rapidminer

Product Quality Prediction and Optimization in Steel Manufacturing

European Commission

ArcelorMittal

UPMC SORBONNE UNIVERSITÉS

PREDICT

IVA

BFI

Predictive Sensor Data Mining for Product Quality Improvement (PRESED)

This e-book focuses on the data science aspects of a completed research project called PRESED (Predictive Sensor Data mining for Product Quality Improvement), co-financed by the Research Fund for Coal and Steel of the European Commission, that RapidMiner was proud to be a part of.

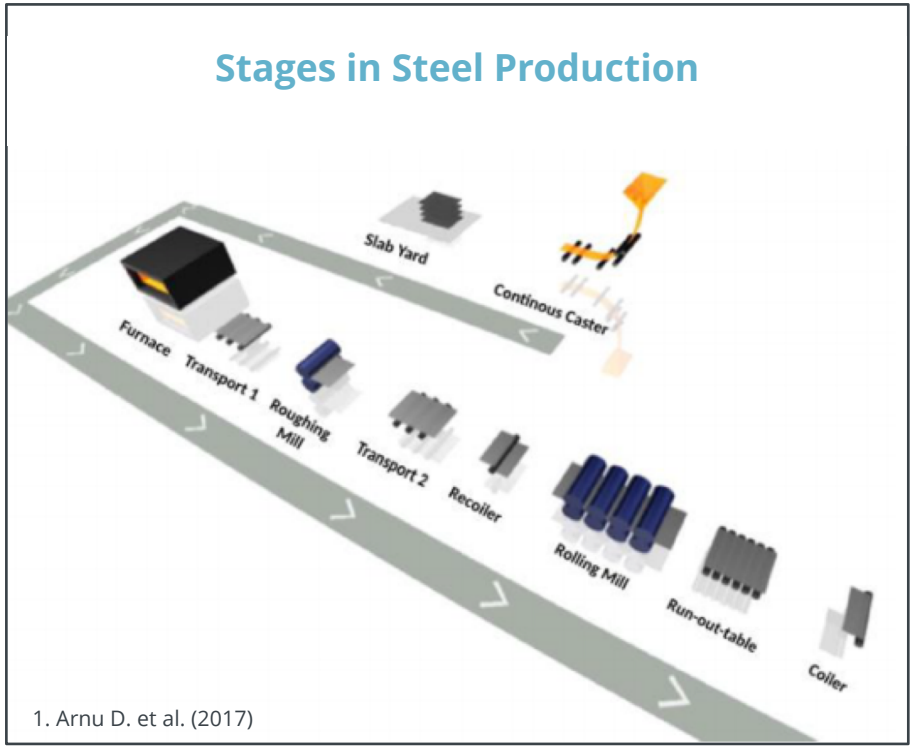
“There is a global increase in demand for steel, but steel manufacturing is a highly sophisticated and costly process where good quality is hard to achieve. Improving the quality remains a major challenge faced by the steel industry. The EU project PRESED (Predictive Sensor Data mining for Product Quality Improvement) addresses this challenge by focusing on widespread recurring problems. The variety and veracity of data, as well as the change in properties of the observed material complicates the interpretation of data....” [1]

1. Arnu D. et al. (2017)

The Steel Manufacturing Use Case

Producing steel is a complicated process, from extracting the iron ore to the final product. The liquified metal that comes out of the furnace (see image to the left) goes through many energy intensive processing steps. From time to time, defects occur to the material. Before this research project, when defects happened, the first reaction was to ask the process experts the cause of it. The experts helped the steel manufacturers understand where the defect came from and enabled the team to find a solution to fix it, but the answer was not always immediate, and it required some investigation. These investigations were what this research project aimed to fast-track and optimize, as the potential upside was huge, greatly reducing waste of energy and resources.

There are heavy machines that consume a lot of energy in steel manufacturing, and thus optimization is key to keeping the cost of final output low and competitive. And in the case of global steel manufacturers, producing millions of tons of steel and iron ore each year, in plants all over the world, optimization techniques that were developed as part of this project have the potential to reduce cost of production by millions of dollars.



1. Arnu D. et al. (2017)

The Challenges

Identifying defects quickly

There are many reasons why quality issues, like cracks and surface defects, happen in in steel manufacturing. Each step in steel processing needs to be very precisely mastered in terms of process conditions (temperature, casting speed, cooling flow rate, etc.), as slight deviations can lead to the occurrence of defects in the product. These conditions can vary from one facility to the next and thus require localized domain knowledge to implement and integrate.

Identifying the occurrence of a defect as early as possible in the production chain is highly valuable, as it spares the cost of going through the rest of the production steps for a defected final product. The application of RapidMiner was critical in conducting time series analysis of the sensory data and predicting defects early on in the manufacturing process.

Metallurgical Defect

Cracks

The Challenges Continued

Big data

In the case of steel manufacturing, there are 100s of sensors on the production line creating TBs of raw data every day. This can be very challenging to store, transform and retrieve in a data science friendly environment for optimal results. Designing a data structure to track the materials over the complete production process was key.

Unstructured data, not enough labeled data

Metallurgical and chemical processes tend to create sensory data that often requires a lot of clean-up and normalization before you can effectively apply machine learning techniques on the data.

Too many process variables

This is the case with most heavy industries, not just steel manufacturing. A lot of the final product, steel coils or sheets in this case, is heterogeneous in dimensions and physical properties. Also, tracking is difficult as the product changes shape (rolling, cutting). This adds a layer of complexity to learn defect patterns and apply machine learning results to identify them early in the process.

