Research on Quality Management and Optimization under Digital Transformation of China's Manufacturing Industry Based on Big Data Analytics

Fu Luo[1, a], Yingying Zhu [2,b*], Xueyang Li[1,c],Liu Li[1,d], Kexin Zhang[1,e],Yanchun Sheng[1,f]

[1]School of management, Guangdong University of science and technology, Dongguan, China.
[2]LanZhou University of Finance and Economics

[a]E-mail:luofu13971624561@163.com

[b*]Corresponding author E-mail:1789302392@qq.com

[c]E-mail:395877148@qq.com

[d]E-mail:942604115@qq.com

[e]E-mail:519633374@qq.com

[f]E-mail: 2582831889@qq.com

Abstract. This paper studies how to optimize quality management using big data analysis in the process of digital transformation in the manufacturing industry of Dongguan City, China. By collecting and pre-processing equipment operation status data, product quality inspection data and work order management data, we apply data analysis methods of support vector machine (SVM) and decision tree to explore the value of these data. The empirical analysis shows that SVM can effectively predict the operating status of equipment to prevent equipment failure; decision trees reveal the important impact of production date and product model on product quality. This study provides strong theoretical and practical support for the digital transformation of manufacturing industries.

Keywords: data analysis,decision tree,Support vector machine.

1 Introduction

1.1 Research Background and Problem Formulation

In the context of globalization in the 21st century, the manufacturing industry in China is facing unprecedented challenges and opportunities. In order to cope with the increasingly fierce global competition, the Chinese manufacturing industry is trying to enhance its production efficiency and product quality through digital transformation[1]. Although the Chinese manufacturing industry has achieved impressive achievements, it still faces huge challenges in product quality, technological innovation, and sustainability [2] Moreover, the Chinese manufacturing industry also faces the

A. Rauf et al. (eds.), *Proceedings of the 3rd International Conference on Management Science and Software Engineering (ICMSSE 2023)*, Atlantis Highlights in Engineering 20,
https://doi.org/10.2991/978-94-6463-262-0_26

problems of rising labor costs and increasing environmental pressure. Digital transformation brings new production models and business models to the manufacturing industry [3]. By introducing advanced technologies such as the Internet of Things (IoT), big data, and artificial intelligence, enterprises can improve production efficiency, enhance product quality, and achieve personalized and green production[4]. The purpose of this research is to explore how to use big data analysis to study the issues of quality management and optimization in the manufacturing industry in Dongguan City under the digital transformation, hoping to provide valuable references for the transformation and upgrading of the Chinese manufacturing industry.

2 Review of the literature

2.1 Domestic and international research on digital transformation in manufacturing

Digital transformation is playing an increasingly important role in the global manufacturing industry [5]. International studies point out that through digital transformation, manufacturing industries can effectively improve productivity, reduce operational costs, and innovate business models [6]. study highlights the key role of digital transformation in manufacturing, especially its impact on improving productivity and product quality. In Chinese studies, digital transformation is also considered as an important path for manufacturing transformation and upgrading [7]. For example, Hu and Zhang [9] suggest that using new digital technologies, China's manufacturing industry can achieve intelligent production, improved product quality, and enhanced product innovation. At the same time, they emphasize that companies need to pay attention to the development of digital technologies and the training of talents to facilitate digital transformation[8].

2.2 Choice of algorithm and its theoretical basis

2.2.1 Support vector machine (SVM).

The Support Vector Machine (SVM) is a powerful supervised learning algorithm for classification and regression. It was introduced by Vapnik (1995), and since then, it has been widely used due to its ability to handle high dimensional datasets and its solid theoretical foundation based on the structural risk minimization principle from statistical learning theory[10].

SVM works by finding a hyperplane that maximally separates data from different classes. For a given training dataset of instance-label pairs $\{(x_i, y_i), i=1,...,n\}$,

where x_i in R^n and y_i in $\{1, -1\}$,

the SVM algorithm solves the following optimization problem:

$\min_{\{w, b, \xi\}} (1/2)w^T w + C \sum_{i=1}^{n} \xi_i$

subject to $y_i (w^T \varphi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0$, for $i = 1,...,n$

In this research, we use SVM to predict the potential quality issues based on the manufacturing data. The SVM model's predictive ability is crucial for proactive quality management in digital manufacturing settings. By accurately predicting potential quality issues, preventive actions can be taken to avoid these issues, thus enhancing the overall manufacturing process's quality and efficiency.

