Definition of innovation problems in organizations using Data Analysis Tasks from Hybrid Sources: Social Networks and Organizational Databases

Ana Gutiérrez[1,2,5], Jose Aguilar[2,3,4,5,*], Ana Ortega[2], Edwin Montoya[2,5]

[1] GIA, Universidad Francisco de Paula Santander, Av. Gran Colombia, Cúcuta, Colombia
[2] GIDITICS, Universidad EAFIT, Medellín, Colombia
[3] CEMISID, Universidad de Los Andes, Núcleo La Hechicera, Mérida, Venezuela
[4] IMDEA Networks Institute, Leganés, Madrid, Spain
[5] Tepuy R+D Group. Artificial Intelligence Software Development. Mérida, Venezuela
*Corresponding author (aguilar@ula.ve; aguilarjos@gmail.com)

Abstract

Nowadays, we can express the experience we have lived with the products we use. Most of the time, we interact with brands and let our likes and dislikes be seen on digital platforms, either by interacting with social networks, filling out satisfaction surveys, or registering requests, complaints, and claims. To get the value of all available customer data of the companies, in this article, we propose to study the data from different sources using an autonomic cycle of data analysis tasks to define innovation problems in an organization. The tasks of the autonomic cycle are filter the customer comments from different sources (e.g., from social networks, PCCS (petitions, complaints, claims, suggestions) systems of organizations, etc.), obtain their keywords, and analyze the patterns of the users to answer the questions of the 5W methodology (what, who, where, when and why?), in order to define innovation problems. Finally, this article analyzes a case study of a fashion company using its PCCS system and comments of Twitter, to identify useful information. Part of the information discovered was the reasons for customer returns, merchandise delivery problems, shipment failures, failure to respond timely to customers, among other things. With this information, the autonomic cycle is able to define customer and organization-oriented innovation problems, in order to respond to these identified problems.

Keywords: Innovation Problems, Autonomic Computing, Data Analysis, Machine Learning,

1. Introduction

Data analysis from different sources can be useful in defining innovation problems by providing detailed information about customer needs and wants, market behavior patterns, and industry trends [1][2]. Analysis of data collected from these sources can help identify common problems and improvement opportunities for innovation. In the business context, it has been used to investigate the factors that influence the innovative behavior of employees [3], to identify innovative strategies that retail companies can use to stay competitive [4][5], to improve processes and services as a factor of competitiveness [6], to study customer experience [7][8][9][10], all previous cases of paramount importance for managers to identifying opportunities for innovation and constant improve.

Managing innovation within organizations implies designing strategies and policies to develop new products, processes, and services [10], improving their ability to compete in a constantly evolving market [11], and generating strategies for the improvement of production [12]. Using big data analysis will allow companies to analyze large amounts of information to identify patterns and trends that can lead to the identification of new business opportunities and improve decision-making [13].

On the other hand, understanding social media data is important for defining innovation problems, as it provides valuable information about customer interests, behaviors and needs, the problems they face, opportunities for product and service improvements, among other things. This data can be used to develop innovative solutions that better meet customer needs [2][4][15]. They can also be used to identify strategies to improve business commercialization and marketing [16]. Another source of data that can be used is the PCCS system since it provides valuable information about problems and opportunities for improvement in the products, services and processes of a company. Therefore, these different data sources can be used to improve customer service, customer satisfaction, and brand reputation, to identify market trends to help develop innovative solutions [17], and to analyze customer complaints to improve services [18], improve the quality of services and identify factors that lead to low levels of customer satisfaction [19]. Thus, these different data sources can help companies to have additional sources of information to identify and define innovation issues.

This paper presents an autonomous cycle of data analysis tasks (ACODAT) for defining innovation problems based on sarcasm analysis in Satisfaction Rating surveys, and customer comments in social networks and PCCS systems. In preliminary works, we have defined the concept of ACODAT [20] [26] [27] [28], designed generic ACODATs for the management of innovation processes in organizations [1], and defined sentiment analysis approaches for the definition of innovation problems [2]. In this article, with the aim of improving the definition of innovation problems, an ACODAT is designed to study sarcasm in customer comments in PCCS systems, surveys and social networks. The main contributions proposed in this paper are:

- Design an ACODAT to define innovation problems based on cynical or sarcasm analysis.
- Define an approach to use different data sources with customer comments, in particular, in PCCS systems, surveys and social networks, for the definition of innovation problems.
- Describe a case study of this ACODAT in the fashion textile industry.

The organization of this work is as follows. Section II describes related works. Section III describes the theoretical framework of this study, in particular, the ACODAT for defining innovation problems. Section IV describes the instantiation of this ACODAT to analyze customer comments from PCCS systems, surveys and social networks. Section V presents a textile case study, including experiments and analysis of results. Section VI provides a comparison of our proposal with other studies, and finally, the conclusions and future works are presented in the last section.

2. Related works

In this section, we describe the main latest papers associated with our approach. First, we present the papers about the use of data analysis to determine innovation problems [1][2] [10][13][14][21] [22]. Second, we present papers about data analysis from hybrid sources [12][6] [23].

2.1 Data analysis to determine innovation problems.

The article [1] of Gutiérrez et al. defines several ACODATs to manage innovation processes in a Micro, small and medium-sized enterprise (MSME), with their multidimensional data models, and the characterization of the AI techniques required in the case of the textile industry. The article [2] evaluates opinions from social networks, which are currently used to establish relationships with customers, to define innovation problems taking into account the feelings expressed in social networks. Jahan, et al. [13] study the impact that technology has had on business innovation. The authors analyze the role of Big Data attributes used in other studies for analyzing business efficiency from operational excellence, financial returns, and customer perspectives.

The paper [14] of Lendel et al. explores how data analysis techniques can identify and solve problems in innovation management. It presents a framework that combines quantitative and qualitative data analysis methods to analyze innovation problems and suggests strategies for addressing these problems. The research includes gathering information through analysis, questionnaires, and interviews, provides insights into the challenges faced in managing innovation processes, and offers recommendations for improving innovation management in companies. The work [21] uses analysis of data to determine problems in companies, in this case, linked to environmental impact. They develop strategies to reduce, reuse and recycle materials to cope with cultural, commercial, and technological changes. In this sense, they present a framework for the definition of innovative processes for material management, considering factors like data analysis, among others.

Lorduy et al. [10] emphasize the impact of innovation on national and regional administration. They analyze the national policies related to science, technology, and innovation from different sources (e.g., national organizations such as the observatories of science and technology, the national administrative departments of statistics, among others). In addition, the researchers apply surveys to understand local business idiosyncrasies actors of the financial, public and university sectors, among others. In the article [22], the objective was to review the existing literature on data-driven innovation (DDI), to understand how companies can use data and analytical tools to facilitate innovation processes. The results showed various aspects, such as innovation models in the market, competitive advantages of the innovation, and the necessity of the democratization of data. However, they highlight that little has been studied on how data helps the innovation process, which would be valuable for companies to learn from their own failures and improve their innovation processes in the future.

2.2 data analysis from hybrid source

Zaman et al. [12] explored various algorithms for prediction with a combination of financial and social media data to predict stock market behavior. Financial data includes historical stock price information and economic data, while social media data captures users' opinions and sentiments about stocks and

companies. This Integrated data analysis enabled the identification of relevant patterns about market trends, and most of the classifiers demonstrated an improved accuracy. Babu et al.[23] examined the impact of big data analytics on manufacturing and service organizations, as well as the ability to drive results and innovation in these companies. They applied data analytics to enhance innovation, explore new products, as well as address intellectual, legal and privacy issues, among other things, which can be shared across the industry to facilitate the innovation processes.

