

# CRISP-DM/SMEs: A Data Analytics Methodology for Non-profit SMEs



Jhon Montalvo-Garcia , Juan Bernardo Quintero   
and Bell Manrique-Losada 

**Abstract** The exponential increase in information due to technological advances and the development of communications has created the need to make decisions based on the data analysis. This trend has opened the doors to new approaches to data understanding and decision-making. On the one hand, companies need to follow data analytic methodologies to manage large volumes of information with big data tools. On the other hand, there are non-profit small and medium-sized enterprises (SMEs) that make efforts to address data analytics according to their different sources and types. They find challenges such as lack of knowledge in methodological and software tools, which allow timely deployment for decision-making. In this paper, we propose a data analytics methodology for non-profit SMEs. The design of this methodology is based on CRISP-DM as a reference framework, is represented by Software Process Engineering Metamodel (SPEM) and is characterized by being simple, flexible, and low implementation costs.

**Keywords** Data analytics · CRISP-DM · Non-profit SMEs

## 1 Introduction

The information processing according to Palmer and Hartley [1] has gradually become the basis for achieving a competitive advantage and, therefore, organizations have to believe that they have the right information at the right time and for the right people. The company's managers should be provided with the appropriate

---

J. Montalvo-Garcia (✉)

Seventh-day Adventist Church of Colombia, Cra 84 #33aa-169, Medellin, Colombia  
e-mail: [jhonmontalvo@gmail.com](mailto:jhonmontalvo@gmail.com)

J. B. Quintero · B. Manrique-Losada

Universidad de Medellin, Cra 87 #30-65, Medellin, Colombia  
e-mail: [juquintero@udem.edu.co](mailto:juquintero@udem.edu.co)

B. Manrique-Losada

e-mail: [bmanrique@udem.edu.co](mailto:bmanrique@udem.edu.co)

tools for the exploitation and data analysis that will allow them to obtain the necessary knowledge in the strategic decision-making process [2]. This is how in the last decade, data warehouses (DW) have become an essential component to achieve competitiveness with modern decision support systems in most companies that handle large volumes of data.

However, these companies only represent a minimal part of the business world. Small and medium-sized enterprises (SMEs) are the dominant form of business organization in all countries of the world, representing more than 95% and up to 99% of the business population [3]. SMEs are considered important at the local, national, and global levels, playing an important role in the national economy [4], and in the social sector; as non-profit companies, who pursue social and community purposes. Although they do not generate profit distribution or enrichment to the partners, strategic and financial decisions involving data analysis are made.

Therefore, non-profit SMEs collect information from different sources and are interested in business intelligence systems [5] and in the trend toward data analytics (DA) that is increasing by technological advance. Despite this interest, the development of data analytics projects is frustrated, since its implementation is usually a complex task due to the generated costs. These costs are associated with technological infrastructure, administrative costs, personnel training, and software tools [6]. Thus, the implementation of analytical projects in non-profit SMEs has few alternatives, since the DA focuses on large companies that have the greater financial capacity [7] and high-volume data management.

To reduce the gap between non-profit SMEs and DA projects, an analytical methodology is proposed according to the needs of these organizations. As a consequence, this paper is organized into four sections, starting with section two, which describes the background of non-profit organizations and the most recognized data analytics methodologies in the industry. Section three presents a simplified CRISP-DM proposal for non-profit SMEs. Finally, in section four, conclusions and future work are presented.

