

Data Perspective of Lean Six Sigma in Industry 4.0 Era: A Guide To Improve Quality

Onur Dogan and Omer Faruk Gurcan

Department of Industrial Engineering

Istanbul Technical University

Macka, 34367, Turkey

odogan@itu.edu.tr, ofgurcan@itu.edu.tr

Abstract

Almost all quality improvement methods require data collection and analysis to solve quality problems. The combination of six sigma and lean manufacturing creates lean six sigma methodology that aims to reach six sigma quality levels, less than 3.4 part per million defectives, by reducing variations and wastes within processes. Achieving the goal depends on data collection to overcome quality problems.

Although many traditional data analysis techniques can be used to develop quality of products and processes, massive data sets collected by industry 4.0 technologies should be mined with powerful data analysis methods that produce meaningful results from big data. It is possible to make effective decisions by utilizing these analysis methods in each step of lean six sigma cycles. The use of data analysis methods at every stage, especially in the measure and analyze stages, has critical importance to make powerful decisions.

The aim of this study is to provide a guide that allows applying lean six sigma to make faster, more reliable and satisfied decisions with data. It contributes to the manufacturing processes with lean six sigma by reducing the lead-time, producing better quality products; on the other hand, it aids to make effective decisions using different mining techniques.

Keywords

Lean six sigma, Industry 4.0, Big data, Data mining, Quality improvement, Process mining

1. Introduction

Since many competitors produce the similar products, quality is a critical problem. Improving quality depends on collection of big data and extraction useful information from it. Quality improvement methods such as inspection, statistical process control, total quality control, zero defects, kaizen and lean six sigma (LSS) need to collect data about quality problems to solve (Köksal, et al., 2011).

LSS emerged initially as different two methods. Similarities and complementary features of them make it useful to use these two philosophies together. It has thus become a powerful and effective tool for sustainable operational outcomes. These two methods creating LSS are six sigma and lean manufacturing. Six sigma has been widely adopted as a discipline that use a systematic problem-solving approach and methodology in a variety of industries (Lee, et al., 2004; General Electric, 2017). Sigma is a statistical term for measuring how a given process deviates from perfection (Angoss Software, 2011). With six sigma, the errors in the manufacturing are reduced to the error level of 3.4 parts per million (ppm) and it is aimed to go to zero defect. Given the increasing customer expectations considerations, it is of great importance for successful companies. Six sigma asserts that this goal can be achieved if the variance in production is within a certain range.

On the other hand, lean manufacturing aims to make production with value-creating operations by removing all kinds of activities that are worthless from the customer's point of view. The cost is reduced by determining and eliminating all kinds of waste and so, the lead-time is shortened. The difference between lean manufacturing and six sigma is to define the basic causes of waste in different ways. While lean manufacturing claims that wastes in the processes are due to non-value added activities to the final product, the six sigma argues that wastes come from the variation in the processes. However, the common goal of both is to create a production system that meets the customer expectations most effectively by producing quality products. Because of its advantages on high quality, LSS has gained worldwide popularity, providing both output and quality-based improvements. This approach, which simplifies systematically manufacturing processes and improves product quality with less than 3.4 ppm errors, has saved significant improvements and cost savings in many companies such as General Electric (2017), Dell Inc. (Lovin ve Yaptangco 2006) and Xerox Corp. (2004).

Approximately 95% of LSS projects follow to improve quality so-called define-measure-analyze-improve-control (DMAIC) approach (Trnka 2012; Kanakana, et al., 2010) to reach six sigma quality levels, less than 3.4 part per million defectives. The Six Sigma system demonstrated by DMAIC is a structured method that helps solve existing problems, see future opportunities and manage projects. DMAIC has proven to be one of the most effective problem-solving methods used up to now, because it forces the teams to use the data to do the following (George, et al., 2004). Some of six sigma projects use DMADV cycle, also known as design for six sigma, to create new processes. The steps DMA (Define, Measure and Analyze) are similar to DMAIC but not the same. DMADV was used for different purposes. For example, Huang et al. (2010) used to improve quality of product. Chen et al. (2005) used improve the assembly efficiency of military products. Table 1 summarizes steps of six sigma both DMAIC and DMADV cycles and gives some sample activities used in the steps.

Table 1. Key steps of Six Sigma for existing and new processes

Phase	Descriptions	Sample activities
Define	Define the purpose and scope of the six sigma project	Define why the project should be done
		Define the targets, goals and scopes of project
		Define the customer requirements
Measure	Measure to determine the current situation	Select the output characteristics
		Assess the performance specifications
		Establish the initial process capability
Analyze	Analyze and determine the actual causes for process improvement	Analyze the current process performance
		Monitor the potential Critical to Process (CTP)
		Analyze what resources will be needed for improvement
Improve	Improve the process by eliminating wasteful causes, removing the problem or reducing the effects of the problem	Improve idea
		Identify optimal operating conditions
		Eliminate wastes
Control	Control the improved process performance	Determine the process capability for CTPs
		Implement the process controls
		Document what you have learned
Design	Design a new process in parallel with customer needs	Develop design to meet customer needs
		Design analyze model
Verify	Confirm the truth of designed model	Create a plan for full implementation
		Validate to model

