Autonomous cycles of data analysis tasks for innovation processes in MSMEs

Ana Gutiérrez
GIA, Universidad Francisco de Paula Santander, Cúcuta, Colombia;
GIDITICS, Universidad EAFIT, Medellín, Colombia and
Tepuy R+D Group, Artificial Intelligence Software Development, Mérida, Venezuela

Jose Aguilar
GIDITICS, Universidad EAFIT, Medellín, Colombia;
CEMISID, Universidad de Los Andes, Núcleo La Hechicera, Mérida, Venezuela;
Dpto. de Automática, Universidad de Alcalá, Alcalá de Henares, Spain and
Tepuy R+D Group, Artificial Intelligence Software Development, Mérida, Venezuela

Ana Ortega
GIDITICS, Universidad EAFIT, Medellín, Colombia, and

Edwin Montoya
GIDITICS, Universidad EAFIT, Medellín, Colombia and
Tepuy R+D Group, Artificial Intelligence Software Development, Mérida, Venezuela

Abstract

Purpose – The authors propose the concept of "Autonomic Cycle for innovation processes," which defines a set of tasks of data analysis, whose objective is to improve the innovation process in micro-, small and medium-sized enterprises (MSMEs).

Design/methodology/approach – The authors design autonomic cycles where each data analysis task interacts with each other and has different roles: some of them must observe the innovation process, others must analyze and interpret what happens in it, and finally, others make decisions in order to improve the innovation process.

Findings – In this article, the authors identify three innovation sub-processes which can be applied to autonomic cycles, which allow interoperating the actors of innovation processes (data, people, things and services). These autonomic cycles define an innovation problem, specify innovation requirements, and finally, evaluate the results of the innovation process, respectively. Finally, the authors instance/apply the autonomic cycle of data analysis tasks to determine the innovation problem in the textile industry.

Research limitations/implications – It is necessary to implement all autonomous cycles of data analysis tasks (ACODATs) in a real scenario to verify their functionalities. Also, it is important to determine the most important knowledge models required in the ACODAT for the definition of the innovation problem. Once determined this, it is necessary to define the relevant everything mining techniques required for their implementations, such as service and process mining tasks.

Practical implications – ACODAT for the definition of the innovation problem is essential in a process innovation because it allows the organization to identify opportunities for improvement.

Originality/value – The main contributions of this work are: For an innovation process is specified its ACODATs in order to manage it. A multidimensional data model for the management of an innovation process is defined, which stores the required information of the organization and of the context. The ACODAT for the definition of the innovation problem is detailed and instanced in the textile industry. The Artificial Intelligence (AI) techniques required for the ACODAT for the innovation problem definition are specified, in order to obtain the knowledge models (prediction and diagnosis) for the management of the innovation process for MSMEs of the textile industry.

1. Introduction

Micro-, small- and medium-sized enterprises (MSMEs) have limited resources, and thus, they must search for efficient ways to do more with less [1, 2], especially in the quarantine economy [3, 4] in light of coronavirus disease 2019 (COVID-19) [5, 6]. Particularly, MSMEs need to innovate and improve their offer of goods, products and services, to respond to the changing needs of the market. Innovation has become the means that allows an MSME to transform and continue to grow to stay in the market, taking advantage of each of the resources available in the organization, human, technological and financial. Several studies have concluded that investment in innovation and technology has an impact on the development of organizations to be more competitive, which leads many times to the introduction of new products and processes [7, 8]. In turn, the return on investment will be reflected in productivity indicators, in good operation and profitability of the organization.

On the other hand, information is becoming more relevant every day for companies to make decisions. Organizations not only need to collect data but also find the right way to analyze it to devise daily actions based on statistics and trends. However, companies currently lack the capacity to use big data and data analytics [9]. Therefore, companies must start using all available data sources, and be able to make the most of data to support decision-making in their organizations. Especially, it is necessary to understand and analyze the different sources of information that will improve the innovation processes with the use of data analytics tasks, to respond to the different phases of them.

Given the importance of the innovation in MSMEs, and the current opportunities that exist to exploit data from the organizations and their contexts can be defined strategies based on data to build data-driven models to guide the innovation processes. One of these strategies is the utilization of the concept of "autonomous cycles of data analysis tasks" (ACODATs) defined in previous works [10–12], which allow generating knowledge models useful for the management of the innovation processes using different data sources. An ACODAT is composed by a set of data analysis tasks to reach a goal for a given problem, where each task has a given role [13–15]: observe the studied system, analyze it and make decisions to improve it. In this way, there are interactions and synergies between the data analysis tasks, to generate the required knowledge with the goal of improving the process under study.

In this paper, we propose several ACODATs for the management of the innovation processes in an MSME. Likewise, in the paper is proposed the specification in detail of the autonomic cycle for the innovation problem definition sub-process, and its application in the textile industry. For the development of the ACODATs, the MetodologIa para el Desarrollo de Aplicaciones de Minería de Datos basados en el aNálisis Organizacional (MIDANO) [16–18] methodology was used, which allows the development of data analytics applications, and especially, the development of ACODATs. The main contributions of this work are:

(1) The specification of ACODATs for the management of innovation processes.

(2) The definition of a multidimensional data model, which stores the required information of the organization and the context for the ACODATs.

(3) The detailed description of the ACODAT for the definition of the innovation problem, which is instanced in the textile industry.

(4) The characterization of the AI techniques required for the ACODAT for the innovation problem definition, in order to obtain the knowledge models (prediction and diagnosis) for the management of the innovation processes for MSMEs of the textile industry.

This work is organized as follows. Section 2 presents the related works. In Section 3, the theoretical framework is presented, specifically, ACODAT, MIDANO and the innovation model used in this work. Section 4 introduces the autonomic cycles proposed, the description of their tasks and their multidimensional data model, using the MIDANO methodology. Section 6 details the case study of the textile industry, and the application of the autonomic cycle for the definition of innovation problems. Finally, the conclusions and future works are presented.

2. Related works
In this section, we present the main recent papers related to our approach, which are the definition of schemes for the automation of innovation processes or the utilization of autonomic cycles in the automation of industrial processes (Industry 4.0).

Ossi *et al.* [19] presented a conceptual framework based on big data and business models to exploit the innovation capabilities. The framework adopted the business canvas model. This framework helps in concentrating on different viewpoints, for example, can create and develop strategies of price based on analytics data. The framework offers ways to organize perspectives for organizational transformation. On the other hand, machine learning (ML) models offer the computational power and functional flexibility required to decipher complex patterns in a high-dimensional data environment [20]. Particularly, in [20] three groups of financial data analysis are identified: (1) portfolio management; (2) financial fraud and distress; and (3) sentiment inference, forecasting and planning.

