Intelligent Steelmaking Based on Advanced Analytics: Reducing Operational Costs of a BOF

Digital technologies are transforming industry at all levels. Steel has the opportunity to lead all heavy industries as an early adopter of specific digital technologies to improve our sustainability and competitiveness. This column is part of AIST's strategy to become the epicenter for steel's digital transformation, by providing a variety of platforms to showcase and disseminate Industry 4.0 knowledge specific for steel manufacturing, from big-picture concepts to specific processes.




Authors

M. Luccini (top row, left)
R&D process engineering, Tenova Goodfellow Inc., Mississauga, Ont., Canada
marco.luccini@tenova.com

V. Scipolo (top row, right)
chief of technology, Tenova Goodfellow Inc., Mississauga, Ont., Canada
vittorio.scipolo@tenova.com

N. Giso (bottom row, left)
data scientist, Tenova S.p.A., Castellanza (VA), Italy
nicolo.giso@tenova.com

G. Bavestrelli (bottom row, right)
digital engineering director, Tenova S.p.A., Castellanza (VA), Italy
giovanni.bavestrelli@tenova.com

Modern manufacturing is characterized by the availability of large sets of data (e.g., sensor data) of different formats, quality and semantics, which is often referred to as big data.[1] The abundance of data creates a vast potential for production process improvement, as well as for increasing product quality and sustainability.[2] The steel manufacturing industry is changing with the introduction of on-line data analysis and the possibilities made available by cloud computing technology. Tenova understands that, in order to maintain a competitive role in the steel business, it needs advanced on-line monitoring and analysis tools, developed on state-of-the-art detection algorithms and leveraging big data solutions.

In this context, data mining, data analytics and machine-learning tools (ML) are becoming more and more important. They allow extracting knowledge from large amounts of data and finding correlations among variables, which can be used to improve process control and decision-making strategies. Moreover, problems such as estimation of time-variant processes and strategy optimization, typical of Information Theory, are part of this approach.

Machine-learning algorithms can be described as computationally viable and robust methods to learn information directly from data without relying on a predetermined equation as a model.[3] ML algorithms come from a wide range of fields, including mathematics, statistics, neuroscience and computer science. These techniques have become increasingly popular over the past decade and are used for multiple tasks such as classification, estimation and regression in a variety of applications in communications, process optimization, production, computer vision and many more.

Today, ML is used in various areas of intelligent manufacturing. For example, learning algorithms are utilized for fault detection in semiconductor manufacturing,[4] estimation of the manufacturing cost of civil jet engines[5] and for energy optimization in marine transportation.[6] It has been shown that ML techniques have a strong ability to handle large, multi-variate data in complex, dynamic and often chaotic environments, while providing a good trade-off between complexity of solutions and resulting precision.[2]

In the specific case of the steelmaking industry, the trend toward digital innovation is accelerating across the sector because companies are much more focused on cost and driving operational efficiencies in an era of relatively low steel prices. Tenova believes that new digital services can have a substantial impact in delivering smarter and more efficient operational solutions.

In this contribution, Tenova proposes some examples of pilot projects that leverage ML techniques and new possibilities achievable with new digital tools. Specifically, control models need to be based on a sound physical understanding of the process but should also account for many uncertainties due to the nature and complexity of the environment in which the process is carried out. As a result, it is crucial to extract useful control information from the raw data stream acquired by industrial sensors. To achieve this objective, Tenova is relying on the promising approaches offered by ML and advanced analytics.

In terms of model optimization, Tenova presents the improvements obtained for steel temperature prediction in the proprietary BOF Static Charge Model. The analysis was carried out on real operational data collected during 2018 in a basic oxygen furnace (BOF) shop in North America. The application of advanced analytics and ML allowed for the identification and resolution of specific cases in which the steel temperature prediction was particularly inaccurate. A correction model was developed on the acquired data and allowed for a significant improvement in the accuracy of predicting the final steel temperature. The obtained results would benefit the client with a lower number of reblows expected.

