Article

Construction of the Invoicing Process through Process Mining and Business Intelligence in the Colombian Pharmaceutical Sector

Jhon Wilder Sanchez-Obando [1,*], Néstor Darío Duque-Méndez [2] and Oscar Mauricio Bedoya Herrera [3]

1 School of Economics and Finance, Universidad de Manizales, Manizales 170001, Colombia
2 Information, Technology and Informatics Department, Universidad Nacional de Colombia sede Manizales, Manizales 170001, Colombia; ndduqueme@unal.edu.co
3 The Department of Systems and Informatics, Universidad de Caldas, Manizales 170001, Colombia; oscar.bedoya@ucaldas.edu.co
* Correspondence: jwsanchez@umanizales.edu.co

Abstract: The invoicing process is critical to the financial management of organizations. However, modeling this process presents challenges such as data updating, information availability, and aligning planned activities with the actual execution of the process. One difficulty is that designing the invoicing process requires extensive knowledge of the activities involved, and process representations based on organizational repositories are not necessarily aligned with the actual invoicing processes in the organization. Process Mining is complemented by the use of dashboards, which are inherent to business intelligence and allow for visual tracking of process behavior. This paper explores how the combination of process mining and business intelligence can enable a new level of process modeling that addresses specific issues in constructing processes that are aligned with real-world activities. To accomplish this, we first propose the Design Science Research (DSR) methodology, which outlines how a researcher or practitioner should approach the task of modeling a specific process using Process Mining augmented with dashboard resources. The research strategy was to identify the most appropriate methodology to construct the actual billing process, which led to the identification of the DSR methodology. This methodology, with its 12-step plan, allowed the construction of an artifact representing the actual invoicing process. Ultimately, the objective of constructing a real invoicing process in the Colombian pharmaceutical sector is achieved through the development of an artifact, complemented by business intelligence dashboards that ensure the alignment of the execution of activities within the process.

Keywords: process mining; processes; pharmaceutical sector; business intelligence; DSR

1. Introduction

Currently, information systems for organizations have high importance due to the impact generated in the execution of all the processes and procedures developed within the organizations [1], in the same way, the compliance of these processes in a manner optimal and efficient within the framework of continuous improvement. Process Mining uses the data recorded in information systems and processes them through algorithms that allow obtaining the structure of the processes while extracting valuable information to understand, verify, discover, and improve organizational processes [2].

Organizational processes are supported by the information stored in information systems, that record data about activities, events, times, and variables associated with the execution of all processes. All this information can be used with process mining techniques to discover how processes are executed and make decisions to improve or automate them. The practical result of this research is the automation of an artifact that represents the model of the real process aligned with the activities executed in real time. In this way, changes in

the order of activities or the timeline would allow the artifact to provide an updated and flexible process in the face of emerging changes [3].

Currently, information systems such as ERP (Enterprise Resource Planning), defined according to SAP v 7.0 [4], play a crucial role in organizations by guaranteeing the real-time measurement of the activities performed in the processes of organizations [5]. These systems not only facilitate the integrated management of business resources but also promote optimal and efficient compliance with processes in the company [6].

Process mining emerges as a fundamental tool in this context by taking advantage of the information recorded in information systems and applying process mining techniques, valuable information is obtained to understand, review, discover, and improve all organizational processes [7]. Given the above, process mining has an impact on ERPs by providing a holistic view of organizational workflows allowing identifying optimization opportunities and making informed decisions to increase efficiency and competitiveness in organizations.

Acquiring and implementing a business intelligence (BI) tool is highly beneficial for organizations such as pharmaceutical companies. Expanding its use can have a positive global impact. However, the pharmaceutical context has some peculiarities that a BI solution should be prepared to respond to. For example, the BI system could lead to the optimization of resources in various departments such as the invoicing department; will improve the state of financial analysis through efficient diagnosis and the identification and application of the best practice protocols for treatment, among others [8]. In this paper, business intelligence facilitates the monitoring of the process derived from process mining through visual dashboards.

