SME Productivity improve using six sigma and data mining tools

Jamblico Pacora, Eng[1], Luigi Flores, Eng[1], Jonatán Rojas, Mg[1], Steffano Reyes, Eng[1], Alexia Cáceres, Mg[1], and Renzo Benavente, Mg[1]

[1]Pontificia Universidad Católica del Perú, Perú, a20120467@pucp.edu.pe, luigi.flores@pucp.pe, jrojasp@pucp.pe, s.reyes@pucp.pe, alexia.caceres@pucp.pe and renzo.benavente@pucp.pe

Abstract– ***This research project aims to ease SME process control by using DMAIC (Lean Six Sigma). Three topics were applied: Lean Six Sigma, Data Mining, and Augmented Reality. All of this is assisted by the most advanced technology to date. SMEs have small processes and this is why Lean Six Sigma was applied to detect problems, measure, analyze and control. The Poka-Yoke tool works as a link between Data Mining tools and Augmented Reality.***

The Data Mining tool is the deep learning neuronal net (with more than one hidden layer). The neuronal net will learn the gauges formula and will show results based on the measurements. This tool is related to Lean Six Sigma as the net will show optimal results with controlled processes only.

Augmented Reality eases data input for the gauges. Input data will be transmitted through AR cards which will be scanned by Smartphones. These AR cards become unique through the use of the Poka-Yoke tool

The expected savings are S/. 6120. The project obtains an NPV of S/. 7811 and a COK of 15.24% which is enough to demonstrate the viability of the project.

Keywords-- SME Productivity, Six Sigma, Data mining, Augmented reality.

I. INTRODUCTION

Micro, Small & Medium Enterprises (MSME) are relevant for today's economies and employment since they represent 50% of the global Gross Domestic Product, with 70% of employees and 90% of the firms and accounts worldwide as reported by the United Nations [1]. Due to information unavailability about non-formal enterprises, this research will only consider formal enterprises. Formal MSMEs hire 7 out of 10 employees and generate up to 40% of emerging economies' GDP [2]. Besides, MSMEs had increased more than 3.8% annually in Latin America and the Caribbean, whereas in the Middle East and North Africa, more than 4%, between 2010 and 2017 according to a World Bank Group study published in 2019 [3].

This research also concluded that the most common MSME sorting among most countries is as follows: an enterprise with less than 10 employees is a microenterprise, small enterprises are those with less than 50 employees; and the medium ones, less than 250 employees [3]. In Peru, the definition includes both the number of employees and the annual sales.

In agreement with this, microenterprises have less than 10 employees and annual sales lower than 150 tax units (UIT in Spanish); meanwhile, medium-enterprises have less than 100 employees and less than 1700 tax units in annual sales [4]. A taxing unit equals 4600 soles (or 1195 USD with an exchange rate of 3.85 PEN/USD).

Micro and small enterprises (MSE) represent 95% of Peruvian enterprises (94.9% and 0.01% respectively) and employ almost 8.23 million people (46.5% of the national labor force) [5]. However, their growth rate did not match the country's growth. Between 2011 and 2017, MSE sales raised from 126000 to 135000 million soles, but their GDP share rate lowered from 24.8% to 19.2% [5]

One of the main reasons for this can be their high mortality rate. According to a National Production Department study, in 2015 and 2016, MSE had a birth rate of 5.3% and 5%, respectively, being considerably higher than the birth rate of the other type of enterprises which did not exceed 0.3% [6]. Their average life span is 6.6 years with a mean of 4 years, which means that near 50% of them stay in the market for less than four years [6].

According to Edgar Alva, the main factors for this situation are little to no advertising, income levels, fierce competition from close businesses, lack of planning and control mechanisms, and the difficulty to acquire the workforce [7].

