Driving Business Value in Industry 4.0: Big Data and Analytics for Steel Process Improvement

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

Marcelo Saparrat
TecnoAp S.A. de CV, San Pedro Garza García, Mexico
msaparrat@tecnoap.com

Ivonne Ibarra
TecnoAp S.A. de CV, San Pedro Garza García, Mexico
ibarra@tecnoap.com

Esteban Lopez
TecnoAp S.A. de CV, San Pedro Garza García, Mexico
elopez@tecnoap.com

Aracely Martinez
TecnoAp S.A. de CV, San Pedro Garza García, Mexico
amartinez@tecnoap.com

World steel producers face new challenges every day due to the need to be competitive in their markets as well as to produce high-quality products to meet the expectations of their most demanding customers.

This situation implies two things:

1. Processes must be well adjusted to maintain product properties within the most stringent tolerances and reduce the dispersion range.
2. It is critical to minimize operational risks that could generate machine downtimes, impacting the overall equipment effectiveness (OEE) of the processes.

For both challenges, the combination of technologies and disciplines such as big data and machine learning makes it possible to have powerful platforms for prediction, exploratory analysis and descriptive analytics, in order to provide:

- Virtual sensors: Predictors for spots where it would be difficult or impossible to measure physically.
- Process optimization in terms of productivity, equipment efficiency and production costs.
- Detection of operational risks that could lead to unplanned machine downtimes.
- Prediction of degradation of equipment performance.
- Detection of risks that compromise the quality of the product.
- Capture of expert knowledge that comes from the experience of experts who are close to retirement.

Quality and process problems can be approached from the perspective provided by data science, developing analytical models to make real-time predictions that can be fed to the control systems or be used as decision support systems for the operation.

Part I — Motivation

Objective — The objective of this work is to provide an introduction and present real cases of one of the disciplines of Industry 4.0 — big data and analytics — showing how data science can contribute to creating value and business benefits for the steel production companies, through the development of analytical models to explain and predict process behaviors.

The expected benefits of the development of these disciplines are:

Operative:

- Reduction of the time in which the process lines operate below their maximum speed.
- Improvement of efficiency by reducing machine downtime.

Quality enhancing:

- Reduction of complaint rates.
- Reduction of tons of retained product.
- Reduction of reprocessing times.
- Reduction of scrap.
- Reduction of the time in which the process lines operate below their maximum speed.
- Reduction of product storage time.
- Improvement of delivery time.

Mathematical Models in the Industry

White-Box Models — The use of mathematical models in industrial environments is not new and has been widely used for different purposes; for example, models for the setup of complex and multi-variate machines such as cold rolling mills. Classic modeling is based on physical models of the process that can be parameterized and, in some cases, use optimization techniques to find optimal operating points. These models generally require few historical data, their maintenance is difficult and they are not easily adaptable to process changes.

Rule-Based Models — Another type of frequently used model is based on production rules, which has inference engines that allow the capture of expert knowledge in each of the rules and functions. This enables the detection of specific process situations, which can trigger actions to interact with other systems or trigger alerts or alarms for the operators. In general, they are easier systems for analysts and processors to deal with, in order to configure the rules and actions of the scenarios that have to be captured, but they do not have the capacity to learn or to automatically modify the rules, generating a strong dependence on the process experts.

Machine Learning (So-Called Black-Box) Models — Machine learning is not a new discipline but together with the computing power of current systems and the ability to handle large volumes of data supported by big data technologies, it has emerged in recent years as an incredibly powerful tool. It is being widely used in complex applications, solving problems that would have been unthinkable before. This synergy opens the door to a new type of mathematical-statistical model: the so-called "data-driven model." This new type of model is built using machine-learning algorithms that are "trained" with historical data. As a result, it is not necessary (although it is very valuable in practice) to know in detail the mathematical relationships of the underlying physics, which in many cases are virtually impossible to model, due to the countless variables involved and the complexity that they entail. The machine-learning algorithms can capture the structures of data relationships and learn in a sense that will be defined in the following section.

Machine-Learning Models

Strictly speaking,[1] a computer program has the ability to learn from experience E, in relation to a set of objective tasks T, by measuring performance P, if the performance at which tasks T are executed improves with experience E.

