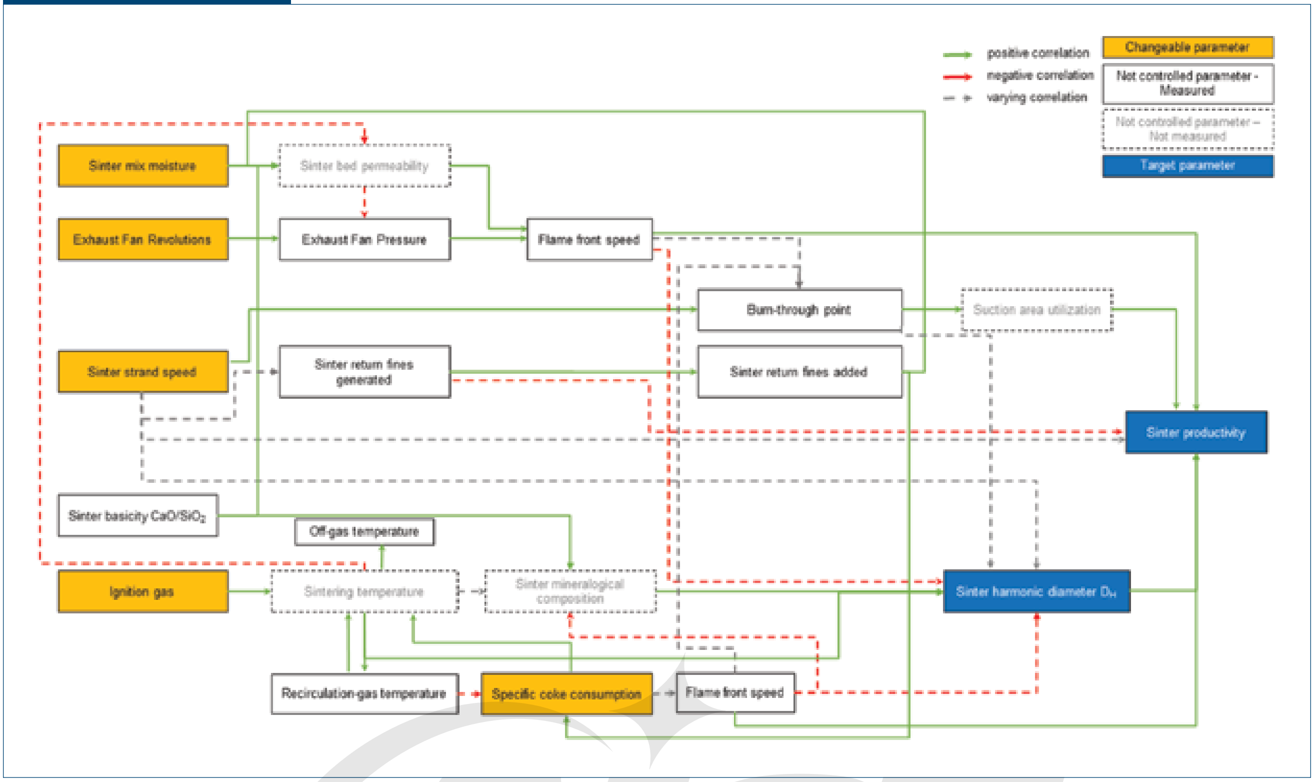


Diagram of influences in the sinter production process.

will also increase, which further results in a positive correlation of the speed of the flame front. On the other hand, negative correlations (2) can be observed for the speed of the flame front and the harmonic diameter where an increase of the flame front speed results in a decrease of the harmonic diameter of the produced sinter. For varying correlations (3), different production behaviors can be observed. For example, a general rule of thumb is that a lower sinter strand speed can increase the harmonic diameter of the sinter but, due to the complexity of the production process, it is not guaranteed.