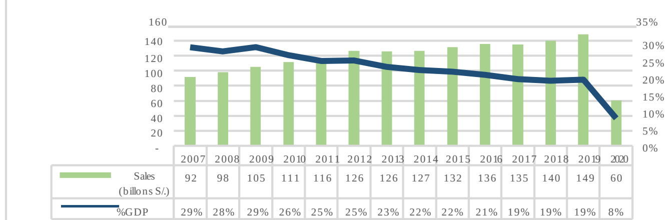

Fig. 1. SME sales (billions of soles S/.) and %GDP in Perú[5]
Source: ComexPeru (2020)

Digital Object Identifier (DOI):
http://dx.doi.org/10.18687/LACCEI2022.1.1.164
ISBN: 978-628-95207-0-5 **ISSN:** 2414-6390

Nowadays there are tools that help with cost reduction such as Lean Six Sigma. This tool uses DMAIC methodology in order to give a structure to the continuous improvement of the system, defect reduction and process variability [8]. There are also Data Mining tools that allow the processing of a great amount of information and pattern identification of said data; all of this in order to predict the behavior of both external and internal factors. This could help with planning activities because it offers a better view of the current situation as well as with the development of adequate control mechanisms by having a grasp of the future situation [9].

II. LITERATURE REVIEW

In this section, the following basic concepts will be described: Lean Six Sigma, Data Mining, Augmented reality.

A. Lean six sigma

Lean Six Sigma is the last generation of business process improvement approaches [10]. It raises process performance rate which results in an increase in client satisfaction rate and better overall results. There are researches that show that Lean Six Sigma can bring benefits for the organizations no matter how big or small they are [11], but there are other researches that these benefits have bigger impacts in great enterprises than in SME [12]. This is the reason why a good number of SME-oriented methodology approaches have been developed [13].

B. Data Mining

The Data Mining technique has three main objectives: description, prediction, and prescription, these are divided into two main categories: supervised and unsupervised learning [14]. Recent computer science technology such as cloud computing and the Internet of Things allows not only the bigger companies but SMEs as well to store and analyze great amounts of data [15]. Unfortunately, the uses of these technologies have had poor efficiency, even though data mining techniques are becoming more popular among SMEs [16]. For example, a simple machine monitoring system could generate around 0.5 Terabytes of data annually [17]. A great number of studies have shown that the enterprises that use analytics have a significant improvement in their earnings when compared to their competitors [16].

C. Augmented reality

Augmented Reality can be defined as the technology that integrates virtual objects into a physical and real environment in a way that creates the notion that everything the user perceives takes part in the same real space they live in [18]. According to Ben-ford et al., an Augmented Reality system should present the following conditions: firstly, combine virtual and real objects on a real background; secondly, all the objects projected must be aligned with each other; finally, the elements should be displayed on real-time as well as in an interactive and tridimensional way [19]. It is important to remark that this type of technology is not limited to sight, but to all of the senses of the user, just as it is not restricted to mounting devices such as virtual reality lenses and the similar. Additionally, the reality-virtual continuum concept proposed by Milgram and Kishino considers Augmented Reality as a link between the real environment and the virtual reality that composed the greater area called mixed reality [20].

Currently, Augmented Reality has been applied in fields such as manufacturing, the military, medicine, robotics, and many more. In these fields, the levels of perception and interaction of the user increase because of the virtual objects that help to perceive sensations and information that under normal conditions a person wouldn't feel. This improvement allows that difficult jobs that before were complicated to the people to complete can be achieved easier now [21].

Although Augmented Reality can help the user to develop specific tasks, there are some limitations in several of its applications, mainly those that require special Augmented Reality equipment, for example, vision devices that may present quality problems or its performance is strongly affected by obstructions. Likewise, another inconvenience is equipment failures that do not allow the user to manipulate the device correctly. These errors can be software bugs, lack of system memory, or excessive battery usage on the equipment. The problems aforementioned are considered the base for further studies that seek to improve the way Augmented Reality is designed and implemented [21].

III. PREDICTIVE MODELS APPLICATION

The scope of this investigation is to combine the basic lean six sigma concepts to find the problems in an MSME to develop a proper interface integrated with Augmented Reality tools that display the predictions of a defined indicator calculated using a neural network.