Kritsadee *et al.* [21] tested a model of factors affecting the innovativeness of small and medium enterprises (SMEs) using the structural equation model (SEM). Data about innovativeness were collected using questionnaires, which were mailed to 283 entrepreneurs. The proposed model determined that learning orientation and proactiveness had direct effects on innovativeness. The analysis addressed the innovation in products, processes, organizational and marketing, and their contribution to the organization's results (e.g. market share, environmental sustainability, profit, etc.). The paper [22] investigated the parameters in the innovation process design that influence the innovation outcomes in the context of smart manufacturing (Industry 4.0), and thus what should be accounted for in the design of innovation processes for smart manufacturing. The research is based on empirical evidence from 18 manufacturing companies and suppliers of manufacturing technology. Finally, the authors of [23] present a systematic literature review about how smart systems have been used to improve the innovation capacities in MSMEs. The results show that there is not an established body of knowledge about how to improve the innovation process based on smart systems.

Sanchez *et al.* [15] defined three autonomic cycles that allow interoperating the actors of manufacturing processes (data, people, things and services). Particularly, they defined a framework for the integration of autonomous processes based on cooperation, collaboration and coordination mechanisms. The framework is composed of three ACODATs that allow the self-configuration, self-optimization and self-healing of the manufacturing process. They implement one of these ACODATs, for the self-supervision of the coordination process mixing it with the theory of multi-agent systems [24]. This ACODAT is implemented and tested using an experimental tool that replays a production process event log, to detect failures and invoke the ACODAT for self-healing when needed. Qin *et al.* [25] proposed a multi-layered framework of manufacturing for Industry 4.0. One of the levels, the intelligence

layer, applies different data analytic tasks to discover useful information from data to improve the manufacturing process. Thus, the intelligence layer creates a knowledge base that serves as a support for the planning and decision-making processes.

Besides, the paper [26] reveals that knowledge management for sustainability research has relied on nine foundational clusters (i.e. informed sustainability practice, social network, firm performance, knowledge sharing culture, green innovation, sustainability assessment framework, global warming, knowledge management and innovative performance) to generate new knowledge. Also, they determine that the method of creating, communicating, disseminating and exploiting shared knowledge is instrumental for firms adopting business practices to enhance firm performance.

The previous studies do not define frameworks and systems for the management of the innovation processes for MSMEs based on the ACODAT concept, neither do they clarify the application of data analytic to improve the innovation capabilities in an organization. These are the main differences in our approach with respect to previous works. On the other hand, the ideas proposed in this work could be used in other areas of an organization, including environmental social governance (ESG) and total quality management (TQM) [27].

3. Conceptual framework

3.1 ACODAT

This research follows the ACODAT concept, which is based on the idea proposed by IBM in 2001 [28]. The ACODAT concept was proposed in [10–12, 29] and has been used in telecommunication [30], education, especially in smart classrooms [11, 12], Industry 4.0 [13–15] and smart cities [31], among other domains. It is based on the autonomic computing paradigm [32], with the purpose of endowing autonomic properties to systems based on a smart control loop.

The main objective of an ACODAT is to extract useful knowledge from data to make decisions [11, 12]. The set of data analysis tasks must be performed together, in order to achieve the objective in the process supervised. The tasks interact with each other and have different roles in the cycle, which are: observing the process, analyzing and interpreting what happens in it and making decisions to reach the objective for which the cycle was designed. This integration of tasks in a closed loop allows solving complex problems. The detailed description of the roles of each task is [11, 12]:

Monitoring: Tasks to observe the supervised system. They must capture data and information about the behavior of the system. Besides, they are responsible for the preparation of the data for the next step (preprocessing, selection of the relevant features, etc.).

Analysis: Tasks to interpret, understand and diagnose what is happening in the monitored system. These tasks allow building knowledge models about the dynamics observed, in order to know what is happening in the system.

Decision-making: Tasks to define and implement the necessary actions based on the previous analyses, in order to improve the supervised system. These tasks impact the dynamics of the system, and their effects are again evaluated in the monitoring and analysis steps, restarting a new iteration of the cycle.

In general, an ACODAT requires:

(1) A multidimensional data model that represents the data collected from the different sources, in order to characterize the behavior of the context, which will be used by the different data analysis tasks.

(2) A unique platform to integrate the different technological tools required by the data analysis tasks to carry out data mining, semantic mining and linked data, among others.

This concept has been successfully proven in different fields, but ACODAT has not been applied in innovation processes.

3.2 MIDANO
MIDANO is a methodology for the development of data analytics-based applications [16, 18], which is made up of three phases:

Phase 1 – Identification of data sources for the extraction of knowledge of an organization: This phase carries out a knowledge engineering process-oriented to organizations/ companies. The main objective of this phase is to know the organization, its processes and its experts, among other aspects, to define the objective of the application of data analysis in the organization. Also, it defines the autonomic cycles and their data analysis tasks.

Phase 2 – Preparation of data: To apply data analysis to a specific problem, it is necessary to have data associated with the problem. This involves performing different operations with the data, with the purpose of preparing them. This process is based on the paradigm ETL: extraction of data from the sources, data transformation and loading of the data in a data warehouse. During this phase are described all the variables of interest and carried out the data processing process (for example: dependency analysis among variables, normalizations, etc.). Also, this phase designs the multidimensional data model of the autonomic cycles, which is the structure of the data warehouse. Finally, it carries out a feature engineering process that consists on transform raw data into features. A feature engineering process includes the tasks of extraction, generation, fusion and selection of variables for the construction of the knowledge models.

Phase 3 – Development of the autonomous cycle: In this phase, the data analysis tasks are implemented, which are going to generate the required knowledge models (e.g. predictive and descriptive models). This stage culminates with the implementation of a prototype of the autonomic cycle. This phase can use existing data mining methodologies for the development of the data analysis tasks. In addition, during this phase, experiments are carried out to validate the knowledge models generated.

3.3 Proposed model of innovation processes
The innovation process is a structured strategy that ensures that the innovation team idealizes an innovation and executes it until its successful implementation. In this section, we explain the innovation process model defined in [23]. According to [23], an innovation process has four sub-processes: problem analysis, ideation, experimentation and commercialization. Each phase (sub-process) is described below.

(1) Problem analysis: The problem must be identified and defined.

- Definition of the problem: This step must indicate and define the problem.

- Specification of needs: It defines a list of requirements necessary to solve it.

(2) Ideation: It defines the concepts to develop.

- Generation of many ideas: In this step are generated ideas. The amount here matters. The more, the better. It can use the technique of brainstorming

- Ideas evaluation: It is the process of comparing and contrasting ideas related to the new product, to select the most promising.

- Selection of the best idea: The idea that best solves the problem is selected.

(3) Experimentation: In this step is generated a version, although not be exact to the initially proposed product.

- Prototype: It is the development of an initial product, which allows deciding if it is feasible.

- Test: The main objective is to validate the creative process.

- Escalation: It transforms a concept (prototype) in a commercial product.

(4) Commercialization: It is the process of launching new products or services to the market.

- Launching: It is oriented to publicize the innovative product and its results.

- Results measurement: It defines the metric to measure the results of the marketing process.

