ANNALS OF FACULTY ENGINEERING HUNEDOARA
– INTERNATIONAL JOURNAL OF ENGINEERING
Tome X (Year 2012) – FASCICULE 1 (ISSN 1584 – 2665)

[1.] Andrej TRNKA

RESULTS OF APPLICATION DATA MINING ALGORITHMS TO (LEAN) SIX SIGMA METHODOLOGY

[1.] FACULTY OF MASS MEDIA COMMUNICATION, UNIVERSITY OF SS. CYRIL AND METHODIUS IN TRNAVA, SLOVAKIA

ABSTRACT: *This paper describes results of our research. This research is focused to (Lean) Six Sigma methodology and Data Mining algorithms. We tried to implement several Data Mining algorithms and methods to (Lean) Six Sigma methodology, especially to DMAIC phases for existing process.*
KEYWORDS: *Data Mining, Knowledge Discovery in Databases, (Lean) Six Sigma*

INTRODUCTION TO (LEAN) SIX SIGMA METHODOLOGY

Six Sigma is a rigorous, focused, and highly effective implementation of proven quality principles and techniques. Incorporating elements from the work of many quality pioneers, Six Sigma aims for virtually error – free business performance. Sigma is a letter in the Greek alphabet used by statisticians to measure the variability in any process. A company's performance is measured by the sigma level of their business processes. Traditionally companies accepted three or four sigma performance levels as the norm, despite the fact that these processes created between 6200 and 67000 problems per million opportunities! The Six Sigma standard of 3.4 problems-per-million opportunities is a response to the increasing expectations of customers and the increased complexity of modern products and processes. [1]

About 95 percent of all Six Sigma projects follow the DMAIC phases. We can use DMAIC cycle for the existing process or DMADV cycle for a new process. The letters in the acronym mean activities in Six Sigma methodology.

- *DMAIC – DEFINE, MEASURE, ANALYZE, IMPROVE, CONTROL*
- *DMADV – DEFINE, MEASURE, ANALYZE, DESIGN, VERIFY*

In our research we improve the existing process, so we focused to the DMAIC cycle of (Lean) Six Sigma methodology.

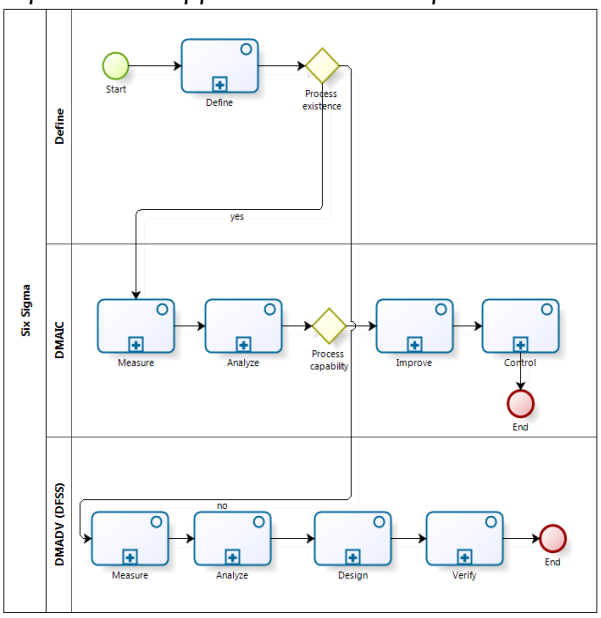

Figure 1: DMAIC vs. DMADV phases

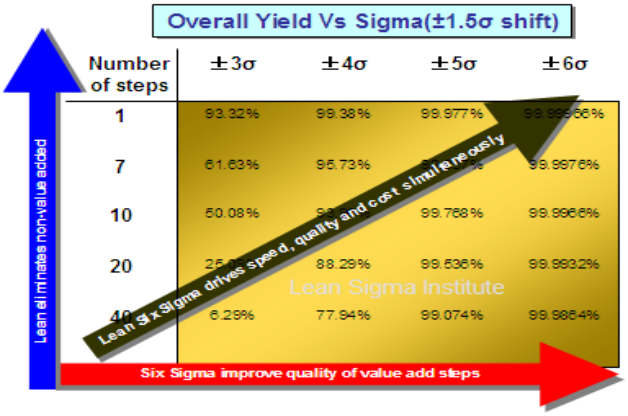

Figure 2: Lean Six Sigma – Steps in the process [3]

The use of Lean Six Sigma has proved to be a powerful and effective way for providing sustained positive operational results in organizations worldwide. Lean Six Sigma is in fact a hybrid philosophy for continuously improving organizations. Lean aims at eliminating waste by creating a culture of improvement where people learn powerful tools for solving problems and continuous improvement based on visual management and standardization that sustains enhancement. On the other hand, Six Sigma is a methodology that aims at reducing variations in production processes in

order to improve quality and meet customers' expectations. The primary premise for Lean is the focus on the creation of value for the customers. The value creation that enhances the organization's overall productivity is done by eliminating non-value added activities using specific sets of tools that optimize the utilization of the people and the processes. [2]. Figure 2 illustrates steps in the process bring down the overall yield at various sigma level.