Furthermore, digital transformation will help not only in terms of continuous improvement of process models, but also with their performance consistency over time by introducing a data infrastructure that allows data accessibility (ubiquitous access, which is a concept well developed in communication and networking), automated advanced monitoring of performance (both for equipment and process models), and multi-plant integration.

Tenova is not alone in this endeavor. Together with Microsoft, a joint program was launched in 2017 with the aim of developing an integrated industrial system to allow a secure and scalable analysis of data gathered in the field through a variety of sensors. This complex architecture is enabled through the Microsoft Azure cloud platform, which offers a trusted and reliable service for data collection, storage and analysis.

Advanced Analytics: Model Optimization and Continuous Improvement

Data Architecture — To process the significant quantity of data gathered from plants, Tenova defined a data science architecture that accounts for both cloud solutions and on-premise scenarios, including hybrid solutions combining the two.

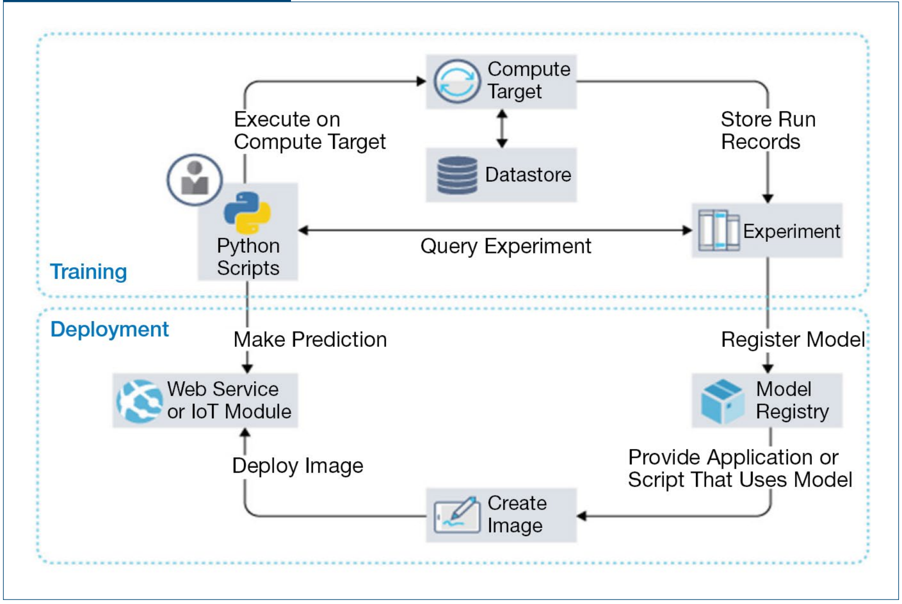
The different options are tailored to customer constraints and needs.

Cloud Architecture: This solution is characterized by a high degree of flexibility. In this context, the training of ML models can take full advantage of cloud resources, scaling as needed, and through the retraining service, the model parameters can be continuously updated. Furthermore, whenever a newer model for the task has been trained, it can be deployed right away on the cloud. This solution facilitates integration with other services such as 3D manuals and Tenova's Internet of Things (IoT) platform. The major limitation of this architecture is that response times are not fast enough for real-time applications; therefore, the models deployed on the cloud can be used to feed data to dashboards, providing useful information to users, but not to control the equipment automatically when low latency is required.

On-Premise Architecture: The main advantage of the on-premise solution is the possibility of automatically controlling the equipment with virtually zero latency. Such architecture is also appropriate for customers who are not ready to trust a cloud-based solution and prefer to host their data in their own plants. In this case, though, retraining and improvement of the models is not applicable as the data is stored locally and is not easily accessible to Tenova's data scientists.