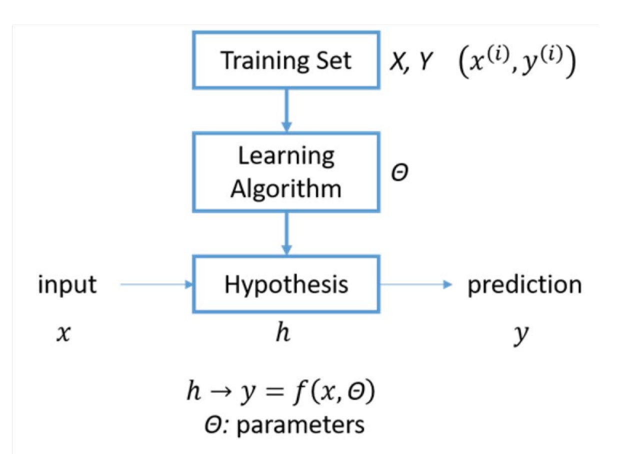
Example: a predictor of mechanical properties can learn from historical data (experience E) to execute task T to predict the variables' yield strength, elongation and tensile strength by measuring the performance of prediction P as the percentage of the population of the test set for which the prediction error is less than or equal to a certain limit. The experience E is called the "training set." The data of the training set is organized in a data set, which is a tabular ordering of the data where each column is a variable that is part of the model and each row is an observation. The variables should be only those that have relevance for the variable to be predicted.

This means that the design of a system based on machine learning requires a series of design decisions that involve the following:

- The choice of training set, which includes the choice of data set variables. Deciding which variables are part of the model and which are left out introduces a bias in the model.
- The "target" function to be learned.
- A representation of this target function.
- An algorithm for learning the target function from the training set.
- A measure of performance to evaluate the goodness of the prediction.

The learning process can be seen as the search in hypothesis spaces, in order to find the hypothesis that best fits the available training examples and other limitations or previous knowledge. Hypothesis spaces contain numerical functions, $y = f(x,\theta)$, with x being an input vector, $x \in X$. The challenge is to know the conditions under which these search methods converge toward an optimal hypothesis; for example, given an approximation function described in Fig. 1.

Figure 1



General scheme of the learning process.

The idea is to find an approximation function $f(x)$, which is fitted using historical empirical samples, denoted by:

$\{(x, y)_i\}$ for $i = 1, \ldots, n$.

Fitting $f(x)$ implies minimizing a certain loss function, which reflects the expected number of mistakes made by $f(x)$ in relation with y:

$$L'\left(f\left(\ \right)\right) = \frac{1}{n}\sum_{i} f\left(x_i\right) \neq y_i \qquad \text{(Eq. 1)}$$

Typically, the squared loss can be a good measure of the fitting performance of $f(x)$:

Squared loss $= L(f(x), y) = (f(x) - y)^2$

The KDD Process (Knowledge Discovery From Data) — Briefly, KDD is the methodological framework to follow to develop predictive or explanatory models and consists of the following stages:

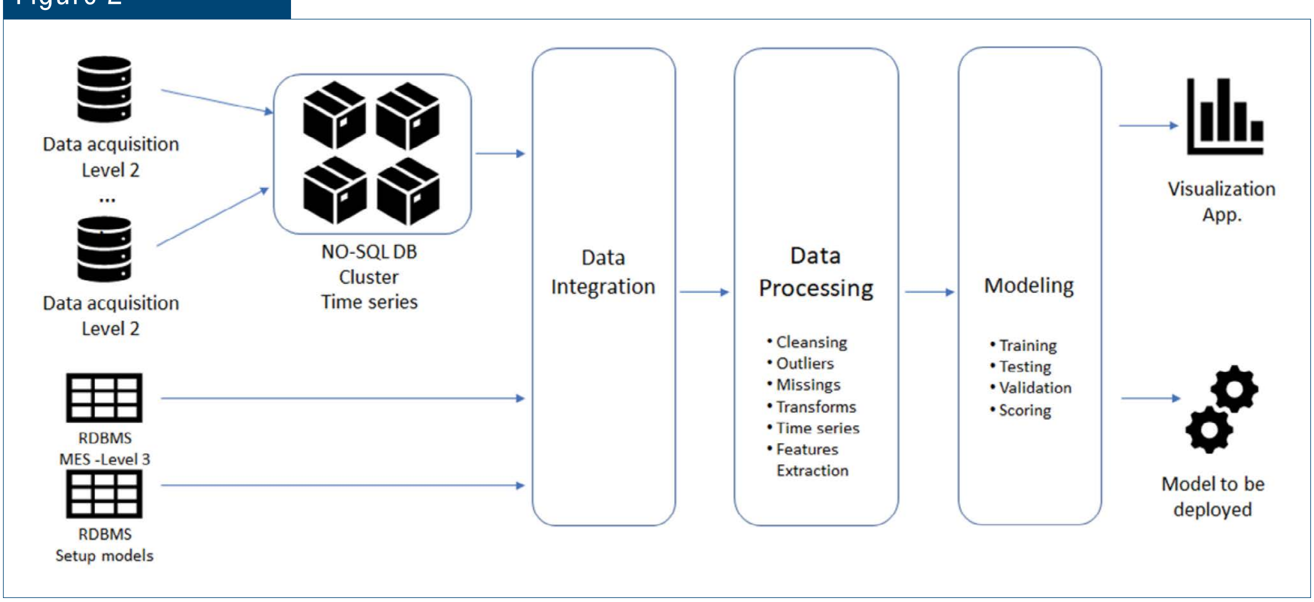
- Data Integration: Integration of the different data sources that will be part of the training, test and validation sets.
- Data Selection: Selection of variables (columns) candidates to be part of the model and selection of observations (rows) that are candidates to be part of the data set.
- Data Cleansing: Process of removal of noise in the data, detection of inconsistencies and solution of missing values.
- Data Transformation: Data preparation in the way that machine-learning algorithms need it.
- Feature Engineering: The creation of new variables calculated from existing ones, or that result from aggregation processes such as average, counting, maximum, minimum, etc. A particular topic at this stage is the treatment of time series, which will also be discussed later.
- Exploratory Analysis: This stage is about learning from the underlying relationships of the data; correlations, causality and limitations are explored, and the extracted knowledge is capitalized for its use in the next stage.
- Model Development: Construction of several models with different machine learning algorithms and selection of the best, after a process of model refinement.
- Validation and Refinement in Production: Once implemented, the model can continue to be improved.

The entire process can be seen in Fig. 2.

Time Series Processing in Industrial Context

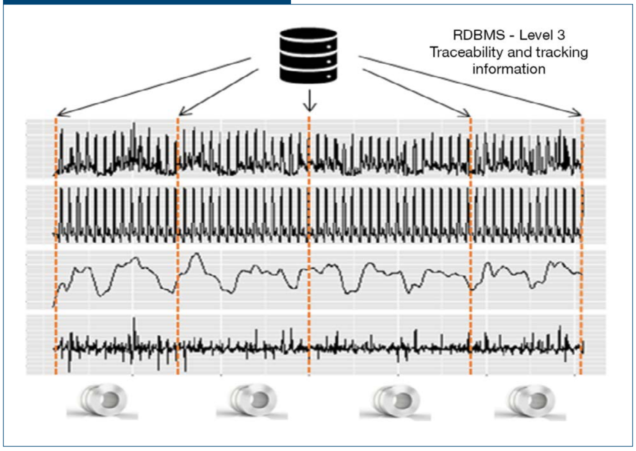
Time series are the most frequently found data type in processes and manufacturing lines, because they are signals from sensors and instrumentation systems.

