Abstract: Data mining in automated production systems provide high potential to increase the Overall Equipment Effectiveness. Nevertheless, data of such machines/plants include specific characteristics regarding the variance and distribution of the dataset. For modelling product quality prediction, these characteristics have to be analysed to interpret the results correctly. Therefore, an approach for the analysis of variance and distribution of datasets is proposed. The evaluation of this approach validate the developed guideline, which identify the reasons for in-consistent prediction results based on two different datasets of the same production system.