Proceedings of the 5th International Conference on Statistics: Theory and Applications (ICSTA'23)
Brunel University, London, United Kingdom – August 03-05, 2023
Paper No. 118
DOI: 10.11159/icsta23.118

Product Quality Control Forecast Using Machine Learning Algorithms: A Case Study

Rosa Arboretti[1], Elena Barzizza[2], Nicolò Biasetton[2], Riccardo Ceccato[2], Livio Corain[2], Marta Disegna[2], Luca Pegoraro[2], Luigi Salmaso[2], Andrea Vinelli[2], Pietro Barbieri[3], Luca Bortolan[3], Matteo Canale[3], Marco Giada[3], Davide Longi[3], Fabio Napol[3], Danny Paganin[3]
[1]Department of Civil Environmental and Architectural Engineering, University of Padova
Via Marzolo, 9, Padova, 35131, Italy
rosa.arboretti@unipd.it
[2]Department of Management and Engineering, University of Padova
Stradella S. Nicola, 3, Vicenza, 36100, Italy
elena.barzizza@phd.unipd.it; nicolo.biasetton@phd.unipd.it; riccardo.ceccato@unipd.it; livio.corain@unipd.it; marta.disegna@unipd.it; luca.pegoraro@unipd.it; luigi.salmaso@unipd.it; andrea.vinelli@unipd.it
[3] Leading multinational company in the fashion industry

Abstract **-** In today's highly competitive global market, industries require near-perfect quality to be profitable. Therefore, the prompt prediction of a defect product has become an issue of paramount importance and an opportunity for manufacturing companies to move quality standards forward. This article presents a real case study based on Machine Learning algorithms suggested to create a knowledge-based intelligent supervisory system to predict defect products on a fashion industry. Defect detection is formulated as a binary classification problem and several Machine Learning algorithms have been compared for a clearer understanding of which algorithm is best suited to the available data. The predictions have been obtained performing the algorithms on pre-processed dataset which was made available by the industry. The Random Forest, LightGBM and C5.0 algorithms performed similarly well on the data. However, the Random Forest algorithm has been selected as the best one since it allows to reduce the rate of false negative (i.e. the proportion of defected-free products wrongly classified as defected ones), that is the goal of the analysed industry.

Keywords**:** Machine Learning, defect prediction, manufacturing, quality control

1. Introduction

Nowadays, quality represents a strategic factor in all industries contributing to classify companies in profitable and non-profitable ones. It can be considered as an index of internal performance, as it allows developing diversify product or service differentiation policies. In recent years, quality of products and services has even increased in importance due to the massive spread of online reviews timely written by customers at no cost who want to share their post-use or post-experience feelings and opinions. Therefore, providing defective or poor-quality products can significantly affect both the brand reputation and the customers' loyalty to a certain brand.

The definition of quality management process has significantly changed over the years [1]. The first quality control (QC) process has been developed during the mass production period (1900-1940) and it simply consisted in the inspection of the final products. However, the increasing pressure from the market led to a first major theoretical shift towards the concept of process quality. The idea underlying this theory was that looking for errors was much less efficient than find the source of the errors and remove it. In the early 60s, the quality assurance theory emerged. The process quality, i.e. a posteriori reaction to quality control of either products or processes, was not enough anymore and industries wanted to prevent risks of failure products or services identifying their major causes before their appearance. With the increase of both the complexity of products and the interdependencies with the supply chain, the ISO 9000 certification has been introduced to define quality management standards to help industries ensure they meet customer and other stakeholder basic requirements related to products or services. In the last ten years, the quality management process shifted further from being implemented because of market pressure to being implemented because of the general importance to deliver high-quality results.

Improving products and services quality is one of the main goals of the fourth industrial revolution, or Industry 4.0, [2] and the interconnection among automation, Machine Learning (ML) algorithms, and real-time data is the key to fulfil it.

From a manufacturing perspective, the ability to efficiently collect and analyse huge amount of data using powerful ML algorithms can improve QC [3] revealing hidden patterns in data and thus enabling a broader analysis of the impact of all the factors that contribute to manufacturing processes.

