

# Quality Control Method of Engine Manufacturing Process Using Data Mining Technique

Jung-Hee Kim<sup>1</sup>, Jinhwa Kim<sup>2\*</sup> and Tae-Ho Lee<sup>1</sup>

<sup>1</sup>Graduate School of Business, Sogang University, 35 Baekbeom-ro, Mapo-gu, Seoul 04107, Korea; vada8288@gmail.com, taeholee@sogang.ac.kr

<sup>2</sup>School of Business, Sogang University, 35 Baekbeom-ro, Mapo-gu, Seoul 04107, Korea; jinhwakim@sogang.ac.kr

## Abstract

**Objectives:** This study analyzes the major defect factors that influence the defective percentage of an engine manufacturing processes and to find the processes that influence other defect factors. **Methods/Statistical Analysis:** This study uses the process data from actual manufacturing sites, and analyzes the major defect factors occurring in the manufacturing processes using the data mining techniques. Analysis on the process data has two steps: The first step is to preprocess to reorganize the raw data into accurate data, and the second step is to analyze for each defect factor to classify defects by factors. **Findings:** As a result of the analysis, it found defects of LEAK during operation of the engine, defects in components of the inlet and outlet systems of the engine and defects in components of the electronic system of the engine influence the percent defective. This study also analyzes the main defect types of product using big data from the engine manufacturing process. To solve the problem that was hard to find analytic data or useful information out of the huge data collected in real time, the data mining technique was employed. It is a practical methodology to offer beneficial information for decision making in the database marketing field. In terms of product quality management, estimating the defect types in the manufacturing process stage will shorten the time consumption in resolving the problems. By reducing the defective products in advance, it also lowers the cost accompanied with defect occurrences and raw materials. **Improvements/Applications:** This study suggests methods to control quality in manufacturing processes with an application of preventing defects in advance and further to control the future defects.

**Keywords:** Data Mining Techniques, Decision Tree, Defect Factors, Manufacturing Processes, Quality Control

## 1. Introduction

With rapid changes in the business environment, quality control is working as an important business strategy in enterprises. Most manufacturing companies need to build an automatic system for improving competitiveness and productivity and also need to collect and provide more precise information in the production sites. In addition,

they use statistical processes to reduce the percentage of defects in the manufacturing process with aim to produce products in a uniform quality. The quality control technique focuses on preventive activities through analysis of causes for defects occurring in the present time before they occur<sup>1,2</sup>.

As the automation systems are complex and the sensor data collected in a real time has numerous process

\*Author for correspondence

variables, there are difficulties in finding and analyzing useful information related to process improvement. In order to settle these issues, the data mining technique is used as a methodology to extract information needed for a strategic decision making<sup>3-5</sup>.

### 1.1 Engine Production Process

The production process of an engine can be controlled and tracked by collecting assembly information of parts and information related to production quality through the latest MES (Manufacturing Execution System). The production process comprises: a processing line to produce key parts for an engine, which are the CYL Block and the CYL Head; and an assembly line to assemble the parts composing the engine and to complete the engine finally. When a variety of produced parts are put in the factory warehouses, the products for each work process are released to the assembly line in accordance with the production work order to go through an assembly process for each part and to make the engine finally<sup>2,6</sup>.

The engines assembled always go through a performance measurement process to check if it works normally. The performance measurement process performs a test working of the engine to check if the completely assembled engine meets the standards for the level of required conditions and the performance of the products. Therefore, the fast defect data sets on the defective engine found in the performance measurement process include basic information and details of defects of the pertinent engine, and are used as basic data for quality control and improvement of the process by uploading and making a database through the SAP (Systems, Applications, and Products in Data Processing) QM (Quality Management) module<sup>1,7,8</sup>.

