



# A framework for a structured problem-solving approach with experimental design as the focus for industrial processing environment

WH van Blerk\* and C van der Vyver\*

*Received: 22 September 2022; Revised: 3 April 2023; Accepted: 10 August 2023*

## Abstract

This article focuses on developing a framework for industrial environments to provide a structured problem-solving approach based on experimentation as a basis to assist analysts and management for strategic decision making for process improvement. The process for developing this generic framework was through an analytical process improvement process case study for a company in an industrial environment. The research methodology includes an interpretivist approach followed by a positivist approach and ends with a constructivism worldview due to the Experimental design approach design depicted by this framework. The summarised goals were, expanding Design of Experiments (DOE) as a statistical approach to complement existing methods and methodologies used for Data Mining, validate the integrity of data through a refining process and applying DOE in combination with traditional Data Mining techniques. The importance for developing this framework was to experiment with historical data, based on real process data then to predict future process behaviour, using full and fractional DOE design scenarios which allows the analyst not to have a one-dimensional analytical approach but to evaluate which design fits the data best. No risk of costly process failures due to experimenting into the unknown by utilising historic data by applying experimentation when evaluating different processing scenarios for possible product improvement and to provide an alternative statistical approach for Data Mining in an industrial environment for screening independent and dependent variables for a DOE model.

**Key words:** Data mining (DM); Six sigma (SS); Design of experiments (DOE); Statistical process control (SPC); Regression analysis (RA).

---

\*School of Computer Science and Information Systems, North-West University, Vanderbijlpark Campus, South Africa, email: [william.vanblerk@nwu.ac.za](mailto:william.vanblerk@nwu.ac.za)

<http://dx.doi.org/10.5784/39-2-003>

## 1 Introduction

A need for a structured framework was identified to assist problem solvers in an industrial processing environment to avoid a haphazard problem-solving approach. The framework consists of two main components: one being methodologies, showing the high level non-analytical (theory) portion of the framework, and secondly selected analytical scientific methods. From a research methodology perspective this research includes an interpretivist approach followed by a positivist approach and ends with a constructivism worldview due to the Experimental design approach design. It also gives an alternative structured option to data analytics. In developing a generic framework for analysts and management to follow when an extensive data analysis is considered with the embedding of experimental analysis to serve as an analytical roadmap for process development and improvement is a challenge. The focus of having DOE as the core for industrial datamining for the analytical process forms the basis of this framework.

Through the process improvement process using scientific experimentation through different operating conditions also focused on optimal process conditions with the lowest possible cost impact for the company.

A significant difference in the analytical approach presented by this framework to similar frameworks is that it is based on available historic data to reduce the risk of costly experimentation with untested data which provide a financial benefit to management.

## 2 Methodologies fitting the framework

Four business contexts were researched with the goal of introducing different methodologies, from data sourcing to decision implementation. Business Intelligence (BI), Knowledge Discovery through Data (KDD), Big Data (BD) and Data Mining (DM). The goal was not to exhaust all possible theories and philosophies but only those few that relate to the analytical process improvement process which at the end forms part of the research framework. A theoretical overview and condensed version of the methodologies referenced for the development for this framework is as follows.

### 2.1 Business Intelligence (BI)

BI has evolved into a multi-dimensional data processing environment where technologies, methodologies, processes and different architectures are utilised for value added process analysis for management strategic decision-making. Unless businesses have an integrated system to analyse and control this ever-increasing volume of data, the accumulated data will be only data and nothing else. Modern BI developed from the traditional three-layered data warehouse to a data warehouse that consists of transactional and non-transactional data transformed for querying, reporting and data analysis when needed.

[21] and [22] clearly indicate that BI is a broad category of data analytical methods and database technologies for gathering, storing, analysing, and providing access to data to assist management in all levels of an organisation to improve their business decisions. The

importance of a stable, integrated data warehouse is becoming increasingly critical for BI as a supportive system for management.

[19] emphasizes that BI provides data in a structured layered process by data extraction from operations into a data warehouse then to be analyzed through parametric and non-parametric analytical tools by management for strategic decisions. BI does not only provide a platform for data analysis but also data for predictive analysis.