The proposed research methodology is Design Science Research (DSR), which is a research approach that combines mixed methods research with the design of artifacts in the field of engineering. This methodology results in an artifact that is the actual invoicing process in the case study, in addition to being the framework to apply process mining aimed at determining realistic processes in organizations.

Similarly, in the context of the pharmaceutical sector in Colombia, there is a growing demand for efficiency and transparency in processes related to collections management and billing. The pharmaceutical industry faces unique challenges, such as the need to comply with strict regulations, manage inventory efficiently, and ensure product traceability [9]. In addition, the market is becoming increasingly competitive, prompting companies to look for ways to improve operational performance and stay ahead in a dynamic and competitive environment [10].

By considering the context of the pharmaceutical sector through a case study, the application of process mining in the design and optimization of invoicing processes becomes highly relevant by taking advantage of the information generated by ERP information systems [11], the pharmaceutical sector can identify opportunities to improve efficiency, reduce costs, and improve customer satisfaction [12]. The hypothesis of this research indicates that the construction of the real invoicing process in the pharmaceutical sector can be developed from the creation of an artifact in the DSR methodology, which is validated in a case study in Colombia.

This article is structured by sections, Section 2 deals with the state of the art, application in different sectors of process mining in organizations and related works, we continue with Section 3 in which the methodology to apply process mining in the selected case study is described. Section 4 analyzes the results obtained. In Section 5, the discussion is carried out and in Section 6, the conclusions and future work are presented.

2. State of the Art

Process mining emerged in the early 1990s as a response to the need for a more accurate and complete view of how processes actually work in organizations, and as a discipline derived from the intersection of data mining and business process management (BPM) [13]. Early work in this area focused on developing techniques for extracting useful knowledge from event logs generated by organizations' information systems.

A useful tool for modeling processes are the so-called Petri nets, which are related to process mining as a powerful and versatile tool for modeling concurrent systems [2] and is an active area of research, they emerged in 1962 as part of the doctoral thesis of Carl Adam Petri, a German scientist who was looking for a way to model the behavior of concurrent systems, such as industrial production systems [14].

On the other hand, BPM has its roots in the industrial efficiency movement of the early 20th century. Pioneers such as Frederick Taylor and Henry Ford introduced the idea of standardizing work processes to improve efficiency [15].

Today, BPM is a mature and widely adopted discipline. Organizations of all sizes are using BPM to improve efficiency, quality, and customer satisfaction [16].

Given the above, BPM and Petri nets are valuable tools for modeling and design processes, but they do not always reflect what is happening, because it is a process representation and does not consider real-time operations or changes. sudden or not represented in the processes. This is because process models may be incomplete or inaccurate, processes may change over time without the models being updated, and events that occur in a process may not be recorded in computer systems.

Process mining overcomes these limitations by extracting data from the event logs of computer systems. This data can include information about when each activity starts and ends, who performs it, what errors occur, and so on. From this data, process mining can discover true process models, analyze process performance, identify bottlenecks and areas for improvement, and verify compliance with standards [17].

Process mining is an essential tool for bridging the gap between process models and reality. It allows organizations to gain a deeper understanding of how their processes work and make informed decisions to improve them [18].

2.1. Applications in Various Sectors

Process mining has found applications in a wide range of industries, including retail, finance, healthcare, manufacturing, and government. To gain visibility into and improve their processes, more and more companies are turning to process mining [19], a technology based on artificial intelligence [20], and big data [21] that analyzes data from real-time processes to understand, optimize, and automate them. The practical result of this research is the automation of an artifact that represents the model of the real process aligned with the activities executed in real time. In this way, changes in the order of activities or the timeline would allow the artifact to provide an updated and flexible process in the face of emerging changes.