### 1.2 Quality Control System and Data Mining Technique

In quality control, conversion into the statistical process control, which is a preventive quality control activity, has a very great impact on reduction of quality costs. In general, the statistical process control is a method to collect data from the process, to analyze statistically, and to identify the quality norms and the status of process ability to produce a desired quality of products. The statistical process control system controls process quality by collecting and analyzing quality data in real time and determines the quality standards through analysis of the accumulated

quality data to present a way to realize a device and a method for statistical process control, which constantly increases the level of quality control. This technique not only can reduce time spent for the process of removing cause for defects if defects can be discovered in advance in the step of process, but it also can lower costs for defects and raw materials by detecting defects in advance<sup>9-11</sup>.

As well, recently, process data are collected in real time by using sensors and measurement technologies due to development and spread of automation systems. Since the data collected in real time are huge in quantity with numerous process variables, there are difficulties in finding useful information and analyzing it. The data mining technique is used as a methodology to extract knowledge from a great quantity of data due to quantitative expansion of data and an urgent necessity to turn big data into useful information and knowledge<sup>3,12</sup>.

Data mining has been mainly used for marketing and sales so far, and has been only rarely used for manufacturing. Studies that have used process data for quality improvement include: a case which designed process conditions for specifying relations between process variables and quality characteristics in the process of casting valves for automobiles and reducing the percent defective<sup>13</sup>; and a study which estimated the parameter of abnormal patterns of a control chart using the neural network model<sup>14</sup>.

### 1.3 Decision Tree

The decision tree is an effective analysis method to induce rules for decision making into the structure of a tree and perform classification and estimation. Since the processes of classification and estimation are expressed as rules for inference with the structure of a tree, it can be more easily understood and used compared to other quantitative analysis methods<sup>6,11</sup>. However, it has shortcomings in that it is sensitive to the size of samples for building a model<sup>12,13</sup>.

Studies on the decision tree technique include: a study which estimated customer defections by using the decision tree with data of customers who have effected an online automobile insurance<sup>15</sup>; a study which built a termination model by using the decision tree technique with personal data and transaction data of customers of credit card companies, mobile communication companies, and insurance companies<sup>16</sup>; and a study which have presented an analysis procedure and a model for estimation by defect item by using large-capacity process data in the manufacturing process<sup>5,7</sup>.

## 2. Research Design and Experiments

### 2.1 Data Collection and Variables Description

The data used for this study is quality data including the defect history of the performance measurement process of the engine production process of an industrial machine manufacturing company in South Korea.

The data on product defects used for this study is details of defects that have occurred in the performance measurement process of the engine manufacturing process. The pertinent data are collected from the SAPQM module. The pertinent data used for the analysis are process data including one hundred pieces of defect history made up of 16 variables. The description and properties of each item is in Table 1.

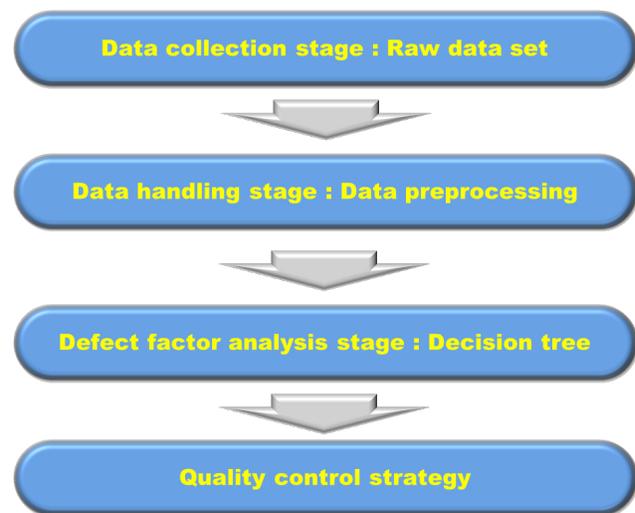
**Table 1.** Variables description in the analysis