[38] describe five important stages to transform data into BI successfully. These are the collection of raw data from business enterprise, data cleaning through search engines and filtering processes, data warehousing, implementation of BI tools, and analysing outputs. Transforming data into value added BI is not easy but must follow a systematic approach and well-designed scientific methods of analysing data for management strategic decision making.

BI is a broad category of applications and technologies for gathering, storing, analysing, and providing access to data to assist management in all levels of an organisation to improve their business decisions. Although BI is the holistic component of the proposed framework for strategic management thinking and decision-making it served as an integrated system that allows users to carry out high-level data analyses with the information in data warehouses.

## **2.2 Big Data (BD)**

The inclusion of statistical techniques to uncover the real value of data in large databases (BD) generally has been underutilised by management in the sense that management predominantly focuses on quick fixes.

[37] summarises conceptual issues in big data science as causality, quality, security, Big Data and uncertainty. For this purpose, he calls it big data science that refers, not to how, but the way big data are analysed. Big data science is an overall philosophy for managing large data sets in the modern manufacturing and services environment as well as the way big data is analysed. Big data (BD) is a consequence of data explosion, experienced by all industries. [35] describe big data as a data mining evolution that became big data as accumulated data grew exponentially. Because of the large amount of data accumulated daily in industry, the traditional DM process had to adopt the growing data explosion.

According to [26], a survey done to identify benefits and challenges showed that the top three biggest benefits of Big Data are to detect product defects to boost quality, improve supply planning and improved defect detection in a manufacturing/production environment.

## **2.3 Knowledge discovery through data (KDD)**

KDD is proposed as a stepwise process and emphasizes the importance of a structured approach for data analysis, through either a methodology, framework or steps. It guides the analyst and management when analysing data. KDD focuses on the development of methods and analytical techniques for making sense of data for strategic management decision making. The KDD process focuses on different approaches for data extraction

by KDD which is software driven and based on statistical analysis that includes probabilistic, statistical, classification, data cleaning and decision trees, but also data mining applications like neural networks and machine learning. [1] describe DM as a knowledge discovery process, it is the analysis step of KDD. These authors also distinguish between DM and KDD in the sense that DM is part of the KDD analytical methodology process.

[17] refers to the primary goal of Knowledge discovery through data (KDD) to be the transformation of data into usable summarised forms for management and data users. In a general sense, this should be the goal of all analysts: to provide data in a summarised, condensed, factual format to allow management to focus on the core issues only, rather than be fragmented amongst various data bases to make sense of available data. [3] focus on the development of methods and analytical techniques for making sense of data for strategic management decision making. Knowledge discovery cannot happen on its own and therefore is based on a structured data driven technique for analytical purposes ([13]; [16]; [31]).

[7] describes KDD as the discovery of new information and knowledge. New information is not necessarily information that has been recently added to a process or business, but can be latent information, never exposed until it has been discovered. Modern KDD evolved into a multidisciplinary activity that utilises techniques such as machine learning, pattern recognition, statistics, data visualisation and high-performance computing with special emphasis on uncovering patterns, identifying outliers through exploratory analysis and structured experimenting (DOE) on databases.

KDD is an interactive and iterative by nature and usually follows a defined structure for analysing data. A typical, generic, KDD process flows as follows: Set goals and objectives before the process starts, quality of data is imperative for fact-based decisions, clean data, design a structure or model prior to analysing data then start the data analysis.

KDD also focuses on the development of methods and analytical techniques for making sense of data for strategic management decision making. Because of this structured analytical approach, the framework reflects a similar process.

## 2.4 Data mining (DM)

Data mining not only concentrates on the manufacturing industry but is also relevant in the service industry. Operating conditions, environment, raw materials, process changes and traditional analytical methodologies will be challenged to validate alternative operating conditions through the DM process. Although DM has grown as a major discipline in IT for analysing industrial data, [39] confirm that DM is not confined to process data but has evolved into all functions of business where the need arises to analyse data. Each of these areas requires specific mining techniques and is developed as new areas are identified for data analysis.