Wang, et al. [24] explored the use of multi-source data for the management of public health emergencies in China. They used the response scenario method, including risk analysis and operational plans at the micro level, as well as strategic decision making at the macro level. They applied the study for monitoring and response to the Ebola virus in China. It was evidenced that it is possible to detect and respond to public health emergencies at very early stages by analyzing data from different sources. In the article [25], Koukaras et al. developed a model to predict stock movement using Twitter and StockTwits data. They collected tweets from these platforms and financial data from Finance Yahoo, and implemented several classification models. The main innovation of this work is the integration of multiple sources to enhance stock prediction accuracy.

Finally, the article [30] focuses on evaluating sustainable development by analyzing normalized economic losses and changes in social structure. They utilize different sources, such as earth observation data, along with data from the national oceanic and atmospheric administrations, to examine the effects of floods, landscape changes, and climate change patterns. By integrating satellite-based with social media data from Twitter, they assessed the sentiment of affected individuals and measure the mental impacts of Hurricane Ida. The study demonstrates that combining of sources provides a more comprehensive understanding of the aftermath of the hurricane (they help understand the mental state of the community, identify affected populations and areas, and model the prevalence of epidemics).

According to the previous works in this section, there are articles related to the use of data analysis to determine innovation problems [1][2] and the use of Big Data to study business innovation and innovation capabilities [13][14]. On the other hand, there are data analysis works from hybrid sources, for example, using organizational data [23], social networks [24][12][25][30], images [30], or financial data [12][25]. They have been applied in different sectors such as manufacturing, health, and finance. This work includes again the concept ACODAT [1][20] to define innovation problems but with the difference that includes different data sources to improve the definition of the innovation problems. Thus, it discovers valuable information about the company's products, services and processes using satisfaction surveys and customers' opinions in social networks and the PCCS systems of the organization, to identify innovation problems.

3. Autonomic Cycle to analyze innovation problems

In this work, we use the ACODAT described in [1]. In that work, the authors defined an ACODAT for the definition of innovation problems, and in [2], the authors instanced this ACODAT using social networks. They apply sentiment analysis on negative Twitter to build innovation problem statements. For the definition of the ACODAT is used the 5W approach (see Figure 1) [30].

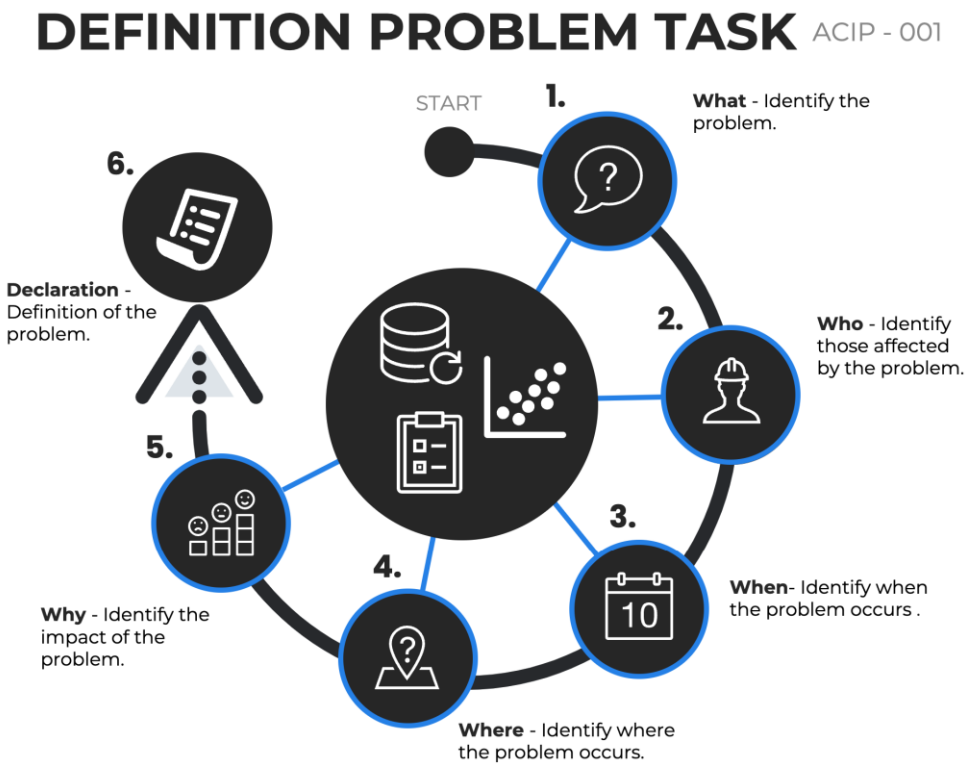

Figure 1. Structure of the ACODAT for the definition of innovation problems [1].

Each of the tasks of this ACODAT is described below:

1. What: In this task, innovation problems are detected. The objective of the task is to determine the innovation problems. For this case, descriptive and detection models can be used.

2. Who: this task identifies those individuals who are impacted by the innovation problems (e.g. customers, organizations, or specific groups). It makes use of descriptive models.

3. When: this task determines the time when the issue will arise, and it uses detection or prediction models.

4. Where: in this task, the location of the issue is determined using models for diagnosis.

5. Why: in this task, the problem's reasons are determined.

6. Declaration: The innovation problems are defined by this task. This task describes the statement of the innovation problems using *Natural language processing* (NLP) techniques.

4. Instantiation of the ACODAT for the definition of innovation problems using hybrid data sources

In this section is defined the instantiation of the ACODAT utilizing different data sources. In this case, we use PCCS systems, social networks, and some market studies based on customer satisfaction interviews. The macro-algorithm in Figure 2 defines the instantiation of the ACODAT for the "Definition of the Innovation Problem" based on [1][2], which is the basis of our SAHDS system (Sarcastic Analysis from Hybrid Data Sources) system.

Input: Dataset
Output: Definition of Problem

1. Analysis of the data sources to Identify sarcasm
    a. Filter the data sources
    b. Preprocess information
    c. Standardize information
    d. Apply Sentiment Analysis:
    Social Network -> Negative Comments
    Surveys -> Text with Positive Sarcasm
    PCCS systems -> Typification
2. Carry out a pattern analysis using clustering approaches
3. Generate the statement problem

Figure 2. SAHDS macro-algorithm

The steps of this macro-algorithm are detailed next:

1. For the first task/step, several sub-steps are required:

    a. Select the sources of the information to be used (surveys, social networks, etc.)

    b. Prepare the data for the analysis. Specifically, Natural Language Processing (NLP) operations are performed, including tokenization, clean character and stopword removal, and stemming/lemmatization.

    c. Normalize the data.

    d. Analyze the sentiment of the text. Then, we check if the polarity is negative and the overall sentiment is negative, suggesting that the text might be sarcastic [31][32][33].

    e. Sort the data to utilize the most pertinent information for our scenario (in our case, only sarcastic positive cases).

2. Perform a cluster analysis based on the 5W model responses.

3. Create an innovation problem statement using the results of previous tasks.

Table 2 shows the modification of the ACODAT for the definition of innovation problems based on the SAHDS-IPO macro algorithm in Fig. 2.