Figure 2



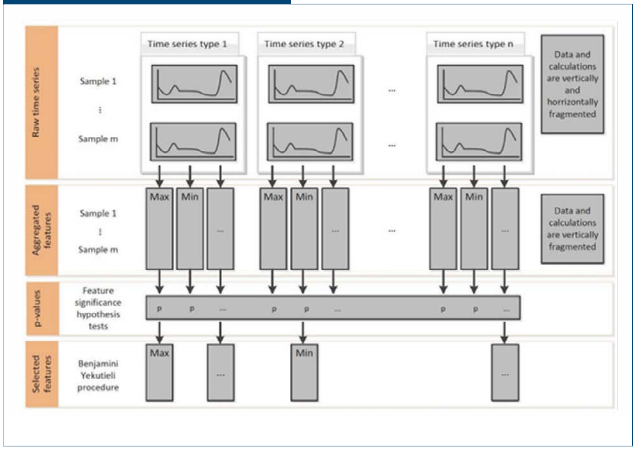
The big picture of data processing pipeline.

Figure 3



Time series slices using traceability/tracking data.

Figure 4



Pre-processing of time series.

In order for the time series to be processed properly, they must be put into context regarding the units that are being manufactured and that are supposed to have ensured and registered identity and traceability. For example, in the flat steel products industry, the manufactured unit is the roll or coil of sheet metal. When the coil goes through a process, for example a hot rolling mill, the time series have a beginning and an end given by the instants in time in which the coil goes through the process.

In the end, variables such as temperature, speed, rolling forces, etc., are pieces of time series for each coil; mathematically speaking, vectors of "n" samples, assuming a uniform temporal sampling.

To obtain the pieces of time series corresponding to each manufactured product (in this case, coils), it is necessary to integrate the traceability and tracking information that resides in the level 3 systems, as shown by Fig. 3. If variables coming from different production lines are to be integrated in the data set, traceability information is fundamental to filter the observations (coils) that followed a certain process path. The outline of this situation is shown in Fig. 4.

The question is how to incorporate variables that represent time series into a model. That is to say, if for the previous case, each observation (row) of the data set corresponds to a coil, each variable of the data set that represents a time series is a variable that is multi-valued, which means it is not a single value but a vector of samples. As it is not possible to treat the data in this way, it is necessary to submit the time series to a pre-processing called feature extraction, by means of mathematical operators whose input argument is a vector and results in a scalar.

There is a great variety of operators for the feature extraction, from the simplest ones (such as average, variance, maximum, minimum) to complex operators (such as fast Fourier transform, wavelets, etc.).[2]

These extracted features become variables of the data set, along with other process variables that are inherently single-valued. However, since time series are typically noisy and contain redundancies, it is necessary to submit them to a filtering process to decide whether it will be included in the data set.

Therefore, the balance must be maintained between the extraction of significant but probably fragile features and robust but probably not significant features. Some features such as the median will not be strongly influenced by outliers, while others such as the maximum value of the time series will be intrinsically fragile.

The Importance of Features for Statistical Anomaly Detection — An anomaly is a pattern that deviates from the expected or normal behavior. Therefore, anomaly detection looks at clues and compares attributes to discover out-of-the-ordinary patterns. Many times, there are multiple anomalies in groups, not just single occurrences.

Anomalies indicate some kind of disruption or abnormal operation of a machine or process. Again, analytics over the signals comes to help with this issue, through the development of models to detect anomalies and generate warnings and alarms to the operator.

Anomaly detection splits into two parts: (1) developing the right features, and (2) feeding these features into a statistical and machine-learning model that detects anomalies in the features. If done correctly, the detected anomalies will have a high correlation with on-site disruptions and can be used to create alerts with a low false positive rate. It may seem like the complexity of such systems focuses on the statistical part. However, it is well known that feature selection is key in real-life applications.

Part II — Real Use Cases

Mechanical Properties Prediction in a Hot Rolling Mill — In the steel industry, the mechanical properties of steel are the main differentiating features among different types of products. Properties such as tensile strength, yield strength and elongation of steel are essential parameters when deciding which material to work with in the construction industry or in the automotive industry. Additionally, these properties are also used to determine the parameters that are used in the steel manufacturing process itself.

Mechanical properties are usually measured in the physical testing laboratory, taking samples of the material at the end of the cold process. This procedure, although reliable, has several disadvantages. First, the sample is usually obtained from one end of the coil, which is where the greatest variability of its indicators is found, due to issues inherent to the process. Second, the mechanical properties can vary along the coil, which is currently impossible to discern. Third, the delivery of laboratory results is not immediate, so many products are retained; this causes an increase in production times and costs.