To obtain accurate results with LSS, which has a systematic effect in reducing waste and cost (Polk 2011), it can be seen that the cycles are based on the data. Since Industry 4.0 technologies provide collecting more and more data, the use of different mining techniques such as big data analytics, data mining and process mining is very important. By mining techniques, decision makers gain time by identifying what is difficult to see at first sight. In addition, the accuracy is high because the decisions made are based on the data. Although many companies have adopted lean manufacturing and six sigma, few are satisfied with results (Guarraia, et al., 2008). Modelling or optimizing is not easy because quality problems contains big data. Data or information is one of the factors for production of goods or services. This forces decision makers to store as much data as possible in order to avoid making wrong decision. In era of industry 4.0, gathered massive data increase the risk for wrong decisions, while they are essential to make decisions. Big data techniques such as text mining, machine learning, deep learning, artificial neural networks and basic data mining techniques such as clustering, association, prediction, classification and process mining algorithms help to reach correct and optimum decisions in various stages of LSS. Therefore, the use of mining techniques in LSS cycles is an indispensable for business decision makers.

2. An Overview to Different Mining Techniques

Since data is critical for LSS, mining data to extract meaningful results gains importance. Developing technology brings new ways to mining data because of collected large amounts of data. Data mining has effective methods for a smaller number of data. However, if data is collected automatically with sensors, big data analytics are more preferable thanks to large amount of data. In addition, process mining methods are nonignorable because LSS focuses on the quality in a process. It should not be forgotten that the point in all analytics is to use data to gain insights, not evaluate whole process. Therefore, analytics can help to confirm different LSS activities such as finding root causes and developing robust solutions. It is only a part of the LSS.

For all mining techniques data preparation and evaluation are common. Because data is mainly redundant, incomplete and inconsistent (Köksal, et al., 2011) data preprocessing step is necessary to overcome this kind of problems for developing data quality before mining. Some of data preprocessing tasks are data cleaning, data transformation, data reduction and discretization (Pyle 1999; Giudici 2003; Witten 2005). Evaluation of mining results depends on comparing different mining techniques. The technique that gives the best result is chosen to interpret. To make optimum and correct decisions, it is clear that knowledge extracted from data sets should be evaluated and interpreted correctly (Dunham 2003).

Except for mentioned in this section, statistics and quality tools used in LSS cycles contain also data-based methods. They are not described because they are more basic than mentioned mining techniques but considered during the study.

2.1. Data Mining Techniques

Data mining (DM) is an exploratory data-analytic process that detects interesting, novel patterns within one or more data sets that are usually large (Nadkarni, 2016). Data mining techniques use the integrated data through large amounts of data stored in databases using statistical and mathematical techniques (Larose, 2005). They can be classified into two groups as descriptive and predictive (Han, et al., 2011). Descriptive DM tasks includes clustering and association while predictive tasks consist of classification and prediction. The most common types of data mining tasks used in LSS cycles:

- **Association analysis** discovers some rules among data. It also named as association rule mining. It tries to find frequent items and generates interesting if-then rules (Hand, et al., 2001).
- **Clustering** is a method of grouping data sets according to similarities between objects in same groups and dissimilarities between objects in other groups (Rahman and Islam, 2014). Unlike classification, the class label of each group is unknown. It calculates distances between data and group center and assigns the data into the nearest group.
- **Classification** aims to classify a data into predefined class. Like clustering, classification technique also uses distance-based algorithms. The class label of each group is known (Talia, et al., 2016).
- **Prediction** associates a data to a quantitative variable and predicts the value of that variable (Talia, et al., 2016).

2.2. Big Data Analytics

Another important and hot topic related to mining techniques is about big data analytics. One of the definition for big data is that it is based on large volumes of extensively varied data that are generated, captured, and processed at high velocity (Günther, et al. 2017). Today, different sources such as sensors, smart devices, organizations and individuals generate data at a very high rate. Gantz and Reinsel form International Data Corporation (2010) found that the created and copied data volume in the world was 1.8 zettabytes (ZB), that 1 ZB equals to 1 trillion gigabyte, and it will be 35 ZB in 2020. It is estimated that this figure will double every other two years in the near future (Gantz & Reinsel, 2011). The more data the more near to perfect decisions on condition analyzing data very well. While big data brings big opportunities, data driven applications need also real-time analysis such as navigation, social networks, finance, biomedicine, astronomy, intelligent transport systems (Oussous, et al., 2017). Thus, big data analytics and efficient methods of data mining are necessary to obtain accurate results. Since big data analytics aims to gain deeper insights into processes, it is a complement of LSS. Nowadays, there are various big data analytical techniques to gain deep insights. Many of them tests the combination of various algorithms and

technologies (Oussous, et al., 2017). In addition, data mining techniques are a part of big data analytics. Some big data analytics that can be used in LSS cycles:

- **Machine learning** aims to objective of machine learning is to discover knowledge and make intelligent decisions (Oussous, et al., 2017). The field of machine learning emerged from within Artificial Intelligence (AI) with techniques such as neural networks (van der Aalst and Damiani, 2015). The difference between data mining and machine learning is not a clear-cut. In this study, machine learning refers to algorithms that learn without being explicitly programmed. Generally, machine learning methods are divided three groups as supervised learning, unsupervised learning, and reinforcement learning. Reader may refer to Qiu et al. (2016) for more details.
- **Text mining** converts large volumes of text data into meaningful summaries, which support evidence-based decision-making (Yazdizadeh and Ameri, 2015). Shortly, text mining refers to techniques that extract information from textual data such as social media sharings, emails and work orders (Gandomi and Haider, 2015). For example, details of a quality problem can be determined using the information extracted by text mining from the information technology (IT) system. The details may be timestamp, person, activity name, resource etc.
- **Video mining** analyze and extract information from video data captured by cameras. Even if the video camera is very rich about the surrounding environment, more effort must be made with advanced mining algorithms to make the videos meaningful (Bu and Chan, 2005). Although video analytics is still in its infancy compared to other types of data mining methods (Panigrahi, et al., 2010), various algorithms have already been developed for processing recorded videos. In manufacturing area, problems can observed with cameras and main causes of the problem can be detected using video mining algorithms.

2.3. Process Mining

Data mining and big data analytics are focus on the data-centered results. Although these results are very important, they are not sufficient to evaluate a process. Process mining (PM), which yields process-centered results, helps to discover, and track real transactions by extracting information from event logs (van der Aalst, 2016). For LSS, which try to eliminate variations and wastes in a process, PM is a useful mining technique. The data collected from information systems are analyzed with PM to see whether deviations and bottlenecks are there from standard procedures with the aim of process enhancement (Rovani, et al., 2015). Therefore, quality problems in a process can be determined and solutions can be generated.

Process discovery, conformance checking and enhancement are three types of PM. Process discovery generates a process model showing flow of whole process without using any prior knowledge. The conformance checking tests by comparing event logs of the process with an existed process model whether process has a variability. In other words, it checks the conformance of the events in the logs with current the model. The goal of the enhancement is to improve a real process by using event logs by comparing the discovered model and event logs (van der Aalst, 2011).

Although both process mining and data mining start from data, data mining techniques are typically not process-centric and do not focus on event data. Therefore, questions related to performance (e.g., bottlenecks in processes) and compliance (e.g., quantifying and diagnosing deviations from some normative model) are not addressed at all. Moreover, none of the traditional data mining tools support process mining techniques in a satisfactory manner (Weijters et al., 2006). Data mining in the narrow sense and process mining are complementary approaches that can strengthen each other. Once discovered and aligned with the event log, process models provide the basis for valuable data mining questions.

Like DM, big data does not also focus on the improvement of end-to-end processes, on contrary to PM. Data is the fuel for all analysis techniques and big data brings advantages about many tasks, one of them is quality. MapReduce and cloud technology are two example of big data analytics for PM (van der Aalst and Damiani, 2015). Big data technology often relies on MapReduce (Dean and Ghemawat, 2008), a basic computational paradigm, which has been remarkably successful in handling the heavy computational demands posed by huge data sets. Conventional process mining tools are often deployed on the process owners' premises. However, cloud-based deployment seems a natural choice for PM.

3. The Suggested Guide to Improve Quality

Figure 1 shows the flow diagram showing LSS stages for existing (DMAIC) and new (DMADV) process. Define phase is almost similar in each cycle. Measure and Analyze phases have some differences but they can be considered similar. The differences are not remarkable. In existing process to make better decisions for improving quality, Improve phase should be used and the process should be controlled. On the other hand, if process does not exist, to create a new process design phases should be used and designed process is validated.