A. Methodology development

The first step is using the DMAIC methodology (from Lean Six Sigma) to enhance the review capacity and the evolution of processes. This tool divides into five points:

Define: The most critical process must be identified and the root causes should be found.

Measure: In this step, the indicators related to the most inefficient process are chosen.

Analyze: This phase is for monitoring the activity in a defined period. A cause-effect diagram could be used to track the results.

Improve: The objectives to accomplish are defined considering the indicators selected previously.

Control: The final step focus on promoting the continuous improvement of the processes and avoiding future errors.

After defining the correct indicators for the critical processes, the model for a neural network that predicts KPI is developed. For the present investigation, the neural network built consists of 4 entries, 2 hidden layers, and 1 exit as shown in figure 2.

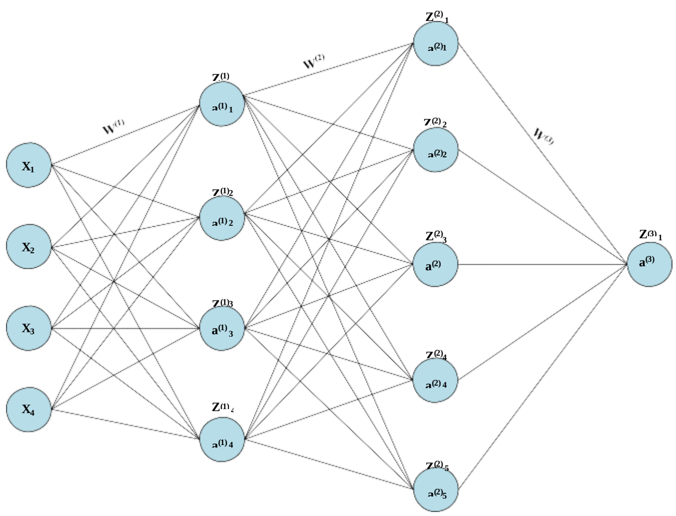

Fig. 2. Neural network

All these concepts are combined to design a proper mobile application that can be used for the workers of any MSME in which the results are displayed through Augmented Reality tools.

B. Six sigma analytics use case

To illustrate correctly the use of the application design, the next cases of use are explained:

B.1 Photocopy stores.

Photocopy centers can use this application to predict the service time they will spend for every request they have according to the time of the day.

This could help especially for those clients who leave a large order to be delivered later on the day.

If photocopy operators could know the approximate time, they will need to complete the order considering other factors such as the workload depending on the time of the day, printers available in the shift, employees available, and others, it would be easier to know if it is possible to complete the order and the time needed to do so. The photocopying process can be seen in figure 3.

In the following case, the photocopy center will consider the next variables: number of pages, type of printer, people available, and time of the day. For this specific case, the variable time of the day will help to know the workload in the photocopy center. In the morning, the demand is considered medium; in the afternoon, low; and in the night, high.