Data mining is an integrated process of various data analysis disciplines and methodologies and is not a stand-alone analytical discipline that provides a business solution to management. [30] describes DM as a methodology which compares to a typical DM approach such as define the problem, get data, transform the data, determine which analytical technique is appropriate to the problem, analyse the data, review results then implement selected

results. DM in general is projected as a framework consisting of four major stages: data accumulation, product family classification, design retrieval and modification [43].

[18] describe DM as an interactive and iterative process of finding knowledge in experimental data sets. This iterative analytical process is typical for DM analysts because since the ninety's analysts have experienced that a once off data analysis process is not effective because the change in operational conditions is dynamic and needs an iterative approach.

[43] describes a data mining framework for DM as selecting and accumulating data for analysis, classification of data into product families, formulating a designed model and an iterative process to change or modify proposed design. Whether the DM process is described as a methodology by [30] or a framework, both follow a structured analytical process.

Although Data mining seems to be the answer for large data analysing for business solutions, one of the issues encountered when analysing data, is missing data. [9] refers to these data as missing at random, data missing completely at random, non-ignorable missing data, and outliers treated as missing data. For this reason, before any data analysis commences, irrespective of the methodology (BI, KDD, DM, BD) followed, data integrity testing is a high priority. Missing data should be part of data cleaning before any data analysis should be attempted.

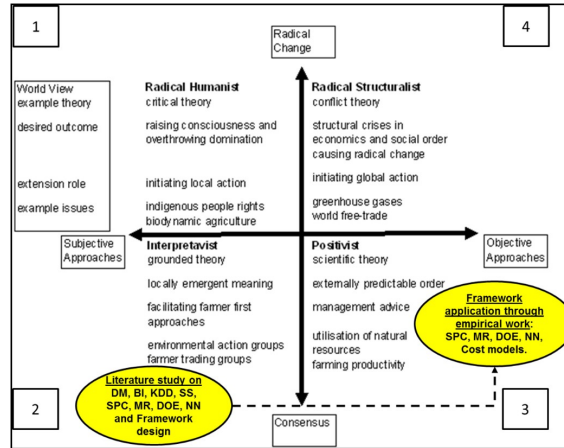
A plethora of DM techniques exist; [30] mentions only a few statistical methods, each with its own purpose in data analysis, but applying Data mining should be structured and methodical. Haphazard analysis will only add to the frustration of analysing data, and in most cases the results are irrational for managers to use. For this reason, the DM application process is more methodology based and less technique driven. Techniques help with analysing data where following a methodology that includes technique application ensures that the analyst stays focused and does not jump to conclusions.

### **3 Research methodology**

From a research methodology perspective this research includes an interpretivist approach followed by a positivist approach and ends with a constructivism worldview due to the Experimental design approach design depicted by the framework. For this reason, not a single research approach is applicable to this research, but positivism seems to be the main approach because of the large empirical portion of the research. This research approach could also be classified as mixed method approach which may be more fitting for this application.

Due to the quantitative basis of this research that flows from identifying different methodologies to a proposal for an experimental design approach for industrial DM not only involves one quadrant, see Figure 1 constructed by [27], adapted for this research.

From an interpretivist approach the different discussions and research to design a framework fit in interpretivist, quadrant 2, a literature study was done that included frameworks (DM, BI, KDD, DM) and analytical methods (SS, SPC, MR, DOE, NN) to improve a the-



**Figure 1:** Dimensions and attributes of extension worldviews with framework design and application.

oretical understanding of selected frameworks and analytical methods for this study. It is still subjective and consensual by nature. The application of the framework through analytical methods (SPC, MR, DOE, cost models) shifted the approach to quadrant 3, a positivist approach. This approach is objective, based on science but still consensual.

The experimental design approach may also fit in the radical structuralist quadrant 4 because the implementation is new and radical to its contribution to change. However, there is no guarantee that long-term implementation benefits are sustainable because the context of the database may change with human intervention, like raw material changes and purchasing policy changes.

## 4 Analytical methods fitting the framework

### 4.1 Six sigma

An industrial process development tool, Six Sigma (SS) also fits well within DOE and covers all stages of a DM methodology. This methodology is very similar to the methodology used for data mining in that both want to find patterns and associations in data not normally detected, for process improvement. The similarities and differences in the DMAIC (Define Measure Analyse Improve Control) approach are relevant to this research because Six Sigma also focuses on process development and streamlining of processes.