Hybrid Architecture: Several hybrid solutions could be developed to leverage both on-premise and cloud where necessary. For instance, when the business case requires real-time predictions, the model can be deployed on-premise with the continuous retraining happening in the cloud. In such a scenario, the model retuned in the cloud is redeployed on-premise at regular intervals or upon request.

Figure 1



Workflow of Azure machine-learning services.

Figure 2

Interface of Tenova's Static Charge Model for the basic oxygen furnace (BOF).

Model Optimization: Applied Case for BOF Temperature Prediction — Tenova has developed its proprietary Static Charge Model for BOF process. The model is a mass-energy balance that takes into consideration the conditions (composition and temperature) of the input material and calculates the amount of hot metal, scrap, additions, and oxygen needed to reach a desired steel composition, weight and temperature at the end of the blow. Thanks to a new mathematical formulation, the Tenova model offers freedom in defining operational parameters such as the optimal basicity ratio, the maximum hot metal ratio, the heat size and the iron oxides in the slag.

The accuracy of the Static Charge Model is fundamental and is at the core of the BOF production: errors in the mass-energy balance will result in sub-optimal use of charged material and lower efficiency in the downstream steel processes, both in terms of energy consumption and time delays. It is particularly important to predict the final temperature of the steel bath: the prediction accuracy affects decisions related to material additions (fuel/cooling) to reach the target steel temperature at the end of blow. Moreover, a good final temperature limits the number of reblow operations that directly affect the heat yield (the ratio between steel weight/total input iron) and the energy consumption of the downstream ladle stations.

One key observation is that whereas the Static Charge Model is based on the physics of the process and the mathematical formulas to model it, in reality there are other elements that are not included in the formulas and that affect the outcome. For instance, the furnace heat losses are the results of very complex and often unmeasurable variables. While not directly measurable, the effect of those variables is noticeable in the collected data used for training the machine-learning models.

The final goal of the presented data science project was to generate a more accurate prediction of the final steel temperature, reducing the error of the initial calculation of the Static Charge Model in order to increase the value for the customer.

The available data for this project was recorded by the programmable logic controller (PLC) and sensors in the plant, and include more than 100 variables. The quality of the input data is essential; therefore, data exploration and data cleaning (i.e., outliers removal) was performed. The model development phase involved the understanding of how the different data relate to each other, by both statistical means (e.g., correlation analysis) and the knowledge of the physics of the process. After this step, the data set had roughly 4,300 heat observations and 17 variables to be used for the prediction (16 variables to be used for the prediction plus the target variable).

First, data were randomly split into training and testing sets while trying different simulations and then, taking into consideration the temporal aspect of the processes, a period of data was used as the training set and a different period as the test set. Multiple ML models were tested, including linear regression, polynomial regression, support vector regression, random forests, extreme gradient boost and neural networks, in order to find the best correction model.

The model was tested on 400 consecutive heats in real operating conditions, which include all the uncertainties of the real BOF process. The results were compared with the performance of the deterministic model relying on mass-energy equations. All the ML models performed better than the deterministic model. The best trade-off between accuracy and complexity was reached with the support vector regression (SVR) model. Further improvements of the models require a greater amount of observation and fine-tuning of the model hyper parameters.

The SVR model error is defined as the absolute value of the difference between the predicted temperature and the real one. Several metrics are used to monitor the quality of the models:

- Median absolute error (MAE): Rare and extreme errors do not influence this metric.
- Root mean squared error (RMSE): Extreme errors influence this metric more heavily.
- Error range (color): The predictions are labeled with respect to the absolute error (to the real value) in the following way:

– Blue: Less than 25°F of error.
– Orange: Between 25°F and 50°F of error.
– Red: More than 50°F of error.

The metric by color is used because, due to the nature of the problem, it is important to have as low errors as possible (under 25°F) and to keep extreme errors under control (above 50°F).