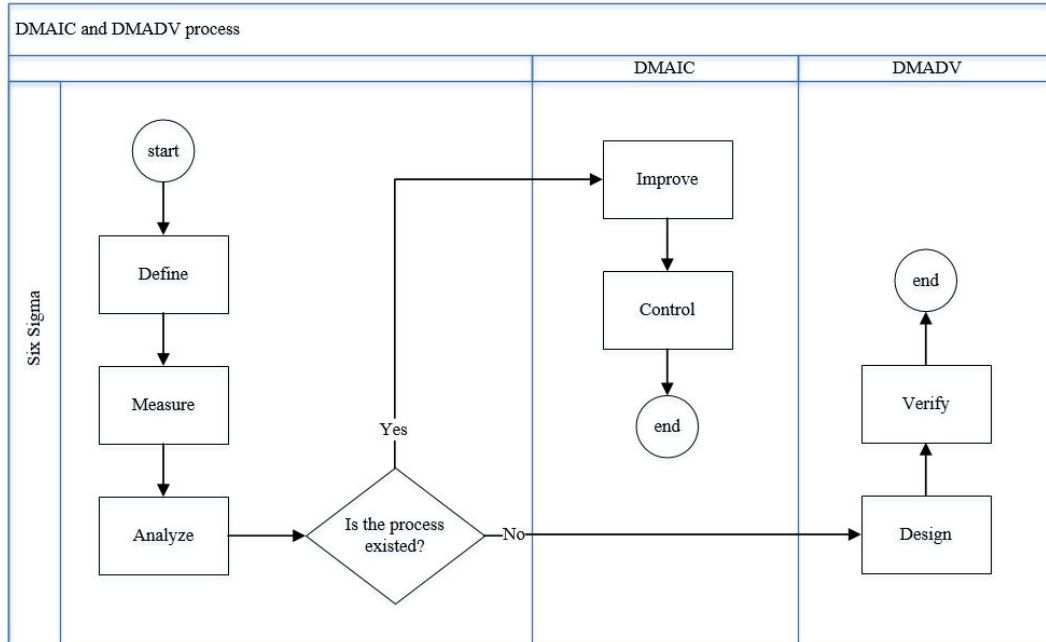


Figure 1. Six sigma stages for existing (DMAIC) and new (DMADV) process

DMADV cycles. Define phase requires to identify problems, process goals with respect to customer needs. Any methods Table 2 demonstrates some methods can be used in both DMAIC and in Define column can be used for this stage. For example, in DMADV cycle, brainstorming and nominal group technique are useful methods to define potential problems. After definition of potential problems, using prioritization matrix or pareto analysis, it is decided that which problem(s) will be solve firstly. To determine process goal according to customer requirements, quality function deployment (QFD) may be useful (Dogan and Cebeci, 2016). For unstructured data such as text and video, big data analytics can be useful to define the problem. With process discovery, status of the process can be shown visually. Figure 2 indicates the suggested roadmap for the Define phase.

Table 2. Methods used in LSS cycles

	Define	Measure	Analyze	Improve	Control	Design	Verify
Statistics	Descriptive	Descriptive, Tally chart, Z-test, Confidence intervals, Predictive	Correlation, T-test, Chi-square test, F-test, Hypothesis tests, ANOVA, Histogram, Predictive	Hypothesis tests, Multivariate Analysis		Descriptive, Predictive	Correlation, Causality
Quality Tools	Brain storming, NGT, Pareto analysis, Matrix diagram, QFD, FMEA, SIPOC, Prioritization matrix, Fishbone analysis,	Pareto analysis, Process sigma	SPC	TRIZ, DOE	FMEA, Control diagram, Standardization, SPC	QFD, DOE	
Data Mining³			Association Rules, Clustering, Classification	Prediction		Market Basket Analysis, Association Rules	
Big Data	Text Mining, Video Mining		Machine Learning, Decision Trees, Text Mining, Video Mining, Artificial Intelligence	Machine Learning, Artificial Intelligence	Machine Learning, Artificial Intelligence		Machine Learning, Artificial Intelligence
Process Mining	Process Discovery	Conformance checking	Process Discovery, Conformance checking	Flow diagrams, Enhancement	Flow diagrams, Conformance checking		Graphing, Visualization

¹ Descriptive includes some statistical methods such as mean, frequencies and standard deviation.

² Predictive includes some statistical methods such as hypothesis testing, analysis of variance and regression.

³ Note that many DM methods are used for big data with advanced algorithms.

ANOVA: Analysis of Variance

DOE: Design of Experiments

FMEA: Failure Mode and Effects Analysis

NGT: Nominal Group Technique

QFD: Quality Function Deployment

SIPOC: Supplier, Inputs, Process, Outputs, Customers

SPC: Statistical Process Control

TRIZ: Theory of Inventive Problem Solving

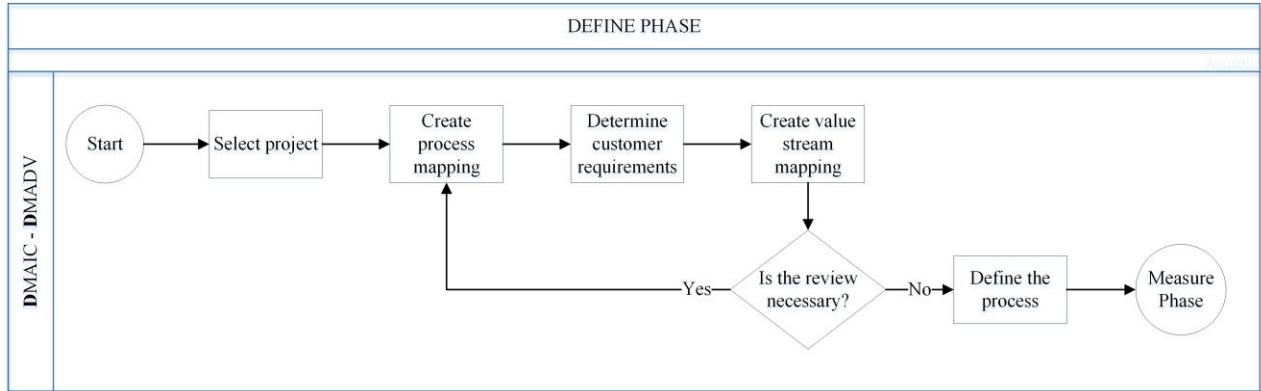


Figure 2. Define Phase

One of the goals of LSS is to eliminate variations in the process. In Measure phase, ANOVA can identify the variation causing quality problems. Tally chart, descriptive and predictive statistics can be used to measure current situation of process. Figure 3 indicates the suggested roadmap for the Measure phase.