In the recent years, we assisted to a growing literature on the use of ML algorithms to classify or predict QC thanks to the exponential increases in computer power, database technologies, optimization methods, new efficient and robust ML algorithms, and big data. ML algorithms can handle high-dimensional, multivariate data massively reducing human efforts and improving product quality [4]. These algorithms can provide suitable solutions to gather fast and reliable information and to understand the implicit relationships existing within large datasets collected in dynamics and complex environments [5]. An example of ML application to optical inspections on finished products adopting a decision tree for the optimal classification of battery separator defects can be found in [6]. ML algorithms have also been adopted to predict final quality of products in the early stages of the manufacturing process (see [7], [8], [9]) and to detect defect products in mature organisations (i.e. organisations who merge different tools to improve manufacturing processes quality, such as lean production, standards conformity, six sigma, design for six sigma) who only generate a few defects per million of opportunities [3]. Several ML algorithms have been trained and evaluated to predict dimensional defects in a real automotive multistage assembly line in [10] while [11] analysed a welding process and they used RF and J48 (a classification algorithm that creates decision trees based on information theory) to correlate arc sound with weld quality. Recently, [12] predicted the quality of drilled and reamed bores of hydraulic valves, by estimating the diameter and the concentricity of the bores, starting from the torque measurements.

This study presents the learning process and pattern recognition strategy for a knowledge-based intelligent supervisory system, in which the main goal is the prediction of defected products. Defect detection is formulated as a binary problem, i.e. OK for defect-free products and KO for defect products. Different supervised ML algorithms to predict binary output have been compared. A real case study of a company working in the fashion industry is presented. The main contributions of this research are: the analysis of the features of an important production phase of the analysed product; the application and comparison of several ML algorithms to predict the QC outcome.

2. Problem statement

The analysed product is a part of a fashion accessory made of acetate. The unit of analysis in this case study is a batch of identical products. The manufacturing process is characterised by a series of five automatic processes that must be completed. At the end of each step, an operator evaluates the final output and decide whether the batch of products can pass to the next step or must re-do the step. Therefore, each batch of products may run each step of the manufacturing process multiple times until the operator is satisfied with the final output. Consequently, the amount of time necessary to complete a process varies according to several subjective factors. At the end of the manufacturing process an operator visually inspects 10% of the batch of products and decide, accordingly to a series of internal quality standard requirements, whether the batch must be classified as a defected or a defected-free one. In this study, 38,743 batches of products have been initially analysed and 28,010 batches made the final ML dataset. Per each batch a set of product features (such as dimension, colour, design) and a set of process features (such as working time per step, external materials used in each step, starting and ending date of the manufacturing process, quantity of products of the batch) have been collected.

3. Predicting product quality control with ML algorithms

Classification is the assignment of an object defined by a set of features to one of several predetermined classes by means of a learning rule. Thus, it belongs to the macro area of supervised ML algorithms aiming at inferring a function to describe labelled training data (e.g. data with classification class) to predict the output categorical variable [13].

Being the outcome variable of this study (QC result) a binary one, the following classification ML algorithms have been implemented:

- Logistic regression
- Multivariate Adaptive Regression Splines (MARS)
- Generalized Additive Model using Splines (GAM)
- Linear Discriminant Analysis (LDA)
- Quadratic Discriminant Analysis (QDA)

- Neural Network (nnet)
- Random Forest (ranger)
- Support Vector Machines (svm)
- Light Gradient Boosting Machine (lightGBM)
- Elastic Net regression (glmnet)
- C5.0 Decision Tree Algorithm

The choice of these algorithms has been driven by the literature review and experience in other real case studies accrued by the authors. The input features used in these algorithms are the ones described in the previous section (both product and process characteristics) so the algorithms must handle both quantitative and qualitative input variables.

The training dataset is used to train a set of candidate algorithms using different tuning parameters. The 10-folds cross validation method is used to evaluate the generalization ability of each candidate algorithm and select the best, according to a relevant model selection criterion. The classification success of the models is measured by several indices based on the confusion matrix (total accuracy, sensitivity, precision, recall, specificity, false negative rate, false positive rate), the Receiver Operating Characteristics (ROC) curve and the Area Under the Curve (AUC). The tuning parameters of the ML algorithms have been optimised in order to minimise the false negative rate (i.e. the proportion of defected products wrongly classified as defected-free ones) that is the aim of the manufacturing process analysed. Indeed, as described in Section 1, from the industry perspective it is worse to classify a product as free from defects when it is not than to misclassify a defect-free product and reprocess it.

4. Results and discussion

The QC outcome, i.e. OK-KO, has been recoded as OK=0 and KO=1. Comparing the ROC curve and the AUC of all ML algorithms described in the previous section it emerges that the Random Forest, LightGBM and C5.0 algorithms performed similarly well on the data (see figure 1). The optimal classification threshold that allows to determine whether a new batch is more probably a defect one or not, is 0.2. Among the three best algorithms, the Random Forest algorithm allows to minimize the rate of false negative (29.24% against 35.0% and 39.71% of the LightGBM and C5.0 algorithms respectively), i.e. the goal of the analysed industry.

