Total Quality Management Through Defect Detection in Manufacturing Processes Using Machine Learning Algorithms

Almira S. Kahya(✉), Selin Şişmanoğlu, Zeynep Erçin,
and Hatice Camgöz Akdağ

Management Engineering Department,
Istanbul Technical University, Istanbul, Turkey
{kahyaal, sismanoglu, ercinz, camgozakdag}@itu.edu.tr

Abstract. Total Quality Management is the new raising value of all industries. The more it is revealed that TQM is one of the key success factors for the companies, the more it is being absorbed by the industries. This study aims to analyze TQM approaches considering its history and development worldwide while observing manufacturing industry with machine learning applications in order to identify the defects in the process before completed which contributes continuous improvement to the system. Also, different descriptions of quality according to the customer satisfaction will be examined.

Keywords: Total Quality Management · Quality · Manufacturing · Customer satisfaction · Machine learning applications · Continuous improvement · Problem solving · Defect detection

1 Introduction

In todays' highly competitive markets, firms have been striving to ensure their production process without failure while satisfying customers' expectations with defect free product. Profit margins and the product quality are highly fragile elements and affected by defect ratios in manufacturing due to nonconforming internal failure costs. Rosyidi, Nugroho, Jauhari, Suhardi, and Hamada [1] state that the organization must invest in various resources which identify with achieving quality objectives to diminish the quality costs, so as to accomplish the desired quality improvement. In this paper, a manufacturing process of a product in detergents sector will be observed while using Machine Learning (ML) Algorithms to identify the root causes and prevent the system from possible unwanted cost and loss of time with comparing different computer-intensive methods that are implemented through R Studio Programme. Logistic regression, decision tree, and linear discriminant analysis, to name a few.

Improvements in technology, increasing competition environment made people to create new things. The number of entrepreneurs is highly augmented in recent years. Everybody is seeking to make new products, implement new processes. On the other hand, existing organizations should develop their products or service quality to not to scratch the competition in global market. Like the all products, services and processes;

N. M. Durakbasa and M. G. Gençyılmaz (Eds.): ISPR 2019, LNME, pp. 508–516, 2020.
https://doi.org/10.1007/978-3-030-31343-2_44

the new products or new process ideas are assessed in terms of their qualities. The feature, performance, reliability and durability are the main factors which form quality. According to Shewhart and Levitt, quality is defined as conformance to specifications; quality of conformance relates to the degree to which a product meets certain design standards [2]. However, quality has not a unique definition. Juran [3, 4] states that quality is fitness for use, Leffler [5] asserts that quality is product desirable attributes, Taguchi emphasizes that quality is loss avoidance and Ryall and Kruithof [6] indicate that quality is meeting customer expectations. All of the products or processes can have some deficiencies which can be decreased by effective quality management methods. The main purpose should be to have at least deficiency. Based on the official web site of the ASQ Quality wordbook, quality management is defined as a system which main goal is to accomplish customer desires during development in the processes with minimum cost, its components are quality planning, quality assurance (defect prevention), quality control (which includes product inspection and other elements, such as competence), and quality improvement [7]. These four parts are very important to determine errors and minimize it.

This paper begins with introduction part and then continue with literature review which is explained total quality management and applications in life. In the following parts, methodology & application in machine learning which will be used in quality to find deficiencies will be analyzed. The last part which is conclusion part, will give the results of machine learning to resolve the deficiencies.

2 Literature Review

2.1 Problem Solving Tools in Quality

In this competitive market conditions, quality improvement is an inevitable phase in manufacturing. According to Tsou and Chen [8], products that has unsatisfactory quality lead to decrease efficiency, enlarge cost of business operations, and lessen customer satisfaction. PDSA cycle and Six Sigma's DMAIC steps are common methods to identify and understand the current conditions of the system and how to approach an existing problem in a problem solving.