## **2 Background**

### ***2.1 Non-profit SMEs***

Non-profit companies are legal and social entities created to produce goods and services, whose legal status does not allow them to be a source of income for the units that establish, control or finance them [8]. Therefore, they can benefit to associates, third parties or the general public from social projects. Non-profit has taken a great importance in the world, not only as organizations that provide social services but as generators of employment and promoters of economic activity. Because of this, they are grouped at an international level according to their social purpose [8] and

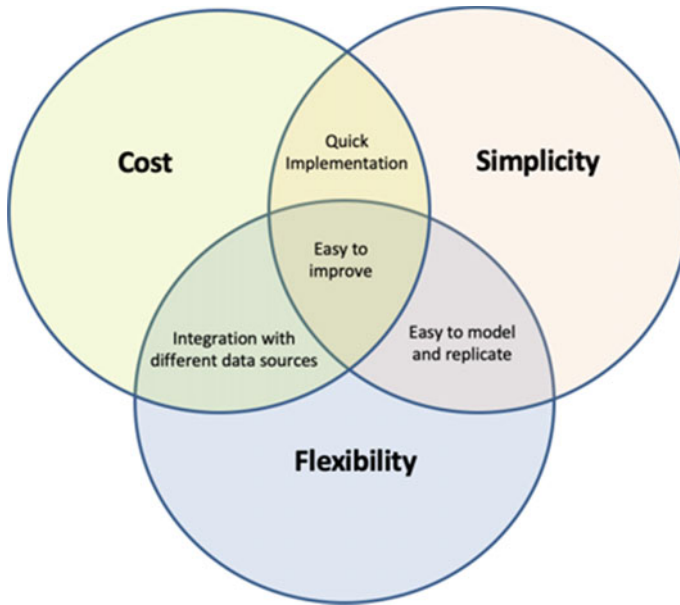
are inspected, monitored or controlled by different state entities such as mayoralities, governorships, and government ministries [9].

Non-profit companies have their main source of income from funds received by natural persons, legal entities or public entities representing countries, through donations or subventions. Regarding the organizational structure, non-profit companies have a high degree of complexity, due to the wide range of organizational possibilities (central, divisional, functional or geographical). Therefore, non-profit companies are recognized within SMEs as foundations, universities, corporations, associations, cooperatives, churches, and among others.

## ***2.2 Data Analytics in Non-profit SMEs***

The ability of SMEs to succeed in the face of larger competitors is centered on personal intuition and the ability to provide superior service. Since big data is changing the business landscape, some big competitors are using big data to improve product quality, marketing operations, and customer relationships. This new efficiency of the largest competitors can be a real threat to the sustainability of the SME business [10], especially for non-profit companies. However, there are other concerns in the SMEs that stand in the way of the growth of the DA and they are the new regulations imposed by the government every year. A clear example of this is the implementation of the International Financial Reporting Standards (IFRS), data protection policies, electronic invoicing, and strict social security regulations, which have increased in a large percentage from 2016 to 2017, as is the case of Colombia [11]. This has put aside the advance in digital transformation, a key aspect in the career of data analytics.

Despite these concerns, an EMC study reveals that 58% of Colombian organizations have current plans to implement big data technologies. The other 42% say that the lack of interest corresponds to the fact that the business culture is not ready yet (46%); but also, because it is very expensive to implement it with respect to the current economic situation (28%), and there is a lack of understanding regarding this trend (25%) [12]. Thus, to minimize the complexity, costs and lack of staff training, non-profit SMEs require that a DA project can be deployed quickly and that it can be easy to model and replicate to other areas of the organization. It also demands that the developed models can be easy to improve in front of any external change that affects the organization and being flexible for the integration of different data sources. Figure 1 shows the practices that non-profit SMEs need when implementing a DA project.



**Fig. 1** Data analytics practices in non-profit SMEs

### ***2.3 Data Analytics Methodologies***

The most referenced models found in the scientific community and proposed for the development of DA projects are as follows: knowledge discovery in databases (KDD), sample, explore, modify, model, assess (SEMMA), and cross-industry standard process for data mining (CRISP-DM). CRISP-DM is the most used in recent years [13]. At the beginning of 1996, the KDD model became the first accepted model in the scientific community that established the main stages of an information exploitation project. Then, from the year 2000, with the great growth that emerged in the area of data mining, two new models were developed that propose a systematic approach to carry out the process: SEMMA and CRISP-DM.