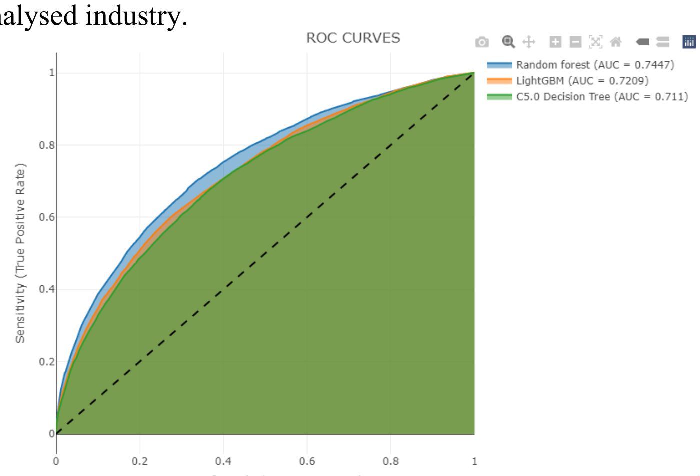

Fig. 1: ROC curves and AUC of the three best ML algorithms

Since defect products occur rarely in the analysed industry (26.89%), the dataset is unbalanced. Therefore, the ML algorithms have been tested on a balanced dataset obtain using the undersampling method. The previous results are robust confirming the Random Forest algorithm with 0.2 threshold as the best one in terms of rate of false negative (0.7% against 4.3% and 3.9% of the LightGBM and C5.0 algorithms respectively).

Acknowledgements

BIRD 2022 project titled “Fuzzy theory in Unsupervised Machine Learning algorithm and Sentiment Analysis”

References

[1] A. Weckenmann, G. Akkasoglu, T. Werner, “Quality management - History and trends”, *The TQM Journal,* vol. 27, pp. 281-293, 2014.

[2] Y.Z. Ray, X. Xun, K. Eberhard, T.N. Stephen, “Intelligent Manufacturing in the Context of Industry 4.0: A Review”, *Engineering*, vol. 3, no. 5, pp. 616-630, 2017.

[3] C.A. Escobar, R. Morales-Menendez, “Machine learning techniques for quality control in high conformance manufacturing environment.”, *Advances in Mechanical Engineering*, vol. 10, no. 2, pp. 1-16, 2018.

[4] H. Deshmukh, V.S. Jadhav, S.P. Deshmukh, K.K. Rane, “Towards Applicability of Machine Learning in Quality Control – A Review”, *ECS Transactions*, vol. 107, no. 1, 14729, 2022.

[5] J.G. Köksal, İ. Batmaz, M.C. Testik, “A review of data mining applications for quality improvement in manufacturing industry”, *Expert Systems with Applications*, vol. 38, no. 10, pp. 13448-13467, 2011.

[6] J. Huber, C. Tammer, S. Krotil, S. Waidmann, X. Hao, C. Seidel, G. Reinhart, “Method for Classification of Battery Separator Defects Using Optical Inspection”, in *49th CIRP Conference on Manufacturing Systems (CIRP-CMS 2016)*, *Procedia CIRP*, 2016, vol. 57, pp. 585-590.

[7] R. Wagner, B. Haefner, G. Lanza, “Function-Oriented Quality Control Strategies for High Precision Products”, in *15th CIRP Conference on Computer Aided Tolerancing – CIRP CAT 2018*, *Procedia CIRP*, 2018, vol .75, pp. 57-62.

[8] F.C. Neto, T.M. Geronimo, C.E.D Cruz, P.R. Aguiar, E.E.C. Bianchi, “Neural models for predicting hole diameters in drilling processes”, in *8th CIRP Conference on Intelligent Computation in Manufacturing Engineering*, *Procedia CIRP*, 2013, vol. 12, pp. 49-54.

[9] P.G. Mathews, M.S. Shunmugam, “Neural-network approach for predicting hole quality in reaming”, *International Journal of Machine Tools & Manufacture*, vol. 39, pp. 723-730, 1999.

[10] R. S. Peres, J. Barata, P. Leitao and G. Garcia, "Multistage Quality Control Using Machine Learning in the Automotive Industry," *IEEE Access*, vol. 7, pp. 79908-79916, 2019.

[11] A. Sumesh, K. Rameshkumar, K. Mohandas, R.S. Babu, “Use of Machine Learning Algorithms for Weld Quality Monitoring using acoustic Signature”. *Procedia Computer Science*, vol. 50, pp. 316-322, 2015.

[12] S. Schorr, M. Möller, J. Heib, D. Bähre, “Quality Prediction of Drilled and Reamed Bores Based on Torque Measurements and the Machine Learning Method of Random Forest”, *Procedia Manufacturing*, vol. 48, pp. 894-901, 2020.

[13] M. Mohri, A. Rostamizadeh, A. Talwalkar, *Foundations of Machine Learning.* Cambridge MA: MIT press, 2012.