Six Sigma is a management strategy, according to [11] which focuses on improving product quality and streamlining production processes. According to [14] and [15] it is a philosophy, with the business goal of increasing process capability, decreasing process variability, and with the main goal of removing defects from business processes]. It is a framework for process quality improvements according to [12]; a management methodology according to [32], with the goal increase process predictability by eliminating defects in order to improve and sustain quality, eliminate waste and achieve sustainable profits.

The DMAIC (Define, Measure, Analyse, Implement and Control) process according to [4]

is a graphically driven process to present visual interpretations for analysis to management.

This shows that Six Sigma is customer focused through process analytics. Both [20] and [15] describe a new extension of Six Sigma that is Design for Six Sigma (DFSS), which is a strategy with the goal of designing or re-designing a product from the beginning of the process life cycle to develop optimized designs. This extension of SS fits well within the concepts of DOE for experimental design.

The DMAIC methodology fits in well for this research showing the importance of a data-driven methodology that, on a macro basis, is used for process development. Six Sigma is based on the DMAIC methodology, which is a structured data analytical process for process improvement and fits within the proposed framework through the data analysis process.

## 4.2 Statistical process control

SPC is not new but a very powerful analytical technique for industrial process improvement. For this research, it formed a critical part in screening the critical few independent variables for DOE analysis and therefor formed a critical component of the developed framework.

[33] describes Statistical Process Control (SPC) as a statistical tool to help set standards as well as monitor, measure, and correct product quality problems. Track variables through SPC charts and observe for patterns like trends, shifts, clusters and non-normal variation. Then only make process adjustments if necessary. [42] stipulates that SPC has been introduced into the general manufacturing industries for monitoring process performance and product quality and to monitor the general process variation that is caused by a few key process variables.

Because this framework sets a roadmap for analysing large data basis, measuring continuous data variation is an integral part of process analysis.

## 4.3 Design of experiments

Of all the traditional statistical techniques used for Data Mining, Design of Experiments (DOE) are not used as an accepted Data Mining technique. DOE is utilised to scientifically determine how an input affects outputs, and then to use this knowledge to optimize processes. The objective for this framework was to focus primarily on the industrial process applications in a manufacturing environment. The proposed approach is different from the traditional Experimental design approach in that historical data accumulated in traditional databases are used to determine effects of variables on different outputs and then to predict the future process operating levels with minimizing operational costs ([2]; [10]; [23]; [28]; [29]; [36]). [23] describes DOE as: “Experimental design as a scientific approach of purposefully change inputs to a process to evaluate the changes in the outputs”.

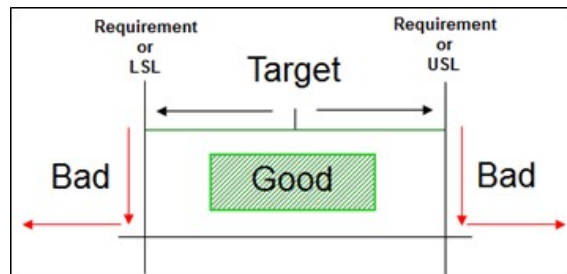
DOE is very useful in assisting in robust design that accommodates most uncontrollable independent variables and fixes them prior to going into production. DOE was popularised by [5] and [6] who primarily discuss designs with many factors that estimate process outcome effects with a minimum number of observations.

DOE is a powerful statistical technique that focuses on evaluating contributions from independent variables (factors) to the effect on output variables (responses). It is a scientific approach by purposefully changing input variables to evaluate the impact on process outcomes in a controlled experimentation environment.

#### 4.4 Cost methods

For this study the associated cost in determining the best cost-effective experimental run is important because not only does each experimental run deviate from the standard, but it also has a cost implication element for each run. [24] discusses some basic elements of quality control that include DOE, as the total loss generated by a product to society will be known. Deviation from the target results is expressed as a loss to the customer expressed in the quadratic quality/loss function, quality and cost of a product are determined by the engineering design and manufacturing process.

The traditional cost model, Figure 2, that is based on the principle that a company only starts to lose money if products are produced outside the process specifications.