The diagram depicted in Fig. 4 contains contextual information about the measurability of different production parameters. Through the discovery interview with the domain experts, four different types of production parameters were defined. The first group of production parameters are the ones that can be directly influenced by the operator through a control variable (e.g., sinter strand speed). The second group of parameters are those that are measured but cannot be directly controlled by the plant operator. These are, e.g., recirculation gas temperature and the flame front speed which are observed, measured and used as an indicator in the plant control strategy. A third group of parameters are those that are not controlled and not measured. The effect of these parameters can be observed through different production indicators but are not measured due to, e.g., absence of the adequate measuring equipment or the sinter plant modus operandi. The final group of parameters are the target values. These values are sinter productivity, as an indication of the sinter plant output, and the sinter harmonic diameter as an indication of sinter quality.

The diagram of influences that was developed through the interviews with the plant operators is defined as the ground truth with possible confounding variables and is defined as the hypothesis. In the next steps this hypothesis is tested through (1) visual analytics approach and (2) analysis of the forecasting model.

Verification Through Visual Analytics — Because setting up ML approaches for verification is often time-consuming, visual analytics seems to be a promising assistance to ML, particularly to do a root-cause analysis.[37] Therefore, to visually assess temporal data and the relation between attributes as well as the related correlation coefficient, two open-source visual analytics applications are used: first, Ordino,[38] an interactive rank-based web application, which is used for data-driven approaches to create, visualize and explore rankings of items. This allows domain experts to rank, sort and/or filter variables from the sinter process to detect outliers or similar pattern over time. Second, further functionality was added by using TourDino[39] to calculate and visualize similarity measures.

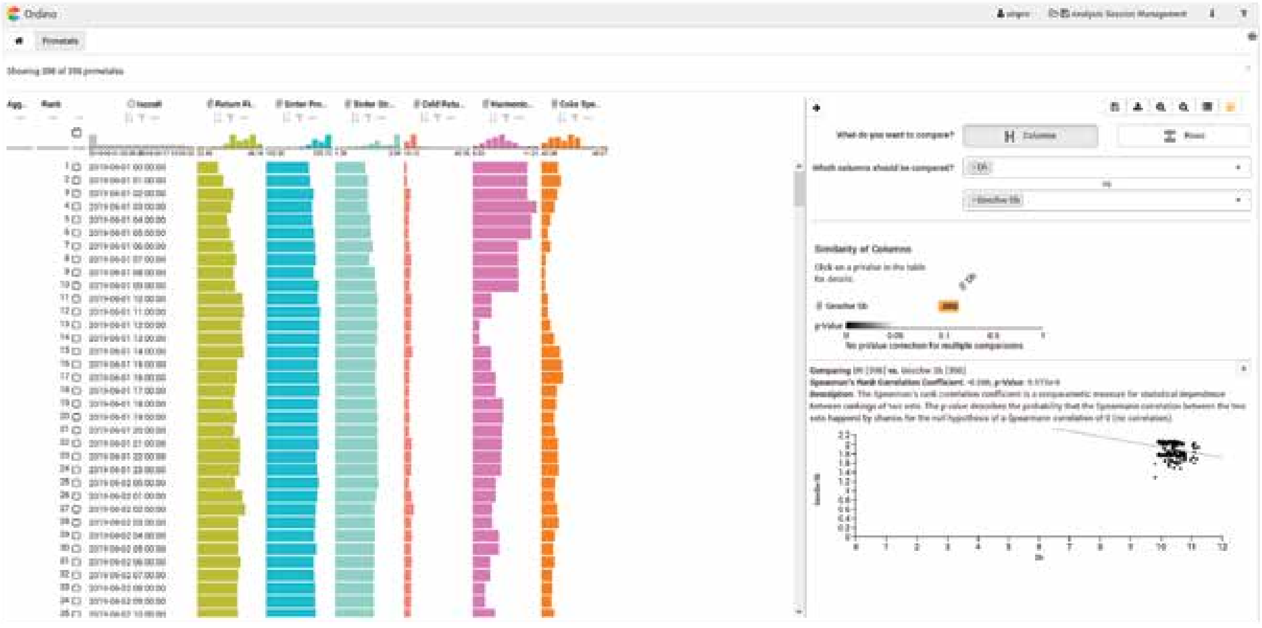
For evaluating the 5-minute data, time stamps were displayed row-wise and attributes column-wise in Ordino. To keep the data analysis process coherent with the forecasting model, the 4-hour time shift for the physical analysis was taken into account, which was not provided by default. As this time delay was not adjusted within the raw data, this feature was additionally implemented in Ordino. This led to a reliable comparison of variables all over the sinter process, where domain experts and process engineers could compare time-adjusted values on demand. This feature was of relevance for the quality target variable, the harmonic diameter. Thus, first insights into the relationship between the time-shifted harmonic diameter and the sinter speed were gained by analyzing the pattern within the tabular view (Fig. 5a).