- Learning cycle: The market will give feedback to know if the idea must be changed, optimized or persevere with it.

- Internal diffusion: It is the communication between the workers. The objective is the utilization of innovation as a positive reinforcement to motivate the organization.

4. Application of MIDANO for the definition of autonomic cycles for an innovation process
In this section is analyzed an innovation process using the MIDANO methodology, in order to define the sub-processes where the autonomic cycles must be defined.

4.1 Sub-processes of an innovation process
An innovation process has different sub-processes, which must be prioritized according to if data analysis tasks can be used. There are 12 sub-processes defined in an innovation process, which are listed in Table 1.

4.2 Prioritization
The criteria to be considered to evaluate the relevance of the sub-processes were defined according to their importance for an innovation process (especially, for a textile organization) and the possibility to carry out data analysis tasks. Thus, these values determine the level of

Processes	Sub-processes	Abbreviation
Problem analysis	Definition of the problem	DDP
	Specification of needs	EDN
Ideation	Generation of ideas	GDI
	Ideas evaluation	EDI
	Selection of the best	SDM
Experimentation	Prototyped	PRO
	Pilot test	PPI
	Escalation	ESC
Commercialization	Launch	LAN
	Results measurement	MDR
	Learning cycle	CDA
	Internal dissemination	DIN

Table 1.
Sub-processes of the innovation process

importance of the sub-processes. For example, a process that is not important has a weight of 1 and a process very important has a weight of 5.

The case study is in the textile sector because it is one of the industrial sectors where MSMEs require more continuous innovation processes, to enable them to be competitive over time [23]. Likewise, it is the industrial sector of interest for the context where the project is developed, for which data are available to carry out data analysis tasks to improve it.

For the construction of the prioritization table, 10 experts from the fashion innovation sector and research professors were consulted, who participated by qualifying each of the criteria. In the final result, each of the answers provided by the experts was averaged. Results are shown in Table 2.

From the previous table, the sub-processes "Problem Definition", "Specification of Needs" and "Measurement of Results" were prioritized. The sub-process "Definition of the Innovation Problem" was the one that had the highest evaluation among the sub-processes because, in most of the criteria evaluated by each of the experts, its rating was equal to or greater than 4. It has a very good rating in each group of criteria: about the possibility to apply data analysis tasks in the process, how it impacts the innovation process and its interest in the textile industry. Particularly, in some criteria about its importance in the innovation process, it has the highest score (its impact in the innovation process and in the generation of new products and services, with a rating of 5).

4.3 Analysis of the strategic objectives to be achieved with these sub-processes using autonomic cycles
For the prioritized sub-processes in Table 2, it is required to characterize the current situation in each one. Table 8 in section "Supplementary Material" contains the actors involved in the sub-process, the data sources and activities that are used and the obtained results (goal to be reached). These results now must be reached using data analytic tasks.

5. Definition of the autonomic cycles
This section presents the ACODATs of the prioritized sub-processes, in order to enable autonomic coordination in the innovation processes (ACIP-000, see Figure 1), but particularly, it describes the design of the sub-process of the definition of the innovation problem.

The goal of ACIP-000 is the self-management of the innovation processes. In order to reach this goal, we propose three ACODATs:

ACIP-001 (Innovation Problem Definition): This cycle is responsible for obtaining useful information for the definition of the innovation problem. The goal of this autonomic cycle is the definition of the innovation problem based on the information of the organization and context.

ACIP-002 (Specification of Needs): This cycle is responsible for obtaining the requirements to be covered by the innovation process. The goal of this autonomic cycle is the identification and characterization of the requirements of the innovation problem.

ACIP-003 (Result Measurement): This cycle is responsible for assessing the quality of the results obtained during the innovation process. The goal of this autonomic cycle is the definition of the strategies and metrics to evaluate the results of the innovation process, and the evaluation of the results to determine the quality of the innovation process.

We have proposed three ACODATs according to the sub-processes prioritized in section 4.2 (ACIP-001, ACIP-002, ACIP-003). This prioritization was made according to the relevance of the sub-processes for the innovation processes of an organization and the possibility of automating them using data.

However, it is important to mention that there are other sub-processes in the model of innovation processes defined in section 3.3. They could be specified in the future using

WEIGHT	CRITERIA TO EVALUATE	PROBLEM ANALYSIS		IDEATION PROCESS			EXPERIMENTATION			COMMERCIALIZATION			
		DDP	EDN	GDI	EDI	SDM	PRO	PPI	ESC	LAN	MDR	CDA	DIN
	Relevance in the innovation process												
5	How quality impacts your innovation processes	5	5	4	4	5	4	4	4	4	4	4	4
5	How much depends on the other organizational processes	4	4	4	4	4	4	4	4	4	4	3	4
5	The complexity of the personnel and infrastructure required	4	4	4	4	4	5	5	4	4	4	4	4
5	How the generation of new products or services impacts	5	5	4	4	5	4	4	4	4	4	4	4
5	How it impacts the business processes and the organization	4	4	3	4	4	4	3	4	4	4	4	4
5	How it impacts on marketing strategies	4	4	4	4	4	4	4	4	5	5	4	4
5	How much it encourages creativity	4	4	5	4	5	4	4	4	4	3	4	4
	Relevance for a textile organization												
5	How much it impacts the organization	4	4	3	4	4	4	3	4	4	4	4	4
5	The level of difficulty to be performed in practice	4	4	3	4	4	4	4	5	4	3	4	3
5	How much help in the consolidation of an intelligent Textile Company (Smart Textile Factory)	4	4	4	4	4	4	4	4	3	4	4	4
	Relevance to do Data Analytics tasks												
5	How many internal or external sources of information exist: databases, Excel sheets, reports	4	4	4	4	3	3	3	3	3	4	3	3
5	What level of access do you have to the Information	4	4	4	4	4	4	3	3	3	4	3	3
5	The level of use of computational tools (word, excel, power point, etc.)	3	4	4	4	3	4	4	4	4	5	4	4
5	The level of use of decision-making software (DSS) - (CRM-ERP-MRP-CIM-COGNOS-informaweberp)	4	4	3	3	3	3	4	4	4	4	4	3
5	The frequency of information collection in this phase of the process	4	4	4	3	3	3	3	4	4	5	4	3
5	Total unweighted	62	61	57	57	58	58	57	58	57	61	57	53
5	Weighted total	312	303	287	284	290	291	285	292	283	306	286	264
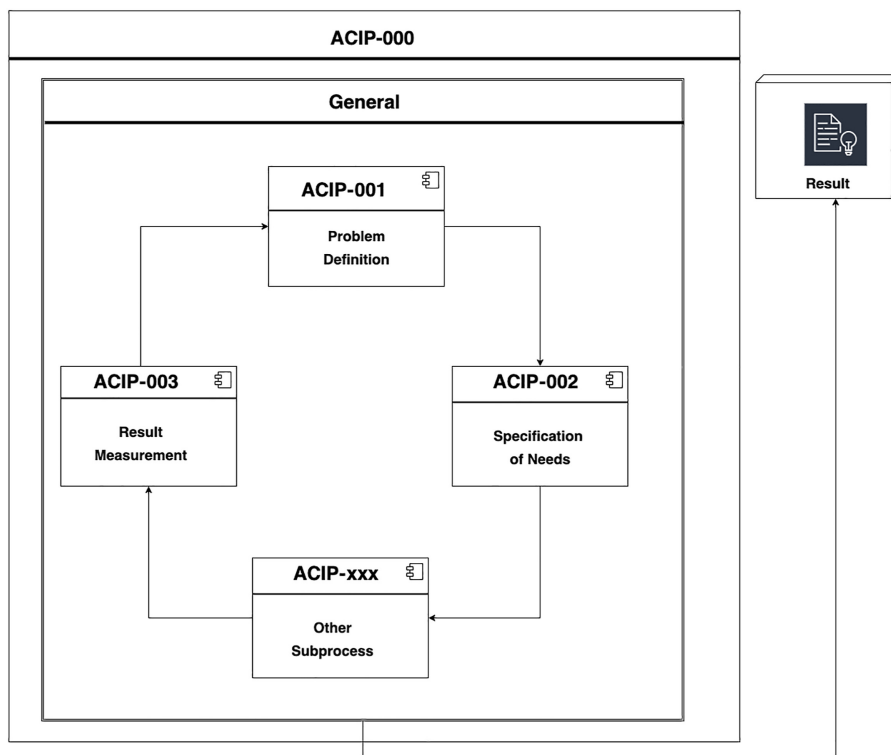
Table 2.
The prioritized sub-processes