**Figure 2:** *Traditional cost diagram.*

Taguchi changed the belief in the traditional cost diagram that money is lost as the process starts to deviate from the set average specification. The three Taguchi quadratic quality loss functions were used for this research, Nominal the best, Smaller the best and larger the best. Each with its comparative signal to noise ratio for each experimental run, for determining the total cost for deviating from the process target standard. The goal was to minimize the variability in the product's performance in response to noise factors that influence the product performance while maximizing the variability in response to noise factors. [40] detailed the application of these loss functions with the resulted implications, using the framework.

#### 4.5 Regression analysis

For this study we focus on linear relationships between multiple independent variables and dependent variables. For this reason, multiple regression serves as a very useful multivariate statistical tool [41]. Multiple regression measures relationships between multiple independent variables and a dependent variable. It sets a platform for measuring the numerical scale for group or individual relationships based on statistical assumptions and measurements [25].



[34] explains that there are various procedures within a wide area of linear regression applications that have a direct implication in quality improvement work, and regression analysis is a complementary statistical approach of analysing data from designed experiments' model outcomes. [8] refers to applying regression analysis effectively as not an easy exercise but adds that it is even more difficult to interpret the results, so that it makes sense in terms of both quantitative and qualitative variables.

Multiple Regression (MR) analysis has been a critical part of statistical techniques used through the years, specifically when trying to find relationships amongst independent variables and a dependent variable. Regression analysis complements designed experiments in predicting the behavior of the dependent variable through selected independent variables. In this research, multiple regression analysis was applied to compare multiple regression analysis to designed experiment model regression.

#### **4.6 Neural networks**

The ability of Neural Networks (NN) to learn by example is one of the many features that enable the analyst to model data and establish accurate rules governing the underlying relationship between various data attributes. Neural network uses training algorithms, which can automatically learn the structure of the data presented by the analyst. This unique analytical feature of neural networks makes it a popular DM technique for analysts as a predictive model. Neural networks consist of three basic stages: Stage 1: Exploration. This stage usually starts with data preparation that may involve cleaning data, data transformations, selecting subsets of records, and, in case of data sets with large numbers of variables, screening of these variables to work only with those variables that add value to the process analysed. Stage 2: Model building and validation. This stage involves considering various models and choosing the best one, based on their predictive performance (i.e., explaining the variability in question and producing stable results across samples). Stage 3: Deployment. This final stage involves using the model selected as best in the previous stage and applying it to new data to generate predictions or estimates of the expected outcome. For this research NN were used to evaluate and compare the MR variation as dependent variables reduces.

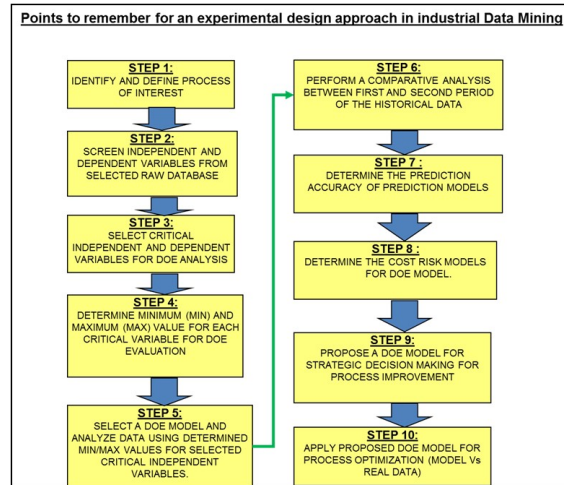
### **5 Summary of results**

While analysing the case study with the main aim of introducing DOE as a DM method an analytical sequential process starts to emerge. Towards the end of the study a pragmatic, efficient and structured problem-solving roadmap showed the importance of a framework for similar studies. The proposed framework is a summarised structured problem-solving roadmap that flowed from a case study for process improvement. The methodologies selected gave a theoretical platform from which data analysis for this case study were referenced from. These methodologies gave a reliable background from different angles how data is stored, managed, and offered for data analysis. These methodologies set the basis for the selection of analytical methods used for process analysis. By applying these analytical techniques, it also become evident that the sequence in which these techniques are used are critical in terms of time, cost, prediction accuracy and data integrity.