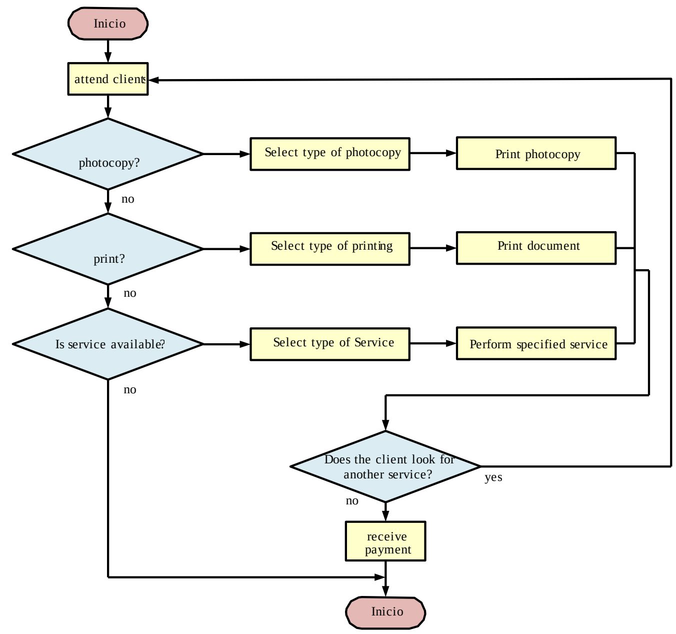

Fig. 3. Protocopy process flowchart

A final result is a number between 0 and 1 that will represent the time needed in hours to attend the order that has arrived. The hour range will vary between 0 and 6 hours. It is important to remark that the result is a prediction; therefore, an error percentage is associated. For the application developed the average error for a model with only a few variables is approximately 0.05%.

Table 1 presents the initial weights for every variable considered in this use case. These weights will feed The Neural Network and will change with every scenario in the learning period. The weights eventually will stabilize when the percentage of error of the predictions is low.

Table 1. Photocopy stores

	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Number of pages	[0-50]	[51-100]	[101-200]	[201-500]	[501-1000]	[1001-1500]	[1501-2000]	[2001-2500]	> 2500
Type of printer	T-31	T-32	T-33	T-34	T-35				
People available	5	4	3	2	1	0			
Time of the day	morning	afternoon	night						

For the application developed, the load of information is imputed by scanning the AR cards that represent each of the variables and weights. When an order arrives, the value of the variables is identified and the AR cards that match those values are selected to be scanned. Considering the following order example:

- *Number of pages:* [51-100].
- *Type of printer:* T-33.
- *People available:* 1.
- *Time of the day:* Afternoon.

The AR cards to scan will be those presented in figure 4.

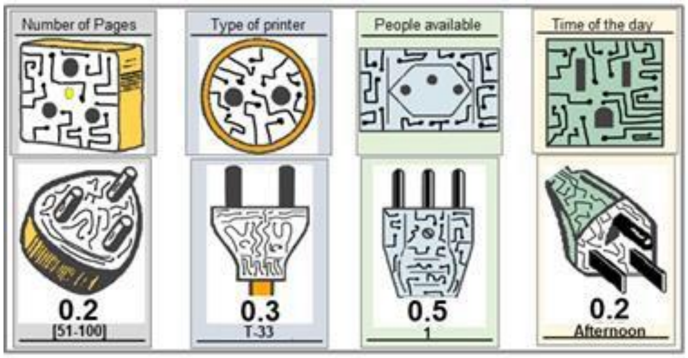

Fig. 4. AR cards and scale indicates [Pacora, Jamblico]

The final result will be displayed on the screen as a c signaling graphic. The green scale indicates a good value for the indicator. For this case of study, the value represents a reliable result (figure 5).

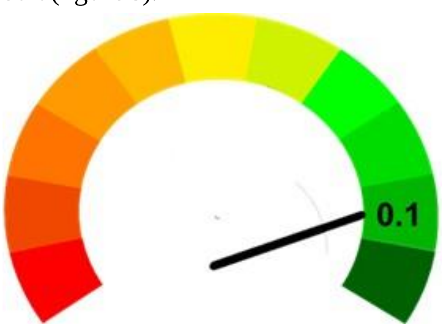

Fig. 5. Scale indicates a good value

B.2 Repair and Maintenance Enterprises.

These enterprises provide technical support or sell spare parts, like mechanical and computer workshops or appliance stores. These businesses can use this application based on the service it provides: for the sale of spare parts, it helps to monitor the demand to predict the restocking time; while for repairs, it allows to control the service times and identification of irregularities in them. The macro process of these enterprises is shown in figure 6. This study was applied in a mechanical maintenance workshop.

Sales Services.