A predictive model for estimating the mechanical properties of the coil at the exit of the rolling mill is therefore advantageous and desirable, since it allows operating downstream in the process and correcting deviations, thus preventing the material from being declassed.

Since the mechanical properties depend greatly on the type of steel, different models were developed for three classes divided according to the chemical composition:

- Niobium steel (NB).
- Carbon steel (C).
- Vanadium steel (V).

For each type of steel, the tensile strength, yield strength and elongation were predicted, so there were nine models in all. The criteria to evaluate the suitability of the model were established as shown in Table 1, setting a tolerance range of deviation of the prediction from the real value and a minimum percentage of the predictions falling in that range for the model to be considered suitable.

The data set used for this case contained process data from a hot rolling mill (HRM) line, comprising a period of approximately 7 months of operation and including variables from the process (temperatures, strip speeds, reductions, etc.) and variables related to the chemical composition of the steel processed. Each observation (row) represented a coil that had been processed by the HRM and whose mechanical properties had been determined in the laboratory. The observations were divided into the three categories described earlier according to the chemical composition of the steel.

Table 1

Criteria for Model Suitability

Mechanical property	Tolerance range	Minimum percentage
Tensile strength	± 20 MPa	85%
Yield strength	± 20 MPa	75%
Elongation	± 3 %	75%

Table 2

Scores of the Models for Each Steel Type

Mechanical property	Score* per steel type		
	CMn	Nb	V
Elongation	83.66%	93.44%	87.50%
Yield strength	82.01%	81.36%	85.16%
Tensile strength	92.41%	87.66%	94.53%

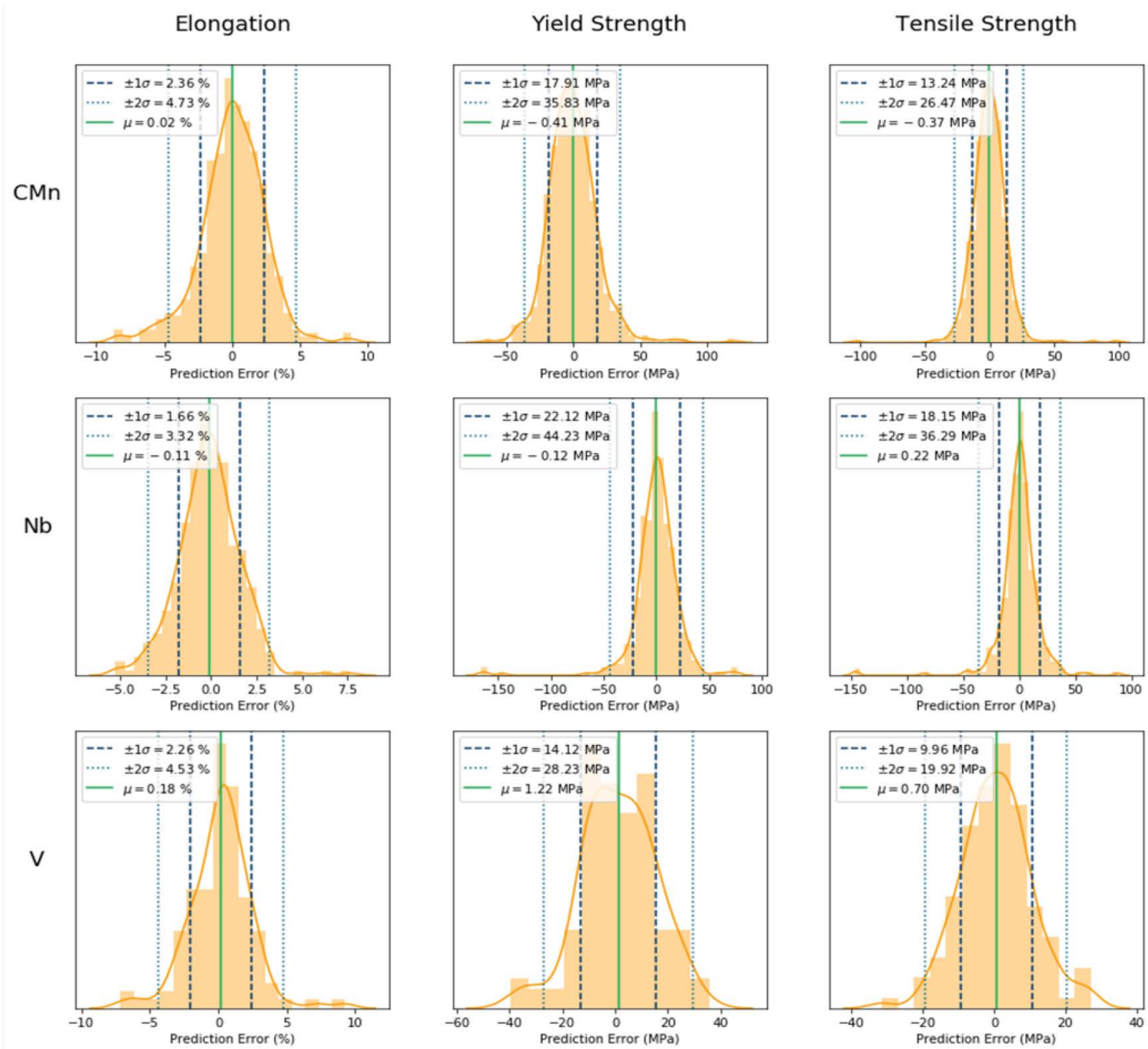
*Percentage of predictions falling inside the tolerance ranges shown in Table 1.