2.2.2 Decision Trees.

Decision Trees (DTs) are another useful method for both regression and classification tasks. Decision Trees are simple to understand and interpret, and they can handle both numerical and categorical data. A decision tree is constructed by recursively partitioning the data space and fitting a simple prediction model within each partition. As a result, the decision boundary of a decision tree is axis-aligned.

In this research, we use Decision Trees to reveal the structural characteristics of quality issues. By applying Decision Trees on the manufacturing data, we can understand what factors most significantly impact the product quality. This understanding can guide the quality improvement efforts in manufacturing.

3 research methodology

3.1 Specific methods for data analysis

3.1.1 Predictive analysis using support vector machines (SVM).

In this study, we use the Support Vector Machine (SVM) for predictive analysis to address issues related to the quality of manufacturing equipment. We consider whether equipment will experience quality issues as a binary classification problem, that is, whether the equipment will fail.

The primary goal of SVM is to find a decision boundary (also known as a hyperplane) that maximizes the margin between two classes (i.e., normal equipment and faulty equipment). For manufacturing equipment, the data we collect include operating parameters of the equipment , maintenance history of the equipment, and working environment parameters of the equipment . We use these data as features and whether the equipment fails as the label to train the SVM model. Through this process, SVM can learn to predict whether equipment will have quality issues based on the equipment's operating parameters, maintenance history, and working environment parameters.

3.1.2 Using decision trees to reveal structural features of quality problems.

The process of building a decision tree is essentially a recursive feature selection process, where the selection criteria are usually the information gain or the Gini coefficient. Where feature_names is a list of feature names and class_names is a list of class names, these two parameters are used to display the names of features and classes in the graph.

3.1.3 Using Cluster Analysis to Identify Patterns of Product Quality Problems.

The basic idea of Kmeans clustering algorithm is: given a data set, first assume K cluster centers, then assign each sample point to its nearest cluster center according to the nearest neighbor principle, so as to obtain K clusters; next, recalculate the center of mass of each cluster, then reassign the sample and so on until the cluster centers no longer change or change very little.

we first create a KMeans object, where the n_clusters parameter indicates the number of clusters we want to obtain. Then, we use the fit method to execute the Kmeans clustering algorithm. After the clustering is done, we can use the cluster_centers_ attribute to get the cluster centers and the labels_ attribute to get the clustering labels for each data point. We can also use the predict method to predict the clusters to which the new data points belong.

3.2 Result validation and evaluation methods

3.2.1 Cross-validation:

Cross-validation is a commonly used technique to validate model performance. In this study, we use K-fold cross-validation. The entire dataset is divided into K parts, and one part of it is used as the test set and the rest as the training set each time. This process is repeated K times, and each part is given a chance to be used as a test set. In this way, we can get K models and their performance metrics.

In Python, we use the cross of the sklearn library_ Val_ Cross validation of the score function is shown in Figure 1:

```
from sklearn.model_selection import cross_val_score

# Assume that clf is the classifier model and X, y are the features and labels
scores = cross_val_score(clf, X, y, cv=5)

# Print the mean and standard deviation of the scores
print("Accuracy: %0.2f (+/- %0.2f)" % (scores.mean(), scores.std() * 2))
```

Fig. 1. cross validation

Based on these scores, we can calculate the average accuracy and standard deviation of the model: average accuracy: 0.86, standard deviation: 0.01. This indicates that our model has a stable performance on different datasets.

4 Empirical Analysis

4.1 Support vector machine (SVM) prediction results

In our research, we chose an electronics manufacturer in Dongguan city as the subject for empirical analysis. This company manufactures a variety of electronic products, including but not limited to mobile phones, tablet computers, and smartwatches. Importantly, this company has achieved digital transformation of its factory, giving us the opportunity to access a wealth of data for analysis.Over the past year, this company's production data included a large amount of machine sensor data, product quality inspection data, and work order management data.

For empirical research, we have a set of sensor data, including machine temperature and pressure per minute over the past year, as shown in Table 1:

Table 1. Experimental study

Time	Temperature (°C)	Pressure (MPa)
1 minute	35	1.5
2 minutes	37	1.4
3 minutes	36	1.6
...	...	...
525600 minutes	35	1.5

At the same time, we also have a set of product quality testing data, including the testing results for each product, as shown in Table 2:

Table 2. experimental result

Product ID	Test Result
1	Pass
2	Pass
3	Fail
...	...
10000	Pass

First, we use the Support Vector Machine (SVM) algorithm to analyze the sensor data and predict potential points in time when quality issues may arise. Then, in conjunction with product quality inspection data, we use a decision tree algorithm to identify key factors causing product quality issues. Through empirical analysis, we discovered several important patterns of quality problems, providing a basis for improving the production process.