Process mining will continue to evolve as new technologies emerge, more data is generated, and more computing power is available for data processing. Pharmaceutical professionals will need to stay abreast of the latest advances in the field in order to take full advantage of its potential [22].

The analysis of the processes through the application of data mining occurred in 1998, the first works on the workflow were evident [23], and research was also carried out in the field of process mining in software engineering [24]. Subsequently, articles with specific applications of process mining were published as they represented important advances [25–28]. To model and discover processes, analyze organizational perspectives, and predict times [29–33]. Thus, it is developed in a cycle composed of phases and discoveries of the processes by obtaining a process map, it is also related to Workflow Management (WFM), which focuses on the automation of business processes, analyzes, identifies, discovers, designs, configures, executes, monitors and evaluates the processes [34].

The first algorithms focused on process discovery, i.e., automatically identifying the steps and flow of a process from event logs without a predefined model. Algorithms such as Van der Aalst's α-algorithm and the heuristic algorithm of Weijters and Van der Aalst laid the foundation for process modeling using the process mining technique [35].

Process mining leverages the information available in ERP information technology systems to visualize and reconstruct process flows while identifying deviations and patterns to address critical issues [2].

It is important to keep in mind that there are different approaches to process discovery techniques, according to [2] some of these approaches are: the general algorithmic approach such as the α-algorithm, the genetic algorithm, the relational mining and discovery techniques, exhibit 1 below lists these approaches with their advantages and disadvantages.

Table 1 shows some process mining algorithms and techniques that have been applied in different areas over time. In addition, it is worth highlighting those researchers have focused on developing and testing algorithms to overcome limitations and problems found in process mining techniques and emphasize the need for more practical studies to test the benefits of process mining in real cases [7].

Table 1. Process mining algorithms and techniques.

Algorithms	Advantages	Disadvantages
Algorithms α	The algorithm uses an evolutionary approach, according to [36] it is one of the first process discoveries and aligned activities with process.	It does not consider the frequency of activities, and it presents difficulties with the treatment of short cycles, and non-local dependencies, among others, with the generation of models that do not adequately adjust to the events [30].
Algorithms Inductive Miner	This algorithm returns a process tree for any event log, it seeks to find the cuts that divide the set of activities into subsets of activities [37].	The inductive mining algorithm is difficult to use on very large event logs considering that sub-logs must be created repeatedly [37].
Algorithms Heuristic Miner	Heuristic Miner is a practical applicable mining algorithm can be used to express the main behavior registered in an event log guarantees to find solutions in reasonable times and is commonly used to solve optimization problems [38].	To use this algorithm the entire process requires the consideration of several factors, such as the degree of mastery of the problem to be optimized, the search space, the alternatives that can be used, the types of calculations that must be performed and finally the quality of the solution. The development and implementation of this algorithm tends to be robust [39]
Relational mining	This algorithm, according to [40] is based on significance and correlation measures to visualize behavior and is accepted as a data exploration technique	One of its disadvantages is that the relational model cannot be transformed into a Petri Net with limitations to comparative evaluation with other process discovery techniques [41].
Discovery techniques	According to [41], they improve the visualization of poorly structured process models and allow the most important processes to be determined in addition to focusing on precision and understandability	According to [40] in discovery techniques, it is difficult to decipher the large number of arches it has.

In this research, the first approach was used because the other process mining approaches involve the design of algorithms and heuristics in first-order logic to understand the relationships and the knowledge base deposited in the ERPs of the organizations, in addition to the flexibility that the alpha mining algorithm allows us to adapt to the needs of real invoicing processes in the case study selected in this research. On the other hand, Business Intelligence (BI) consists in transforming stored information into knowledge, so that accurate data can be provided to a specific user at the right time to support the decision-making process in real time. BI integrates a set of tools and technologies that facilitate the collection, integration, analysis, and dashboard.