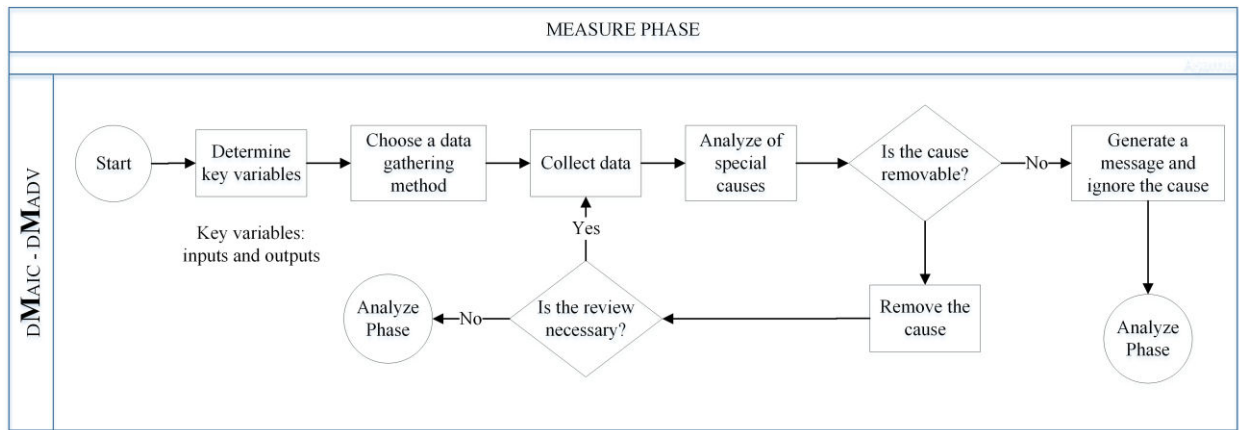


Figure 3. Measure Phase

Data gathered in the Measure phase is used to analyze process. Correlation and regression analysis explain relationship between dependent and independent variables. Process competence and process performance are analyzed using SPC. Association rules, clustering and classification are the most preferred DM methods to analyze process. Machine learning algorithms can help to navigate the user without any learning effort. In addition, if the data include also timestamp, for process analysis, process discovery, a type of PM, can be more efficient technique. Figure 4 shows suggested roadmap for the Analyze phase.

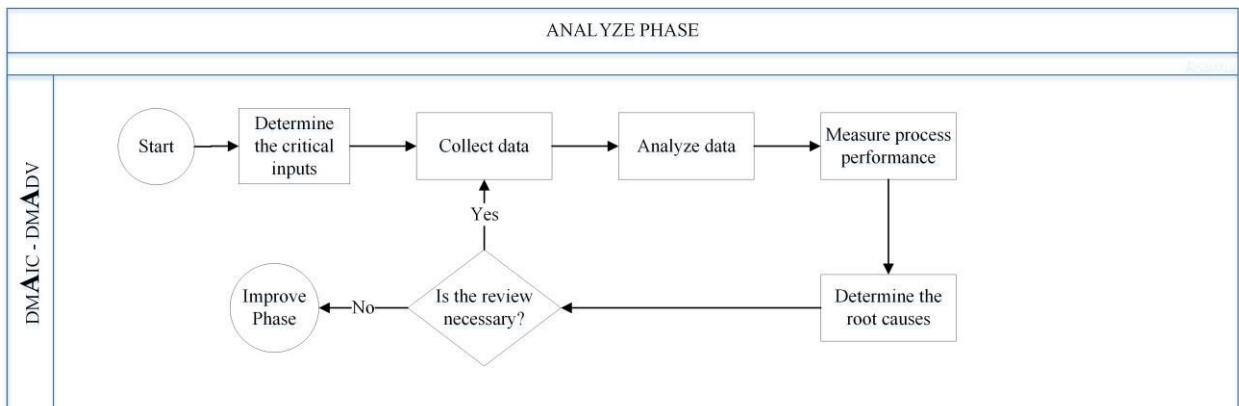


Figure 4. Analyze Phase

Multivariate statistical methods such as clustering analysis and discriminant analysis can be used to improve the process. TRIZ for solving innovatively problems and DOE to find best parameters can be used. Figure 5 demonstrates the suggested roadmap for the Improve phase.