Autonomous Cycle Task	**Macro-algorithm**		**Task improvement**	**Data Source**	
	Social Network	**Database**		**Social Network**	**Database**
Task 1. Identify sarcasm	1.a. Preprocess information 1.b. Standardize information 1.c. Identify Keywo 1.d. Determine sar	1.a. Preprocess information 1.b. Standardize information 1.c Identify Keywor 1.d Determine sarc		Twitter	Satisfaction Rating surveys Customer Comments on PCCS systems
Task 2. What	2. Clustering based keywords	2.a Clustering analy using the customer comments	The comment of the customer is used to identify what is the problem	Twitter	Satisfaction Rating surveys
		2.b Clustering analy based on the typifi of the problem in the PCCS system	The typification (complaint/claim) will allow identifying the reason for the problem		Analysis of PCCS systems
		2.c Clustering analy based on answer-r on market study	Identify if the answer gives a solution to the request		Market Study: Answer_according request Y-N Again_talk_Virtual-Asistent
		2.d Clustering analy based on deficienc	Identify deficiencies in the product as variety, quality, difficulty finding the product		Product : Quality Variety KidnessAtention EasyFindProduct OffeerCompletLook
Task 3. Who	3. Clustering based on id_users	3.a Clustering analysis based on ids in the PCCS systems 3.b Clustering analy based on assigned	Identify the external customer affected by the problem and the internal customer who will solve the problem.	Id users	PCCS system: Id
Task 4. When	4. Clustering analy based on the date	4.a Clustering analy based on the date	Identify the date: PCCS system or survey	Date, month, working	PCCS system: date Satisfaction Survey: date

				day of the Twitter	
Task 5. Where	5.Clustering analys based on localization	5.a Clustering analy based on unity 5.b Clustering analy based on origin	Identify the unit responsible for solving the PCCS request or the place where answering the survey	Localization	PCCS system: Responsible_unity
Task 6. Why	6.Clustering analys based on negative sentim	6.a Clustering analysis based on response times	Identify the waiting time for a response to a PCCS system	Compound	PCCS: response times (1-40 days)
		6.b Clustering analy based on satisfacti rating	Identify the bad level of satisfaction		Satisfaction Rating: The level of satisfaction
		6.c Clustering analy based on recomm brand	Identify how highly would you recommend this brand to others.		Market Study: Recommend brand (1-10)
		6.d Clustering analy based on difficulty	Identify difficulty in handling the request		Market Study: Manage request
		6.e Clustering analy based on customer experience	Identify customer experience (good, bad)		Market Study: customer experience
		6.f Identify the toxi level and sarcasm comments	Identify communication level (easy or difficult)		Market Study: Easy_communication
		6.g Clustering analy based on sarcasm identified using comments	Identify the sarcasm in the comments		Satisfaction Rating: Customer Comments
Task 7. Declaration	7. Definition of the problem	7a. Definition of th Problem	Problem Declaration based on the previous answers	NLP	NLP

Table 2. ACODAT adapted to the SAHDS macro-algorithm

The cycle performs an iterative process to analyze data, identify innovation problems, and generate structured statements about these problems. A summary of the data processing techniques and algorithms employed in the autonomous cycle is presented below. First of all, an integration of multiple data sources was performed, combining data from social networks, surveys, and PCCS systems [17][19]. Therefore, firstly, a standardization of the information was performed, which was the process of converting data from different sources into a common and consistent format to facilitate analysis and comparison [37]. Secondly, for the data preprocessing process, a cleaning and transformation of raw data into a format suitable for further analysis was performed [36]. Next, tasks such as tokenization, stopword

removal, special character cleaning, and lemmatization were performed. Also, relevant features (e.g., date, time, location) were extracted.

Regarding the techniques used in the tasks, for sentiment analysis [2][12][33], the vaderSentiment library was used to determine polarity and sarcasm. To do this, the keywords in the texts were determined using the TF-IDF metric on the verbs and nouns present in them [2], in such a way that the texts could be classified as positive, negative, neutral or compound [34]. For clustering, the K-means algorithm and the silhouette metric were used to determine the optimal number of clusters [2][33]. In particular, various clusterings were made based on different criteria (users, dates, locations, etc.) to identify patterns and trends in the data. Finally, for the generation of problem statements, predefined templates were used as rules to follow. Each task of the adapted ACODAT is described in detail below.

Task 1. Identify sarcasm. This task is defined by different processes to analyze the data sources:

Information preprocessing: In this step, the comments from the organizational databases or social networks are transformed. Specifically, the text is tokenized, stop words are removed, special characters are sanitized, and the text is stemmed/lemmatized. Figure 3 shows a sub-algorithm (a typical NLP technique) for preprocessing comments. With the PCCS data is not necessary to do this step, because the typification is normalized when it is predefined during the design of the PCCS system. The typification predefines the types of registers (request, complaint, claim, suggestion) of a client in the PCCS system.

Input: Text Comment
Output: Clean Text Comment
1. Input Text Comment
2. Remove Stop Words
3. Remove URLs
4. Remove special characters
5. Remove hashtags
6. Tokenization
7. Lemmatization

Figure 3. Information preprocessing sub-algorithm

Standardize the information: this step normalizes the information to be used. Figure 4 shows the sub-algorithm that defines the steps to extract the information from the survey dataset. For example, the year is extracted from the date value in the survey and the PCCS systems. Similarly, it is done with the rest of the information of interest contained in the comments or PCCS systems to answer the questions of the 5W model (e.g., responsible to solve the problem, localization of the origin of the petition, etc.). In the case of the day, if the hour is less than noon (12) then the day column has a value of 1 (it means morning), else it will be 2 (it means afternoon).

Input: Dataset
Output: Standardize Dataset
1. Standardize Information
   a. Extract from the date field the value of the year
   b. Extract from the date field the value of the month
   c. In the survey, If rating <= 3:
      a. Include = 1
   else
      b. Include = 0
   d. In the social network
      If hour < 12:00pm : a. Working day = 1
      else b. Working day= 2
   e. Extract from the data the text
2. Verify the location of each user
   a. In the PCCS system identify the source from which the request, complaint or claim was made.
   b. Identify the place where did make the survey.
   c. In the social network identify the location of each user
      If user == company_user : a. user = 1
      else b. user = 2

Figure 4. Standardize information sub-algorithm

Keyword acquisition: In this step, the comments are preprocessed to obtain the keywords. In the social network are used the user's comments, in the satisfaction surveys, the answers of the users, and in the case of the PCCS system, the classification established in the company for the comments. With this data, keywords are determined using the TF-IDF metric [2].

Apply Sentiment Analysis: According to the keywords previously identified by text, then we use vaderSentiment library [29] in order to identify the sarcasm. It's important to mention that Sarcasm is a form of figurative language, where the speaker expresses his/her thoughts in a sarcastic way [34]. This library defines word polarities and classifies them as positive, negative, neutral, or compound (hybrid sentiment). Additionally, this library defines a composite score value for each text. It represents the average polarity of all words in the text. Figure 5 shows the macro algorithm used to perform sarcasm analysis on the dataset. If the composite score is less than 0.1, then the sarcasm is positive, otherwise, the sarcasm is negative. The next step is to filter the comments. In the social network, the best keywords are the negative words, while in the surveys the keywords are the sarcasm words. The PCCS uses the typification. This new subset of sarcastic comments serves as a source of information for continuing the next task. By analyzing and categorizing the keywords, organizations can identify trends, problem areas and opportunities for improvement. This allows them to take concrete steps to address recurring problems, implement solutions and strengthen their relationship with customers.

Input: Initial Dataset
Output: Sarcasm Text
1. Apply sentiment analysis
If Compound_Score < -0,1
    a. Sarcasm=Positive
    else
    b. Sarcasm=Negative
If Negtive_Score < 0
    a. Comment= Positive
    else
    b. Comment=Negative.
3. Filter the text comments

Figure 5. Sarcasm Analysis sub-algorithm

Task 2. What: Identify the problem: This task requires the definition of groups that characterize the innovation problems. Figure 6 shows a sub-algorithm that defines the clustering problem based on the keywords obtained previously. For this task, the silhouette metric with the K-means clustering algorithm was used to determine the numerical value of k [29, 35]. Silhouette values typically range from -1 to 1, with high values indicating that the data fits its own cluster well. In general, silhouette values greater than 0.5 imply good clustering results [29, 35].

Input: Keywords of the previous phase
Output: Clusters
1. Call the Clustering technique
2. Analyze the Clusters

Figure 6. Keyword-driven Clustering sub-algorithm

Finally, this task analyzes the centroid of each cluster to determine the answer to the “what” of the 5W model using the keywords [30]. In summary, Figure 7 shows the main steps to determine the 'what'.