KNOWLEDGE DISCOVERY IN DATABASES AND DATA MINING

Knowledge Discovery in Databases (KDD) has been defined as the non-trivial extraction of implicit, previously unknown and potentially useful information from data. The KDD process (Figure 3) is iterative and interactive, consisting of nine steps. The process is iterative at each step, meaning that moving back to previous steps may be required. The process starts with determining the KDD goals and ends with the implementation of the discovered knowledge. [8], [9]

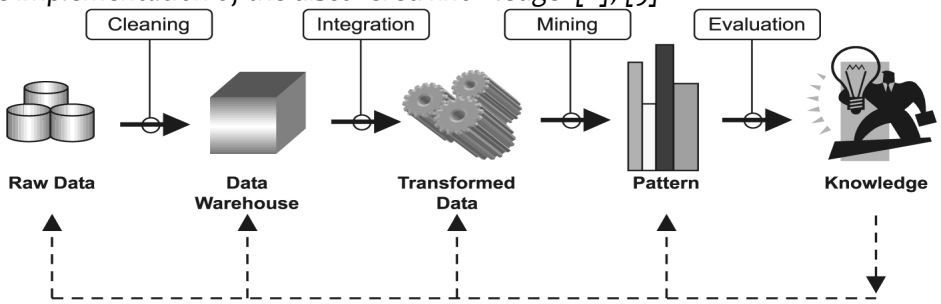

Figure 3: KDD process [7]

It is very important to note, that Data Mining is only a part of KDD process. Data mining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques. [4], [10] Data mining is defined as the process of discovering patterns in data. The process must be automatic or (more usually) semiautomatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities.

In data mining, the data is stored electronically and the search is automated – or at least augmented – by computer. Even this is not particularly new. Economists, statisticians, forecasters, and communication engineers have long worked with the idea that patterns in data can be sought automatically, identified, validated, and used for prediction. What is new is the staggering increase in opportunities for finding patterns in data. The unbridled growth of databases in recent years, databases on such everyday activities as customer choices, brings data mining to the forefront of new business technologies or methodologies. [4], [11]

RESULTS OF RESEARCH

In our research we tried [4] to implement a few Data Mining methods to Six Sigma methodology and improve results with them. References [5] and [6] describe possibility of utilization Data Mining algorithms in industry, generally. We improved each phase of DMAIC with Data Mining algorithms and methods:

Define:

- *Identifying promotional target customers (with Regency-Frequency-Monetary analysis)*
- *Making offers to customers (with Self-Learning Response model)*

Measure:

- *Preparing data for analysis*
- *Feature selection of fields*
- *Modelling for a binary target*
- *Modelling for a numeric range target*

Analyze:

- *Condition monitoring of machines*

Improve:

- *Market basket analysis*

Control:

- *Fraud detection*
- *Defects prediction*

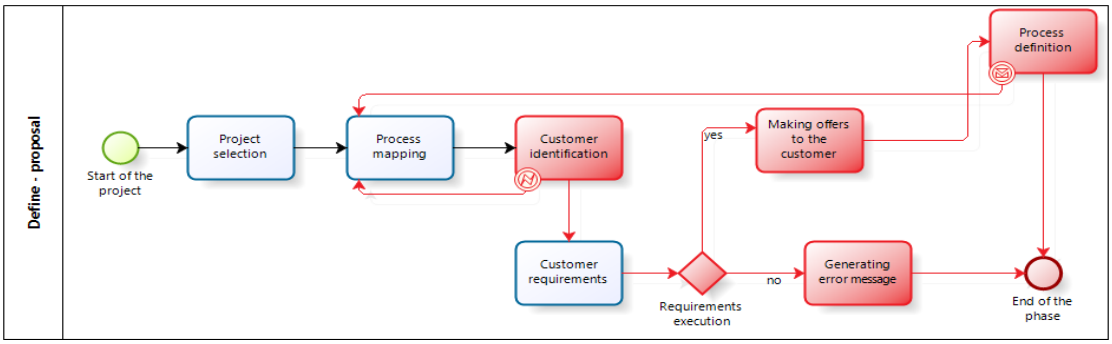

Figure 4: Proposed D phase

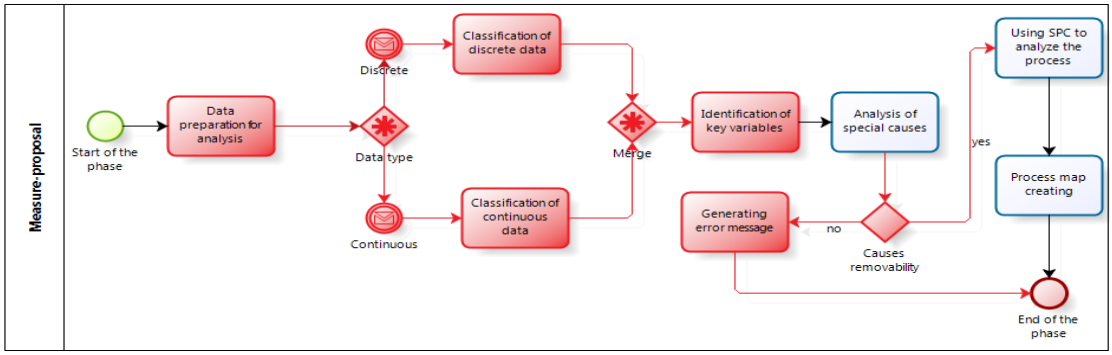

Figure 5: Proposed M phase

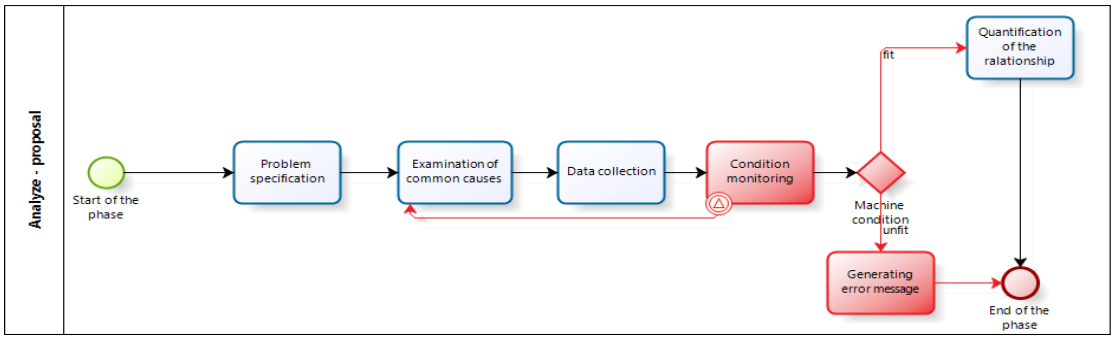

Figure 6: Proposed A phase

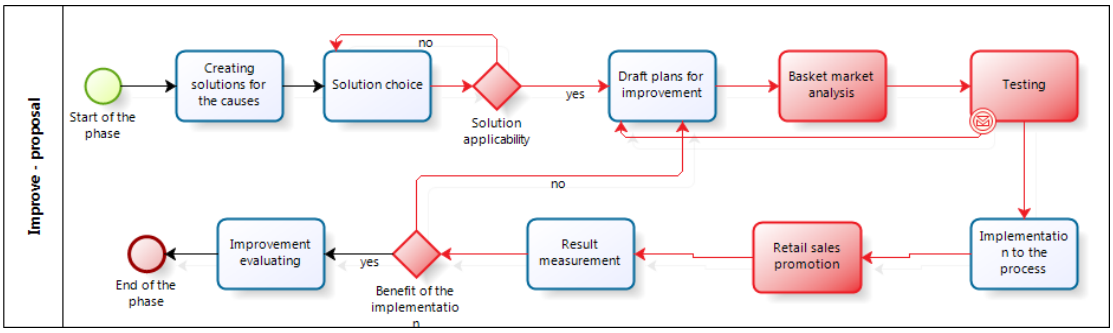

Figure 7: Proposed I phase

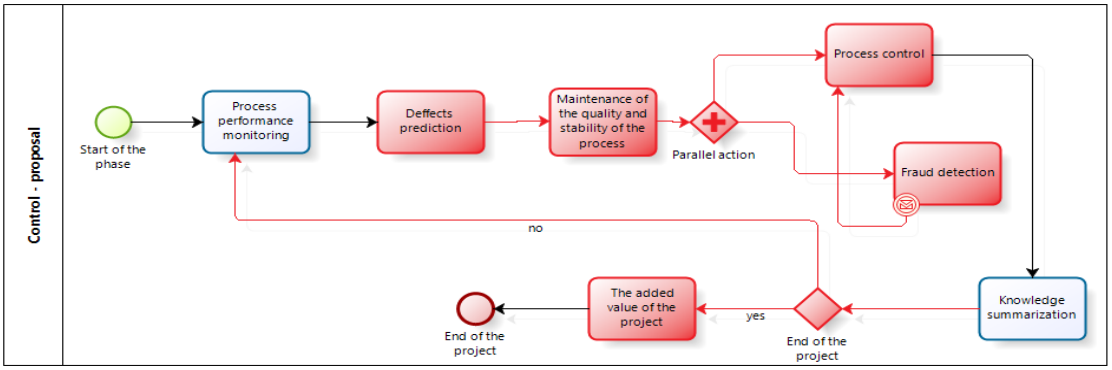

Figure 8: Proposed C phase

CONCLUSION

Figure 4 – Figure 8 shows our improvement in each phase of DMAIC. The red elements represent our benefits. This benefit of using Data Mining algorithms and methods should be multiplied with utilization of Data Warehouses and ETL process. [13] We suggest using simulation [12] for each proposed Data Mining model.

ACKNOWLEDGMENT

This paper was supported by project VEGA 1/0214/11. Grateful acknowledgment for translating the English edition goes to Juraj Mistina.

REFERENCES

[1] *Thomas Pyzdek, Paul Keller: The Six Sigma HandBook: A Complete Guide for Green Belts, Black Belts, and Managers at All Levels, The McGraw-Hill 2010, ISBN 978-0-07-162337-7*

[2] *Issa Bass, Barbara Lawton: Lean Six Sigma – Using SigmaXL and Minitab. McGraw-Hill 2009, 362 p., ISBN 978-0-07-162621-7*

[3] *Lean Six Sigma Institute: What is Lean Six Sigma. online: <http://www.sixsigmainstitute.com/leansigma/index_leansigma.shtml>, cited 20.6.2011*

[4] *Andrej Trnka: Proposal of Using Selected Data Mining Methods in Six Sigma Methodology. In: Journal of International Scientific Publication: Materials, Methods & Technologies. Info Invest Ltd. Vol. 4, Part 2, 2010, pp. 409-417, ISSN 1313-2539*

[5] *Michal Kebisek, Michal Elias: The Possibility of Utilization of Knowledge Discovery in Databases in the Industry. In: Annals of MTeM for 2009 & Proceedings of the 9th International Conference Modern Technologies in Manufacturing, 2009 October 8-10, Cluj-Napoca, Romania. Cluj-Napoca: Technical University of Cluj-Napoca, 2009. pp. 139-142 ISBN 973-7937-07-04*