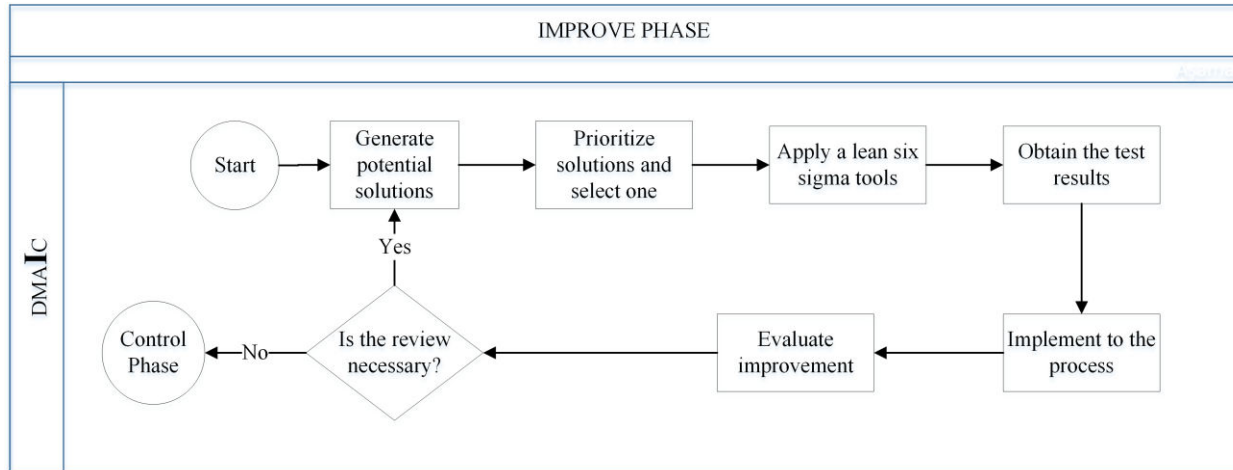


Figure 5. Improve Phase

In the control phase, FMEA to avoid potential risks and control diagrams to check whether the process is in control or not can be used. An unsupervised learning algorithm can be created using machine learning techniques to give an alarm when the process is out of control. At the same time, conformance checking can be used to check created model from event logs and actual process. Figure 6 shows the suggested roadmap for the Control phase.

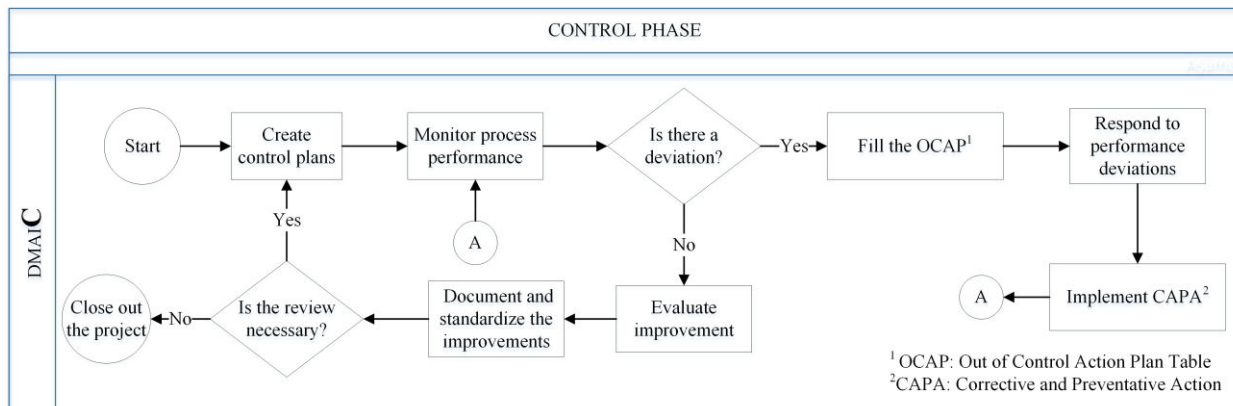


Figure 6. Control Phase

To create a new process, in the Design phase can utilize QFD to consider customer requirements. Descriptive and predictive statistics methods can be useful for basic information about the design. Market basket analysis and association rule mining as a DM method identifies products and content that go well together. Figure 7 indicates the suggested roadmap for the Design phase.

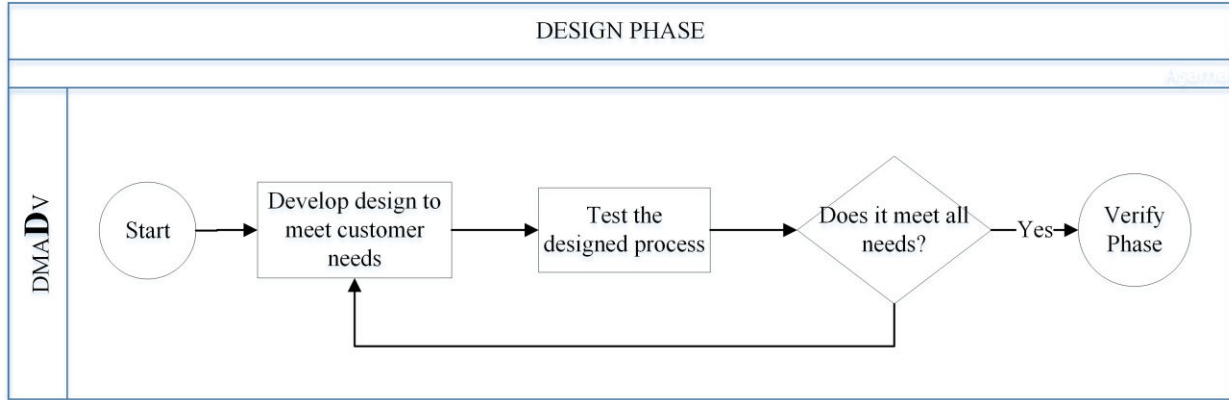


Figure 7. Design Phase

In the last step of DMADV cycle, some metrics are created to keep and a pilot run is developed to verify the new process. Results can be visualized using Graphing or Visualization as PM methods to see whether there is any potential failures affecting quality. Like in control step, a proper machine learning algorithm can be used to give an alarm when the process is out of control. Figure 8 shows the suggested roadmap for the Verify phase.