There are seven quality tools which are particularly Flow Chart, Control Chart, Check Sheet, Histogram, Pareto Diagram, Cause and Effect Diagram, Scatter Diagram to identify the nonconformance in the process and try to solve the problems in the quality management. The common goal of those tools is to identify the deficiencies in the process and improve the system eventually. The techniques and components of total quality management is used in many areas like Six Sigma, lean management departments. Thus, the relationship among the customer and company develops which results to reduce deficiencies in manufacturing process. Ertosun, Zehir and Müceldilli [9] assert that the main factor to have competitive advantage is changed from quality to novelty with changing world conditions. Thanks to novelty which enables to create new products, enter new markets for companies and permits to modify themselves rapidly. Moreover, WHY-WHY Diagram technique is a milestone for this case, since it looks for the root causes of the problem by asking "Why" five times.

2.2 Machine Learning Approaches

Statistical applications of quality such as Six Sigma that is a part of Statistical Process Control tries to prevent defects in the products or services by practicing statistical methods to control the processes. Like Six Sigma, Machine Learning is one of the arising approaches that use statistical applications. Therefore, Machine Learning is one of the new computer-based approaches, which can act and learn like humankind, can be used for finding deficiencies in the process to set an example. It is more accurate than traditional rules. Sebastiani [10] describes that, thanks to this methodology, interferences are not required from expert witnesses for the development of distributing an alternate arrangement of classifications which will be resulted to a significant reserve funds as far as master labor and accomplishment by human specialists. Moreover, machine learning applications can be used to determine the deficiencies in the process not only after the process completed but also it is possible to specify the root cause of the problems like WHY-WHY Diagram technique. There is a huge advantage of machine learning is a time saving approach and the computers do the procedure without faults unlike individuals who have to use analytical thinking to improve the overall system with minimum defects and maximum quality to ensure total quality management standards. Abdaziz [11] explains that, all association staff individuals during the time spent covering clients' desire through using critical thinking strategies to improve the nature of every hierarchical item and services are connected by total quality management. It is used in every area like automotive, manufacturing process such as detergent, cancer screening, sales analysis. The methods which are used in machine learning are differentiated according to dataset characteristics. Decision Trees, Logistic regression, Boosted Trees, Linear Discriminant Analysis etc. are classification methods which can be used in this case to classify whether a product is defective or non-defective. Since we have only 2 classes as defective and non-defective whose values are 1, 0 respectively. "As Escobar and Morales-Menendez [12] stated, A learning process (LP) and pattern recognition (PR) strategy for a knowledge-based (KB) ISCS is presented, aimed at detecting rare quality events from manufacturing systems." The defect detection is formulated as a binary classification problem, in which the l1-regularized logistic regression (LR) is used as the learning algorithm. The most relevant properties are shown by the product of ideas which is used for forecasting model. Data visualization is vital manner for both the owner of the facility and client to make the data more clear and easy to interpret. Therefore, the machine learning techniques should not include unnecessary variable that leads to prevent model to be fit and absence of full comprehension which will cause the inconsequential properties. Decision trees is also used to promote product and determine errors and reduce them in total quality management with acceptable visual support. Yussupova, Kovács, Boyko1 and Bogdanova1 [13] describe that the fundamental things for the client which are the most critical item perspectives are distinguished and displayed from decision trees.

The four level of total quality management which are quality control, quality assurance, continuous quality improvement, quality award models will be completed and will reveal deficiencies and causes. By using several Machine Learning models, quality control will be completed easily and defects will be reduced before production is completed which improves the production process with a higher yield. However, it should be considered as Nakajima [14] stated, in ML-based systems any minor changes in training dataset, can lead to a major influence on the test results which has to be considered carefully in order not to end up accusing an important variable as useless. Therefore, this delicacy might rise a new objection to quality tools. In this study validation set approach is used to determine the samples precisely, since there is a huge observation set.

3 Method

3.1 Logistic Regression

Logistic Regression is a type of classification method and can use logistic function formula simply as follows:

$$\mathrm{P(X)} = \frac{e^{\beta_0 + X\beta_1}}{1 + e^{\beta_0 + X\beta_1}} \tag{1}$$

While using maximum likelihood method to fit the non-linear model. Here, the formula looks for the probability of an observation that comes from that class, not for the values.

Therefore, the goal of the formula given in the Eq. (2) is to find the maximum value of the beta that is the regression coefficient.

$$\mathrm{L}(\beta_0, \beta_1) = \prod\nolimits_{i:y_i=1} p(x_i) \prod\nolimits_{i':y_i} (1 - p(x_{i'})) \tag{2}$$

R will be doing this procedure behind it.