To implement a BI platform in Oracle, it is necessary to follow a series of intermediate steps that are common in the development of this type of software tool, such as the construction of a data warehouse (DW). Some of these steps are considered fundamental for the successful implementation of a BI system, including planning tasks and defining expected results, determining the architecture that the BI system will follow, and selecting

and installing the tool. of BI, the construction of the dimensional model of the data warehouse, the extract, transform, and load (ETL) process, and finally the development of the BI application [42].

2.2. Related Works

There are several works and researches in the field of process mining and business process optimization that deal with topics like the one proposed in this article. Among them, the studies of [43–45] stand out, which have investigated the application of process mining in improving financial processes in different business contexts.

In Smith, A., J. B., & B. C. (2018) [43], process mining techniques are used to analyze and optimize invoicing processes in a telecommunications company. Although their approach allowed us to identify some areas for improvement and optimization, its main limitation was the lack of integration with specific information systems such as SAP and it did not consider the type of data repository available according to the size of the organization, which limited the depth of analysis and implementation of improvements in real time.

On the other hand, Garcia, C. L. (2023) [45] studied the application of process analysis techniques in the banking sector to improve the efficiency of invoicing processes. Although they were able to identify some patterns and trends focused on fraud risk, their focus was on the identification of deviations and not on the detailed modeling of processes, which limited their ability to provide specific recommendations in the provision of financial services in the banking sector.

Finally, Pérez, L., R. M., and G. R. (2019) [46] studied the implementation of business process management systems (BPMS) to improve the efficiency of invoicing processes in a service company. His approach offered greater automation of processes and its main disadvantage was the complexity in configuring and maintaining the systems, as well as the difficulty in adapting to changes in business processes. The practical result of this research is the automation of an artifact that represents the model of the real process aligned with the activities executed in real time. In this way, changes in the order of activities or the timeline would allow the artifact to provide an updated and flexible process in the face of emerging changes.

This article has great potential to positively impact the Colombian pharmaceutical sector by offering practical, evidence-based solutions to optimize invoicing processes, key processes in successful organizational management of the sector. The sector-specific focus, the detailed application of process mining, and the in-depth analysis of the results differentiate it from other research and make it a valuable contribution to the research field of process modeling using process mining. Process mining is an underexplored tool in the Colombian pharmaceutical sector. Our research contributes to its adoption and demonstrates its value in optimizing key processes such as invoicing.

According to Marella, D. A. (2016) [47], process modeling and aligned activities are an essential activity in business process management that consists of representing the various steps, activities, and decisions that make up a process. The main objective of process modeling is to capture the reality of the business in an understandable and structured way, which facilitates the identification of areas for improvement and the making of informed decisions to optimize the efficiency and effectiveness of organizational processes.

According to Mayorga, H. S. A. (2018) [48], the objective of process modeling is the actual discovery of the process and the different possible variants or paths in the actual execution of the process to verify compliance with the procedures, policies and business rules.

According to Pavón-González, Y., Ortega-González, Y. C., Infante-Abreu, M. B., Souchay-Alzugaray, S., and Cobiellas-Herrera, L. M. (2021) [49], to model a process, it is necessary to identify the modeling needs to delimit the scope, the processes that you want to model, and where each one begins and ends, it is also necessary to define the purpose of modeling, you must determine what part of the process you want to represent and guide on what conceptual aspects should be considered for the analysis, likewise,

the functional domains involved in the process, the organizational sector, its context, are identified. the reference models that will be considered and the tools that will be used.

In Sanchis, R., Poler, R., and Ortiz, Á. (2009) [50] mention that the most common techniques in process modeling are the flow diagram, which represents a sequence of processes, as well as the data flow diagram, which allows you to see how data flows through the organization, the transformation of this data and its finalization. IDEF (Integrated Definition for Function Modeling) techniques that represent and model processes and data structures related to business processes.

Based on the previous related research, the approach proposed in this research work integrates the alpha process mining algorithm with a Business Intelligence Dashboard that contributes to the knowledge gap related to the actual construction of invoicing processes in organizations that generate an impact a disjunctive paradigm in the construction of organizational processes, by proposing a methodology that builds an invoicing process through process mining.