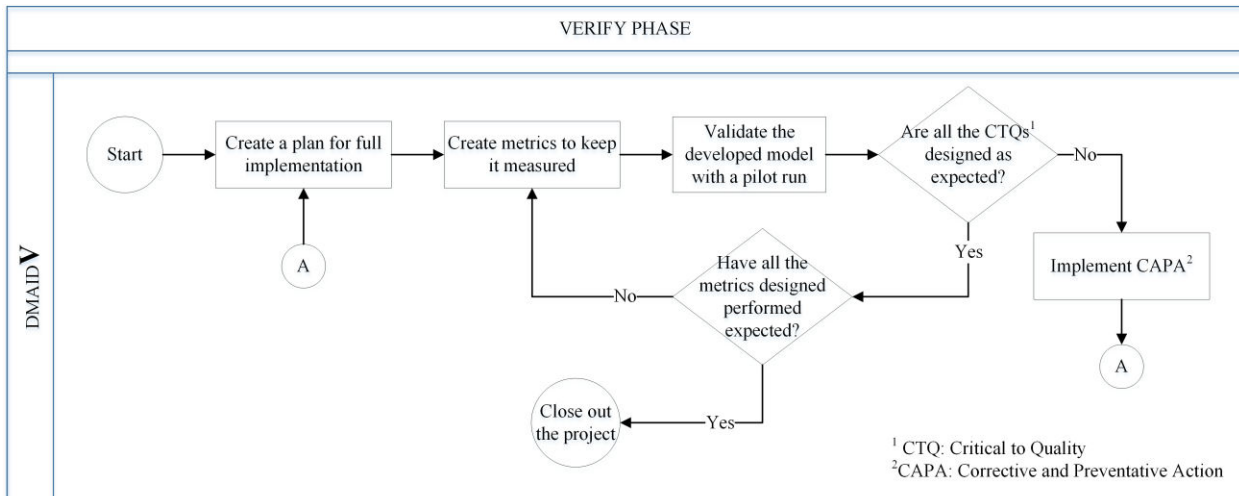


Figure 8. Verify Phase

4. Conclusion

There are various methods such as inspection, statistical process control, total quality control, zero defects, kaizen and lean six sigma to improve quality. In this study, lean six sigma as a method quality improvement is chosen. It aims to reduce variations and wastes within processes using different techniques. Lean six sigma collects data to achieve its goal. Collected data should be analyzed to make optimum and correct decisions. However, industry 4.0 technologies make possible to collect enormous amount of data. Therefore, traditional data analysis techniques are not sufficient because they require more time and cost. It is possible to benefit from advanced techniques, which are suitable for big data, such as big data analytics and process mining in addition to traditional techniques to make effective decisions for quality problems. Each step of lean six sigma cycles, DMAIC (define-measure-analyze-improve-control) and DMADV (define-measure-analyze-design-verify), contains various techniques. They were shown in Table 2 and flowcharts for each phases are drawn separately.

The aim of this study is to provide a guide that makes easier, faster, more reliable and satisfied decisions with data for improving quality in processes. As a further research, the suggested guide will be used in a manufacturing project that intends to solve quality problems using lean six sigma methodology.