Table 1 summarizes how the ML model (SVR) is able to perform better than the deterministic model. The MAE index decreases from 25°F to 19°F (24% reduction). More importantly, the distribution of the temperature error is significantly improved: blue points (number of heats with temperature error <25°F) increase from 49% to 61% and red points (number of heats with error >50°F) decrease from 19% to 10%. Those results are visualized in Fig. 3 in the form of a scatter plot of the predictions vs. real values.

If the model were to be always accurate, the points would be exactly on the black line. The different colors reflect the absolute error between the real value and the prediction, as outlined in the metrics by color. The fact that the cloud of points is much more tightly distributed around the black line for the SVR rather than the deterministic model confirms that the former is consistently more accurate than the latter.

Fig. 4 presents a graphical comparison of the temperature error of the two models on a heat-by-heat basis for the subset of 400 heats that compose the test set. It is evident how the error of the SVR (blue line) shows a more limited variance than the deterministic model (green line): SVR temperature prediction error is much more compact on the lower values and does not suffer from the few extreme error peaks that affect the deterministic model.

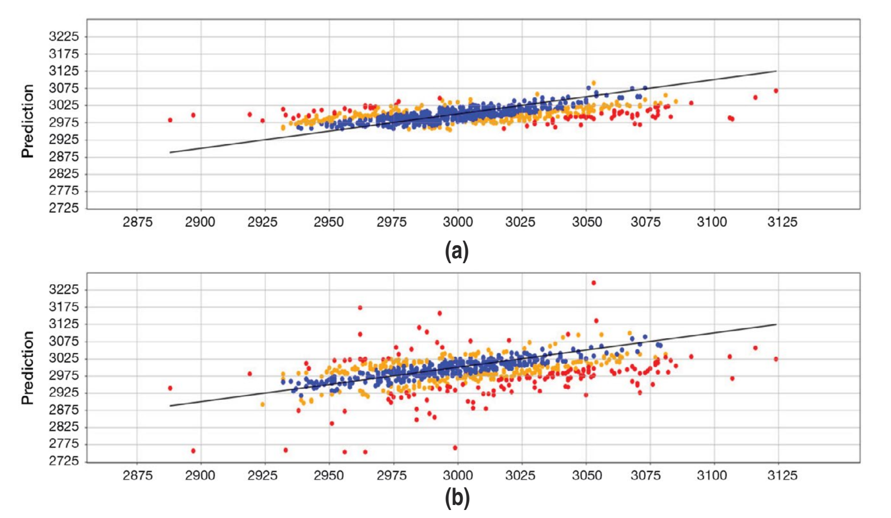
In conclusion, from the data collected during a few months of operation, it was possible

Table 1

Comparison Summary Between Support Vector Regression (SVR) Model and Deterministic Model

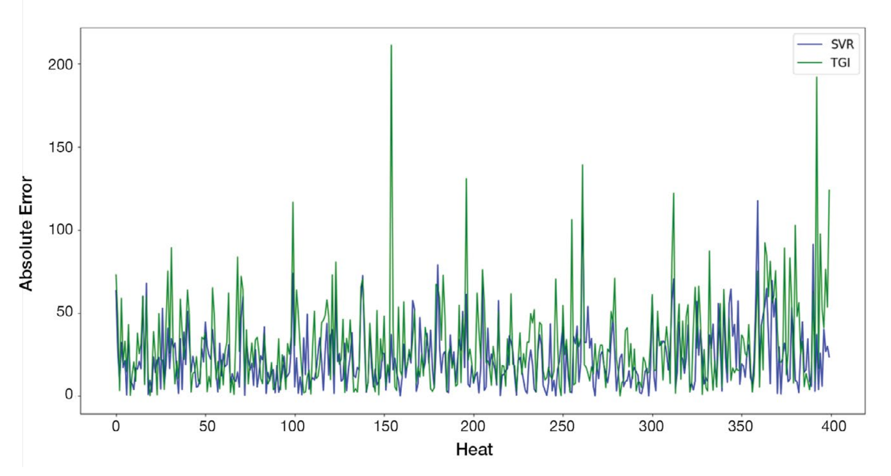
	SVR	Deterministic
MAE (°F)	19.314	25.600
RMSE (°F)	30.626	45.530
Blue (% heats)	61.500	49.400
Orange (% heats)	27.700	30.800
Red (% heats)	10.800	19.800

Figure 3



Scatter plot of temperature prediction: SVR model (a) and deterministic model (b).