However, as the list of important variables provided was large and cognition was limited in correctly guessing the correlation coefficient between the variables,[40] TourDino[39] was integrated into the Ordino application. TourDino helps in seeking relationships and patterns in data and provides an overview of the statistical significance of various attribute comparisons without losing the existing ranking. It indicates correlations with a Spearman's Correlation coefficient in addition to a scatterplot.

With a Spearman's Rank correlation coefficient of 0.741, the hypothesis regarding a strong positive relationship between the sinter return fines (kg/metric ton) and the specific coke consumption (kg/metric ton) ($n = 398$, $p < 0.001$) was visually confirmed (see Fig. 5). Thus, the positive relation from the forecasting model was verified based on domain knowledge by using visual analytics. Similarly, there was a negative correlation between the sinter strand speed and the harmonic diameter ($R_s = -0.266$, $n = 398$, $p < 0.001$). The correlation coefficient only holds true for the selected time period because, according to the domain experts, the sinter strand speed and the harmonic diameter show varying correlations. To test this variation, the selected time period was changed and a positive correlation between these variables ($R_s = 0.063$, $n = 744$, $p < 0.001$) was obtained.

Verification Through Forecasting Model Analysis — Following the verification through visual analytics, a verification of the diagram was performed using the analysis of the forecasting model. At this stage, the diagram of influences, developed in a model-driven way, was compared to the data-driven forecasting model. The previous verification approach using visual analytics targeted single relationships between different production variables. Verification of these relationships is crucial for confirmation that the

Figure 5



Rank-based Ordino shows the variable Isozeit in ascending order and displays the other variables respectively (a). TourDino shows a negative correlation between sinter strand speed (Sinter Strand Speed) and harmonic diameter (Dh) (b).

effects observed in the production are also visible through data. However, unlike the previous step, here the focus is on a more holistic view of the production; that is, the quantification of the impact of different production variables to the target value (harmonic diameter).

The forecasting model, developed in the first step, is a regression model that uses different data representations (features) to perform a prediction. These features range in their complexity from simple (e.g., maximum and minimum value of the time series) to more complex (e.g., continuous wavelet transformation parameters, linear trend parameters, etc.). These features are used as input to the forecasting model and once the model is trained and its hyperparameters tuned, the most influential features regarding the predicted value can be extracted.

In the analysis, the top 100 features from the forecasting model were used. This number of features was selected based on the performance analysis of the model, i.e., the observed performance of the forecasting model degraded with the introduction of more features. Once the features were extracted from the trained model, they were analyzed with respect to the production variables from the diagram of influences.

During analysis of the forecasting model, the existence of the measured and controlled production variables were identified from the diagram of influences in the most important features. Some of the production variables in the model could not be directly identified, but a form of their representation was found in another variable. For example, the exhaust fan pressure variable was not present in the list of features; however, a variable that represents the flap opening that directly influences the exhaust fan pressure was identified. Another example is the sinter bed permeability, which was represented through the standard deviation of one suction box reading. The burn-through point was represented through the temperature readings in several suction boxes toward the end of the sinter strand, etc. Through this analysis, the previously defined hypothesis could be confirmed. Furthermore, the analysis of the model and the most important features enabled the discovery of new insights from the production.