3.2 Decision Tree

After plotting the tree, we examined the confusion matrix to find out the error rate of classification with this formula as given in the Eq. (3):

$$E = 1 - argmax_c(\hat{\pi}_{mc}) \tag{3}$$

Those variables increase the explanatory power of the model, even though they are not directly affecting the class determination by the model.

3.3 Linear Discriminant Analysis

LDA is less direct approach in comparison with Logistic Regression Model, however; in some cases, it can be more reliable. Therefore, it uses the Bayes' Theorem to assign the variables to the classes with the formula given in the Eq. (4):

$$P_r(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^{K} \pi_i f_i(x)} \tag{4}$$

4 Application

4.1 Dataset and Variables

In this study, there are mainly 9 variables. For example, date is represented as date in the fit model for dataset, on the other hand; experience variable has been shown in the model as Exp_1, Exp_2 and Exp_3 according to corresponding values in data cleaning process of the study. The model is fit according to 14 variables as indicated in the Table 1. The output is Defective referring to defect information of coloring. This variable will be the target variable in prediction stage. Additionally, Defective is a qualitative response and could have tried to clarify by those 9 variables as stated below table.

Table 1. Variables and dataset

Variable	Type of variable in dataset	Explanation of variables in the dataset
Date	Date	Data of measurement
Shift	Shift_1, Shift_2, Shift_3	Measurement shift period
Experience	Exp_1, Exp_2, Exp_3	Experience information of measurement (a metric related with the measurement, not related with operator)
Percentage of dye (%)	DyePercentage	Percentage of coloring
Flow rate	'Flow rate'	Flow rate of coloring
Teflon	Teflon	Teflon rate of part
Thickness	Thickness	Thickness rate of coloring
Part	Part_a, Part_b, Part_o	Which part of product is colored

4.2 Logistic Regression

To classify either the product is defective or non-defective we fit a logistic regression model with glm() function. This fitted model explains the deficiency through the variables of date, time and shift of the production,

condition of the environment with experience categorical variable, dye percentage, flow rate of coloring, teflon rate of part, thickness rate of coloring, colored part of the product.

After fitting binomial logistic regression model, it is examined that thickness variable has the highest explanatory power of the deficiency. As seen in the Variable Importance graph below which is created by varImp() function (Fig. 1).

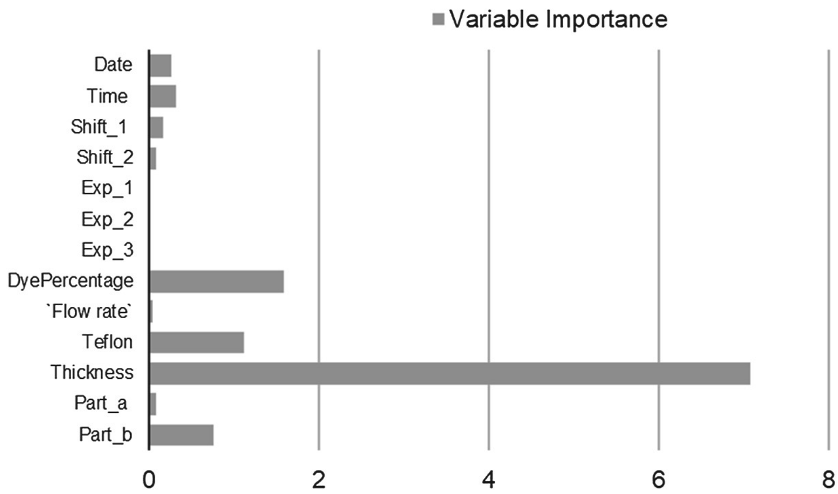

Fig. 1. Importance of the variables. Shift 3 and Part_o are not included in the chart because R found atomic importance in the model.

Although Teflon rate and Dye percentage seem to have a significant effect in the model, they do not tend to improve accuracy. We cannot say that other variables are useless, however to detect defection thickness should be examined firstly. With this machine learning algorithm, the defects are identified after the production which has a non-conformance cost for the firm (Fig. 2).