### ***2.4 Reference Model for Data Analytical Project in Non-profit SMEs***

CRISP-DM is considered the standard and most referenced methodology to develop data mining and knowledge discovery projects [14]; it is flexible and can be easily adapted to each analytical task in terms of DM processes [15]. CRISP-DM goes into greater detail about the tasks and activities to be carried out in each phase of the data

mining process, while KDD and SEMMA provide only a general guide of the work by each phase.

For this proposal of data analytics for non-profit SMEs, CRISP-DM was taken as reference model for the following reasons: (i) it is the model most referenced by its wide acceptance; (ii) all its phases and activities are properly organized, structured and defined; and finally, (iii) it facilitates the understanding and revision of a project. An adaptation of CRISP-DM was made, excluding some tasks to be simple to implement, improve and replicate; as also flexible to adapt and pay for the non-profit organization, in order to reduce the effort of personnel, time, and costs in the development of DA projects.







### 3 Methodological Proposal

CRISP-DM/SME’s methodology is proposed as a result of the analysis of data analytics methodologies and the needs of non-profit SMEs. The methodology includes roles of data science, descriptions by phase, activities required in each phase, work products, guidance, and tools. To represent the methodology, a diagram built in SPEM is used since it is a standard language for the modeling of software development processes oriented to work products. The methodological proposal uses the notation represented in Fig. 2 [16].

#### 3.1 Graphic Representation

The methodological proposal is represented in the SPEM diagram, in Fig. 3. The diagram shows that the sequence of the phases is not strict and can interact between

Fig. 2 SPEM elements used

Stereotype	SPEM 2.0
Role Use	
Phase	
Activity	
Work Product Use	
Guide	
Tool Definition	

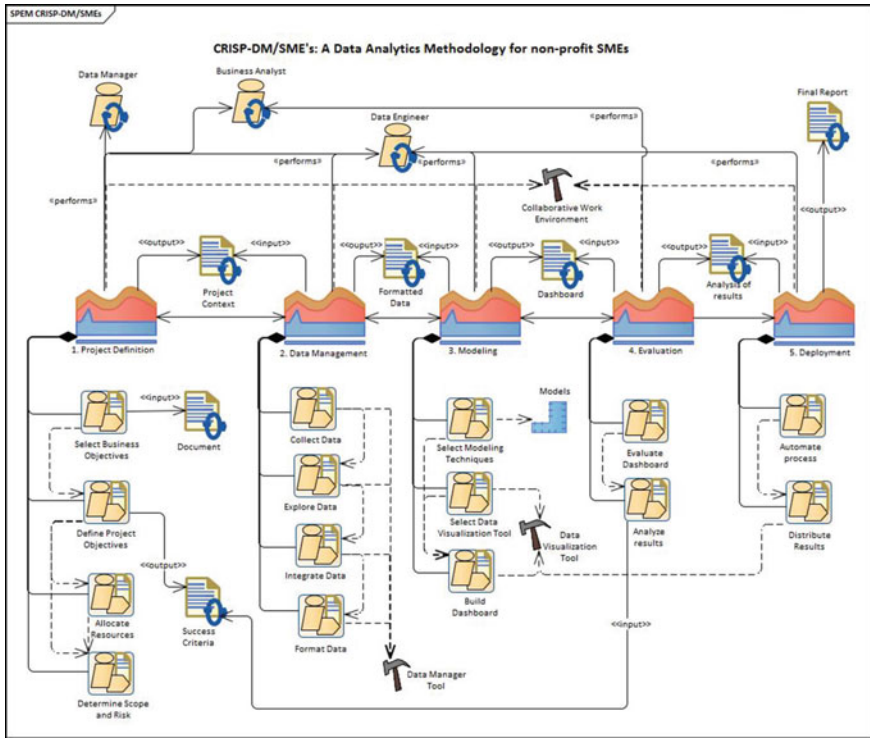


Fig. 3 CRISP-DM/SMEs

each phase. Therefore, project execution can move forward or backward if necessary. The phases described in the diagram contain a set of tasks necessary to guarantee the quality of the project and produce the work products. In addition, technological tools, guides, and models can be used to develop activities. Activities are executed through one or several roles that can be internal or external to the organization.