The alpha algorithm, also known as α-miner or α-algorithm, is a fundamental method in process mining [51]. It is used to discover process models from event logs, which are detailed records of activities performed within a system or process [52]. The alpha algorithm works iteratively, building a process model step by step from event log data, starting with the identification of the initial activities of the process, i.e., the activities that have no preceding activities. For each initial activity, the algorithm searches each event log trace for the next possible activity. If a consistent next activity is found, it is added to the model, the process is repeated until no more next activities are found to add to the model, and finally the Alpha algorithm can handle concurrency by identifying groups of activities that can be executed in any order [2].

2.3. Definitions

The next subsections show some definitions used throughout the paper.

Definition 1. *(Event, property, property function) A range of attributes can be found in an event in light of process mining, such as a timestamp, an object that shows the administrator, and relevant costs, to name a few. Let the universe of events as $\xi = \{z_1, z_2, \ldots z_m\}$, and the universe of properties as $\rho = \{r_1, r_2, \ldots r_n\}$. In that case, $x \in \xi \to \tau$. For instance, $x(z1) = A$ and $x(z2) = B$.*

Definition 2. *(Set of Sequences) S * illustrates the collection of all limited loops on a set S, for a given set S.*

Definition 3. *(Case, Trace, Log) From initiation to end of the activities can be considered as a case. A trace $\theta \in \varepsilon*$ is an orderly occurrence. Let the collection of all traces is Γ. It should be noted that while multiple cases may follow the same track, they are all distinct from one another. A collection of traces is an event log [53].*

The description of the miner algorithm is as follows:
Let L be an event log over T for alpha algorithm. $\propto$ (L) is defined as follows.

1. $T_L = \{t \in T \mid \exists_{\sigma L}\, t \in \sigma\}$,
2. $T_I = \{t \in T \mid \exists_{\sigma L}\, t = \text{first}\,(\sigma)\}$,
3. $T_0 = \{t \in T \mid \exists_{\sigma L}\, t = \text{last}\,(\sigma)\}$,
4. $X_L = \{(A, B) \mid A \subseteq T_L \wedge A \neq \emptyset \wedge B \subseteq T_L \wedge B \neq \emptyset \wedge$
   $\forall_{a\, \in A}\, \forall_{b\, \in B}\, a \to_L b \wedge \forall_{a_1,\, a_2\, \in A}\, a_1\, \#_L\, a_2 \wedge \forall_{b_1,\, b_2\, \in B}\, b_1\, \#_L\, b_2\}$,
5. $Y_L = \{(A, B) \in X_L \mid \forall_{A' B \in\, X_L}\, A \subseteq A' \wedge B \subseteq B' \Rightarrow (A, B) = (A', B')\}$,
6. $P_L = \{P_{(A,\, B)} \mid (A, B) \in Y_L\} \cup \{i_L, 0_L\}$,
7. $F_L = \{(a, P_{(A,\, B)}) \mid (A, B) \in Y_L \wedge a \in A\} \cup \{(P_{(A,\, B)}, b) \mid (A,B) \in$
   $Y_L \wedge b \in B\} \cup \{\, (i_L, t) \mid t \in T_I\,\} \cup \{\, (t, 0_L) \mid t \in T_0\,\}$, and
8. $\alpha(L) = (P_L, T_L, F_L)$.

Based on the previous algorithm, it can be observed that the first step allows to visualize all the transitions (t) or activities that appear in the event log. The second step allows visualizing the traces (A, B) and relationships between activities. Steps 4, 5 and 6 allow pairing the activities, eliminating unrelated pairs and determining the position of each pair of activities. Step 7 represents the relationships between the positions (P) of the activity pairs, and step 8 displays these relationships in a Petri net diagram (P, T and F).