ACODATs to automate them as well. Thus, ACIP-xxx refers to ACODATs for the other innovation sub-processes, such as generation of many ideas, ideas evaluation, selection of the best idea, among others.

Finally, the alerts module is an information system on the execution status of an innovation process (started, executed, finished), and additionally, it would inform about which of the sub-processes would be running.

In this article, we detail the ACIP-001, which was the one that obtained the highest evaluation in the prioritized processes.

5.1 Specification of the autonomic cycles for the "definition of the problem"
The Autonomous Cycle for the Innovation Problem Definition (ACIP-001 – Problem Definition) has as its main objective the characterization of the innovation problem, i.e. the statement of the problem. In general, this autonomic cycle is defined by a set of data analysis tasks, which use everything mining techniques to get useful information to create the statement of the innovation problem. We use the 5Ws model to define this cycle because it allows defining what the problem is and not the solution (see Figure 2). The 5Ws model was established by the Greek rhetorician Hermagoras of Tendon, from where it has evolved [33]. In the 5W model, each question must obtain an answer based on specific data.

Table 3 shows the general description of each task of this autonomic cycle.
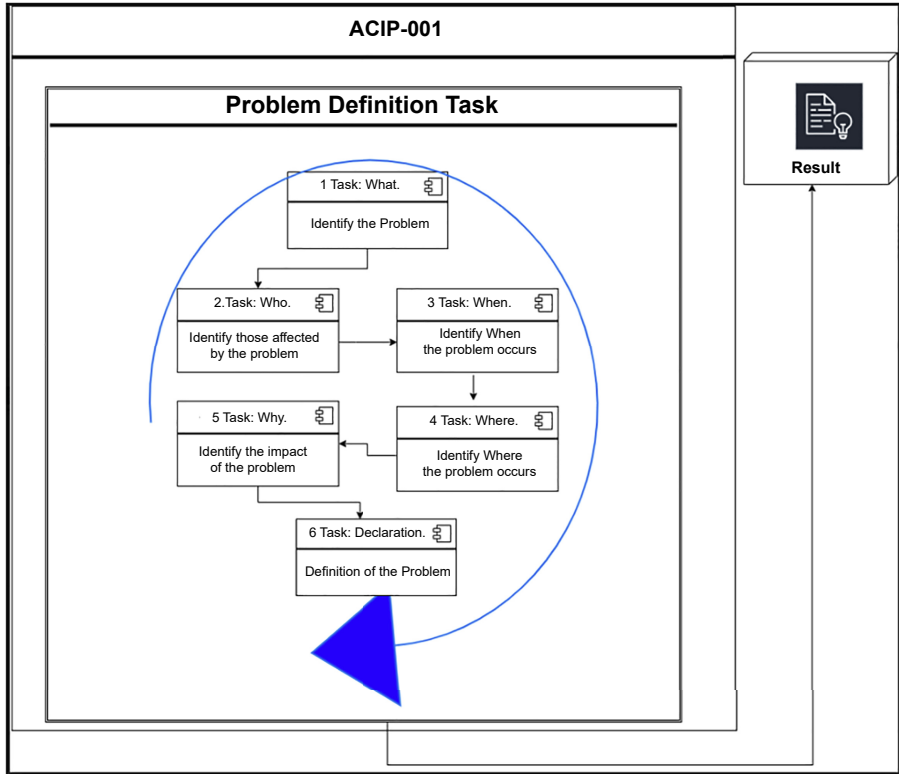
Now, we describe each task.

ACIP-001

Problem Definition Task

1 Task: What.
Identify the Problem

2.Task: Who.
Identify those affected by the problem

3 Task: When.
Identify When the problem occurs

5 Task: Why.
Identify the impact of the problem

4 Task: Where.
Identify Where the problem occurs

6 Task: Declaration.
Definition of the Problem

Result

Figure 2.
Structure of the
ACIP-001

Task name	Knowledge models	Data sources
1. What: Identify the problem	Descriptive model	Market studies (customer opinions, satisfaction surveys, etc.), internal databases (CRM[1], PQRS[2])
	Detection model	Social networks (Instagram, Facebook, etc.)
2. Who: Identify those affected by the problem	Descriptive model	Market studies (customer opinions, satisfaction surveys, etc.), internal databases (CRM[1], PQRS[2]) Social networks (Instagram, Facebook, etc.)
3. When: Identify when the problem occurs	Detection model	Market studies (customer opinions, satisfaction surveys, etc.), internal databases (CRM[1], PQRS[2])
	Predictive model	Social networks (Instagram, Facebook, etc.)
4. Where: Identify where the problem occurs	Diagnostic model	Market studies (customer opinions, satisfaction surveys, etc.), internal databases (CRM[1], PQRS[2])
	Predictive model	Social networks (Instagram, Facebook, etc.)
5. Why: Identify the impact of the problem	Diagnostic model	Market studies (customer opinions, satisfaction surveys, etc.), internal databases (CRM[1], PQRS[2])
	Predictive model	Social networks (Instagram, Facebook, etc.)
6. Declaration: Definition of the problem	Natural Language Processing (NLP)	

Note(s): [1]CRM: Customer Relationship Management
[2]PQRS System (Requests, Complaints, Claims and Suggestions)

Table 3.
Description of the tasks
of ACIP-001

(1) *Task 1. What: Identify the problem*: The first step identifies the problem through the data obtained. Some examples of data sources can be quality problems, customer complaints or derived from competitive surveillance activities. Its objective is to determine the occurrence of an innovation problem (i.e. it is necessary to create an original solution). This task uses detection and descriptive models to identify the problem.