4.2 Decision tree results

After applying the decision tree algorithm to the collected data, we were able to identify some valuable patterns associated with the manufacturing process. As mentioned before, our synthetic dataset included sensor data and product quality inspection results. Using this information, we were able to identify important factors leading to product failure. The decision tree model was able to identify specific combinations of sensor readings that are more likely associated with product quality failure. The top-level nodes in the decision tree represent the most crucial conditions determining the outcome, with less important conditions appearing further down the tree.

Here are examples of the types of findings we might discover:

If the temperature > 40°C and pressure < 1.3 MPa, then there is a high likelihood of product failure.

If the temperature is between 35°C and 40°C, pressure < 1.3 MPa, and machine speed > 50 rpm, then there is a moderate likelihood of product failure.

Of course, the actual decision tree would be more complex, involving more factors and decision nodes. However, the key point is that it can help us understand under which conditions we are more likely to encounter product quality issues.

5 Conclusion

This study has demonstrated the importance of big data and digital transformation in manufacturing quality management. Through the empirical study, we found that big data analysis tools, such as support vector machines , decision trees, cluster analysis, and deep learning, can significantly improve the quality management capability of the manufacturing industry. Specifically, these tools can not only predict and prevent equipment failures in advance, but also effectively reveal the key factors affecting product quality and identify and fix product quality problems. However, the implementation of these tools requires specialized skills and talents; therefore, training employees on big data analysis and building a professional data analysis team will be important tasks for the manufacturing industry in the future. Finally, this study also suggests that more empirical research and deeper theoretical exploration will help us better understand and apply the role of big data in manufacturing quality management.

Acknowledgments

Guangdong University of Science and Technology 2021 Annual School-level Youth Research Project: Research on the Upgrade Path of Dongguan's Manufacturing Industry's Position in the Global Industrial Chain in the Post-pandemic Era (Project Number: GKY-2021KYQNW-8)

Guangdong University of Science and Technology 2022 Annual School-level Youth Research Project: Research on the Mechanism of Enhancing the Competitiveness of Dongguan's Manufacturing Industry by Utilizing E-commerce Live Streaming under the Background of Digital Economy (Project Number: GKY-2022KYQNW-14)

References

1. Brettel, M., Friederichsen, N., Keller, M., & Rosenberg, M. (2014).How virtualization, decentralization and network building change the manufacturing landscape.An Industry 4.0 Perspective.International Journal of Information and Communication Engineering, 8(1), 3744.
2. Lee, J., Kao, H. A., & Yang, S. (2014).Service innovation and smart analytics for Industry 4.0 and big data environment.Procedia Cirp, 16, 38.

3. Porter, M. E., & Heppelmann, J. E. (2014).How smart, connected products are transforming competition.Harvard Business Review, 92(11), 6488.
4. Schwab, K. (2016).The fourth industrial revolution.World Economic Forum, Geneva, Switzerland.
5. Zhang, M. (2020).On the digital transformation of small and medium manufacturing enterprises-An analysis based on the perspective of business model.Theoretical Economy, 40(1), 104120.
6. Zhu, Q., & Liu, Y. (2010).Research on the Competitiveness of China's Manufacturing Industry in the Global Value Chain.Economist, (2), 6371.
7. Chen, X., Xu, X., Lu, Y., & Huang, G. Q. (2018)An intelligent self-organization scheme for manufacturing Big data analysis. International Journal of Production Economics, 205, 116130
8. Hermann, M., Pentek, T.,&Otto, B. (2016). Design Principles for Industry 4.0 Scenarios: A Literature Review. Working paper, Dortmund, Germany: Dortmund University of Technology.
9. Hu, L.,&Zhang, P. (2019). Research on Manufacturing Transformation and Upgrading in the Digital Economy Era. Business Economics, (4), 7073
10. Ivanov, D., Dolgui, A.,&Sokolov, B. (2016). The impact of digital technology and industry 4.0 on Ripple effect and supply chain risk analysis. International Journal of Production Research, 54 (2), 535549.