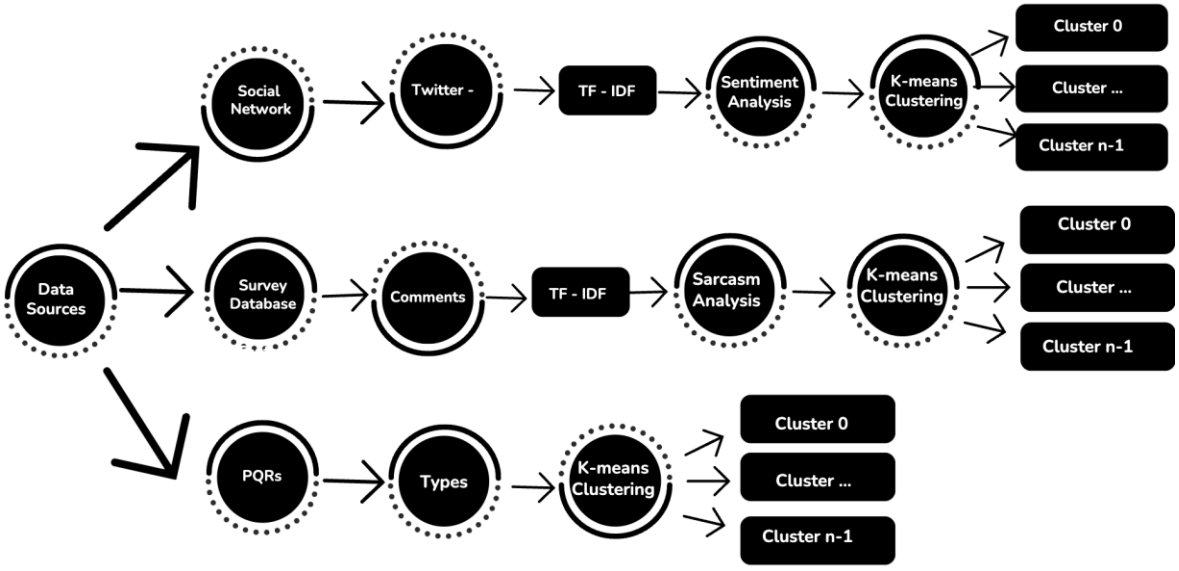

Figure 7. Determination of the "What" of the 5W model

Task 3. Who: Identify who is affected by the problem: this task uses the information from the previous task to identify who is affected by the problem. For this purpose, for each cluster obtained in Task 1, a new clustering is performed (See Figure 8) based on comments or aspects to be improved that the clients made in the surveys. In the PCCS system, it is also possible to identify who is affected by the problem. In this case, the entity or dependency that will resolve the petition, complaint, claim or request. For social networks, an external client is identified who writes the text or comment, or an internal client who responds to the comment. In this case, an individual may be an employee of the business, a physical or online consumer, or an entity of the business that participates in multiple requests by responding to different requests.

Input: Comments Clusters and user-Ids or unity responsible clusters
Output: Who
1. Call the Clustering technique based on
    a. User information from social network comments
    b. User or the responsible unity in the PCCS system
    c. User from surveys database
2. Analyze the Clusters

Figure 8. User-driven Clustering sub-algorithm

Task 4. When: Determine when the problem occurs: This task uses the information collected in task 1 to identify when the problem occurs. In this case, a new group is created for each cluster of the first task using the date as the similarity criterion (see Figure 9).

Input: Comments Clusters and date and PCCS Clusters date
Output: When
1. Call the Clustering technique based on date information
    a. Publication in the social network
    b. Completion of the survey
    c. Registration in the PCCS system
2. Analyze the Clusters

Figure 9. Date-driven Clustering sub-algorithm

Task 5. Where: Identify where the problem occurs: This task uses the information collected in task 1 to identify where the problem occurs. Again, in the social network, it is the place where the comment was published (geolocation). In other cases is used the customer's stores where they answer the survey. This information can be determined when they visit a physical store in the malls, or outlet stores, or maybe they call the phone line. In the PCCS system is the origin from where is registered the petition (see Figure 10).

Input: Stores and origin petition Clusters
Output: Where
1. Call the Clustering technique based on origin information
    a. Geolocation from the social network comment
    b. Place where is answered the survey
    c. Origin of the petition in the PCCS system
2. Analyze the Clusters

Figure 10. Site-based clustering sub-algorithm

Task 6. Why: Identify the impact of the problem: This task uses the information collected in task 1 to identify the impact of the problem. In the social network is used the compound score value of the task of sentiment analysis to determine the impact of the problem. According to this value, the impact can be low, medium or high. The survey is based on questions about satisfaction with the products, experience, quality, kindness in the attention, etc. (the scores in these aspects). In the other case, we use the answer time in the PCCS system, which indicates whether the stores are effective in meeting the customer's request (see Figure 11).

Input: Questions Clusters and answer time PCCs
Output: Why

1. 1. Get compound_score value from social network
   if compound_score is between [-0,01 , -0,33) then
      a. Why = High
   else if compound_score is between [-0,33 , -0,66):
      b. Why = Medium
   else c. Why = Low
2. Get answer value from the survey
   if the score is between [1 , 3):
      a. Why = High
   else if the score is between [4, 6):
      b. Why = Medium
   else c. Why = Low
2. Get time value PCCs
   if time is between [1 , 5):
      a. Why = Low
   else if the score is between [6, 10):
      b. Why = Medium
   else c. Why = High

Figure 11. Impact Identification sub-algorithm

Task 7. Declaration: Definition problem: In this task, we use all the results of the previous tasks to define the innovation problem. Specifically, in this work, several sentences are created, one oriented to the Petitions, Complaints and Claims of the clients, and others taking into account the comments where the clients expressed their thoughts in the surveys or social networks. Below are defined these sentences of the innovation problems:

Sentence -> using Social Networks:

<Who(user)> + "published" + <When(date)> + "from" + <Where(localization)> + "about" + <What(keywords)> + "such that the impact is" + <why><why(compoundSa)>

Sentence -> using the PCCS system of the organization:

<Who(assigned agent_pcc)> + "received" + <When(date_pcc)> + "request from" + <Where(Ticket)> + "about" + <What(typifaction_pcc)> + "such that the impact is" + <why(time_pcc)>

Sentence -> using Customer opinions in the surveys:

<Who(id_cliente> + "published" + <When(date survey)> + "from" + <Where(SurveyName)> + "about" + <What(list of aspects)> or <What(answer-request)> or <What(deficiencies)> + "such that the impact is" <Why(The level of satisfactionSurvey)> or <why(RecommendBrandSurvey)> or <Why(difficulty)> or <Why(customer experience_survey)> or <Why(communication)>

These statements can end up in a single statement that adds the results of all.

5. Experiment and Analysis of Results

This section presents the experimental context for the instantiation of the ACODAT for the Definition of the Innovation Problem using different data sources (social networks, surveys, and a PCCS system).

5.1 Case Study

In this study, we used datasets from the OFFCORSS company, which provided datasets of surveys applied to customers who responded in physical stores or outlets or made queries or requests through the virtual VoiceBot. We were also provided with a database containing the records of the PCCS registered by customers. Finally, the information from the company's Twitter account was used to determine the customers who followed it and gave their opinions about it. Table 3 summarizes the information used from each source.

Data Source	Data
Social Network	Twitter Comments, id_user, localization, date
Database Surveys	Satisfaction Rating: Customer Comments, Answer_according request Y-N, Again_talk_Virtual-Asistent, Quality, Variety, KidnessAtention, id, date, EasyFindProduct, OffeerCompletLook, The level of satisfaction
PCCS system	Typification, id, date, Responsible_unity, response time, TicketOrigin

Table 3. Data sources and data

The study focuses solely on OFFCORSS, a Colombian fashion company, since the data comes exclusively from this company's sources (surveys, PCCS system, Twitter), which allowed a deep and detailed analysis of the specific company, providing a concrete scenario to implement the ACODAT autonomic cycle in a real context. Thus, the results were interpreted taking into account the specific context of OFFCORSS. The specific findings may not be directly applicable to other companies in the fashion industry, different brands may have different customer profiles, market strategies and specific challenges, so caution should be exercised when extrapolating the findings to other companies without additional validation. However, it is important to note that the study serves as a basis for broader future research involving multiple companies in the industry where comparative studies could be conducted to validate the applicability of the method in different contexts.