(2) *Task 2. Who: Identify those affected by the problem*: This task identifies who are affected by the problem (e.g. specific groups, organizations, customers). This task uses descriptive models.

(3) *Task 3. When: Identify when the problem occurs:* This task identifies when the problem occurs or will occur, for which it can use detection or prediction models.

(4) *Task 4. Where: Identify where the problem occurs:* This task identifies where the problem is occurring, for which it uses diagnosis models.

(5) *Task 5. Why: Identify the impact of the problem:* This task identifies the importance of the problem, for this, it seeks to answer questions such as: What impact does it have on the business? What impact does it have on all stakeholders (i.e. employees, suppliers and customers)?

(6) *Task 6. Declaration: Definition of the problem*: This task aims to define the problem statement. For this, it uses NLP techniques to define the narrative.

Finally, the results module is a dashboard to report the execution status of this ACODAT, in particular, the results of its tasks. For example, when task 1 finishes, then it shows the information of the negative twitters; or when task 6 finishes, then it reports the problems that have been defined.

5.2 Multidimensional data model
The multidimensional data model for the previous ACODATs is defined in this section.

The model in Figure 3 includes different data sources, from market studies (e.g. customer opinions, satisfaction surveys), organizational databases (e.g. CRM, PQRS), until social networks (e.g. Instagram, Facebook). Data from each source are included in a different dimension in the data model, according to its characteristics. The main dimensions are the following:

Customers: It stores customer data such as age, gender, marital status, occupation, income, level of education, nationality, direction, country, department, municipality, neighborhood and stratum.

Market study: Stores general market study information, such as the objective, hypothesis, kind of investigation, type of analysis and conclusions. Also, it is linked to other dimensions like:

(1) Product satisfaction: It stores the satisfaction rating data of a product resulting from surveys that answer questions such as, what do you like the most, changes to improve, characteristics of other products that you would like in this product, product comfort, user experience, etc.

(2) Product price: It stores product price sensitivity data such as if you know the product, would you pay more or less to get it? Product units that you would buy taking into account the reduction or increase in price? Money willing to pay, a reasonable price, brand trust, factors that influence the purchase decision and what you like best about the product?
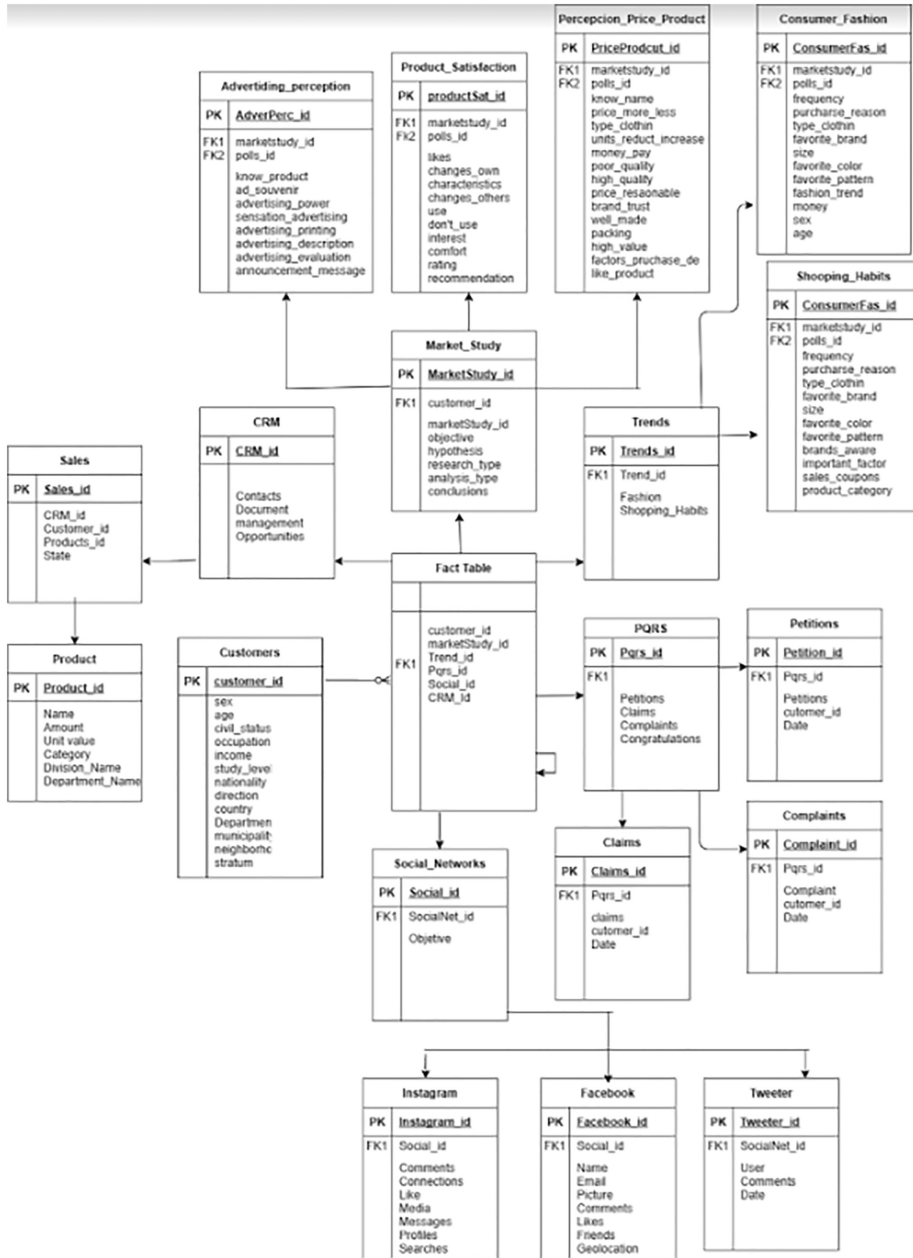
Figure 3.
Multidimensional
data model

(3) Advertising perception: It stores data on the perception of advertising, such as product knowledge, recall of the ad, evaluation of the power of advertising, feeling you have when you see an advertisement, the impression that the advertising gives,

how would you evaluate the advertising in comparison with other publications of the competition? and what would be the main message of the advertisement?

Trends: it describes information about current fashion.

(4) Consumer fashion: It stores data on fashion and its consumers, resulting from surveys that answer questions such as frequency of buying new clothes, reasons for shopping, clothing type, favorite brand, favorite color, favorite pattern, trend tracking, gender and age.

Social networks: It describes information about the social networks (Twitter, Instagram, etc.). For example,

(5) Instagram: It stores data of this social network, such as connections (contacts), likes, etc.