Focusing on the sales service, we identified that the products with the peak demand are the filters, which is why this was the subject of the study. For this, we consider two

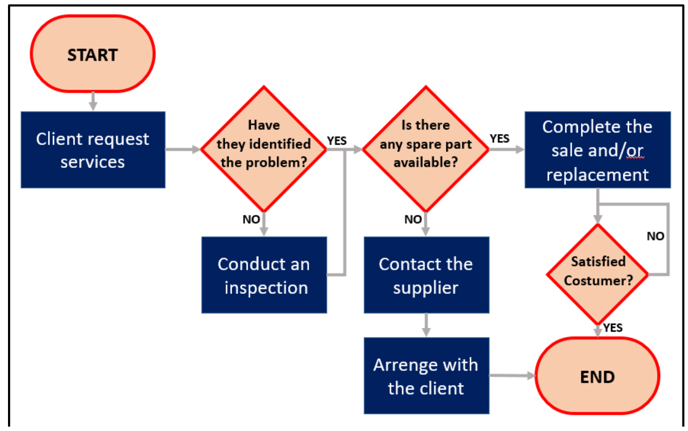

Fig. 6. Repair/Sale process flowchart

variables: the type of filter (oil, gasoline, air) and its subtypes. Then four attributes were assigned: the first one (X1) is the filter type, the second and third (X2 y X3) are the subtype filters, and the fourth one (X4) is the quantity of each sold on the day, as shown in the table 2.

Table 2. Weight Assignation

X1: Filter type	Oil	Gasoline	Air							
Weight	0.1	0.2	0.3							
X2: Filter subtype	**1**	**2**	**3**	**4**	**5**	**6**				
Weight	0.1	0.2	0.3	0.4	0.5	0.6				
X3: Filter subtype	**7**	**8**	**9**	**10**	**11**	**12**				
Weight	0.1	0.2	0.3	0.4	0.5	0.6				
X4: Filters sold	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Weight	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1

The subtype filters are divided into two attributes to enhance the weight distribution. Since both represent the same variable, they are mutually exclusive and shouldn't be used simultaneously. Also, the number of subtypes corresponds to the maximum within the three types of filters.

Before using this application, it is necessary to order and classify the inventory of the items, so it is suggested to use the first 3 of the 5's. It is proposed to label each filter as follows: FXY, where F stands for Filter; X, for the filter type; and Y, for the filter subtype. The inventory for each product is shown in table 3, because the app should notify when the security stock level is reached to restock.

Table 3. Filter Inventory Level

Oil						
Subtype filters	**1**	**2**	**3**	**4**	**5**	**6**
Stock	30	40	45	80	120	60
Subtype filters	**7**	**8**				
Stock	50	20				
Gasoline						
Subtype filters	**1**	**2**	**3**	**4**	**5**	**6**
Stock	20	80	70	40	100	30
Subtype filters	**7**	**8**	**9**	**10**	**11**	**12**
Stock	25	70	80	15	70	30
Air						
Subtype filters	**1**	**2**	**3**	**4**	**5**	**6**
Stock	60	80	45	40	45	20

Then, all the permutations of the attributes are submitted in the application, and the AR cards are labeled as shown in table 2 (Figure 3). The stock indicator is calculated according to the next formula:

$$\text{Available stock} = 1 - \frac{\text{Quantity of sold Items}}{\text{Stock items}}$$

Fig. 7. Indicator learning process and labeled AR cards

Although this indicator is functional, this only keeps track of the sales in a day, which wouldn't help to notify when the security stock level is reached. Therefore, the app also stores the accumulated sales, as shown in figure 8.

Fig. 8. Daily and Accumulated Indicator

IV. Results

Since both businesses have similar profits, we would consider the results as the same. For the MYPE model to be evaluated a total of 10 workers was considered, this is the maximum number of people for microenterprises and the minimum amount for small ones. It was also assumed that a minimum of two users will acquire the application because people in Peru are not used to buying apps.

They will have 4 sets of the AR Cards and the 10 workers will be trained to be White Belts in Lean Six Sigma, that is, they will be aware of the basic tools for process improvement such as the 5's, Poka Yoke, standardization and other methodologies. The cost of the training was estimated from the Global Lean course.

