Big Data

This article is part of the Digitalization Applications 101 learning module, which provides a comprehensive understanding on the basic concepts of digitalization terminologies, technologies and its applications in the steel industry. The course was developed by the Digitalization Applications Technology Committee as an introductory course to educate industry personnel in digitalization.

Author

Michael F. Peintinger
Managing Director, QuinLogic (SMS group), Cincinnati, Ohio, USA
michael.peintinger@quinlogic.us

The vision of a highly flexible smart factory producing customer-specific products at low additional cost in a short time to market is becoming a reality. Big data is the fuel of this fourth Industrial Revolution. The driving force behind this is the ever-increasing capability of analyzing data and the interaction with cyber-physical systems for commercial gains. This article defines the term in context of the steel industry and explores challenges as well as potential benefits.

What Is Big Data?

Everything we do is increasingly leaving a digital trace. When we browse the internet or partake in online shopping, our location and payment information is tracked and recorded, creating a profile of who we are and what we do. The same is true for material we produce. During production, a vast amount of data is captured from sensors generating a digital twin of the physical piece of material. Relating data from individual process steps generates even bigger data sets describing not only the current state but also the entire genealogy of the product. Considering the huge number of products that are manufactured, the amount of data aggregated over a given timeframe is larger than what can be analyzed by humans or commonly used software tools, and this is when the label "big data" is used.

Definition: Big Data – Big data describes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage and process data within a tolerable elapsed time.

Benefits of Big Data Analytics

The main benefit that can be gained from big data analysis is the detection of patterns and a better understanding of correlations and dependencies as well as the derivation of predictive models. Applications in the steel industry are, e.g., the root-cause analysis of defects detected by a surface inspection system at the hot mill and tracing it back to events at the caster. Training artificial intelligence (AI) algorithms on historical big data also enables predictive analytics. Monitoring incoming data in real time can trigger alarms and allows for corrective action once such a pattern is detected again. High-speed networks and integrated long-term data storage make plantwide big data integration feasible.

Types of Data

Structured data is located in a fixed field within a defined record, e.g., in a spreadsheet or a relational database. Order, customer and financial data are examples. As the name suggests, this kind of data is usually stored according to a predefined data model and this kind is also used in traditional data analysis. Unstructured and semi-structured data and its analysis is one of the main characteristics of the term big data. An estimated 80% of business-relevant information is unstructured. Examples are images, videos, uncategorized websites and documents. Another way of categorizing is by data that the business currently owns or generates and therefore has and controls access to, denoted