Conclusions and Outlook

This paper presented a case study in which a data-driven approach was used to discover influencing parameters in the sinter production process. Within a case study, a forecasting model was developed to predict the harmonic diameter as a central quality parameter indicating the grain size distribution of the finished sinter. Due to the complexity of the model, an approach for the increase of the explainability of the complex (black box) forecasting model was developed, enabling easier discovery of new insights and control strategies. Five important steps were identified, in which a forecasting model in the sinter production scenario was developed, followed by development of the diagram of influences through interviews with the domain experts. Once developed, the diagram underwent verification through visual analytics and forecasting model analysis. Finally, results were communicated to the end users, domain experts and the applicants of the model. The visual analytics approach proved to be a promising assistance to visually analyze relations between variables for the sinter production due to an easy detection of patterns over time. Particularly by evaluating correlations between given data and time-adjusted (time shift) data from the physical analysis, a valid characterization that corresponds to the diagram of influences based on expert interviews could be achieved. Summarizing the insights from the visual analysis, it can be stated that visual analytics can be applied as a promising alternative or additional support to complex and time-consuming ML approaches. Overall, it is the authors' opinion that the approach provides cognitive decision support[41] by fulfilling the transparency and the accountability requirements of deciders in the use case. For future work, the authors want to expand their research into the domain of causal discovery and causal inference to enable easier verification of domain knowledge and use this input to further develop ML models in the domain of sinter production.

Acknowledgment

This work has partially been supported by the FFG, Contract No. 854184: "Pro2Future is funded within the Austrian COMET Program Competence Centers for Excellent Technologies under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry for Digital and Economic Affairs and of the Provinces of Upper Austria and Styria. COMET is managed by the Austrian Research Promotion Agency FFG."

References

1. Spanaki, K.; Gürgüç, Z.; Adams, R.; and Mulligan, C., "Data Supply Chain (DSC): Research Synthesis and Future Directions," *International Journal of Production Research,* Vol. 56, 2018, pp. 4447–4466.
2. Rüßmann, M.; Lorenz, M.; Gerbert, P.; Waldner, M.; Justus, J.; Engel, P.; and Harnisch, M., "Industry 4.0: The Future of Productivity and Growth in Manufacturing Industries," Boston Consulting Group, Vol. 9, No. 1, 2015, pp. 54–89.

