# Smart Data Analytics for Manufacturing Processes

by

Fabian Mohr

Submitted to the Department of Chemical Engineering
on August 15, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Chemical Engineering Practice

In times of Big Data is common in manufacturing processes to store as much information as possible to gain better insights and to support the decision making process with the goal of safe and efficient operations. However, the large amounts of data are often of little use without the suitable tools to process and analyze it. Data analytics can be described as the process of applying techniques with the purpose of discovering useful relationships among variables of interest. Those techniques can include, but are not limited to data transformations, variable predictions, classifications or fault detection. For each of those objectives many powerful tools for data analytics and machine learning have been developed in the last decade. Becoming an expert in the selection and application of all of these methods is challenging, as methods come from many disciplines, including applied statistics, analytical chemistry, pattern recognition, operations research, and computer science.

This multitude of methods can often be overwhelming for anybody who is trying to use those tools without being an expert in the field. No single method produces the best results for all real-world datasets. This situation has motivated the development of smart data analytics, that is, a decision tree that automatically selects the best methods based on a systematic interrogation of the characteristics of the dataset. This approach enables a user who is not an expert in the field of data analytics to find and apply the most suitable algorithm based on the nature of the investigated problem.

This thesis introduces two different smart data analytics approaches for the objectives of supervised classification and fault detection, as they arise for example in the context of process monitoring schemes for chemical manufacturing processes. For both objectives, a visual representation of the method selection process is presented in form of a data analytics triangle. The necessary interrogation framework for the model selection process is introduced and the overall approaches for both objectives are demonstrated in case studies showing great performance over a variety of different supervised classification and fault detection problems.

Additionally, smart data analytics predictive modeling techniques are applied to industrial end-to-end biomanufacturing datasets for two monoclonal antibody products to predict critical quality attributes. While these methods are known to perform well on second-order tensorial data, they do not allow for the utilization of higher-order tensorial data. Consequently, different approaches considering both second-order and third-order tensorial data combined as possible inputs to the predictive modeling problem are proposed. It is shown that the utilization of the proposed methods is capable of significantly improving the prediction performance if the dataset is analyzed correctly beforehand.

Thesis Supervisor: Richard D. Braatz

Title: Edwin R. Gilliland Professor of Chemical Engineering