Items	Description	Properties
Serial No.	Engine number	ID
EG suffix	Engine model	
PR/O Mat. Des	Detailed engine model	
Issued Date	Date of occurrence of defect	
Issued Week	Week of occurrence of defect	
Defect Contents	Details of defect (text)	
Part(L1)	System 1 comprising the defective part (L1)	
Part(L2)	System 1 comprising the defective part (L2)	
Damage	Defect Phenomenon Code	
Damage1	Defect Phenomenon Code 1	TARGET
Defect Mat. Des(1)	Part with the defect	
Responsibility	Responsibility for the defect	
Cause	Cause for the defect	
Cause Contents	Cause for the defect (text)	
Team in Charge	Team responsible for the defect	
Action Contents	Details of actions	

### 2.2 Method

Analysis on the process data comprises two steps: the first step is to preprocess to reorganize the raw data into accurate data, which is easily used for analysis, by processing missing values and removing outliers; and the second step is to analyze for each defect factor to classify defects by factors and analyses the process pattern for the classified defects<sup>3</sup>.

The step in preprocessing of data is to clean and convert data in order to reorganize the data into data easily analyzed before applying the data mining technique. Since data used for data mining have great uncertainty, it needs preprocessing. The step of analysis for each defect factor is to classify defects by factor using the decision tree technique for the processed data. The decision tree technique is very effective to conduct classification and to clearly express the defect pattern. The research procedures in this study are shown in Figure 1.



**Figure 1.** Procedure for research method.

## 3. Experimental Results

With the process data including the defect history, it analyzed the defect pattern by using the decision tree technique. In the engine manufacturing process, defects are largely divided into three types: 1. The defect of LEAK, 2. The defect in the inlet and outlet systems, and 3. The defect of the electronic system. The following is the result of the decision tree analysis for finding the defect patterns in Figure 2.

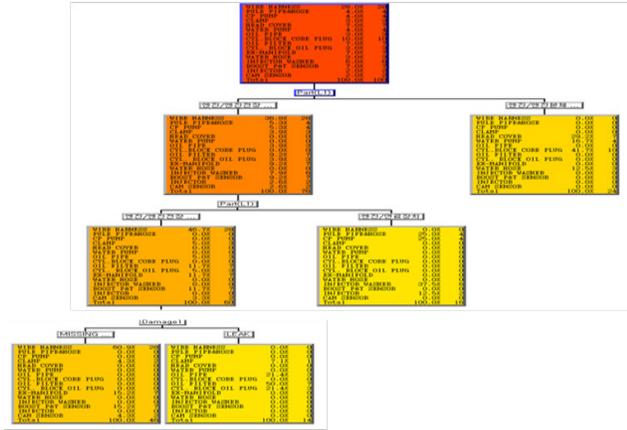


Figure 2. Results of decision tree analysis.

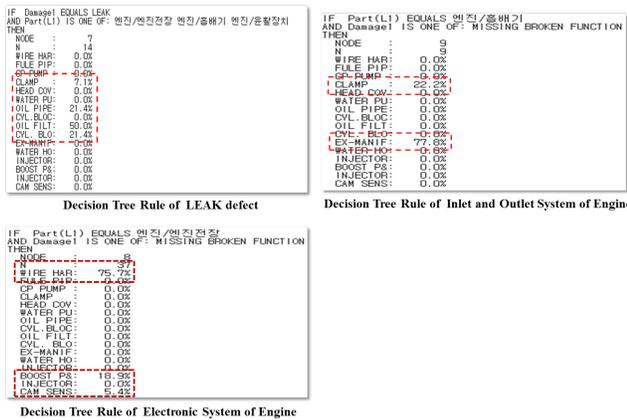


Figure 3. Decision tree rules of the major defect factors.

Firstly, the manufacturing process having the greatest impact on defect factors is the item of the defect of LEAK including oil leak (water leak/gas leak) occurring during operation of the engine as shown in Figure 3. Figure 3 also shows that the significant parts that serve as the major factors for the defect of LEAK are found to include the Oil Filter (50%), the Cyl Block (21%), and the Oil Pipe (21%). Pre-control is required to establish fool-proof plans to prevent the defect of LEAK especially among the manufacturing quality guarantee methods for the pertinent parts.