5.2 ACODAT in the case study

This section describes each task of the ACODAT for defining innovation problems.

Task 1. Sarcasm analysis. For the task, the first step is the PLN techniques to determine the information to process (in our case, the keywords). An example of keywords obtained from the different sources are

Data Source	Keywords
Social Network	Assistance, robbery, vigilance
Surveys	Suggestions, ¨Need help¨
PCCS system	TV-devolution, TV-Return Request, VD-Data update, VD-Payment Verification,

Table 4. Example of keywords by source

For the surveys, we take the keywords of the comments and apply sentimental analysis using vaderSentiment Library. Then, we get the value for each text according to the compound value, and particularly, the sarcasm is true when the value is lower than -0.1. Table 5 presents a partial view of the results of the sarcasm analysis for this dataset.

ITEMID	ANSWERVALUE	ORIGEN	compound_Score	Negative_Score	Positive_Score	Neutro_Score	sarcasmo
200	1no resolvieron Nada. No es comprensible lo qu...	37	-0.2960	0.155	0.0	0.845	True
200	Hola	37	0.0000	0.000	0.0	1.000	False
200	Excelente!	37	0.0000	0.000	0.0	1.000	False
200	Hola	37	0.0000	0.000	0.0	1.000	False
200	Fosl	37	0.0000	0.000	0.0	1.000	False
...	...	...	...	...	...	...	...
616	Disponibilidad de zapatos de tallas 24	54	0.0000	0.000	0.0	1.000	False
616	El stock para todas las tallas, es decir si se...	54	-0.2960	0.073	0.0	0.927	True
200	Necesito que me ayuden con un paz y salvo y no...	37	-0.5267	0.196	0.0	0.804	True
200	buenas tardes,_x000D_\n_x000D_\naun no tenido ...	37	-0.5267	0.112	0.0	0.888	True
200	Necesito completar mi compra, pero aun no teng...	37	-0.2960	0.115	0.0	0.885	True

Table 5. Results of the Sarcasm Analysis step

In the PCCS system, the predefined typification is used. Table 6 presents some of the predefined examples of PCCS typification's.

	ID	State	Affair	Responsible	Required	Month	Date	Assigned	Update	Typification
0	746861	Cerrado	TV-Eventos promocionales	Centro De Experiencia	2021-12-06	Diciembre	2021-12-07	Servicio al cliente TV	2022-01-08	TV-Eventos promomocionales
1	748684	Cerrado	TV-CancelaciÃ³n de Pedido	Centro De Experiencia	2021-12-11	Diciembre	2022-01-03	Servicio al cliente TV	2022-01-08	TV-Cancelación de pedido
2	749790	Cerrado	TV-CancelaciÃ³n de Pedido ADDI	Centro De Experiencia	2021-12-15	Diciembre	2021-12-21	Servicio al cliente TV	2022-01-08	TV-Cancelación de pedido
3	750162	Cerrado	TV-Solicitud DevoluciÃ³n AutomÃ¡tica	Centro De Experiencia	2021-12-16	Diciembre	2021-12-24	Ventas Tienda Virtual	2022-01-08	TV-Solicitud Devolución
4	751229	Cerrado	TV-Solicitud DevoluciÃ³n AutomÃ¡tica 2*1	Centro De Experiencia	2021-12-20	Diciembre	2021-12-29	Ventas Tienda Virtual	2022-01-08	TV-Solicitud Devolución

Table 6. Results of examples of PCCS typification's.

The social network uses negative Twitter comments. In this step, for sentiment analysis is used the vaderSentiment library in each keyword in the Twitter register (column of text) to determine each person's mood. This library calculates the Compound_score of each text. Table 7 provides a partial view of the sentiment analysis results for this dataset. Note that tweets are negative when Compound_score < 0.

TweetE	compound_Score	Negative_Score	Positive_Score	Neutro_Score
@DanielSamperO @fdbedout there is no gentleman...	-0.2960	0.115	0.000	0.885
RT @Dantestereo Putting the former prisoner in...	-0.9231	0.425	0.000	0.575
Putting the ex-prison in prison for bribing wi...	-0.8689	0.380	0.000	0.620
@OFFCORSS Perhaps if today in the world and in...	-0.9074	0.197	0.046	0.757

Table 7. Results of the Sentiment Analysis for Twitter

Task 1. What: Identify the problem: This task defines clusters that identify innovation issues related to comments in the surveys, social networks or PCCS systems. The keywords in the comments or typifications are used by the K-means clustering algorithm. In the case of surveys, five clusters were found according to the values of the silhouette metric (see Table 8).

N clusters	Average silhouette
0	0.70
1	0.72
2	0.75
3	0.80
4	0.87

Table 8. Clusters based on the keywords in the case of surveys.

The cluster centroids defined by this task are shown in Table 9. It shows the keywords for each cluster that should be used to define the innovation problem.

Cluster centroids				
Cluster 0:	**Cluster 1:**	**Cluster 2:**	**Cluster 3:**	**Cluster 4:**
super	order	does	dont	did
suggestions	received	solution	answer	solve
moment	address	request	improve	concern
brand	able	payment	sizes	answer
boxes	resolved	attention	think	thank
aisles	time	good	sent	help

Table 9. Keywords of the centroid of each cluster for the surveys

For social networks have been used the same process explained in the article [2]. Thus, we use the keywords in negative tweets to apply the K-means clustering algorithm. The results are shown in Table 10.

Cluster centroids			
Cluster 0:	**Cluster 1:**	**Cluster 2:**	**Cluster 3:**
hotline	negative	worst	terrible
stolen	violated	answer	service
minutes	makes	happened	bought
waiting	responding	email	bad
follow	rating	assistance	refund
resources	platform	robbery	requested
gentleman	timely	say	claim
size	habeas	vigilance	WhatsApp
small	report	bonus	requesting
price	data	consumer	process

Table 10. Keywords of the centroid for some clusters for the social network

Finally, for the PCCS system is used its typification. Some examples of typification are TV-Return Request, VD-Data update, VD-Payment Verification, etc. The typification identifies the main words that describe the problem (see in Table 11 some clusters).

Cluster centroids		
Cluster 0 TV-Return Request TV-Order status VD-Data update TV-Order with new delivery	**Cluster 1** VD-Payment verification VD-Modification of quota VD-Request for payment VD-Request for Order Cancellation	**Cluster 2** VD-Incentive validation query VD-Affiliation status VD-Right of request VD-Information on how to place an order via WEB

Table 11. Centroids for some clusters for the PCCS system

In the next task, a grouping process using different similarity criteria (user, date, location) is performed for each cluster obtained in this phase.

Task 2. Who: Identify the ones affected by the problem: In this task, we can observe who was affected by the problem. In the case of social networks, they can be external customers who post on the network, or employees who respond to customer comments. In the case of the surveys, they were always customers who visited the store or made a purchase in the physical or virtual store. In the case of the PCCS, it was possible to identify the area to which the request was assigned in order to provide a response and solution to the customer who submitted the request. In this case, the units responsible for the PCCS were standardized. The assigned units were as follows: 1 -> Experience Center, 2 -> Virtual Store Sales, 3 -> CTS Transportation, 4 -> Portfolio, 5 -> Returns Area, 6 -> Customer Service TV, 7 -> Customer Service, 8 -> Legal, 9 -> Treasury, 10 -> Affiliation Area, 11 -> Administrators, 12 -> Marketing and Sales (See Table 12).