2.3.1. Graphs

In this subsection, Equation (1) is presented, which allows to visualize the order of occurrences in the whole Petri net. The ordering relationships between the occurrence in the whole operation portrayed graphically in the Directly Follows Graph. It operates by building a directed graph that displays the ordering relationships and the rate of every incident in the log [53]. It does not offer a comprehensive process model, but it is a quick and effective approach to seeing how the process behaves. The DF equation is as follows where E be an event log [54]:

$$DF(G, H) = \sum <\theta \in E> cnt(g \to h\ in\ \theta) \tag{1}$$

Equation (1) cnt(g, h) counts the countable occasions event c happens in a trace of the event E immediately after event d. Each event E is reflected as a node in the graph, ordered edges unites the nodes that indicate the direct follow links between pairs of events. This is how the DFG was created. The value of the DF(g, h) equation, which measures the frequency of the direct follow relationship between g and h in the event log E, determines edge's weight among two nodes g and h.

2.3.2. Heuristic Miner

The following section explains the five steps of the mining heuristic of the mining algorithm, which begins by constructing a Petri net model with the minimum number of elements necessary to accurately represent the features of the event log. The process involves breaking down the log into three phases: mining the dependency graph, merging relationships, and achieving dependent interactions.

1. $f(g) = |\{\theta \in E : g \in \theta\}|$
2. $DF(g, h) = \sum <\theta \in E> cnt(g \to h\ in\ \theta)$
3. $PAR(g, h) = \sum <\theta \in E> cnt(g \to h\ in\ \theta) \times cnt(h \to g\ in\ \theta)$
4. $CHO(g, h) = \sum <\theta \in E> cnt(g \to \perp \to h\ in\ \theta)$
5. $L(g) = \sum <\theta \in E> cnt(g \to g\ in\ \theta)$

On the other hand, the accuracy of the mining algorithm can be evaluated by determining whether the process model allows only those actions that are directly related to the events recorded in the event log. If the process structure does not allow any traces that deviate from the characteristics described in the event log, it is considered to be fully accurate [55]. Let enm(k) $\subseteq$ H be the batch of actions in the workflow in #state(k). Let #hist (k) $\subseteq$ H+ be the background of k, i.e., the chain of incidents initiated for the corresponding scenario until k. #hist (k) eliminates the newest activity surroundings to k, nonetheless the string of circum- stances resulting in k. enl(k) = {#activity(k') | k' $\in$ K $\wedge$ #hist (k') = #hist (k)} $\subseteq$ H is the batch of operations that were taken by incidents with the identical background. The assumption is that activity is being linked on to similar stage having same background, #hist (k') = #hist (k) implies #state(k') = #state(k). The above notation applies to most general structure. The framework fails to account for allowing a variety of character- istics beyond noticed if precision is good. Hence, |enm(k)| $\approx$ |enl(k)|. The framework considers allowing a variety of characteristics beyond noticed if precision is less. Hence, |enm(k)| $\gg$ |enl(k)|. The following depicts the expression of precision [56]:

$$precision(l, m) = \frac{1}{|k|} \sum_{k \in K} \frac{|en_l(k)|}{|en_m(k)|} \tag{2}$$

Equation (2) it evaluates if a workflow structure can generalize the sample attitudes seen in the log of event or enables characteristics not found in the log of event but connected to that found inside it.

3. Methodology

The methodology of this research focuses on the DSR methodology, which aims at innovative solutions to real-world problems and serves as a guide for designing projects, managing projects, identifying and mitigating risks in projects, and building theories from projects and the publication of results [57]. Likewise, DSR is also a methodological approach that deals with the design of artifacts that serve human purposes, as a form of scientific knowledge production aimed at solving real-world problems and making a scientific contribution [58]. This article aims to build the invoicing process in the Colombian pharmaceutical sector with the use of process mining validated in a case study. To do this, two fundamental questions will be addressed: What is the appropriate methodology for building the invoicing process through process mining and business intelligence, a case study of the Colombian pharmaceutical sector? How to validate the methodology for building an invoicing process in the Colombian pharmaceutical sector? The analysis will allow not only to identify bottlenecks, but also to validate existing process models and verify the compliance with the actions and tasks established within the framework of the process model developed in the organization.