References

- Angoss Software, *Key Performance Indicators, Six Sigma and Data Mining*, Canada, 2011.
- Bu, F., and Chan, C.-Y., Pedestrian Detection in Transit Bus Application: Sensing Technologies and Safety Solutions, *IEEE Intelligent Vehicles Symposium*, Las Vegas, NV, 6 – 8 June, 2005.
- Huang, C.-T., Chen, K.S., and Chang, T.-C., An application of DMADV methodology for increasing the yield rate of surveillance cameras, *Microelectronics Reliability*, vol. 50, no. 2, pp. 266 – 272, 2010.
- Cheng, Y.H., *The improvement of assembly efficiency of military product by Six-Sigma*, NCUT Thesis Archive, Taiwan, 2005.
- Dean, J., and Ghemawat, S., MapReduce: Simplified Data Processing on Large Clusters, *Communications of the ACM*, vol. 51, no. 1, pp. 107 – 113, 2008.
- Dogan, O., and Cebeci, U., A Methodology For New Product Development by Using QFD, FMEA and Its Application in Metal Plating Industry, *16th International Symposium on Manufacturing Researches*, Istanbul, Turkey, 12 – 14 October, 2016.
- Dunham, M. H., *Data mining introductory and advanced topics*, Prentice Hall/Pearson Education, New Jersey, 2003.
- Gandomi, A., and Murtaza H., Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management* vol. 35, pp. 137–144, 2015.
- Gantz, J., and David R., *The Digital Universe Decade – Are You Ready?*, EMC Corporation, 1 – 12, 2010.
- General Electric, The Roadmap to Customer Impact, Available: <http://www.ge.com/sixsigma/SixSigma.pdf> (Accessed 04.02.2018).
- General Electric, What is Six Sigma: the roadmap to customer impact?, Available: <http://www.ge.com/sixsigma/SixSigma.pdf> (Accessed 04.01.2018)
- George, M.L., Rowlands, D., and Kastle, B., *What is Lean Six Sigma*, McGraw-Hill, New York, 2004.
- Giudici, P., *Applied data mining: Statistical methods for business and industry*, John Wiley, New York, 2003.
- Guarraia, P., Carey, G., Corbett, A., and Neuhaus, K., *Lean Six Sigma for manufacturing*, Bain&Company, Los Angeles, 2008.
- Günther, W.A., Mohammad H.R.M., Marleen H., and Frans F., Debating big data: A literature review on realizing value from big data, *The Journal of Strategic Information Systems* vol. 26, no. 3, pp. 191-209, 2017.
- Han, J., Kamber, M., and Pei, J., *Data mining: Concepts and techniques*, 3rd Edition, Morgan Kaufmann, San Francisco 2011.
- Hand, D.J., Mannila, H., and Smyth, P., *Principles of data mining*, MIT Press, Cambridge, 2001.
- Kanakana, M.G., Pretorius, J.H.C., and Van Wyk, B., Lean six sigma framework to improve throughput rate, *IEEE 17th International Conference on Industrial Engineering and Engineering Management (IE&EM)*, Xiamen, China, 29 – 31 October, 2010.
- Köksal, G., Inci B., and Murat C.T., A review of data mining applications for quality improvement in manufacturing industry, *Expert Systems with Applications* vol. 10, no. 38, pp. 13448–13467, 2011.
- Larose, D.T., *Discovering Knowledge in Data: An Introduction to Data Mining*, John Wiley & Sons, Hoboken, New Jersey, 2005.
- Lee, Y.-H., Kwang G. M., Chonghun H., Kun S. C., and Tae H. C., Process improvement methodology based on multivariate statistical analysis methods, *Control Engineering Practice* vol. 8, no. 12, pp. 945-961, 2004.
- Lovin, C., and Yaptangco, T., Best Practices: Measuring the Success of Enterprise Testing, Available: <http://www.dell.com/downloads/global/power/ps3q06-20060252-Yaptangco.pdf> (Accessed 03.01.2018).
- Nadkarni, P., Core Technologies: Data Mining and “Big Data”, in *Clinical Research Computing: A Practitioner's Handbook*, Academic Press, Cambridge, 2016.
- Oussous, A., Benjelloun, F.-Z., Lahcen, A.A., and Belfkih, S., Big Data technologies: A survey, *Journal of King Saud University – Computer and Information Sciences*, In Press, Corrected Proof, 2017.
- Panigrahi, B.K., Abraham, A., and Das, S., *Computational intelligence in power engineering*, Springer, Verlag, Berlin, Heidelberg, 2010.
- Polk, J. D., Lean Six Sigma, Innovation, and The Change Acceleration Process Can Work Together, *American College of Physician Executives* vol. 37, no. 1, pp. 38-42, 2011.
- Pyle, D., *Data preparation for data mining*, Morgan Kaufmann, San Francisco, 1999.
- Rahman, M. A., and Zahidul M. I., A hybrid clustering technique combining a novel genetic algorithm with K-Means, *Knowledge-Based Systems* vol. 71, no. 1, pp. 345-365, 2014.
- Qiu, J., Wu, Q., Ding, G., Xu, Y., and Feng, S., A survey of machine learning for big data processing, *EURASIP Journal on Advances in Signal Processing*, 1–16, 2016.

- Rovani, M., Maggi, F. M., Leoni, M., and van der Aalst, W.M.P., Declarative process mining in healthcare, *Expert Systems with Applications* vol. 42, no. 23, pp. 9236-9251, 2015.
- Talia, D., Trunfio, P., and Marozzo, F., Introduction to Data Mining, in *Data Analysis in the Cloud Models, Techniques and Applications*, Elsevier, Amsterdam, 2016.
- Trnka, A., Results of application data mining algorithms to (lean) six sigma methodology, *Annals of Faculty Engineering Hunedoara - International Journal of Engineering*, 2012.
- van der Aalst, W.M.P., *Process Mining: Data Science in Action*, 2nd Edition. Springer, Dordrecht, 2016.
- van der Aalst, W.M.P., *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer, Berlin, 2011.
- van der Aalst, W.M.P., and Damiani, E., Processes Meet Big Data: Connecting Data Science with Process Science, *IEEE Transactions on Services Computing* vol. 8, no. 6, pp. 810 – 819, 2015.
- Weijters, A.J.M.M., Van der Aalst, W.M.P, and Medeiros, A.K., *Process Mining with the Heuristics Miner Algorithm*, BETA working paper series 166, Eindhoven University of Technology, 2006.
- Witten, I.H., *Data mining: Practical machine learning tools and techniques*, 2nd edition, Morgan Kaufman, Boston, 2005.
- Xerox Corp., Lean Six Sigma Leads Xerox, Available: https://www.xerox.com/downloads/usa/en/n/nr_SixSigmaForumMag_2004_Aug.pdf (Accessed 03.01.2018).
- Yazdizadeh, P., and Farhad A., A Text Mining Technique for Manufacturing Supplier Classification, 35th Computers and Information in Engineering Conference, Boston, Massachusetts, USA, 2 – 5 August, 2015.

Biographies

Onur Dogan graduated from the Sakarya University with a Bachelor's Degree in Industrial Engineering in 2010 and received a Master's Degree in Management Engineering from the Istanbul Technical University in 2013. He is a Ph.D. candidate in the same university. He studied on intelligent decision support system, lean manufacturing and quality approaches such as QFD, FMEA or DOE during the master thesis. His PhD research interests include process mining, data mining and customer relationship management.

Omer Faruk Gurcan graduated with a BS from Selçuk University in 2006, a MS from the İstanbul Technical University and now continues PhD in Industrial Engineering Program in İstanbul Technical University. His research interests are focused on knowledge management, statistical analysis, data mining and machine learning.