Chapter 15
A Tutorial to Developing Statistical Models for Predicting Disqualification Probability

Ilmari Juutilainen
University of Oulu, Finland

Satu Tamminen
University of Oulu, Finland

Juha Röning
University of Oulu, Finland

ABSTRACT

Different industries utilize statistical prediction models that predict the product properties in process planning, control, and optimization. An important aim is to decrease the number of disqualifications. The model can prevent disqualifications efficiently if the disqualification probability is predicted accurately. This study gives step-by-step instructions for developing, validating, comparing, and visualizing models that predict the disqualification probability with high accuracy. The work summarizes industrially applicable statistical modeling methods that are most suitable for the development of accurate predictors for the disqualification probability. Currently, the information on such statistical methods, e.g. quantile regression, modeling of distribution shape, and joint modeling of mean and deviation, is scattered in the existing literature. The main contribution of this work is that it pulls together this methodology into a unified framework which allows the comparative analysis of probability predictors that are based on the different approaches. The proposed modeling procedure (ProPred) is demonstrated using three manufacturing industry applications. In the case applications, the predictors generated using the ProPred procedure are 10-30% more efficient in avoiding disqualifications by means of process planning and control operations than the baseline predictors.

DOI: 10.4018/978-1-4666-0128-4.ch015

INTRODUCTION

As the competition in product quality and demand for high-quality and energy-efficient products grow, manufacturing industries will meet more quality challenges. The quality requirements are becoming stricter and the cycle of new product development is shortening. Many of these quality challenges can be solved in the terms of statistical methodology (Hahn & Doganaksoy, 2008; Montgomery, 2009). The importance of the reduced variation in product properties was initially lifted to everyday practice of process improvement by Taguchi (Taguchi, 1987). Later, several another approaches to reducing variation in product properties has been proposed, e.g. (Steiner & Mackay, 2005). Also, the business process improvement practice Six Sigma pays attention on the reduction of the costs associated with the variation (Pyzdek & Keller, 2009). The objective is to meet the quality requirements with optimized cost. The disqualification rate can be decreased by accurate process control and well-done planning of production, products and processes. As the numerous applications have proved, statistical models that predict the dependence of the conditional distribution of product properties on the source information and process variables are useful tools to improve the planning and process control practices in manufacturing industries. The modeling of tail distribution behavior has become an important tool for risk management (Crouhy, Galai & Mark, 2005).

Rapidly growing amounts of data are measured from the production process and stored into process databases. In addition to the quality assurance measurements on product properties, the process data contain measurements on variables that cause the desired properties and the variation in it. Typically, process data bases include at least the amount and properties of the process input materials, the planned process settings, the target values for the process variables and real-time measurements on the realized process variables. (Miletic, Quinn, Dudzic, Vaculik, & Champagne, 2004).

Designed experiments have a long history on the primary data source of statistical quality improvement and for the estimation of prediction models for the utilization in planning and control (Montgomery, 2009). In comparison to the designed experiments, the advantage of process databases is the larger number and cheaper prize of data points. Thus, the production data provide a very competitive data source for the estimation of statistical models for the needs of control and planning (Harding, Shahbaz, Srinivas & Kusiak, 2006). This chapter is a methodological guide that relies on experience on several applications in which process data is utilized for the development and estimation of prediction models which are then taken into action in the planning and control in different manufacturing industry applications.

This chapter describes a step-by-step approach to the estimation and selection of statistical models that predict accurately the probability of disqualification. The aim is to provide a guide which helps to develop models to be used for predicting and simulating the effect of source information on the risk of disqualification. Models like this can give significant benefits when utilized by integrating them into software that are used in process control and in the planning of products, production and processes (Choudhary, Harding, & Tiwari, 2009). The benefits are realized as the decreased number of disqualifications, decreased variation in the product properties and savings in the consumption of raw materials and energy.

Our claim is that disqualifications related to failing with quality requirements can be decreased in many industries by the following ways:

- By utilization of statistical models that predict the quality properties of end products. The models can do decisions on behalf of humans or give additional in-