Within this research, it is worth mentioning relevant data of the company from which the data sample is taken, belonging to the pharmaceutical sector, the Colombian pharmaceutical company based in Bogotá, dedicated to the research, development and marketing of generic and branded medicines. Its constant focus on quality has allowed it to offer medicines that make a difference to patients over the years. It has a presence in 10 Latin American countries, manufacturing and marketing a broad portfolio of medicines that cover the main health needs, it has a complete portfolio with more than 150 commercial products and more than 25 international bioequivalence and pharmaceutical equivalence studies that support the efficacy and safety of your medicines.

The method for conducting research in the DSR consists of 12 major steps or stages which are illustrated in Figure 1.
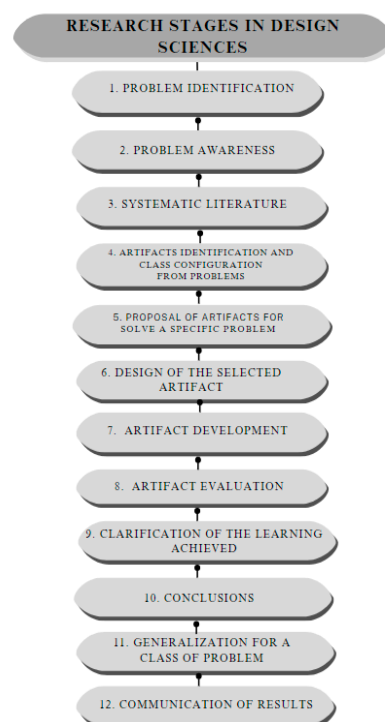

Figure 1. Stages of design science research. Source: Adapted from [58].

From Figure 1, the stages of research in design sciences illustrated in Figure 1 begin with the identification of the problem in stage 1, which must be relevant to be able to structure the formalized research question, in stage 2, awareness of the problem, it is important to understand the problem and investigate to understand the context of the problem and its causes and also consider the functionalities of the artifact, its expected performance and its operational requirements [58].

In stage 3 of the systematic literature review, it is essential to consult various technical literature databases such as Scopus, WoS, IEEE, Google Scholar, Science Direct. This helps the researcher to justify the importance of building and developing the artifact and solving the problem [58].

For the identification of artifacts and configuration of problem classes in stage 4, it is important to identify artifacts that address similar problems that allow the researcher to use best practices and the configuration of the problem class defines the scope of the contributions of the researcher. Artifact, in stage 5, the proposal of artifacts is related to solving a specific problem, the researcher proposes the artifacts and considers their reality, context and feasibility to find possible solutions. In stage 6, the design of the selected artifact, the entire context in which the artifact operates, and the satisfactory solutions to the study problem must be considered. It is important to describe all procedures for the construction of the artifact and its evaluation [58].

In stage 7 of artifact development the construction of the artifact is carried out. In stage 8 of artifact evaluation, the behavior of the artifact is observed and measured to provide a satisfactory solution to the problem. In stage 9, with the clarification of the learning achieved, the factors that contributed to the success of the research are explained along with the elements that failed. Stage 10, the conclusions, shows the results of the research and the decisions made during its implementation, and also indicates the limitations of the research that may give rise to future studies [58].

Stage 11 of generalization for a class of problems allows the knowledge to be applied in similar situations by other organizations and ends with stage 12 of communication of the results, which broadly contributes to a great advance of knowledge; this communication can be done in journals, seminars, conferences, among others [58].