Data Source	**Identify Who**	**Clusters**
Social Network	Clothing customers	External users or internal users
Surveys	Clothing customers	Physical store customers, Online store customers, or Outlets customers
PCCS system	Units responsible	Experience Center, Virtual Store Sales, CTS Transportation, Returns Area, Customer Service TV

Table 12. Identify Who

Task 3. When: Identify when the problem occurs. In this task, it became evident that the PCCS was filed in the months of November and December 2021, and the responses to the surveys answered by customers were carried out in 2021. In the case of social networks, the Twitters were in July and August 2021.

Task 4. Where: Identify where the problem occurs. In this task was used the place where they publish (see Table 13). In the case of social networks, geolocation was used in order to identify the city of the publication. In the case of surveys, directions of stores and the virtual Voicet IP were used to identify the

place where customers were located. For the PCCS, the place was taken from the value indicated by the system where the request was received.

Data Source	Identify Where
Social Network	Geolocation
Surveys	Virtual Voicet IP, Store location
PCCS system	Website or Chat IP, or social network used to connect to the PCCS system

Table 13. Identify Where

Task 5. Why: Identify the effect of the problem. This task took into account the information on the time used to respond to the PCCS, representing this measure the impact that receiving a timely or delayed response has on the client according to the time spent by the unit provided. Some results for time are 0 days, or 7 seven days, with 0 days, implies that a timely response was given, and the impact is low. While the responses took between 0 and 7 days, which implies a medium impact, considering that the organization has up to 15 days to respond to the PCCSs received. For the answers analyzed by the surveys, the sensations received from 0 to 3 were taken into account for some of the following questions: about satisfaction in general, of the products, experience, variety, quality, kindness in the attention. Finally, in the case of social networks, a compound value from (-0.66 to -1] corresponds to a weak impact, from (-0.33 to -0.66] to a medium impact, and from (-0.01 to -0.33) to a strong impact.

Task 6. Declaration: Definition of the problem: This task uses all the results of the previous task to define innovation problems. According to the patterns defined in section IV, Table 14 shows the interpretation of the information for cluster 2 of the social network, Table 15 shows the Interpretation of the information for cluster 3 of the surveys, and Table 16 shows the interpretation of cluster 0 of PCCS.

5.3 Discussion of the results

This section presents the results for each data source (social network, PCCS system, and surveys).

5.3.1 Results for cluster 2 of social networks.

Task.	Results
1. What	The worst answer, happened, robbery, say.
2. Who	Clothing customers and one internal user
3. When	In the month of July 2021 in the afternoon session.
4. Where	The publications are from Tunja, Medellin, and Bogotá.
5. Why	The impact is Medium

Table 14. Interpretation of the information for cluster 2 of the social network

Problem definition:

Using the pattern of the Sentence for social networks, the statement of the innovation problem is:

*Clothing customers **published** in July 2021 in the afternoon session **from*** Tunja, Medellin, and Bogotá and ***about** Worst answer, happened, robbery, say. **The impact is** Medium.*

In accordance with the previous clause, non-compliance in the delivery of goods, as well as purchases that are returned or do not come, may constitute a violation of consumer rights. This suggests that in order to enhance customer service in general, innovation in the logistics and transportation process must be considered. For example, intelligent fleets can be designed with the use of sensors and data analysis to optimize routes, delivery times and vehicle maintenance. It is also possible to explore the use of drones to deliver packages in areas of difficult access or high traffic congestion.

5.3.2 Results for cluster 3 of the Customer survey

Task.	Results
1. What	*don't, answer, improve, size, think, sent*
2. Who	Customers
3. When	Year 2021
4. Where	Outlets Stores
5. Why	Acceptable level with respect to the variety of products in the store, of the other variables there are no ratings less than or equal to three.

Table 15. Interpretation of the information for cluster 3 of the customer survey

Problem definition:

Using the pattern of the sentence for customer surveys, the statement of the innovation problem is:

*Customers **published** in the year 2021 **from** Outlets Stores*, **about** *don't, answer, improve, size, think, sent* **such that the impact is** *acceptable.*

In accordance with the previous clause, customers generally complain about lack of responsiveness, shipping failures, and lack of guidance and advice on marketing products, among other things. For example, innovation in product offering and size availability must be defined, so garments could be designed to be adjustable and adaptable. Thus, garments could have features such as adjustable elastic waistbands, variable length straps, or detachable parts. Innovation in garment sizing has the potential to transform the shopping experience and improve customer satisfaction by providing garments that fit more precisely and comfortably to a wide range of bodies.

5.3.3 Results for cluster 0 of the PCCS system

Task.	Results
1. What	VD- *Data update*
2. Who	*Affiliations area*
3. When	*November and December 2021*
4. Where	Zendesk
5. Why	0 – 7 days

Table 16. Interpretation of the information for cluster 0 of the PCCS

Problem definition:

Using the pattern of the sentence for PCCS systems, the statement of the innovation problem is:

The affiliations area **received** in November and December 2021 **requests from** Zendesk **about** VD- Data update, **such that the impact** is medium.

According to the previous statement, customers in general are placing requests to change data, update personal data, quantity of items in the order, product sizes. Improving data updating can be approached from different perspectives, from process optimization to the adoption of advanced technologies. Innovation in this process seeks to achieve faster, more accurate and efficient updates that contribute to successful operations and decision making.

Definition of the innovation problem integrating the previous definitions

Proposing an innovative project to solve the problem of customer services and the lack of timely response, leads to the search for solutions to answer the following question: What to do?, First create mechanisms that lead to effective and efficient communication with the customer to build trust through digital platforms accessible, improving customer service in terms of online shopping, returns, complaints, claims and payment methods.

Also, our innovation proposal must focus on three key areas to enhance customer experience: efficient data updating, revolutionary garment sizing, and smart logistics management. By addressing these challenges, we aim to create a seamless, personalized, and reliable journey for customers, ultimately driving customer satisfaction, loyalty, and sustainable growth. Thus, our proposal must take advantage of the opportunity offered by the technologies by making changes in the marketing structure through an analysis of the external and internal environment of the company, reviewing and analyzing new strategies at technical, commercialization, and financial level, which must be frequently monitored and evaluated.

6. Comparison with previous works

In this section, we propose several criteria against which to compare our work with previous works. We define the following criteria:

Criterion 1: They use an ACODAT to define innovation problems.
Criterion 2: They use different data sources to define innovation problems.
Criterion 3: They manage innovation through Big Data in order to define innovation problems.

Table 17 shows the quality assessment of related works based on these criteria.

Work	Criterion 1	Criterion 2	Criterion 3
[1]	✓	✗	✗
[2]	✓	✗	✗
[6]	✗	✗	✗
[13]	✗	✗	✓
[22]	✗	✗	✓
This work	✓	✓	✓

Table 17. Comparison with preceding works.

For criterion 1, Gutiérrez et al. [1] define ACODAT for definite innovation problems and [2] carries the instantiation of this ACODAT. For criterion 2, Gutiérrez et al. [2] use social networks for the instantiation, but this work uses different data sources such as from social networks, surveys and the PCCS system, in order to define innovation problems. For Criterion 3, Sakila et al. [13] and Samarasinghe et al. [22] manage innovation through Big Data to investigate the impact on innovation. Particularly, the article [2] discovers problems that motivate improving product and service innovation, while in this work, had the opportunity of discovering problems not only with products or services but also oriented to improve processes, to think about innovations in marketing and organizational level.

As shown in Table 17, current papers do not satisfy all three criteria. On the contrary, this work fulfills all the criteria because it uses ACODAT and different sources of data to improve the quality of the results obtained. With this work was possible to generate innovation proposals for different problems, such as purchases that are returned, shipment failures, etc. The integration of the different data sources gives a holistic vision of the problems that allow improving both the issue identification and the innovation proposal. The combination of data sources allows for cross-validation of information. When insights are consistent across multiple sources, it increases the confidence in the accuracy of the findings. Also, can reveal hidden patterns that may not be apparent when examining data in isolation. Patterns that emerge from the synergy of multiple sources can offer deeper insights into customer behaviors and preferences.