### 3.2 Roles

CRISP-DM involves three roles in the data science industry [17], whose profiles are necessary for the development of tasks. Data engineer is the main role within DA projects in the non-profit SMEs. Data manager is responsible of leading the DA team and providing the financial, human, and software resources for the project. Business analyst is responsible for improving business processes and is the intermediary between the data engineer and data manager.

### 3.3 Phases

**Phase 1: Project Definition.** Select business goals that will impact the data analytics project. This information is obtained from the strategic plan of the non-profit SME or another similar organizational document. The objectives of the data analytics project are defined and aligned with the business. As a result of this task, the success criteria of the DA project are obtained. Human and financial resources are estimated and allocated for the realization of the project. And finally, the scope and risks are determined.

**Phase 2: Data Management.** The data is collected from the different sources to be explored. Then, it is integrated with the appropriate format according to the project. The data engineer uses a tool to manage the data and the phase will obtain the formatted data.

**Phase 3: Modeling.** The selection of the model is closely related to the visualization tool. The tool performs the analysis of the data applying the models chosen for the DA project. The dashboards are built according to the requirements defined in the first phase and contemplated in the success criteria. Samples are taken with a group of small data before proceeding with the evaluation phase, in order to validate the expected results.

**Phase 4: Evaluation.** The dashboard must be submitted to the stakeholders of the project for their respective assessment and acceptance. The results obtained with the model and the dashboard are evaluated according to the success criteria and the selected business objectives. If the dashboard does not meet the success criteria, then the business analyst will decide if the DA project should go back to an earlier phase to improve the model or the dashboard.

**Phase 5: Deployment.** This phase consists of automating the data source for the dashboard. A solution is built to integrate, update, and format the data for use in the model and the dashboard. Then, the dashboard is shared with the stakeholders of the project, through protected applications or links.

### 3.4 Work Product

A work product is the result of each phase of the DA project, which can be used in another phase. The project context is the main document for the other phases of the project, which must contain the business objectives related to the project, the specific objectives of the DA project, the resources allocated, and the scope and possible risks. The criteria for success are also elaborated as a work product. At the end of all tasks in the data management phase, a work product called formatted data is obtained and will be the source of origin for the modeling phase. This work product will have identified the size, fields, attributes, errors, and data types. You will also have the list of the different sources of data for the project. The dashboard will have graphics of descriptive or predictive analysis and it will be the work product that is shared

with all the stakeholders of the DA project in the deployment phase. The analysis of results is the work product that defines if it is necessary to redefine the project, the success criteria or the objectives and if it is necessary to return to an earlier phase or continue with the deployment. Finally, a final report is generated, documenting the effort made in the DA project with the implementation of the CRISP-DM/SME's methodology in order to contribute to the scientific and business community through success cases.

### **3.5 Tools and Guidance**

A collaborative work environment tool is used by the team involved in the development of the project so that each one can control the assigned tasks. On the other hand, the data engineer uses the necessary tools to manage the data and build the dashboard. Each tool is defined in the project context. Additionally, guidance is used to select the models to follow for the data representation.

## **4 Conclusions and Future Work**

The proposed methodology seeks to reduce the effort of implementing data analytics projects in non-profit SMEs, to improve the collection, storage, processing, and analysis of data. It induces the creation of awareness in decision-making based on the exploration of information. Having an analytical methodology allows to minimize complexity, costs, and helps to have trained personnel for the implementation of DA projects. It also facilitates rapid deployment, improvement, and integration into other projects.

The roles integration in the methodology helps assign responsibilities and tasks to each of the stakeholders who participate in the DA project, involving the skills of professionals in the data science industry. Additionally, the use of a dashboard is a decisive work product for data analysis since a dashboard graphically represents the result of the project. Finally, the usage of SPEM language for the representation of the methodological proposal allows an abstract description of the fundamental elements of the process of data analytics and also the description of how they are related to each other.