The Solution

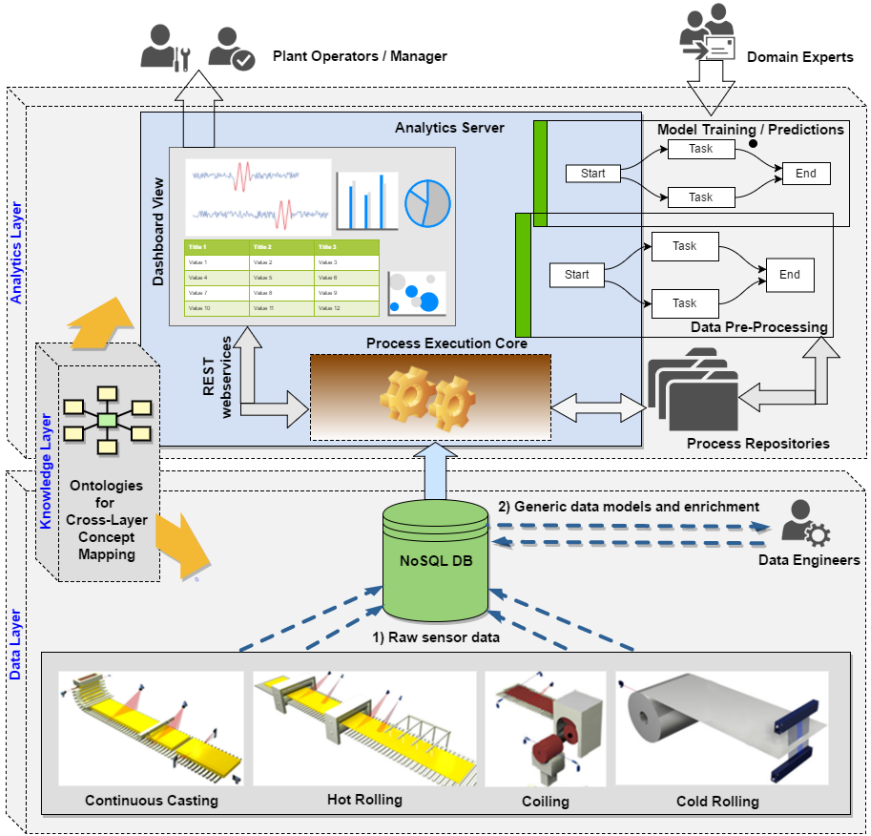
NoSQL Architecture

Not all data science projects are about deploying cutting-edge machine learning algorithms. Sometimes getting the fundamentals right (in this case time series sensor data storage and retrieval) is challenging.

At the high level, the new design for managing the sensory data better splits everything into three planes: a data layer, an analytics layer and a knowledge layer as shown.

The PRESED architecture serves the needs of three major stakeholders:

1. **The Data Engineers:** They model the raw data into a generic model. This is a preparatory activity which may be done infrequently.
2. **The Domain Experts:** They author the data mining processes on the now unified data model.
3. **The Plant Operators and Managers:** These actors execute the previously created processes to seek advanced insights on product quality and potential defects. The results are presented visually to assist human interpretation and enable timely corrective measures.

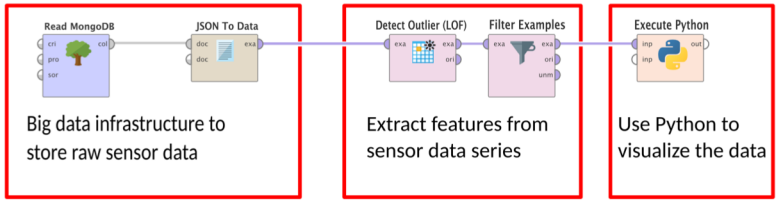


1. Arnu D. et al. (2017)

The Solution Continued

Visualization using RapidMiner Server and custom dashboards

Then an analytics server was built using RapidMiner and other tools, which enabled the team to improve model construction and management.



Applying Machine Learning to predict defects

The project partners have identified huge potential for the application of machine learning to predict product defects early in the production line. Some of these ideas are currently being deployed and tested in plants across the world. Testing ideas like Outlier Detection, detecting deviant observations that can help to identify possible problems in the production line early on, or the application of unsupervised learning to detect anomalous behavior of one or more continuous casting process variables.

The PRESED research project was a big step forward in unlocking the power of data sience in complex, physical manufacturing processes. RapidMiner is honored be the data science partner that helped turn this idea into a deployed solution, driving real business impact in steel manufactuing across the world.

Next Steps

Try time series analysis in RapidMiner Studio. We have a lot of interesting sample data sets and corresponding template process.

Try the RapidMiner Studio sample process analyzing German gas price data. See how the Extract Aggregates and the Differentation operators from the Time Series Extension are used to calculate the mean, minimum, maximum and standard deviation of the gas prices series and the differentiated gas price series for each day.

Reference

1. Arnu D. et al. (2017) A Reference Architecture for Quality Improvement in Steel Production. In: Haber P., Lampoltshammer T., Mayr M. (eds) Data Science – Analytics and Applications. Springer Vieweg, Wiesbaden; https://link.springer.com/chapter/10.1007/978-3-658-19287-7_12

RapidMiner brings artificial intelligence to the enterprise through an open and extensible data science platform. Built for analytics teams, RapidMiner unifies the entire data science lifecycle from data prep to machine learning to predictive model deployment. 400,000+ analytics professionals use RapidMiner products to drive revenue, reduce costs, and avoid risks.

RapidMiner Inc.

rapidminer.com
@rapidminer