Innovation is crucial for SME growth. Integrating insights from various platforms helps SMEs identify innovation opportunities that are rooted in real customer needs, reducing the risk of developing solutions that miss the mark. SMEs can quickly identify emerging issues and trends, enabling them to adapt their strategies and offerings promptly to meet changing customer demands.

Now well, the implementation of our approach may have several roadblocks. A possible roadblock is the variability in data formats across different sources, which complicates their integration and joint analysis. integrating data from corporate systems, such as PCCS, with data from social networks can present technical difficulties. This integration often requires substantial technological resources, and personnel with specialized expertise in data analysis and machine learning. In turn, the quality of data from social networks can be inconsistent, with potential issues such as noise, and spam, among others. Moreover, the collection and use of customer data, particularly from social networks, encounter both legal and ethical challenges concerning data privacy. Finally, there are also significant technical and logistical challenges associated with handling large volumes of data in real time.

7. Conclusions

The main contribution was to be able to use different data sources to improve the definition of the innovation problems in MSMEs. In this case, we have used sentiment analysis to determine if a text is sarcastic or not, and depending on the result, we worked with those texts that had a positive score for sarcasm. The PCCS data sources were also used, which proved to create and improve the problem sentences and their grammatical structure. As a result, it is improved the quality of the innovation proposal based on the problem that was defined. It confirms that different data sources can help companies to have an additional source of information that can identify and define innovation problems.

Thus, improvements in innovation processes are possible using data analysis tasks. The use of different data sources can discover valuable information about the company's products, services, and process pain points and areas for improvement. The combination of diverse data sources brings a synergistic effect, offering deeper, more accurate, and more actionable insights than analyzing each source in isolation. This approach leverages the strengths of each data source to provide a more holistic understanding of customer sentiments and innovation challenges. Combining real-time data from social networks and online platforms with internal data sources like PCCS systems or surveys enables the organization to gain insights and identify emerging issues in real-time. This proactive approach facilitates swift problem resolution.

The main limitation in the development of this work was that the databases used were at the local level, with the information provided by the studied organization, which meant using static data in a date range. Another limitation is the incomplete datasets, not all customers express their opinions online, which could lead to an incomplete representation of customer sentiments. This can skew the analysis and conclusions drawn. This work used data from different sources such as social networks, and PCCS systems. Thus, other limitations of relying on specific sources could include potential biases or incomplete representation of customer opinions if certain demographics or platforms are not well-covered. Also, data is constantly evolving. Thus, the clustering models need to adapt to new information, features, and trends to maintain their effectiveness over time.

Also, another limitation of the work is that the specific findings are based on OFFCORSS, but the insights gained from this study are valuable for understanding the application of the ACODAT autonomic cycle and are informative for other companies. Particularly, the unique context of OFFCORSS served as a practical

example, but the methodology and specification of data analysis tasks are valid for any application context. Thus, the findings provide a basis for further research. For example, future studies involving multiple companies across different contexts could validate and refine the applicability of the ACODAT autonomic cycle, extending its relevance beyond the initial case.

One work in the future is the implementation of an online tool with a user interface that can be intuitive, easy to use and visually attractive to carry out the whole process of defining the problem with the use of these different tasks of data analysis. Another future work is to give continuity to the implementation of the stages of the innovation process, having as the source of data and inputs the results of this first sub-process of problem definition. The idea is to apply the ACODAT methodology to the other sub-processes in order to incorporate autonomous cycles of data analysis tasks. Finally, another future work is to implement the automatic formulation of the problem verifying if the formulation is appropriate, and correct, according to the context in which it is being used.

Finally, social media data may not be fully representative of the entire client base, as it only reflects the opinions of active users on these platforms. Clients with limited or no internet access, such as the elderly or low-income groups, may be excluded from the analysis based on social media data. Additionally, some social networks are more popular among specific demographic groups (e.g., youth), which may skew the results, leading to over- or under-representation of certain segments. Also, PCCS systems tend to capture primarily the opinions of customers who have experienced problems or complaints, which can result in a biased perspective that emphasizes negative feedback and omits positive or neutral experiences. Satisfaction surveys may also exhibit self-selection bias, where only certain types of customers are likely to respond. Furthermore, extreme opinions—whether very positive or very negative—are often overrepresented in social networks and complaint systems. Fluctuations in activity levels on these platforms can also lead to uneven representation over time. To address these issues, future work should diversify data sources by incorporating traditional surveys, focus groups, and in-depth interviews. Implementing stratified sampling techniques will help ensure the representation of diverse demographic groups, while weighting methods can adjust for data representativeness. Additionally, cross-validation with other market and industry data sources will enhance the robustness of the findings. In conclusion, future work should analyze different strategies to manage biases.

List of abbreviations

PCCS: petitions, complaints, claims, suggestions
ACODAT: autonomous cycle of data analysis tasks
MSME: Micro, small and medium-sized enterprise
SAHDS: Sarcastic Analysis from Hybrid Data Sources
NLP: Natural Language Processing
TF-IDF: Term frequency – Inverse document frequency

Declarations

Conflicts of interest/Competing interests: The authors declare no conflict of interest to declare.

Availability of data and material: The data will be available if requested, with a justification of what its use will be.

Funding: Ana Gissel Gutiérrez Buitrago is supported by a PhD grant financed by Universidad EAFIT.

Authors' contributions: *A Gutiérrez:* Methodology Formal analysis, Experiments, codification, Writing; *J. Aguilar:* Conceptualization, Methodology Formal analysis, Simulation, Writing, *A. Ortega:* Formal analysis, Writing; *E. Montoya:* Formal analysis, Writing, Funding acquisition

Acknowledgements: All the authors would like to thank the "Vicerrectoría de Descubrimiento y Creación",

References

[1] Gutiérrez, A., Aguilar, J., Ortega, A. and Montoya, E. "Autonomous cycles of data analysis tasks for innovation processes in MSMEs", Applied Computing and Informatics, 2022, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/ACI-02-2022-0048

[2] Gutiérrez, A., Aguilar, J., Ortega, A. and Montoya, E. “Sentiment analysis in the social networks for the definition of innovation problems in organizations.” Submitted to publication, 2022

[3] Fitriana, A. and Satrya, A. “Employee Service Innovative Behavior in the Public Sector: The Role of Proactive Personality and Work Engagement,” in Proceedings of the 19th International Symposium on Management (INSYMA 2022), 2023, pp. 423–432. https://doi.org/10.2991/978-94-6463-008-4_53.

[4] Bae, Y., Choi, J., Gantumur, M., Kim, N. Technology-Based Strategies for Online Secondhand Platforms Promoting Sustainable Retailing. Sustainability. 2022, Vol. 14. https://doi.org/10.3390/su14063259

[5] Luamba, D., Blye, D., Williams, D., and Chagadama, J. “Innovative Strategies for Small Retail Companies’ Sustainability,” International Journal of Business and Management Research, 2021, Vol. 9, no. 3, pp. 330–338. https://doi.org/10.37391/ijbmr.090311.

[6] Saldarriaga, M., Guzmán, M., and Concha, E. “Innovación Empresarial: Factor de competitividad y calidad de vida en Popayán, Colombia”. Revista Venezolana de Gerencia, 2019, Vol. 2, pp. 151-166. ISSN: 1315-9984.

[7] Hopkins, J. “An investigation into emerging industry 4.0 technologies as drivers of supply chain innovation in Australia,” Computers in Industry, 2021, Vol. 125, https://doi.org/10.1016/j.compind.2020.103323.