The framework is therefore a culmination of the theoretical methodologies and analytical techniques constructed as part of the results of a case study to assist analysts in following a structured approach in problem solving.

### 5.1 Flowchart representing the analytical process

When reflecting back to the analytical process followed during the research, a few generic points are summarized in a flow chart format, figure 3, to keep in mind when an analyst or management need to embark on a study of a similar nature. These points are expressed in sequential “steps” because they follow a generic roadmap of the case study for this research (Van Blerk 2016). This roadmap was specific in keeping experimental design as the main theme which also complements the development of the proposed framework through this research.



**Figure 3:** Flowchart for an industrial data mining analysis [40].

## 6 Proposed framework overview

The advantage for following the proposed sequential framework, Figure 4, is that it will assist the analyst to systematically approach and analyse data. With specific reference to DOE the experimenter will know the probable impact of changing inputs on the process changes before any dynamic in-line process changes are made. In addition, changes can be introduced beyond the test model’s parameters with a higher degree of confidence than when developing a DOE process development model based on guesswork, gut feel, experience, and the high probability of high-cost implications associated with trial-and-error experimental runs.

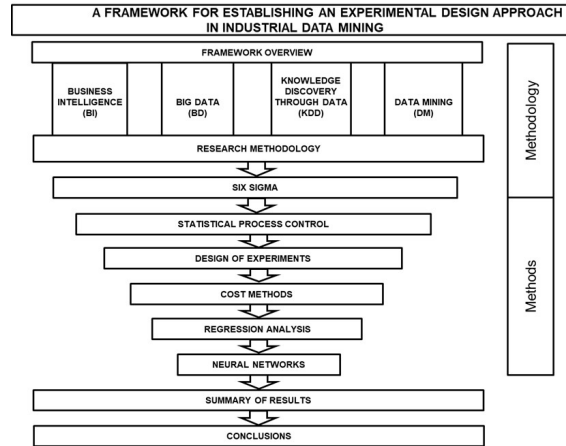


Figure 4: Proposed framework [40].

## 7 Discussion on developing the framework

This framework shows that empirical analysis should follow a bottom up and not a top-down approach. This means that the analytical process should start with analysing individual variables, then progress to multi-variate analysis, if necessary, not the other way around. It saves time and a better understanding of the process data will be realised. The challenge for this study was not the empirical analysis, but to follow the framework process, objectively validating the results, refrain from explaining deviations, and jump to conclusions based on the accepted norm and personal experience to fit your understanding. From a pragmatic industrial operating perspective, the funnelled sequential empirical analytical process keeps analysts and management focused on a prescriptive analytical roadmap to restrain from including non-value-added issues that could clutter the objective of the study.

## 8 Future research

Because the framework was developed near the end of the case study analysis, the results for each analytical method is not part of this article. Future articles will share the results for each section of framework which should put the framework in context in terms of how data was analysed and the effect of following the proposed framework.

## 9 Conclusions

The analytical approach initially started with a top-down approach using MR analysis to evaluate the significance of independent variable contribution within a MR model. During the analytical process, the need for a bottom-up approach became evident for individually evaluating the significance of each independent variable to the process output. By doing both approaches, a holistic analytical approach was achieved that led to multidimensional model options for process improvement.

The developed framework for this study, is a generic framework for analysts and management to follow when an extensive data analysis is considered. This framework with the embedded experimental analysis design serves as an analytical roadmap for process development and improvement. Proposing DOE as the core for industrial data mining focuses the analytical process on process improvement through scientific experimentation through different operating conditions to determine optimal process conditions with the lowest cost impact for the company.