3. Thalmann, S., "Data-Driven Decision Support," *IT — Information Technology,* Vol. 60, No. 4, 2018, pp. 179–181, https://doi.org/10.1515/itit-2018-0017.
4. Bettinger, D.; Widi, M.; Schürz, B.; Stohl, K.; Ehler, W.; and Zwittag, E., "VAiron Sinter — Integrated Optimization of Ore Preparation and Sintering," *Ironmaking and Steelmaking,* Vol. 34, No. 5, 2007, pp. 372–377, https://doi.org/10.1179/174328107X203994.
5. Olden, J.D., and Jackson, D.A., "Illuminating the "Black-Box: Understanding Variable Contributions in Artificial Neural Networks," *Ecological Modeling,* Vol. 154, 2002, pp. 135–150.
6. Wang, S.H.; Li, H.F.; Zhang, Y.J.; and Zou, Z.S., "A Hybrid Ensemble Model Based on ELM and Improved AdaBoost. RT Algorithm for Predicting the Iron Ore Sintering Characters," *Computational Intelligence and Neuroscience,* 2019, https://doi.org/10.1155/2019/4164296.
7. Shigaki, I., and Narazaki, H., "A Machine-Learning Approach for a Sintering Process Using a Neural Network," *Production Planning and Control,* Vol. 10, No. 8, 1999, pp. 727–734, https://doi.org/10.1080/095372899232551.
8. Kawaguchi, T.; Sato, S.; and Takata, K., "Development and Application of an Integrated Simulation Model for Iron Ore Sintering," *Tetsu-to-Hagané,* Vol. 73, No. 15, 1987, pp. 1940–1947, https://doi.org/10.2355/tetsutohagane1955.73.15_1940.
9. Xiaohui, F.; Xuling, C.; and Wang, Y., "Expert System for Sintering Process Control," *IntechOpen,* 2010, https://doi.org/10.5772/7073.
10. Jäger, M.; Nadschläger, S.; and Küng, J., "Concepts for Trust Propagation in Knowledge Processing Systems — A Brief Introduction and Overview," *Proceedings of the 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications (IEEE TrustCom 2018),* 2018, https://doi.org/10.1109/TrustCom/BigDataSE.2018.00212.
11. Jäger, M., and Küng, J., "Introducing the Factor Importance to Trust of Sources and Certainty of Data in Knowledge Processing Systems — A New Approach for Incorporation and Processing," *Proceedings of the 50th Hawaii International Conference on System Sciences,* Vol. 50, 2017, https://doi.org/10.24251/HICSS.2017.521.
12. Uluskan, M., "Artificial Neural Networks as a Quality Loss Function for Six Sigma," *Total Quality Management and Business Excellence,* 2018, pp. 1–18, https://doi.org/10.1080/14783363.2018.1520597.
13. Wang, J.; Qiao, F.; Zhao, F.; and Sutherland, J.W., "A Data-Driven Model for Energy Consumption in the Sintering Process," *Journal of Manufacturing Science and Engineering, Transactions of the ASME,* Vol. 138, No. 10, 2016, pp. 1–12, https://doi.org/10.1115/1.4033661.
14. Pasha, S.; Ritz, C.; Stirling, D.; Zulli, P.; Pinson, D.; and Chew, S., "A Deep Learning Approach to the Acoustic Condition Monitoring of a Sintering Plant," *APSIPA ASC 2018,* https://doi.org/10.23919/APSIPA.2018.8659486.
15. Wu, M.; Chen, X.; Cao, W.; She, J.; and Wang, C., "An Intelligent Integrated Optimization System for the Proportioning of Iron Ore in a Sintering Process," *Journal of Process Control,* 2014, https://doi.org/10.1016/j.jprocont.2013.11.012.
16. Wu, T.; Liu, Y.; Tang, W.; Li, X.; and Yu, Y., "Constraint Genetic Algorithm and Its Application in Sintering Proportioning," *IOP Conference Series: Materials Science and Engineering,* Vol. 231, No. 1, 2017, https://doi.org/10.1088/1757-899X/231/1/012022.
17. Sun, L.; Wu, X.; Wang, C.; and Shi, Z., "Research on the Optimization Algorithm of Sintering Ingredients," *IOP Conference Series: Materials Science and Engineering,* Vol. 382, No. 2, 2018, https://doi.org/10.1088/1757-899X/382/2/022089.
18. Vannocci, M.; Colla, V.; Pulito, P.; Zagaria, M.; Di Mastromatteo, V.; and Saccone, M., "Fuzzy Control of a Sintering Plant," *FCTA 2014 — Proceedings of the International Conference on Fuzzy Computation Theory and Applications,* 2014, pp. 68–75, https://doi.org/10.5220/0005082400680075.
19. Giri, B.K., and Roy, G.G., "Mathematical Modeling of Iron Ore Sintering Process Using Genetic Algorithm," *Ironmaking and Steelmaking,* Vol. 39, No. 1, 2012, pp. 59–66, https://doi.org/10.1179/1743281211Y.0000000037.