The costs of the implementation materials were divided among the most used tools in improvement projects, the first is 5's that include costs of classification cards, communication boards, signage materials, shelves, bins, trays and stationery.

The second is Poka Yoke, that contain sensors (movement and / or temperature), alarms and tools manufactured and/or purchased. The other implementation costs are standardization (process training and templates to standardize jobs). It should be noted that the training of the application will be free through YouTube tutorials.

For the economic savings generated with process improvements, they will be reflected in the decrease of the time the worker spends. There are other types of savings such as losses and inventories; but time will be analyzed because it is the most direct impact factor. For example: defects and waste lead to a rework and that is a waste of time for the worker. Three different scenarios were examined on the possibilities of the impact of this factor. The saving time for an 8-hour workday will be approximately 20 (4%), 30 (6%) and 40 (8%) minutes for a day from the pessimistic scenario to the optimist respectively.

Table 4. Implementation costs

Investment	Costs (S/.)
App	23.60
AR Card	240.00
White Belt Training	1,921.04
Implementation materials - 5`S	2,615.00
Implementation materials - Poka Yoke	2,150.00
Implementation materials - Others	3,615.00
Total	10,564.64

***Referential exchange rate 1 dollar = S/. 3.85**

The cost per hour of the minimum wage in Peru (S/ .850 at the time) was used as a reference and the assumption of the 10 workers is maintained. The cost per hour was 5.06 based on 21 monthly days and an 8-hour day. The saving per man hours of each scenario will be indicated in table 5.

In order to accept the project in all scenarios, it is required to have a Discount Rate lower than the IRR. Both the normal and the optimistic scenario, the COK turned out to be lower than the IRR as well as the WACC. On the other hand, in the pessimistic scenario only the COK was lower than the Internal Rate of Return.

Table 5. Cost scenarios

Stage	Optimistic	Normal	Pessimistic
Daily day (h)	8	8	8
Proposed time (h)	7.36	7.52	7.68
Saving hour (h)	0.64	0.48	0.32
S/./h (workers)	5.06	5.06	5.06
Workers	10	10	10
Saving	32.38	24.29	16.19

Monthly savings (S/.)	680	510	340
Annual savings (S/.)	8,160	6,120	4,080

With a discount rate of 15.24% (COK) a NPV of S/.14,611, S/.7,811 and S/.1,011 was reached for the optimistic, normal and pessimistic scenarios respectively. In the same way, with a rate of 24.56% (WACC) a VAN of S/.8,041, S/.2,505 and -S/.3,031 respectively was achieved.

Table 6. COK and WACC

	COK	WACC
IRR Optimistic	0.658	0.718
IRR Normal	0.437	0.402
IRR Pessimistic	0.192	0.029

V. CONCLUSIONS

The economic evaluation of the project resulted in an estimated net present value (NPV) for the normal scenario of S/. 7,811 and when the COK and WACC discount rate are applied the NPV becomes S/.2,505. The project guarantees estimated annual savings of S/. 6,120 (normal scenario); unlike a Lean Six Sigma improvement project, control methods can be automatize using indicator-based management in mobile apps. Any improved process without automatization is only manuals and words. This is the reason why the application will have the objective of preserving continuous improvement throughout the years with the help of AR cards.

Artificial Neural Networks are still an experimental model and because of this, there is no defined architecture for their design. However, through trial and error, a model with two hidden layers was created which is capable of simulating and learning any problem related to the twelve proposed indicators. Besides, Augmented Reality helped by simplifying new data gathering through AR card scanning (similar to a bar code reader) and the visualization of the obtained results through color-coded gauges.