## References

- [1] ALAZMI AR & ALAZMI AR, 2012, *Data mining and visualization of large databases*, International Journal of Computer Science and Security, **6(5)**, pp. 295-314.
- [2] ANTONY, J, 2014, *Design of Experiments for Engineers and Scientists. 2nd ed.*, Elsevier, London.
- [3] AYOBAMI A & RABI'U S, 2012, *Knowledge discovery in database: A knowledge management strategic approach.*, Paper presented at the 6th Knowledge Management International Conference, Johor Bahru, Malaysia, July. <http://ssrn.com/abstract=2088965>. Date of access: 4 Jun 2015.
- [4] BOSLAUGH S., 2013, *Six Sigma: quality control standard*, Salem Press, Encyclopedia. <http://eds.a.ebscohost.com.nwulib.nwu.ac.za/eds/detail/detail?sid=0d11d5d0-af59-4b16-8a5d-55378b394047%40sessionmgr4004&vid=0&hid=4110&bdata=JnNpdGU9ZWRzLWxpdmU%3d#db=ers&AN=89677627>. Date of access: 16 Sep 2014.
- [5] BOX GEP, HUNTER WG & HUNTER JS, 1978, *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Wiley, New York.
- [6] BOX GEP & DRAPER NR, 1969, *Evolutionary Operation: A Statistical Method for Process Improvement*, Wiley, New York.
- [7] BREŠIĆ B, 2003, *Knowledge acquisition in databases*, Management Information Systems, **7(1)**, pp. 32-41.
- [8] BRIGHTMAN HJ, 1978, *Data Analysis in Plain English with Microsoft Excel*, Brooks/Cole: Duxbury Press.
- [9] BROWN ML & KROS JF, 2003, *Data mining and the impact of missing data*, Industrial Management and Data Systems, **103(8)**, pp. 611-621.
- [10] CANO EL, MOGUERZA JM & REDCHUK A, 2012, *Six Sigma with R: Statistical Engineering for Process Improvement*, Springer, New York.
- [11] DE CARVALHO G, DA SILVA CE & COSTA K, 2014, *Application of Six Sigma methodology in improving of the industrial production processes*, Applied Mechanics and Materials, **627**, pp. 327-331.

- [12] DESALE SV & DEODHAR SV, 2013, *Lean Six Sigma principal in construction: a literature review related to abstract*, Journal of Information, Knowledge and Research in Civil Engineering, **2(2)**, pp. 133–139.
- [13] BOX GEP, HUNTER WG & HUNTER JS, 2003, *Data Mining: Introductory and Advanced Topics*, Prentice Hall, New Jersey.
- [14] EI-HAIK BS & SHAOUT A, 2011, *Software Design for Six Sigma: A Roadmap for Excellence*, Wiley, Hoboken.
- [15] ENOCH OF, SHUAIB A & HASBULLAH AHB, 1999, *Applying P-diagram in product development process: an approach towards design for Six Sigma*, Applied Mechanics and Materials, **789-790**, pp. 1187-1191.
- [16] FAYYAD U, PIATETSKY-SHAPIO, G & SMYTH P, 1996, *From data mining to knowledge discovery in databases.*, AI Magazine, **17(3)**, pp. 37–54.
- [17] FELDKAMP N, BERGMANN S & STRASSBURGER S, 2015, *Knowledge discovery in manufacturing simulations*, Proceedings of the 3rd ACM SIGSIM Conference on principles of advanced discrete simulation. <http://dl.acm.org/citation.cfm?id=2769468>. Date of access: 12 Apr 2016.
- [18] HORNÍKOVÁ A, DURAKBASA, NM, GÜCLÜ E & BAS G, 2011, *Data mining: novel statistical method*, Paper presented at the 8th International Conference, Smolenice, Slovakia.
- [19] JHA R, 2014, *Association rules mining for business intelligence*, International Journal of Scientific and Research Publications, **4(5)**, pp. 1–5.
- [20] KAI Y & BASEM SE, 2009, *Design for Six Sigma: A Roadmap for Product Development. 2nd ed.*, McGraw-Hill, New York.
- [21] KOČEKOVÁ, A., KOPČEK, M. & TANUŠKA, P., 2013, *Business Intelligence in process control*, Research papers Faculty of Materials Science and Technology Slovak University of Technology, **21(33)**, pp. 43–53.
- [22] BABU AR & BABU NR, 1999, *Industrial intelligence: a business intelligence-based approach to enhance manufacturing engineering in industrial companies*, Procedia CIRP, **12**, pp. 384–389.
- [23] LAUNSBY RG & SCHMIDT RS, 1978, *Understanding industrial designed experiments*, Air Academy Press, Colorado.
- [24] LOCHNER RH & MATAR JE, 1990, *Designing for Quality: An Introduction to the Best of Taguchi and Western Methods of Statistical Experimental Design*, ASQC Quality Press, New York.
- [25] MOEINADDINI M, ASADI-SHEKARI Z & SHAH MZ, 2014, *Analysing the relationship between park-and-ride facilities and private motorised trips indicators*, Arabian Journal for Science and Engineering, **39(5)**, pp. 3481–3488.