20. Ding, Y.; Yan, F.; Yang, G.; Chen, H.; and Song, Z., "Quantitative Analysis of Sinters Using Laser-Induced Breakdown Spectroscopy (LIBS) Coupled With Kernel-Based Extreme Learning Machine (K-ELM)," *Analytical Methods,* Vol. 10, No. 9, 2018, pp. 1074–1079, https://doi.org/10.1039/c7ay02748f.
21. Yang, G.; Han, X.; Wang, C.; Ding, Y.; Liu, K.; Tian, D.; and Yao, L., "The Basicity Analysis of Sintered Ore Using Laser-Induced Breakdown Spectroscopy (LIBS) Combined With Random Forest Regression (RFR)," *Analytical Methods,* Vol. 9, No. 36, 2017, pp. 5365–5370, https://doi.org/10.1039/c7ay01389b.
22. Laitinen, P.J., and Saxén, H., "A Neural Network Based Model of Sinter Quality and Sinter Plant Performance Indices," *Ironmaking and Steelmaking,* Vol. 34, No. 2, 2007, pp. 109–114, https://doi.org/10.1179/174328107X155312.
23. Lei, Q.; Zhao, G.; Chen, X.; and Wu, M., "Sinter Product Rate Forecasting Based on T-S Fuzzy Model," *Chinese Control Conference,* 2017, pp. 2144–2148, https://doi.org/10.23919/ChiCC.2017.8027673.
24. Zhang, J.; Xie, A.; and Shen, F., "Multi-Objective Optimization and Analysis Model of Sintering Process Based on BP Neural Network," *Journal of Iron and Steel Research International,* 2007, https://doi.org/10.1016/S1006-706X(07)60018-1.
25. Webster, N.A.S.; Pownceby, M.I.; Ware, N.; and Pattel, R., "Predicting Iron Ore Sinter Strength Through Partial Least-Squares Regression (PLSR) Analysis of X-Ray Diffraction Patterns," *Powder Diffraction,* 2017, pp. 1–4, https://doi.org/10.1017/S0885715617001038.
26. Kumar, V.; Sairam, S.D.S.S.; Kumar, S.; Singh, A.; Nayak, D.; Sah, R.; and Mahapatra, P.C., "Prediction of Iron Ore Sinter Properties Using Statistical Technique," *Transactions of the Indian Institute of Metals,* Vol. 70, No. 6, 2017, pp. 1661–1670, https://doi.org/10.1007/s12666-016-0964-y.
27. Fan, X.H.; Li, Y.; and Chen, X.L., "Prediction of Iron Ore Sintering Characters on the Basis of Regression Analysis and Artificial Neural Network," *Energy Procedia,* Vol. 16 (Part B), 2012, pp. 769–776, https://doi.org/10.1016/j.egypro.2012.01.124.
28. Umadevi, T.; Naik, D.K.; Sah, R.; Brahmacharyulu, A.; Marutiram, K.; and Mahapatra, P.C., "Studies on Parameters Affecting Sinter Strength and Prediction Through Artificial Neural Network Model," *Transactions of the Institutions of Mining and Metallurgy, Section C,* Vol. 125, No. 1, 2016, pp. 32–38, https://doi.org/10.1179/1743285515Y.0000000020.
29. Song, L.; Qing, L.; Xiaojie, L.; and Yanqin, S., "Synthetically Predicting the Quality Index of Sinter Using Machine Learning Model," *Ironmaking and Steelmaking,* Vol. 9233, May 2019, https://doi.org/10.1080/03019233.2019.1617573.
30. Wang, R.; Wang, A.M.; and Song, Q., "Research on the Alkalinity of Sintering Process Based on LS-SVM Algorithms," *Advances in Intelligent and Soft Computing,* 2012, pp. 449–454, https://doi.org/10.1007/978-3-642-30126-1_71.
31. Song, Q., and Wang, A.M., "Simulation and Prediction of Alkalinity in Sintering Process Based on Grey Least-Squares Support Vector Machine," *Journal of Iron and Steel Research International,* Vol. 16, No. 5, 2009, pp. 1–6, https://doi.org/10.1016/S1006-706X(10)60001-5.
32. Song, Q.; Wang, A.M.; and Zhang, Y.S., "The Combination Prediction of BTP in Sintering Process Based on Bayesian Framework and LS-SVM," *TELKOMNIKA Indonesian,* 2013, https://dx.doi.org/10.11591/telkomnika.v11i8.3087.
33. Xu, J-X., and Hou, Z-S., "Notes on Data-Driven System Approaches," *Acta Automatica Sinica,* Vol. 35, No. 6, 2009, pp. 668–675, https://doi.org/10.1016/s1874-1029(08)60090-7.
34. Christ, M.; Braun, N.; Neuffer, J.; and Kempa-Liehr, A.W., "Time Series FeatuRe Extraction on Basis of Scalable Hypothesis Tests (tsfresh – A Python Package)," *Neurocomputing,* Vol. 307, 2018, pp. 72–77, https://doi.org/10.1016/j.neucom.2018.03.067.
35. Breiman, L., *Machine Learning,* Vol. 45, No. 1, 2001, pp. 5–32, https://doi.org/10.1023/a:1010933404324.
36. Lepri, B.; Oliver, N.; Letouzé, E.; Pentland, A.; and Vinck, P., "Fair, Transparent, and Accountable Algorithmic Decision-Making Processes: The Premise, the Proposed Solutions, and the Open Challenges," *Philosophy and Technology,* Vol. 31, No. 4, 2018, pp. 611–627, https://doi.org/10.1007/s13347-017-0279-x.