[6] *Pavol Zavackz, Maximilian Stremy, Jaroslav Geci: Application of Genetic Algorithms in Combined Discrete Dynamic Systems. In: Process Control 2010: 9th International Conference. Kouty nad Desnou, 7.-10. 6. 2010, Pardubice: University of Pardubice, 2010, pp. C052a-1-7. ISBN 978-80-7399-951-3*

[7] *Chia-Chen Chen, An-Pin Chen: Using Data Mining Technology to provide a Recommendation Service in the Digital Library. In: The Electronic Library, Vol. 25 Iss: 6, pp.711 – 724, ISSN 0264-0473*

[8] *Max Bramer: Principles of Data Mining. Springer, 2007, 343 p., ISBN 1-84628-765-0*

[9] *Oded Maimon, Lior Rokach: Introduction to Knowlege Discovery in Databases. Data mining and Knowledge Discovery Handbook. Springer 2005, edited by Maimon O., Rokach L., pp. 1-17, ISBN 0-387-24435-2*

[10] *Daniel Larose: Discovering Knowledge in Data: An Introduction to Data Mining. John Wiley 2005, 222 p., ISBN 0-471-66657-2*

[11] *Soumen Chakrabarti, et. al: Data Mining: Know it all. Elsevier Inc. 2009, 460 p., ISBN 978-0-12-374629-0*

[12] *Peter Schreiber, Pavol Vazan, Pavol Tanuska: Production Optimization by using Genetic Algorithms and Simulation Model of Production System. In: Annals of Daaam for 2008 & Proceedings of the 19th International Daaam Symposium - Intelligent Manufacturing & Automation: Focus on Next Generation of Intelligent Systems and Solutions; 2008 October 22-25; Trnava, Slovakia. Daaam Int Vienna, Vienna Univ Technology, pp. 1229-1230, ISBN 978-3-901509-68-1*

[13] *Robert Halenar: Contribution of Near Real Time ETL. In: 2011 International Conference on Database and Data Mining (ICDDM 2011): 25-27, March, 2011 : Proceedings / editors: Steve Thatcher and Liu Guiping. Sanya: IEEE, 2011, pp. 243-247, ISBN 978-1-4244-9610-5*