The multidimensional data model depicted in Figure 3 includes all the data required by the ACODATs. It describes all the variables of interest, which will be used as data sources to build the knowledge models (descriptive, predictive, among others) defined in each of the tasks of the ACODATs. This will allow having the necessary information to apply the different data analysis techniques to reach the goal of each ACODAT.

6. Case study
This section presents the experimental context for the instantiation of ACIP-001 (Innovation Problem Definition).

6.1 Experimental context
In this case study, we used data from the "Ramara Jeans" store, in Cucuta, Norte de Santander - Colombia. The store is dedicated to the manufacture, sale and marketing of all kinds of jeans, pants, shorts and skirts. Its objective is to provide the best service and quality in the products it offers, becoming a leader in the production of comfortable, versatile garments with competitive prices in the market.

The store currently has social networks on Facebook like Ramara Cúcuta, and Instagram like Ramara Jeans and on WhatsApp a line 313-8092414. It also has a team dedicated to virtual sales of products nationwide to attend to all requests, doubts and questions from its customers. The dataset used in this instantiation is from Instagram.

6.2 Instantiation of the ACIP–001: definition of the problem
At the beginning of the innovation process, it is necessary to define the problem. In this section, we describe how the ACIP–001 is instantiated in this case study.

(1) *First task:* This task can use descriptive and detection models to group and detect potential customer problems according to the client behaviors on the web, customer complaints on social networks, etc. Table 4 shows an example of a log file in an

Store id	Customer id	Age	What problem
Store-1	000110	35	Long waiting or delivery times
Store-1	000111	35	They would not recommend the brand
Store-1	000112	29	Product return
Store-1	000113	41	Low product quality
Store-1	000114	26	Abandonment of the purchase

Table 4.
"What": Information generated by the first task

organization, which can be built from a social network (using NLP techniques) or a PQRS database. The last column describes the results of the reported information by the clients.

Also, we can carry out a sentimental analysis to determine the negative sentiments in the social network (maybe due to a problem). For example, we can analyze the client's tweets (see Table 5). If a tweet is negative, it could be a complaint or the presence of a problem. For this task, the priority is to analyze the negative tweets (sentiment = 0) to identify the problem.

For this task, it is necessary to execute an NLP process to detect the problem in the negative tweets, which is composed of the next tasks: tokenize, remove stop words, clean special characters and stemming/lemmatization.

(2) *Second task:* It uses the information collected in the previous step to identify the person who is affected by the problem. In this case, this person could be an online customer face-to-face client consumer, etc. We can use a descriptive model that groups the clients according to the problem, in order to determine the type of clients affected by this problem.

For example, in Figure 4 are shown three different clusters (groups of customers) for three different problems. In this case, one of them are well-differentiated (cluster 1, which has only loyal customers). Cluster 2 (green impulsive customers) has some overlap with cluster 3 (customers by necessity).

(3) *Third task:* This task identifies when occurs the problem, which may occur before the purchase, due to some damage to the garment, or after the purchase. Examples are that the garment is very small or large, that the texture is very bad, etc. In Table 6, the column "when" represents the results of a predictive model about when the problem occurs: (0) before the purchase and (1) after the purchase. Also, we can use a detection model in order to detect in real-time a problem.

(4) *Fourth task*: This task identifies where the problem occurs. In this case, it is very important to identify the context of the problem, for which can be used a diagnosis model. In Table 6, the column "where" shows the results of a predictive model to determine where the problem occurs: (0) according to the customer's perspective or (1) into the organization. Also, it is possible to use a diagnosis model for the same problem.

(5) *Fifth task*: This task identifies the importance of solving the problem. For that, it can diagnose or predict the impact of the problem. In Table 6, the column "impact" shows the results of a predictive model about the impact of the problem. The value (0) is low impact, (1) is medium impact and (3) is high impact.

| Id	Tweets	Sentiment	Sentiment_ description
1613	@cruuzzy All the products are so beautiful . . .	1	Positive
5048	@radicalj Marvellous - not. How very thwartin. . .	0	Negative
9955	@annajudithk Delivery times are too long:(	0	Negative
1318	The cloth is always nice:-)	1	Positive
6739	@Bett_Homes there are not many products in the store. . .	0	Negative
4179	I love the blue and black ones:)	1	Positive
4084	@MarkBreech Not sure it would be a good thing 4 . . .	1	positive
1798	I loved them, I want another two	1	positive
3892	thank you for your response to my question	1	positive
3027	Hey @Indie_Shell Thanks For Following:) \n\n#. . .	1	positive

Table 5.
Identify negative
tweets
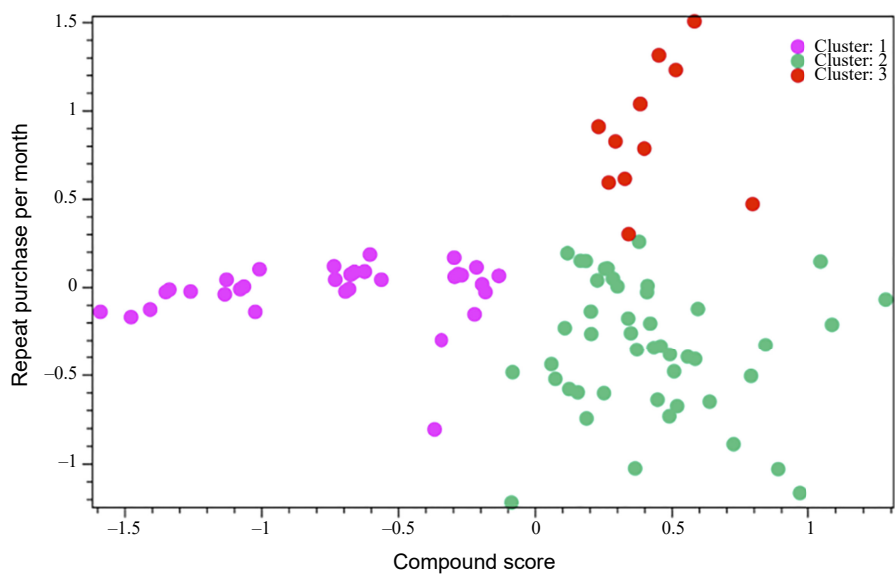
Figure 4.
Who: Clusters to
determine
customer types

Store id	Problem id	Description	When	Where	Impact
Store-1	000001	Long waiting or delivery times	1	1	3
Store-1	000002	They would not recommend the brand	0	0	1
Store-1	000003	Product return	1	0	1
Store-1	000004	Low product quality	1	1	1
Store-1	000005	Abandonment of the purchase	0	0	3

Table 6.
When and where:
Predictions generated
by the third, four and
fifth tasks

(6) *Sixth task*: It defines the problem taking into account the results of each of the previous tasks. In this task can be used NLP to define the statement of the problem in order to combine the what, who, when, where and why results. Additionally, we can add more information on the context using data from the reviews, tweets, etc. For example, we can use the information of the negative tweets (e.g. the keywords of their texts, determined by metrics such as TF-IDF) [34]. Some examples of statements of a problem, in this case study, are:

- "Long waiting or delivery times" is a "problem with high impact" "after the purchase"

- "Abandonment of the purchase" is a "problem with high impact" "before the purchase"

- "Long waiting or delivery times" because "Delivery times are too long"

- "They would not recommend the brand" is a "problem according to the customer's perspective"

7. Results discussion
The main result of this work is the definition of different ACODATs for the management of the innovation processes in an organization and the detailed description of the autonomous

cycle for the sub-process of innovation problem definition. For this, the data analysis tasks of the cycles were defined and the data sources were identified. Each task builds an appropriate knowledge model using the respective data sources to accomplish its specific objective. For example, in the case study, the first task carried out a sentiment analysis on tweets to identify the problem, and the second task carried out a clustering model to identify the types of users for each problem.