37. Suschnigg, J.; Ziessler, F.; Brillinger, M.; Vukovic, M.; and Mangler, J., et al., "Industrial Production Process Improvement by a Process Engine Visual Analytics Dashboard," *HICSS 53,* 2020, https://doi.org/10.24251/HICSS.2020.163.
38. Streit, M.; Gratzl, S.; Stitz, H.; Wernitznig, A.; Zichner, T.; and Haslinger, C., "Ordino: Visual Analysis Tool for Ranking and Exploring Genes, Cell Lines, and Tissue Samples," *Bioinformatics,* 2019, https://doi.org/10.1093/bioinformatics/btz009.
39. Eckelt, K.; Adelberger, P.; Zichner, T.; Wernitznig, A.; and Streit, M., "TourDino: A Support View for Confirming Patterns in Tabular Data," *EuroVis Workshop on Visual Analytics,* 2019, https://doi.org/10.2312/eurova.20191117.
40. Rensink, R., and Baldridge, G., "The Perception of Correlation in Scatterplots," *Comput. Graph. Forum,* Vol. 29, 2010, pp. 1203–1210, https://doi.org/10.1111/j.1467-8659.2009.01694.x.
41. Thalmann, S.; Gursch, H.G.; Suschnigg, J.; Gashi, M.; Ennsbrunner, H.; Fuchs, A.K.; and Lindstaedt, S., "Cognitive Decision Support for Industrial Product Life Cycles: A Position Paper," *Cognitive 2019.*

Disclaimer

The information (including, e.g., figures and numbers) provided in this document may contain general descriptions or characteristics of performance based on estimates and assumptions that may not be verified. It is no representation, does not constitute and/or evidence a contract or an offer to enter into a contract to any extent, and is not binding upon the parties. Any obligation to provide and/or demonstrate respective characteristics shall only exist if expressly agreed in the terms of a contract. Estimates and assumptions must be analyzed on a case-by-case basis and might change as a result of further product development.

Primetals Technologies excludes any liability whatsoever under, or in connection with, any provided information, estimates and assumptions. The information, estimates and assumptions provided shall be without prejudice to any possible future offer and/or contract. Any use of information provided by Primetals Technologies to the recipient shall be subject to applicable confidentiality obligations and for the own convenience of, and at the sole risk of, the recipient. ✦

This paper was published in the AISTech 2020 Conference Proceedings. AIST members can access the AISTech 2020 Conference Proceedings in the AIST Digital Library at digital.library.aist.org.