Figure 4



Temperature error over 400 consecutive heats: SVR model (blue) and deterministic model (green).

to develop a model to correct and increase the accuracy of steel temperature prediction of the BOF Static Charge Model, thanks to ML methods. Simulated results on real heat data showed that the number of heats that exhibit a final steel temperature error greater than 50°F is drastically reduced, which would translate to an important reduction of reblow operations and, consequently, would result in both higher yield (reduced iron oxidation) and important energy savings in the downstream ladle furnace stations.

Model Monitoring and Continuous Improvement — Continuous improvement and monitoring of model performance is one key advantage of cloud data infrastructure, which offers new ways of delivering and making accessible essential information about model performance in real time. Raw data are processed and transformed in the most useful summary information to characterize a process status or classify an anomalous behavior. Meaningful information can then be accessed or delivered to multiple clients in an autonomous way, in order to have the best possible "status" of the equipment or model performance at the current moment, or alternatively in a recent time period. In this way, performance of process models can be easily tracked and maintained consistent in time by retuning operations when necessary.

The process model can be considered a "digital twin" of the real process, where the model is capable of simulating and describing the process evolution in time according to specific key performance indicators (KPIs) and within certain operational boundaries. This example refers to Tenova's model for carbon prediction in an argon oxygen decarburization (AOD) plant. This model estimates the carbon content at the process turndown based on input information at the start of the heat and continuous information (i.e., offgas and oxygen consumption) during the decarburization process. The metric of interest is the estimated carbon, which is compared to the measured carbon content obtained by physical measurement. The estimation error is defined as the difference between the measured and the estimated values. Within a certain error threshold, the model is considered to be precise and reliable. From a process control point of view, it is significant to know how precise the model is now and how it has been in the recent heats to identify potential error trends. Moreover, it is important to know how well the operators are following the model (compliance indicator).

The Model Control dashboard shown in Fig. 5 displays those metrics in real time, summarizing the model performance. The dashboard is fully accessible via a normal web browser on any device. Fig. 5, box A, shows the model accuracy during the last six heats (right) and its average value per day in the last 6 days (left). The graph on the right of box A shows the accuracy threshold of the model: red points correspond to an estimation error if above the accuracy threshold limit, green points if within the accuracy limit. A model with all green estimation is performing well, assigning a positive "trust score" to the model; on the other hand, a model with all red estimation would be unreliable with a trust score close to zero. The average score, normalized on the number of heats per day, is shown in the graph on the left in box A.

Fig. 5, box B, shows the model score for the current day, which describes how good the model prediction is from 0 to 100% based on the estimation error in each heat of the last day. The lower indicator shows how much the operators are compliant to the model indication: higher values correspond to operators following the indications of the model about the stop-blowing value. High score and high compliance mean the model accuracy is high; therefore, the model can be trusted. Low score and low compliance mean that it is not possible to evaluate the model reliability since the operators are not following the model indications, therefore deviating from known conditions.

Finally, Fig. 5, box C, shows the dynamic tuning portion, in which the estimation model parameters are tuned to the last available data to improve the model estimation. The updated model results are shown in the top graph (blue dots) to assess the improvements in the last six heats. The operator can choose at this point to update the model estimation parameters to increase the model accuracy or to keep the current one if model score is high enough.