In particular, this autonomous cycle defines the fundamental input for the model of innovation processes proposed in section 3.3: the possible problems that are sources of innovation. Some of these identified problems will later be converted into an innovative product following our model. For example, in the case study, "Long waiting or delivery times" identifies a problem in the final delivery of the product that should lead to innovation in the purchase delivery processes. Another example is "Abandonment of the purchase", which identifies the disinterest shown by customers when they are about to buy a product. This may imply requiring innovation in product presentation/marketing strategies.

Another important result to highlight is the prioritization of sub-processes. To do this, the potentially automatable sub-processes of the innovation model proposed in section 3.3 were first analyzed using the organization and environment data. Subsequently, using the opinion of the experts, it was determined which of them is more relevant (priority) to automate in an initial process of automation of the management of the innovation processes in an organization. For this, the MIDANO methodology was used (see sections 3.2 and 4), which also, allowed defining the ACODATs and designing the autonomous cycle for the first prioritized sub-process (see section 5).

Another result is the definition of the data multidimensional model to be used by the ACODATs. It identifies the set of variables that must be used by the tasks of the ACODATs. With them, the data analysis tasks can build the different knowledge models (predictive, descriptive, etc.) which later are used to reach the goal of each autonomous cycle.

Finally, in the case study is instanced the first autonomous cycle, whose main objective is the identification of problems that potentially will be sources of innovation processes in the organization. In particular, it defines a sentiment analysis task to identify twitters that potentially describe a problem. It then groups those tweets by customer types. It then uses predictive models to determine when and where these problems occur, and their impacts. Finally, it performs a PLN process to formulate the sentences of these problems and potential sources of innovation processes.

It is the first step in demonstrating that it is possible to apply artificial intelligence techniques to improve innovation processes. It is a challenge to implement the rest of the ACODATs, but the preliminary results encourage the continuation of the application of these techniques in the innovation processes in the organizations.

8. Comparison with previous works

In this section, we propose criteria to compare our proposition of autonomic cycles to automate the innovation processes with other works. We define the next criteria:

(1) *Criterion 1:* they automate one of the sub-processes (e.g. definition of the innovation problem) of the innovation processes.

(2) *Criterion 2:* they use everything-mining techniques in the analysis of the innovation processes.

(3) *Criterion 3:* they study the definition of the innovation problem from the customer's or organization's perspectives.

(4)   *Criterion 4:* they consider different aspects of the problem (impact, where occurs, etc.)

In Table 7, a qualitative comparison with related works is made, based on previous criteria.

As shown in Table 7, current papers did not satisfy all the criteria. Specifically, in criterion 1, our research is the only one that automates the innovation processes, in this case, using the ACODAT concept. For this automation, paradigms such as multi-agent systems can be used in conjunction with our ACODAT architecture to model the entire innovation process [24].

For criterion 2, Ossi *et al.* [19], Qin *et al.* [25], Garcia *et al.* [35] worked on the innovation based on data mining. The basis of our proposal is autonomous decisions based on knowledge models from the data extracted from market studies, internal databases, social networks, etc. Thus, this work is based on everything mining techniques. Similarly [13, 14] present autonomic cycles for self-configuration, self-optimization and self-healing during the manufacturing process based on everything mining techniques.

For criterion 3, Kritsadee *et al.* [21] tested a model of factors affecting the innovativeness of SMEs. They analyze products, processes, as well as organizational and marketing innovation. Stoettrup *et al.* [22] investigated those parameters in innovation processes and, in particular, their influence on innovation outcomes in the context of smart manufacturing. Our paper is the only one that proposes the automation of the innovation problem definition using autonomic cycles.

Finally, for criterion 4, our proposal is the only one that evaluates different aspects of an innovation problem, such as its impact on an MSME, among other aspects.

9. Conclusion

This paper proposes the automation of the innovation process in MSMEs, through the definition of ACODATs. Also, the paper applies one of the ACODAT (for the definition of the innovation problem) in an MSME, in the "RAMARA jeans" store. Our ACODATs use different data sources to build knowledge models about the innovation process (e.g. predictive and descriptive models). Through the use of our ACODATs in the innovation process, it is possible to generate knowledge for the organization, not only to identify a problem, but also, to identify where it happened, when it happened and the impact it has on the organization. Particularly, the ACODAT for the definition of the innovation problem is essential in an innovation process because it allows the organization to identify opportunities for improvement.

On the other hand, there are many data sources that companies have but do not know how to use and get the most out of them. Specifically, the multidimensional data model defined for the ACODATs determines the required information from the organization and the context. With this information, it is possible to analyze it in real time to support the decision-making process based on data, and generate useful information for the organization.

Work	Criterion 1	Criterion 2	Criterion 3	Criterion 4
[13]	✗	✓	✗	✗
[14]	✗	✓	✗	✗
[19]	✗	✓	✗	✗
[21]	✗	✗	✓	✗
[22]	✗	✗	✓	✗
[25]	✗	✓	✗	✗
[35]	✗	✓	✗	✗
This work	✓	✓	✓	✓

Table 7.
Comparison with
previous works

The results of the case study allow concluding that it is feasible to use our ACODATs to automate the model of the innovation process proposed in section 3.3. The preliminary results show its utility to identify the problems that potentially will be sources of innovation processes in the organization. These preliminary results encourage the continuation of the application of the rest of ACODATs, in order to automate the innovation processes in the organizations using artificial intelligence techniques.

For future works, it is necessary to implement all ACODATs in a real scenario to verify their functionalities. To do this, it is necessary to do a detailed design of the rest of ACODATs. Also, it is important to determine the most important knowledge models required in the ACODAT for the definition of the innovation problem. Once determined, it is important to define the relevant everything mining techniques required for their implementations, such as data and process mining tasks.

References

1. Lim W, Gupta S, Aggarwal A, Paul J, Sadhna P. How do digital natives perceive and react toward online advertising? Implications for SMEs. J Strateg Marketing. 2021: 1-35.

2. Rao P, Kumar S, Chavan M, Lim W. A systematic literature review on SME financing: trends and future directions. J Small Business Management. 2021: 1-31.