REFERENCES

[1] United Nations. MSMEs: Key to an inclusive and sustainable recovery. https://www.un.org/en/observances/micro-small-medium-businesses-day

[2] The World Bank. Small and Medium Enterprises (SMEs) Finance. The World Bank Group, 1818 H Street NW, Washington, DC 20433. https://www.worldbank.org/en/topic/smefinance

[3] International Finance Corporation. Micro, Small and Medium Enterprises - Economic Indicators (MSME-EI) Analysis Note December 2019. International Finance Corporation, The World Bank and SME Finance Forum. https://www.smefinanceforum.org/data-sites/msme-country-indicators (2019)

[4] Superintendencia Nacional Tributaria - SUNAT (2008). MYPES: Características de la Micro y Pequeñas Empresas. http://www.sunat.gob.pe/orientacion/mypes/caracteristicas-microPequenaEmpresa.html

[5] ComexPerú. Situación de las MYPE en 2017: muchos retos en el camino. Semanario 948 - Economía. Published on august 03, 2018. Last accessed 06/22/2018. https://www.comexperu.org.pe/articulo/situacion-de-las-mype-en-2017-muchos-retos-en-el-camino (2018).

[6] Ministerio de la Producción. Estadística MIPYME: Micro, Pequeña y Mediana Empresa. Last accessed 09/12/2018. http://ogeiee.produce.gob.pe/index.php/shortcode/estadistica-oee/estadisticas-mipyme (2018).

[7] Alva, E. La desaparición de las microempresas en el Perú: Una aproximación a los factores que predisponen a su mortalidad - Caso del Cercado de Lima. Economía y Desarrollo, vol. 158, núm. 2, julio-diciembre, 2017, pp. 76-90. Universidad de La Habana. https://www.redalyc.org/pdf/4255/425554493005.pdf (2017).

[8] Tenera, A., & Pinto, L. C. A Lean Six Sigma (LSS) Project Management Improvement Model. Procedia - Social and Behavioral Sciences, 119, 912–920. https://doi.org/10.1016/j.sbspro.2014.03.102 (2014).

[9] Han, J., Kamber, M. & Pei, J. Data Mining: Concepts and Techniques. Elsevier Inc. All rights reserved (2012).

[10]Snee, R. Lean Six Sigma–getting better all the time. International Journal of Lean Six Sigma (2010).

[11]Kumar, M., Antony, J., & Douglas, A. Does size matter for Six Sigma implementation? Findings from the survey in UK SMEs. The TQM journal, 21(6), 623-635 (2009).

[12]Deshmukh, S. & Chavan, A. Six Sigma and SMEs: a critical review of literature. International Journal of Lean Six Sigma, 3(2), 157-167 (2012).

[13]Felizzola Jiménez, H., & Luna Amaya, C. Lean Six Sigma en pequeñas y medianas empresas: un enfoque metodológico. Ingeniare. Revista chilena de ingeniería, 22(2), 263-277 (2014).

[14]Vadim, K. Overview of different approaches to solving problems of Data Mining (2018).

[15]Soroka, A., Liu, Y., Han, L., & Haleem, M. S. Big data driven customer insights for SMEs in redistributed manufacturing. Procedia CIRP, 63, 692-697 (2017).

[16]Packianather, M. S., Davies, A., Harraden, S., Soman, S., & White, J. Data mining techniques applied to a manufacturing SME. Procedia CIRP, 62, 123-128 (2017).

[17]Mourtzis D, Vlachou E, Milas N, Industrial Big Data as a Result of IoT Adoption in Manufacturing. Procedia CIRP 2016;55: 290-295.

[18]Van Krevelen, D. W. F., & Poelman, R. A survey of augmented reality technologies, applications and limitations. International journal of virtual reality, 9(2), 1 (2010).

[19]S. Benford, C. Greenhalgh, G. Reynard, C. Brown, and B. Koleva. Understanding and constructing shared spaces with mixed-reality boundaries. ACM Trans. on Computer-Human Interaction, 5(3):185– 223, Sep. 1998

[20]P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. IEICE Trans. on Information and Systems, E77-D (12):1321–1329, Dec. 1994

[21]Mekni, M., & Lemieux, A. Augmented reality: Applications, challenges and future trends. Applied Computational Science, 205-214 (2014).