The Model Control dashboard represents an evolution of a normal human-machine interface (HMI) in terms of:

- Accessibility: Being web-based, it is easily accessible by almost any device, including tablets.
- Visualization: Advanced analytics are performed on the cloud allowing for the display of highly valuable information instead of raw data.
- History: The virtually unlimited space on the cloud allows for the retrieval of large amounts of data to calculate significant trends at the most suitable time resolutions (hours, days, weeks...).

Technology Alignment

The possibility of a technological alignment among different production sites certainly represents an important achievement in the context of operational efficiency. Tenova's proposed solution relies on the cloud infrastructure to enable and simplify the sharing and accessing of valuable process models from multiple locations.

With this strategy, the benefits from updating and improving a single control model shared by all the production sites and processes can be reaped immediately. In the example of the BOF Static Charge Model, the latest release of the "trim" algorithm with improved final temperature calculation will be immediately accessible by all the plants sharing the model. If a master model is available on the cloud, immediately all the users of the model would benefit from the updated version at once, instead of needing to update the installation at each production site.

In addition, different plants would share a similar data format concerning the model use, which would allow for easier metric comparisons among locations as well as a better identification of the most significant variables impacting the process.

Tenova describes the advantages of this solution in terms of "horizontal integration" among the different production sites. In this scenario, each improvement in a single plant can potentially be shared with minimum effort to other production sites within the same company, generating a cross-learning system.

The cloud infrastructure represents the basis on which an effective horizontal integration and alignment can be built, leaving to each client the possibility of taking advantage of the shared information or to remain more local, in a fully scalable solution according to the client's needs.

Figure 5

Model Control dashboard for C level prediction in the argon oxygen decarburization process.

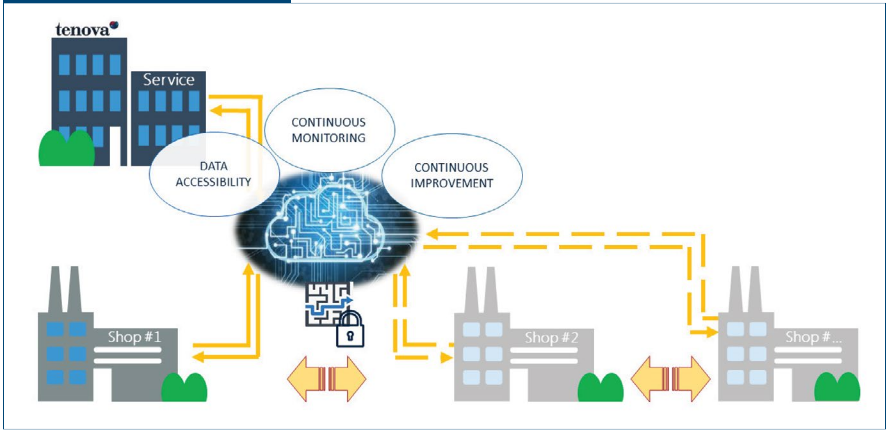
The Strategic Partnership With Microsoft

At the initial stages of the digital transformation in Tenova, it was realized that the breadth of the technological and cultural changes to put into effect would benefit from a strategic partnership with a leading technology provider. After a broad investigation, Microsoft was selected. At the beginning of 2017, a partnership with Microsoft was signed, involving strategic consulting as well as the adoption of their leading cloud platform, cognitive services and mixed reality devices. For the past two years, Tenova has worked closely with Microsoft to develop its IoT infrastructure based on Microsoft Azure, implement a predictive maintenance solution and a data science pipeline to process incoming data from customer plants, develop ML models for optimizing processes in customer plants, and redefine its customer service initiatives.

Microsoft Azure has become a cornerstone in Tenova's digital offerings, as it is a well-established industrial cloud platform compliant with the strictest security and privacy regulations, such as the General Data Protection Regulation (GDPR). The Azure cloud provides state-of-the-art digital services as well as unlimited scalability on demand.