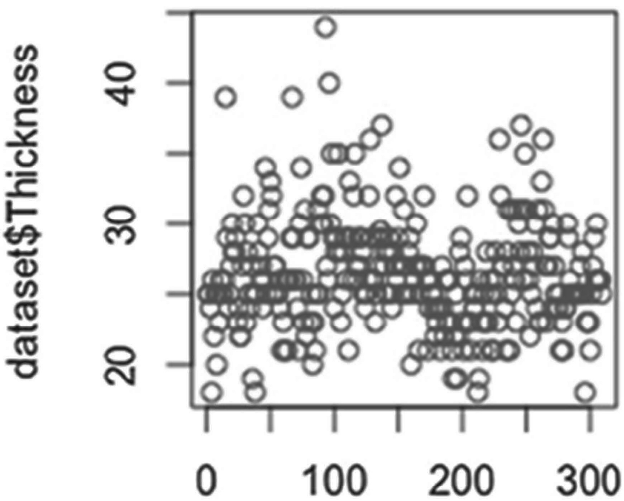

Fig. 2. Scatter plot of thickness variable

4.3 Decision Tree

We also plotted this data in a decision tree to examine if another possible solution or relations occur. But the results are the same as below (Fig. 3):

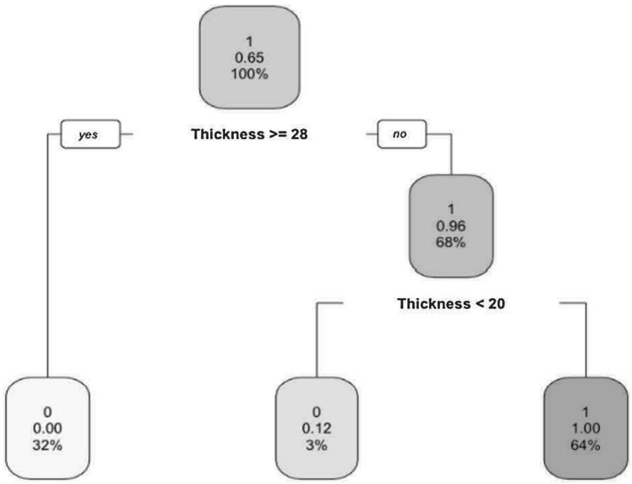

Fig. 3. Decision tree of thickness variable

Even though our number of observation is not that high to see the prediction accuracy of our model we split the dataset into test and training samples with the proportion of 80:20 and applied our model to the training data. Then, we test the rest to reveal whether to ensure the model can predict with low error rate which means model explains the data convincible. The prediction results and the actual states of the deficiencies are shown in the confusion matrixes with their test prediction accuracy rate (Table 2).

Table 2. Accuracy of alternative models

GLM with all variables: Accuracy : 87,09%

	Non-Defective	Defective
Non-Defective	42	2
Defective	1	17

GLM with Thickness variable: Accuracy : 88,70%

	Non-Defective	Defective
Non-Defective	43	7
Defective	0	12

Lda with all variables: Accuracy : 95,71%

	Non-Defective	Defective
Non-Defective	19	1
Defective	2	48

Lda with Thickness variable : Accuracy : 98,57%

	Non-Defective	Defective
Non-Defective	21	1
Defective	0	48

As seen in the accuracy tables linear discriminant analysis (lda) outperforms the other methods. James, Witten, Hastie and Tibshirani [15] stated if the number of observation is relatively smaller to number of predictors, and the predictors in each class has Gaussian distribution, linear discriminant analysis offer more reliable model than logistic regression. Also in lda.fit, model performs more accurate if the p = 1 and variable is thickness. Below lda.fit graphs are attached to reveal the distribution of defected (group 1) and non-defected (group 0) products (Fig. 4).