All data sets were subjected to a cleansing process: variables unrelated to the mechanical properties were identified and discarded, as well as variables with a high proportion of missing values; subsequently, observations with missing values were either removed or filled with an arbitrary value (e.g., column mean), depending on the nature of the variables. The distributions of the variables in the resulting data set were examined to identify outlier values and it was determined whether they represented measurement errors, record errors or true values, and they were treated accordingly.

The data sets were divided into training subset and test subset. For each steel type, a machine-learning algorithm called Gradient Boosting Regressor[4] was applied to predict each mechanical property. The algorithm was trained and cross-validated with the training subset, and then evaluated using the test subset. After tuning the parameters of the algorithm, the resulting scores of all the models surpassed the established thresholds (see Table 2). The distributions of the prediction errors for each model can be seen in Fig. 5.

Front-End Bending Prediction in a Reversing Hot Rolling Mill — During the hot rolling process, the thickness of the slab is reduced up to 99%. In the particular line where this project was carried out, most of this reduction is performed in a 4-high reversing roughing mill. In the roughing mill, the steel slab coming from the furnace passes through a set of horizontal

Figure 5



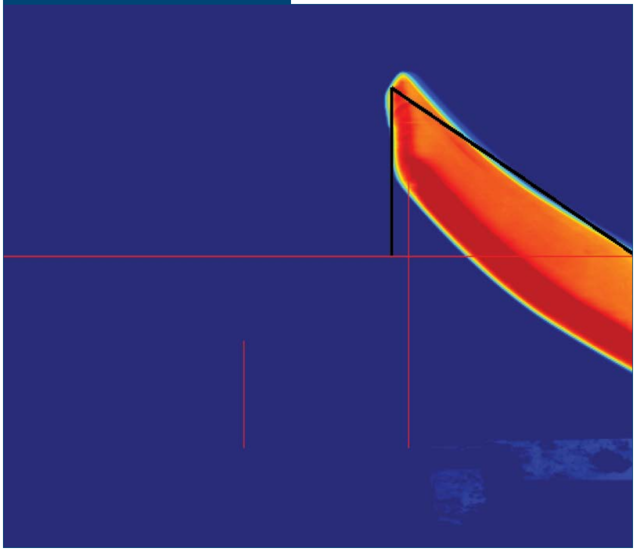
Distribution of errors in the predictions of mechanical properties.

and vertical rollers in a reversible fashion, i.e., passes several times through the same set of rollers. During the passes of the slab in the reversing mill, there is sometimes a curvature defect in the tip of the sheet being processed, as can be seen in Fig. 6. This defect, due to the form it presents, is colloquially known as SKI or ski. Depending on its direction and magnitude (height), it can cause the slab to hit the frame of the next stage or to be inserted into the conveyor rollers, with its consequent damage to the equipment and production stoppage.