[8] Dekimpe, M., “Retailing and retailing research in the age of big data analytics,” International Journal of Research in Marketing, 2020, Vol. 37, no. 1, pp. 3–14, https://doi.org/10.1016/j.ijresmar.2019.09.001.

[9] Trivedi, J. “Examining the customer experience of using banking chatbots and its impact on brand love: The moderating role of perceived risk”. Journal of internet Commerce, 2019, Vol. 18, no 1, p. 91-111, https://doi.org/10.1080/15332861.2019.1567188

[10] Lorduy, O., and Rangel, J. “Análisis de indicadores de innovación para una muestra de empresas manufactureras de Montería (Colombia),” Semestre Económico, 2019, Vol. 22, no. 52, pp. 49–73, https://doi.org/10.22395/seec.v22n52a3.

[11] Shayah, H., and S. Zehou, “Organizational culture and innovation: A literature review”, in Proceedings of the 2019 3rd International Conference on Education, Culture and Social Development (ICECSD 2019), 2019, https://doi.org/10.2991/icecsd-19.2019.58.

[12] Zaman, N., Ghazanfar, M., Anwar, M., Lee, S., Qazi, N., Karimi, A., et al. “Stock market prediction based on machine learning and social sentiment analysis”. TechRxiv. Preprint. https://doi.org/10.36227/techrxiv.22315069.v1

[13] Sakila, J., and Mesbaul, S., “Innovation Management: Is Big Data Necessarily Better Data?”, Management of Sustainable Development, 2022, Vol. 14, no. 2, pp. 27–33. https://doi.org/10.54989/msd-2022-0013

[14] Lendel, V., Hittmar, S., and Siantova, E., “Identification of the Main Problems in the Management of Innovation Processes and the Draft of Appropriate Recommendations”, Lecture Notes in Business Information Processing, 2015, 10th International Conference, Maribor, Slovenia, https://doi.org/10.1007/978-3-319-21009-4_17

[15] Zhou, F., Jiao, R., and Linsey, J., “Latent customer needs elicitation by use case analogical reasoning from sentiment analysis of online product reviews”, Journal of Mechanical Design, Transactions of the ASME, 2015, Vol. 137, no. 7, https://doi.org/10.1115/1.4030159

[16] Gutiérrez, A., Aguilar, J., Ortega, A. and Montoya, E, “Intelligent Systems in the Innovation Processes Of Textile Msmes-A Systematic Literature Review,” International Journal of Entrepreneurship, 2021, vol. 25, no. 5.

[17] Joung, J., Jung, K., Ko, S., and Kim, K., “Customer complaints analysis using text mining and outcome-driven innovation method for market-oriented product development,” Sustainability (Switzerland), 2019, Vol. 11, no. 1., https://doi.org/10.3390/su11010040

[18] Roper, S., and Bourke, J., “Innovating into trouble: When innovation leads to customer complaints,” Res Policy, 2022, Vol. 51, no. 10, https://doi.org/10.1016/j.respol.2022.104593

[19] Ghazzawi, A., and Alharbi, B., “Analysis of Customer Complaints Data using Data Mining Techniques”, Procedia Computer Science, 2019, Vol. 163, pp. 62-69, https://doi.org/10.1016/j.procs.2019.12.087

[20] Vizcarrondo, J. Aguilar, J. Exposito E., Subias A., "MAPE-K as a service-oriented architecture," *IEEE Latin America Transactions*, 2017, Vol. 15, No. 6, pp. 1163-1175, https://doi: 10.1109/TLA.2017.7932705.

[21] Chembessi, C., Beaurain, C., and Cloutier, G., “Analyzing technical and organizational changes in Circular Economy (CE) implementation with a TOE framework: insights from a CE project of Kamouraska (Quebec).”, Circ.Econ.Sust. 2022, Vol. 2, 915–936, https://doi.org/10.1007/s43615-021-00140-y

[22] Samarasinghe, S., and Lokuge, S., “Data-driven Innovation: Understanding the Direction for Future Research.”, ACIS 2022, 2022, Proceedings. 16, https://doi.org/10.48550/arXiv.2212.03061

[23] Babu, M. M., Rahman, M., Alam, A., and Dey, B. L. “Exploring big data-driven innovation in the manufacturing sector: evidence from UK firms,” Annals of Operations Research, 2021, https://doi.org/10.1007/s10479-021-04077-1

[24] Wang, X., Liu, Y., Zhang, H., Ma, Q., & Cao, Z. “Public Health Emergency Management and Multi-Source Data Technology in China”. Intelligent Automation and Soft Computing, 2018, Vol. 24, no. 1, pp. 89–96. https://doi.org/10.1080/10798587.2016.1267246

[25] Koukaras, P., Nousi, C., Tjortjis, C., “Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning”. Telecom, 2022, Vol. 3 no. 2, pp 358-378, https://doi.org/10.3390/telecom3020019

[26] Sánchez M., Exposito E., Aguilar J., “Implementing self-* autonomic properties in self-coordinated manufacturing processes for the Industry 4.0 context”, Computers in Industry, Vol 121, 2020, https://doi.org/10.1016/j.compind.2020.103247.

[27] 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: New Advances in Information Systems and Technologies (Rocha, Á., Correia, A., Adeli, H., Reis, L., Mendonça Teixeira, M. (eds)), 2016, Vol 444. pp 993–1002. https://doi.org/10.1007/978-3-319-31232-3_94

[28] Aguilar J., Garcès-Jimènez A., Gallego-Salvador N., De Mesa J., Gomez-Pulido J., Garcìa-Tejedor À. "Autonomic Management Architecture for Multi-HVAC Systems in Smart Buildings," *IEEE Access*, 2019, Vol. 7, pp. 123402-123415, https://10.1109/ACCESS.2019.2937639.

[29] Morales, L., Ouedraogo, C.A., 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”. *Service Oriented Computing and Applications* 2019, Vol. 13, pp. 199–219.

[30] Koukaras, P., Nousi, C., Tjortjis, C., “Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning”. Telecom, 2022, Vol. 3 no. 2, pp 358-378, https://doi.org/10.3390/telecom3020019

[31] Fornieles, R., “De Lasswell a Gorgias: los orígenes de un paradigma”, Estudios sobre el Mensaje Periodístico, 2012, Vol. 18, no. 2, https://doi.org/10.5209/rev_esmp.2012.v18.n2.41043

[32] Singh L., and Singh, S., “Empirical study of sentiment analysis tools and techniques on societal topics,” Journal of Intelligent Information Systems, 2021, Vol. 56, no. 2, pp. 379–407, https://doi.org/10.1007/s10844-020-00616-7

[33] Oad, A., Koondhar, I., Butt, P., Lei, H., Oad, A., Khuhro, M., and Bhutto, S., “VADER Sentiment Analysis without and with English Punctuation Marks,” International Journal of Advanced Trends in Computer Science and Engineering, 2021, Vol. 10, no. 2, pp. 1483–1488, https://doi.org/10.30534/ijatcse/2021/1371022021

[34] Sánchez, J., “Humor, ironía y sarcasmo en las redes sociales: creación léxica a partir del formante grecolatino -itis”, Revista de Investigación Lingüística, Vol. 25, pp. 259–283. https://doi.org/10.6018/ril.482871

[35] Oliveira, G., Coutinho, F., Campello, R., and Naldi, M. “Improving k-means through distributed scalable metaheuristics,” Neurocomputing, 2017, Vol. 246, pp. 45–57, https://doi.org/10.1016/j.neucom.2016.07.07

[36] Singh, A. “A Comprehensive Survey on Machine Learning” Journal of Management and Service Science (JMSS). 2021, Vol. 1, pp. 1–17, https://doi.org/10.54060/JMSS/001.01.003

[37] Salama M, Kader HA, Abdelwahab A. “An analytic framework for enhancing the performance of big heterogeneous data analysis”. International Journal of Engineering Business Management. 2021, Vol. 13, pp. 1–11, https://doi.org/10.1177/1847979021990523