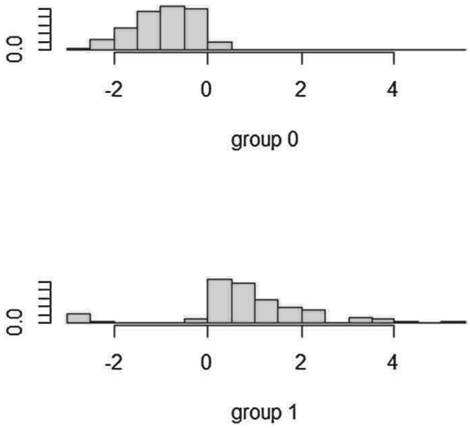

Fig. 4. Distribution of defected (group 1) and non-defected (group 0) products

5 Conclusion

In conclusion, TQM is related with all the phases of a production starting from planning to customer satisfaction. Since any failure that is not detected will reduce customer satisfaction and will bring non-conformance costs, it is crucial to capture failure possibilities in production processes. Amiri, Noghondarianb and Safaeic [16] explained that even though implementing such a system may lead to a rise in administrative costs, total cost of quality can be balanced with lower costs of failure. In order to reduce cost of failure, machine learning algorithms can be installed to the processes. In our example, before starting the manufacturing process, if the part of the production which some preventive action should be taken to control the products' color thickness. This will lead less failure at the end of the manufacturing, therefore an increase in the probability of customer satisfaction.

References

1. IEEE: Quality improvement by variance reduction of component using learning investment allocation model. In: 2016 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM) Industrial Engineering and Engineering Management (IEEM), p. 39 (2016)
2. Elshaer, I.: What is the Meaning of Quality? MPRA, pp. 3–7 (2012). https://mpra.ub.uni-muenchen.de/57345/1/MPRA_paper_57345.pdf
3. Juran, J.M.: Quality Control Handbook. McGraw-Hill, New York (1974)
4. Kemp, S.: Quality Management Demystified. McGraw-Hill, New York (2006)
5. Leffler, K.B.: Ambiguous changes in product quality. Am. Econ. Rev. **72**(5), 956 (1982)
6. ISO 9000:2008: The Quality Management System. Fundamentals and vocabulary
7. Asq.org: Quality Assurance vs Quality Control: Definitions & Differences | ASQ (2019). https://asq.org/quality-resources/quality-assurance-vs-control. Accessed 24 Feb 2019
8. Tsou, J., Chen, J.: Case study: quality improvement model in a car seat assembly line. Prod. Plan. Control. **16**(7), 681 (2005). https://doi.org/10.1080/09537280500249223

9. Zehir, S., Zehir, C., Mücedilli, B., Ertosun, Ö.: Total quality management practices' effects on quality performance and innovative performance. Procedia Soc. Behav. Sci. **41**, 274 (2012). https://www.sciencedirect.com/science/article/pii/S1877042812009111
10. Sebastiani F.: Machine Learning in Automated Text Categorization (2001). https://arxiv.org/pdf/cs/0110053.pdf
11. Al-Qahtani, A., Abdaziz, N.: The impact of Total Quality Management on organizational performance. Eur. J. Bus. Manag. 1 (2015). ISSN 2222-1905 https://www.researchgate.net/publication/294886200_The_impact_of_Total_Quality_Management_on_organizational_performance
12. Escobar, C.A., Morales-Menendez, R.: Machine learning techniques for quality control in high conformance manufacturing environment (2018). https://doi.org/10.1177/1687814018755519
13. Yussupova, N., Kovács, G., Boyko, M., Bogdanova, D.: Models and methods for quality management based on artificial intelligence applications. Acta Polytechnica Hungarica **13** (2016). https://www.uni-obuda.hu/journal/Yussupova_Kovacs_Boyko_Bogdanova_67.pdf
14. Nakajima, S.: Quality assurance of machine learning software. In: 2018 IEEE 7th Global Conference on Consumer Electronics (GCCE), Nara, Japan, pp. 601–604. IEEE (2018)
15. James, G., Witten, D., Hastie, T., Tibshirani, R.: An introduction to Statistical Learning with Applications in R, pp. 138–139. Springer, New York (2013)
16. Amiri, F., Noghondarian, K., Safaei, A.: Evaluating the performance of variable scheme X-bar control chart: a Taguchi loss approach. Int. J. Prod. Res. **52**(18), 5385–5395 (2014). https://doi.org/10.1080/00207543.2014.906762