There are solutions in the literature that have addressed the problem of SKI through finite element modeling and simulation, such as References 6 and 7. According to their reports, the SKI height depends on mainly the temperature difference between the upper and lower face of the material, the speed difference between the rollers, the friction coefficients between the rollers and the sheet, as well as the difference in diameter between the rollers.

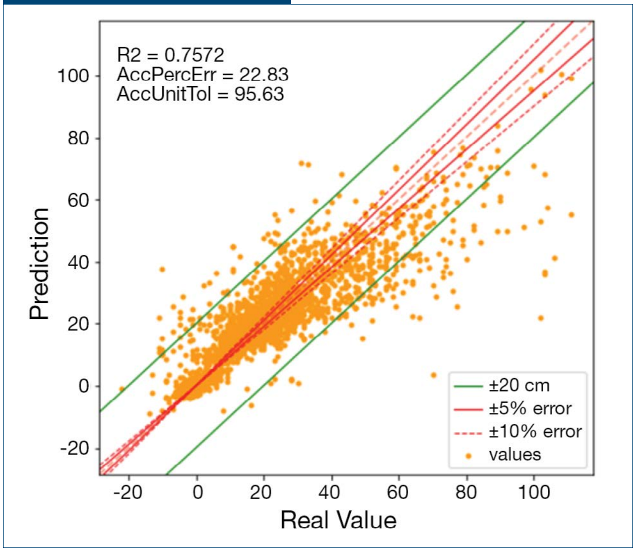
For this particular project, the goal was to determine the process variables that influence the presence of the defect and to be able to determine which slabs are more susceptible to present SKI by predicting the SKI height or front-end bending magnitude. The data set of raw process features consisted of variables from different stages of the process and product characteristics, mainly the heating furnace and the

Figure 6



Upwards front-end bending of the tip of a steel sheet during the fourth pass.

Figure 7



Scatter of the predicted ski values against the real values.

reversing mill setup data. An infrared thermography system (IRT) provided the magnitude, direction and angle of the SKI, which were used as targets. The variable groups used for modeling were: slab thermal profile, rollers, reversing mill setup, threading roller speed and slab characteristics.

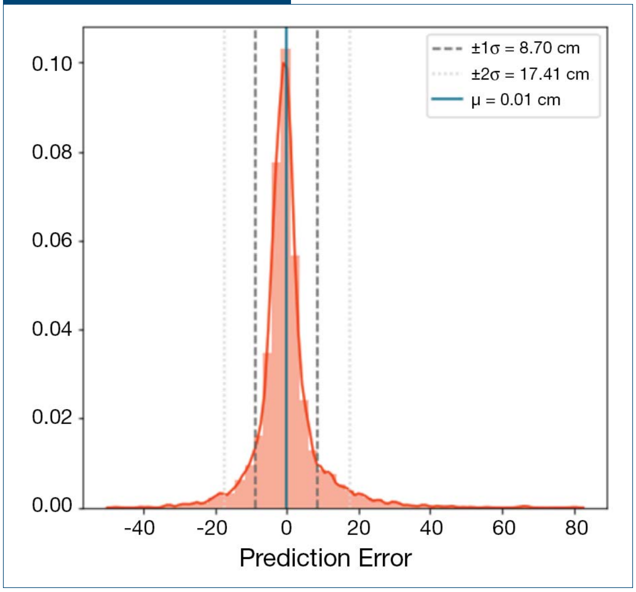
An exhaustive collection of data was done to minimize the probability of having insufficient data or not completely describing the phenomenon. After the stage of data cleansing and variable engineering, it was found that the phenomenon could not be explained with only a small number of variables. Principal component analysis was performed on the data, which showed that 95% of variance was explained by 21 main components from a data set with 223 variables, of which 85 variables are calculated.