Secondly, among the parts composing the inlet and outlet systems of the engine, the Ex-Manifold (77%) and the Clamp (22%) is found to be the parts where major defect phenomena of the types of missing,

broken, and malfunction. Since the two parts account for over 90% of the pertinent defects, it is needed to take actions to remove the causes for the defects through intensive audit activities for quality improvement by performing intensive control of the pertinent parts and checking if there is any factor causing the defect in the process.

Finally, among the parts composing the electronic system of the engine, the Wire Harness (75%) and the P&T Sensor (19%) are found to be the parts where the major defect phenomena of the type of missing, broken and malfunction have occurred. Quality control of the Wire Harness, which is the biggest component, is found to require more intensive control, and a higher level of quality control is demanded.

## 4. Conclusions

This study aims to analyze defect data in the manufacturing process of engines, to control the percentage of defective products, and to secure competitive edge of companies in accordance with improvement of quality and productivity in manufacturing. To achieve this purpose, this paper presents an analytic method to analyze the defect types by using the defect data collected from the engine manufacturing processes and to find the major factors having the greatest impact each other. The processes classified as the major defect factors require an intensive care in the field. As well, in order to reduce the difficulties in collecting analytic data from a large-capacity database, the time spent for removing the cause for defects in the process can be reduced by estimating the defect factors in the manufacturing process using the data mining technique, which is a practical methodology to provide useful information. Finally, the suggested method in this study is expected to be used as one of the methods to control quality in a similar manufacturing process in the future for further expansion with diverse applications of preventing defects in advance.

## 5. References

- Hancock W, Yoon J, Plot R. Use of ridge regression in the improved control of casting process. Quality Engineering. 1998; 8(3):395-403.

2. Montgomery D, Peck E, Vining G. Introduction to linear regression analysis, 4th ed. USA: Wiley; 2007.
3. Banks D. Pre-analysis of super large industrial data sets. *Journal of Quality Technology*. 1992; 24:115–92.
4. Sivakumar S, Venkataraman S, Selvaraj R. Predictive modeling of student dropout indicators in educational data mining using improved decision tree. *Indian Journal of Science and Technology*. 2016; 9(4):2–5.
5. Undavia J, Dolia P, Patel A. Customized prediction model to predict post-graduation course for graduating students using decision tree classifier. *Indian Journal of Science and Technology*. 2016; 9(12):2–7.
6. Lee K, Lee H. A study on the combined decision tree (C4.5) and neural network algorithm for classification of mobile telecommunication customer. *Korea Journal of Intelligent Information Systems*. 2003; 9(1):139–55.
7. Byeon S, Kang C, Sim S. Defect type prediction method in manufacturing process using data mining technique. *Journal of the Society of Korea Industrial and Systems Engineering*. 2004; 27(2):10–6.
8. Goodlin B, Boning D, Sawin H, Wise B. Simultaneous fault detection and classification for semiconductor manufacturing tools. *Journal of the Electrochemical Society*. 2002; 150(12):778–84.
9. Berry M, Linoff G. Data mining techniques: for marketing, sales, and customer support. USA: John Wiley and Sons; 1997.
10. Galit S, Nitin R, Peter C. Data mining for Business Intelligence. USA; Wiley; 2009.
11. Michael J. Data mining techniques. USA: John Wiley and Sons; 1997.
12. Ture M, Tokatli F, Omurlu I. The comparisons of prognostic indexes using data mining techniques and cox regression analysis in the breast cancer data. *Expert Systems with Applications*. 2009; 36(4):8347–254.
13. Montgomery D. Introduction to statistical quality control. USA: John Wiley and Sons; 1996.
14. Guh R, Hsieh YA. Neural network based model for abnormal pattern recognition of control charts. *Computer and Industrial Engineering*. 1993; 36:97–108.
15. Lim S, Her Y. Customer churning forecasting and strategic implication in online auto insurance using decision tree algorithms. *Korea Journal of Information Systems Review*. 2006; 8(8):125–34.
16. Chang N. Improving the effectiveness of customer classification models: A pre-segmentation approach. *Korea Journal of Information Systems Review*. 2005; 7(2):23–40.