- [26] NEDELCO B, 2013, *About big data and its challenges and benefits in manufacturing*, Database Systems Journal, **4(3)**, pp. 10–19.
- [27] PARMINTER T, BOTHA N & SMALL B, 2003, *Appreciating the influence of our own and others' world views upon extension strategies*, <http://www.regional.org.au/au/apen/2003/refereed/095parmintertg.htm>. Date of access: 3 Jun 2015.
- [28] PETERKA P, 2008a, *The six sigma method and design of experiments*, <https://www.statease.com/pubs/sixsigma&DOE.pdf>. Date of access: 13 Sep 2014.
- [29] PETERKA P, 2008b, *The six sigma method and design of experiments*, <http://www.6sigma.us/six-sigma-articles/the-six-sigma-method-and-design-of-experiments/> Date of access: 16 Sep 2014.
- [30] PETRE R, 2013, *Data mining solutions for the business environment*, Database Systems Journal, **4(4)**, pp. 21–29.
- [31] PUROHIT N, PUROHIT S & PUROHIT R, 2012, *Data mining, applications and knowledge discovery*, International Journal of Advanced Computer Research, **2(6)**, pp. 458–462.
- [32] RAJKUMAR S. & RAMESH S, 2014, *Impact of implementing Six-Sigma on the workplace efficiency in service industries in Karnataka*, International Journal of Production Research, **1(13)**, pp. 65–79.
- [33] RENDER B, STAIR MR & HANNA ME, 2012, *Quantitative Analysis for Management. 11th ed.*, Pearson Prentice Hall, Harlow.
- [34] RYAN TP, 1989, *Statistical methods for quality improvement*, Wiley, New York.
- [35] SOWMYA Y, RATNA M & BINDU C, 2015, *A review on big data mining, distributed programming frameworks and privacy preserving data mining techniques*, International Journal of Advanced Research in Computer Science, **6(1)**, pp. 121–126.
- [36] SUNDARARAJAN K, 2010, *Design of experiments: A primer*, <http://www.isixsigma.com/tools-templates/design-of-experiments-doe/design-experiments-%E2%90%93-primer/>. Date of access: 20 Sep 2014.
- [37] SWAN, M, 2015, *Philosophy of Big Data: Expanding the human-data relation with big data science services*, Paper presented at the IEEE First International Conference on Big Data computing service and applications. [www.melanieswan.com/documents/Philosophy\\_of\\_Big\\_Data\\_SWAN.pdf](http://www.melanieswan.com/documents/Philosophy_of_Big_Data_SWAN.pdf) Date of access: 20 Sep 2014.
- [38] TEMBHURKAR MP, TUGNAYAT RM & NAGDIVE AS, 2014, *Overview on data mining schemes to design business intelligence framework for mobile technology*, International Journal of Advanced Research in Computer Science, **5(8)**, pp. 128–133.
- [39] THOTA LS & RAO AA, 1999, *Overview of empirical data mining research*, International Journal of Advanced Research in Computer Science, **4(10)**, pp. 49–55.

- [40] VAN BLERK WH, 2016, *A framework for establishing an experimental design approach in industrial data mining*, Vanderbijlpark: NWU (Thesis – PhD).
- [41] WEISBURG GS, 1985, *Applied Linear Regression*, Wiley, New York.
- [42] XIE L & KRUGER U, 2012, *Advances in Statistical Monitoring of Complex Multivariate Processes: With Applications in Industrial Process Control*, Wiley, Hoboken.
- [43] YU L & ZHANG ZF, 2014, *The framework of product definition based on data mining techniques*, *Applied Mechanics and Materials*, **687-691**, pp. 4874–4877.