As future work, we propose the realization of a Web platform that allows the repository of each of the data analytics projects of non-profit SMEs, with the respective phases, tasks assigned to roles, work product, and guides on how to perform each task. This platform should be within the reach of the non-profit SMEs and should contribute to the reduction of effort in the process of development of the DA project.



## References

1. A. Palmer, B. Hartley, *The Business and Marketing Environment*, 3rd edn. (McGraw-Hill, London, 2000)
2. E. Soto, La información como recurso estratégico generador de conocimientos, *Soportes audiovisuales e Informaticos* (Universidad de la Laguna, Spain, 2004), pp. 58–60
3. P. Pytel, A. Hossian, P. Britos, R. Garcia-Martinez, Feasibility and effort estimation models for medium and small size information mining projects. *Inf. Syst.* **47**, 1–14 (2015). <https://doi.org/10.1016/j.is.2014.06.004>. (Elsevier, Argentina)
4. R. Mullins, Y. Duan, D. Hamblin, P. Burrell, H. Jin, Z. Ewa, B. Aleksander, A web based intelligent training system for SMEs. *Electron. J. e-Learn.* **5**(1), 39–48 (2007). (ERIC, Germany, Poland, Portugal, Slovakia, United Kingdom)
5. G. Lawton, Users take a close look at visual analytics. *Computer* **42**(2), 19–22 (2009). <https://doi.org/10.1109/mc.2009.61>. (California)
6. H. Florez, Inteligencia de negocios como apoyo a la toma de decisiones en la gerencia. *Rev. Vincul.* **9**(2), 11–23 (2012). (Bogota)
7. T. Guarda, M. Santos, F. Pinto, M. Augusto, C. Silva, Business intelligence as a competitive advantage for SMEs. *Int. J. Trade Econ. Financ.* **4**(4), 187–190 (2013). <https://doi.org/10.7763/ijtef.2013.v4.283>. (Portugal)
8. W. Franco, D. Samiento, G. Serrano, G. Suarez, *Entidades sin animo de lucro* (Consejo Tecnico de la Contaduria Publica, 2015). <http://www.ctcp.gov.co>
9. Conferencia Colombiana de ONG, *Quiénes conforman el sector de las Entidades Sin Ánimo de Lucro ESAL en Colombia* (2016). <http://ccong.org.co/>
10. B. Ogbuokiri, C. Udanor, M. Agu, Implementing bigdata analytics for small and medium enterprise (SME) regional growth. *IOSR J. Comput. Eng. Ver. IV* **17**(6), 2278–2661 (2015). <https://doi.org/10.9790/0661-17643543>. (Nigeria)
11. Sinnetic, *PYMES se desaceleran en transformación digital e innovación por responder a múltiples requerimientos estatales*, vol. 18 (Sinnetic, Colombia, 2017), pp. 1–3
12. A. Gonzalez, *Big data y analítica en Colombia: A un paso de despegar* (2014). <https://searchdatacenter.techtarget.com>
13. KDnuggets, *What Main Methodology are you Using for Your Analytics, Data Mining, or Data Science Projects?* (KDnuggets, 2014) <https://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html>
14. H. Achmad, V. Sabur, A. Pritasari, H. Reinaldo, Data mining and sharing to create usable knowledge, implementation in small business in Indonesia. *Sains Humanika* **2**, 69–75 (2016). (Indonesia)
15. M. Dittert, R. Härtling, C. Reichstein, C. Bayer, *A Data Analytics Framework for Business in Small and Medium-Sized Organizations*, vol. 73 (Springer, Germany, 2018), pp. 1–13. <https://doi.org/10.1007/978-3-319-59424-8>
16. V. Menéndez, M. Castellanos, Software process engineering metamodel (SPEM). *Rev. Latinoam. Ing. Softw.* **3**(2), 92–100 (2008). <https://doi.org/10.18294/relais.2015.92-100>
17. DataCamp, *The Data Science Industry: Who Does What* (2018). <https://www.datacamp.com/community/tutorials/data-science-industry-infographic>