3. Lim W. The quarantine economy: the case of COVID-19 and Malaysia. In: Lim W, Cheong H, Kaur S, editors. COVID-19, business, and economy in Malaysia. New York: Routledge; 2022. p. 3-23.

4. Mello S, Tomei P. The impact of the COVID-19 pandemic on expatriates: a pathway to work-life harmony?. Glob Business Organizational Excell. 2021; 40(5): 6-22.

5. Bretas V, Alon I. The impact of COVID-19 on franchising in emerging markets: an example from Brazil. Glob Business Organizational Excell. 2020; 39(6): 6-16.

6. Lim W. History, lessons, and ways forward from the COVID-19 pandemic. Int J Qual Innovation. 2021; 5(2): 101-8.

7. Castela B, Ferreira F, Ferreira J, Marques C. Assessing the innovation capability of small- and medium-sized enterprises using a non-parametric and integrative approach. Management Decis. 2018; 56(6): 1365-83.

8. Bagheri M, Mitchelmore S, Bamiatzi V, Nikolopoulos K. Internationalization orientation in SMEs: the mediating role of technological innovation. J Int Management. 2019; 25(1): 121-39.

9. Ortiz M, Joyanes L, Giraldo L. The marketing challenges in the big data age. E-Ciencias de la Información. 2015; 6(1): 16-45. 2015.

10. Sánchez M, Aguilar J, Cordero J, Valdiviezo P. Basic features of a reflective middleware for intelligent learning environment in the cloud (IECL). proceeding Asia-Pacific Conference on Computer Aided System Engineering. (APCASE); 2015.

11. Sánchez M, Aguilar J, Cordero J, Valdiviezo-Díaz P, Barba-Guamán L, Chamba-Eras L. Cloud computing in smart educational environments: application in learning analytics as service. In: Rocha Á, Correia A, Adeli H, Reis L, Mendonça Teixeira M, editors. New advances in information systems and technologies. Advances in intelligent systems and computing, 444; 2016. p. 993-1002.

12. Aguilar J, Buendia O, Pinto A, Gutiérrez J. Social learning analytics for determining learning styles in a smart classroom. Interactive Learn Environments. 2022; 30(2): 245-61.

13. Aguilar J, Garces-Jimenez A, R-Moreno MD, García R. A systematic literature review on the use of artificial intelligence in energy self-management in smart buildings. Renew Sustainable Energy Rev. 2021; 151.

14. Lopez C, Aguilar J, Santorum M. Autonomous VOs management based on industry 4.0: a systematic literature review. J Intell Manuf; 147: 2021.

15. Sánchez M, Exposito E, Aguilar J. Implementing self-* autonomic properties in self-coordinated manufacturing processes for the Industry 4.0 context. Comput Industry. 2020; 121.

16. Pacheco F, Rangel C, Aguilar J, Cerrada M, Altamiranda J. Methodological framework for data processing based on the Data Science paradigm. In: proceeding XL Latin American Computing Conference. (CLEI); 2014.

17. Puerto E, Aguilar J, López C, Chávez D. Using multilayer fuzzy cognitive maps to diagnose autism spectrum disorder. Appl Soft Comput. 2019; 75: 58-71.

18. Aguilar J. A general ant colony model to solve combinatorial optimization problems. Revista Colombiana De Computación. 2001; 2(1): 7-18.

19. Ossi Y, Jukka S, Porras J, Vesa H. Innovation capabilities as a mediator between big data and business model. J Enterprise Transformation. 2019; 1: 1-18.

20. Goodell J, Kumar S, Lim W, Pattnaik D. Artificial intelligence and machine learning in finance: identifying foundations, themes, and research clusters from bibliometric analysis. J Behav Exp Finance. 2021; 32: 100577.

21. Kritsadee P, Sanguan L, Somnuk A. Factor affecting innovativeness of small and medium enterprises in the five southern border provinces. Kasetsart J Social Sci. 2017; 38(3): 204-11.

22. Stoettrup M, Heidemann A. Design parameters for smart manufacturing innovation processes. Proced Coll Int pour la Recherche en Productique. 2020; 93: 365-70.

23. Gutiérrez A, Aguilar J, Montoya E, Ortega A. Intelligent systems used in the Micro, medium and small enterprises to improve innovation capabilities in the textile industry – a systematic literature review. Int J Entrepreneurship; 25(5S): 2021.

24. Aguilar J, Bessembel I, Cerrada M, Hidrobo F, Narciso F. Una Metodología para el Modelado de Sistemas de Ingeniería Orientado a Agentes Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artif. 2008; 12(38): 39-60.

25. Qin J, Liu Y, Grosvenor R. A categorical framework of manufacturing for industry 4.0 and beyond. Proced Coll Int pour la Recherche en Productique. 2016; 52: 173-8.

26. Chopra M, Saini N, Kumar S, Varma A, Mangla S, Lim W. Past, present, and future of knowledge management for business sustainability. J Clean Prod. 2021; 328: 129592.

27. Lim W, Ciasullo M, Douglas A, Kumar S. Environmental social governance (ESG) and total quality management (TQM): a multi-study meta-systematic review. Total Qual Management Business Excell. 2022; 1-23.

28. Kephart J, Chess D. The vision of autonomic computing. Computer. 2003; 36(1): 41-50.

29. Riofrio G, Encalada E, Guamán D, Aguilar J. Business intelligence applied to learning analytics in student-centered learning processes. In: proceeding 2015 Latin American Computing Conference. (CLEI); 2015.

30. Morales L, Ouedraogo C, Aguilar J, Chassot C, Medjiah S, Drira K. Experimental comparison of the diagnostic capabilities of classification and clustering algorithms for the QoS management in an autonomic IoT platform. Serv Oriented Comput Appl. 2019; 13(3): 199-219.

31. Aguilar J, Garcès-Jimènez A, Gallego-Salvador N, De Mesa J. G, Gomez-Pulido J, García-Tejedor J. Autonomic management architecture for multi-HVAC systems in smart buildings. IEEE Access. 2019; 7: 123402-15.

32. Vizcarrondo J, Aguilar J, Exposito E, Subias A. MAPE-K as a service-oriented architecture. IEEE Latinoamerica Trans. 2017; 15(6): 1163-75.

33. Fornieles Sánchez R. De Lasswell a Gorgias: los orígenes de un paradigma. Estudios sobre el Mensaje Periodístico. 2012; 18(2): 739-55.

34. Gaind B, Varshney N, Goel S, Mondal A. Identifying short-term interests from mobile app adoption pattern. Computación y Sistemas. 2019; 23(3): 829-39.

ACI

35. Garcia J, Delsing J. Autonomous production workstation operation, reconfiguration and synchronization. Proced Manufacturing. 2019; 39: 226-34.

Supplementary material
Supplementary material is available at: https://github.com/gistag/Supplementary-Material/tree/main

Corresponding author
Jose Aguilar can be contacted at: aguilar@ula.ve

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**