Erroneous observations (e.g., wrong dates arising from manual operation of the line) in the data set were removed prior to modeling. The data set was then randomly split into train and test subsets, in an 80:20 ratio, having 12,782 and 3,364 observations, respectively. The train observations were then checked for outliers, which were removed from the data. Collinearity was also investigated and the variables with the highest collinearity were transformed into ratios to enhance their differences and improve their predictive power.

A meta-model of four individual models rendered the best performance in different tests. Such individual models were: a deep neural network, an XGB model,[4] a random forest model and an extra tree model. The predictions of the individual models were then averaged to get the final prediction. Modeling was made with the Python libraries sklearn[3] and keras,[5] while all data pre-processing was carried out in R.

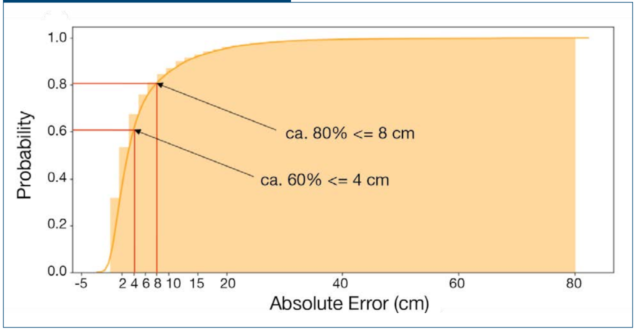
The best model had an R^2 of 0.75 on the test set, with 95% of the predictions having an absolute error below 20 cm and 61% having an error under 4 cm. The cumulative distribution function of the absolute errors is shown in Fig. 7. This model is currently being implemented, with the objective to predict possibly dangerous setup configurations that may lead to unsafe magnitudes of front-end bending. The

Figure 8



Distribution of the prediction error.

Figure 9



Cumulative distribution function of the absolute errors of the model.

prediction system will be fed with setup data and will raise an alarm if such an unsafe situation is detected. The distribution of the error is shown in Fig. 8, and the cumulative distribution function of absolute errors in Fig. 9.

Conclusions

As could be seen in the examples, the use of industrial analytics to predict process behaviors is a fact, not just a theoretical formulation. If the required conditions to develop analytical models are met, that is, historic data availability, data quality, relevant process variables part of the process instrumentation, and the availability of experienced domain experts to work in a multi-disciplinary team with data scientists, it is possible to create value through analytics-based innovations.

It is not an easy path, but the potential business value to be generated implies a substantial return on investment and could enable organizations to develop new differentiating strategies in the steel market.

References

1. T. Mitchell, *Machine Learning,* McGraw-Hill Inc., New York, N.Y., USA, 1997.
2. M. Christa, A. Kempa-Liehrb and M. Feindta, “Distributed and Parallel Time Series Feature Extraction for Industrial Big Data Applications,” October 2016.
3. F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *JMLR 12,* 2011, pp. 2825–2830.
4. T. Chen and C. Guestrin, “Xgboost: A Scalable Tree Boosting System,” *Proceedings of the 22nd ACM International Conference on Knowledge Discovery and Data Mining,* ACM, 2016, pp. 785–794.
5. F. Chollet et al., Keras, GitHub repository, 2015, available on https://github.com/fchollet/keras.
6. D. Anders et al., “A Dimensional Analysis of Front-End Bending in Plate Rolling Applications,” *Journal of Materials Processing Technology,* Vol. 212.6, 2012, pp. 1387–1398.
7. T. Kiefer and A. Kugi, “Modeling and Control of Front-End Bending in Heavy Plate Mills,” *IFAC Proceedings,* Vol. 40.11, 2007, pp. 231–236.

✦

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

Did You Know?

Researchers Devise Way to Print Better Martensitic Steel Parts

Researchers from Texas A&M University and the U.S. Air Force Research Laboratory have developed an optimized process framework that allows for defect-free 3D printing of components from martensitic steels.

“Strong and tough steels have tremendous applications but the strongest ones are usually expensive — the one exception being martensitic steels that are relatively inexpensive, costing less than a dollar per pound,” said Ibrahim Karaman, head of the university’s Department of Materials Science and Engineering. “We have developed a framework so that 3D printing of these hard steels is possible into any desired geometry and the final object will be virtually defect-free.”

Their work appears in *Acta Materialia.*