World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:16, No:10, 2022

Evaluation of the CRISP-DM Business Understanding Step: An Approach for Assessing the Predictive Power of Regression versus Classification for the Quality Prediction of Hydraulic Test Results

Authors : Christian Neunzig, Simon Fahle, Jürgen Schulz, Matthias Möller, Bernd Kuhlenkötter

Abstract : Digitalisation in production technology is a driver for the application of machine learning methods. Through the application of predictive quality, the great potential for saving necessary quality control can be exploited through the data-based prediction of product quality and states. However, the serial use of machine learning applications is often prevented by various problems. Fluctuations occur in real production data sets, which are reflected in trends and systematic shifts over time. To counteract these problems, data preprocessing includes rule-based data cleaning, the application of dimensionality reduction techniques, and the identification of comparable data subsets to extract stable features. Successful process control of the target variables aims to centre the measured values around a mean and minimise variance. Competitive leaders claim to have mastered their processes. As a result, much of the real data has a relatively low variance. For the training of prediction models, the highest possible generalisability is required, which is at least made more difficult by this data availability. The implementation of a machine learning application can be interpreted as a production process. The CRoss Industry Standard Process for Data Mining (CRISP-DM) is a process model with six phases that describes the life cycle of data science. As in any process, the costs to eliminate errors increase significantly with each advancing process phase. For the quality prediction of hydraulic test steps of directional control valves, the question arises in the initial phase whether a regression or a classification is more suitable. In the context of this work, the initial phase of the CRISP-DM, the business understanding, is critically compared for the use case at Bosch Rexroth with regard to regression and classification. The use of cross-process production data along the value chain of hydraulic valves is a promising approach to predict the quality characteristics of workpieces. Suitable methods for leakage volume flow regression and classification for inspection decision are applied. Impressively, classification is clearly superior to regression and achieves promising accuracies.

Keywords : classification, CRISP-DM, machine learning, predictive quality, regression
Conference Title : ICPEAD 2022 : International Conference on Production Engineering and Automation Development
Conference Location : Istanbul, Türkiye
Conference Dates : October 20-21, 2022