The work with Microsoft is led by Tenova's digital team, a team of data scientists and software developers whose mission is to work closely with process

Figure 6



The cloud infrastructure will be pivotal to enable and simplify technology alignment among multiple plants.

engineers and business developers within Tenova to identify, promote and facilitate a coordinated implementation of new services and business opportunities, as they are made available by emerging digital technologies.

Conclusions

In steelmaking, process control models need to be based on a sound physical knowledge of the process but should also account for many uncertainties due to the nature and complexity of the environment in which the process is happening. In such a context, it is fundamental to transform raw data acquired in the plant into information useful from a process control, maintenance and safety point of view. Data analytics and machine-learning tools are at the core of the digital transformation and are becoming more and more important to improve process control and decision-making strategies.

In this contribution, Tenova presented the results achieved in a pilot project aimed to optimize the steel temperature prediction of its BOF Static Charge Model. Based on a few months of operational data collected, machine-learning models were explored to obtain a more accurate temperature prediction, which would result in direct benefit to the client by reducing the number of reblow operations, therefore improving the process yield and reducing energy consumption in the downstream stations.

The optimization of process models is not the only possibility offered by digital transformation. Tenova recognizes other important advantages that could be unlocked by relying on cloud data infrastructure and advanced analytics, such as ubiquitous information accessibility, continuous process improvement and monitoring, and technology alignment to pursue horizontal integration among different production sites.

Tenova is fully involved in developing and implementing the opportunities offered by digital transformation, whose benefits have already become noticeable. The centralized digital platform developed by Tenova will provide more informative support, so as to offer a broad portfolio of services according to the company's strategic objectives: offering process optimization and plant monitoring for improved safety and lower environmental impact, as well as personalized spare parts management and remote assistance to achieve a truly effective predictive maintenance.

In this endeavor, the partnership with Microsoft is strategic in leveraging the Azure data infrastructure, which offers fully scalable and secure solutions to protect the client's valuable data and make it accessible.

The manufacturing plant of the future appears more and more as a place where competitive value is produced through the integration of the enabling technologies of Industry 4.0 into the workplace. For the metals industry, ubiquitous information accessibility granted by connectivity and data sharing will play a significant role in facing the challenges posed by fluctuating markets and the potentially hazardous working environment. As operators explore new ways of working, Tenova is willing to design and embrace a digital strategy that represents the most effective solution for its clients.

References

1. J. Lee, E. Lapira, B. Behrad and H-A. Kao, "Recent Advances and Trends in Predictive Manufacturing Systems in Big Data Environment," *Manufacturing Letters,* 2013, pp. 38–41.
2. T. Wuest, D. Weimer, C. Irgens and K-D. Thoben, "Machine Learning in Manufacturing: Advantages, Challenges, and Applications," *Production & Manufacturing Research: An Open Access Journal,* Vol. 4, No. 1, 2016, pp. 23–45.
3. O. Simeone, "A Brief Introduction to Machine Learning for Engineers," 17 May 2018, https://arxiv.org/abs/1709.02840.
4. D. Kim, P. Kang, S. Cho, H-J. Lee and S. Doh, "Machine Learning-Based Novelty Detection for Faulty Wafer Detection in Semiconductor Manufacturing," *Expert Systems With Applications,* Vol. 39, No. 4, 2012, pp. 4075–4083.
5. J-L. Loyer, E. Henriques, M. Fontul and S. Wiseall, "Comparison of Machine Learning Methods Applied to the Estimation of Manufacturing Cost of Jet Engine Components," *International Journal of Production Economics,* Vol. 178, 2016, pp. 109–119.
6. J.H. Ang, C. Goh, A.A.F. Saldivar and Y. Li, "Energy-Efficient Through-Life Smart Design, Manufacturing and Operation of Ships in an Industry 4.0 Environment," *Energies,* Vol. 10, No. 5, 2017, pp. 1–13.

✦

This paper was presented at AISTech 2019 — The Iron & Steel Technology Conference and Exposition, Pittsburgh, Pa., USA, and published in the Conference Proceedings.