Regarding the search equations used in this research, the following list was compiled:

- Google Scholar: process mining and billing or billing case pharmaceutical.
- Scopus: TI-TLE-ABS-KEY (billing AND process AND for AND process AND mining), yielding a total of 81 results.
- Web of Science: TITLE: billing or invoicing process for process mining and case study pharmaceutical, yielding a total of 9260 results.
- IEEE: process mining and billing or invoicing process, yielding a total of 6 results.
- ScienceDirect: invoicing or billing process for process mining, yielding a total of 675 results.

3.1. Problem, Literature and Extract of Event Log

In this article, the construction of the invoicing process of the pharmaceutical sector in Colombia is carried out through process mining and business intelligence through a case study. In this paper, business intelligence facilitates the monitoring of the process derived from process mining through visual dashboards. It is important to mention that the pharmaceutical industry aims to research and develop drugs that improve some health problems and quality of life. The pharmaceutical industry is also a business sector dedicated to the manufacture, preparation and marketing of medicinal chemical products [59]. Likewise, the pharmaceutical market is made up of private and public institutions that have the responsibility of contributing to the development of the delivery of pharmaceutical drugs [60].

When reviewing the databases, some of the work related to the topic aborted in the case study is found in Smith, A., J. B. and B. C. (2018) [43], process mining techniques are used. to analyze and optimize billing processes in a telecommunications company. On the

other hand, Pérez, L., R. M., and G. R. (2019) [46] studied the implementation of business process management systems (BPMS) to improve the efficiency of billing processes in a service company. Their approach offered greater automation of processes. In addition, García, C. L. (2023) [45] studied the application of process analysis techniques in the banking sector to improve the efficiency of billing processes.

Consequently, the construction of the invoicing process is developed from the 6 activities related to the processes, which are invoices (RV), cancellations (RF), cash receipts (DZ-DG), value credit notes (NC), and credit notes for an inventory of an organization in the pharmaceutical sector in Colombia. The construction of the invoicing process, a case study of the pharmaceutical sector in Colombia, through Process Mining and Business Intelligence is carried out with information from the SAP S4Hanna v 7.0 information system, information related to the invoicing of cash receipt documents, inventory credit notes, value credit notes, invoices and cancellations in the period between April 2022 and March 2024.

The file extracted from SAP S4Hanna v 7.0 contains a variety of essential data that provides a detailed view of the transactions of the invoicing processes. Among the fields included in this file are:

Document ID: This field uniquely identifies each accounting document recorded in SAP S4Hanna, which provides a means to track and manage financial transactions.

Entry Time: Indicates the time at which the accounting document was recorded in the system, which allows you to analyze the timing of transactions and identify possible activity patterns.

- User Name: This field records the name of the user who entered the accounting document in SAP S4Hanna, allowing you to provide information about who performed the transaction and track responsibility in the process.
- Document Type: Defines the type of accounting document, such as Invoices (RV), Cancellations (RF), Cash Receipts (DZ-DG), Value Credits (NC), Inventory Credits (NV), which makes it easier to categorize and specifically analyze each type of transaction.
- Document date: Indicates the date on which the accounting document was created, which allows you to organize transactions in time and analyze trends over time.
- Posting date: This date represents the date on which the document was posted, which can be different from the document date and is relevant for accounting and financial management processes.
- Reference Key (Document Consecutive): Provides an additional identifier for the accounting document that can be used to link it to other related documents or to track specific transactions.
- Voided with: Indicates whether the accounting document has been reversed or canceled in SAP S4Hanna, which is important for managing errors or incorrect transactions.
- Reference: This field provides additional information associated with the accounting document, which may include details about the nature of the transaction, external reference numbers, or any other information relevant to the invoicing process in the case of invoices and credit memos. corresponds to the consecutive one in the DIAN of these documents. With the files extracted from SAP S4Hanna through transaction FB03-view document, a database is created with the information for the construction of the invoicing process of the pharmaceutical sector in Colombia, through process mining and intelligence of business.

3.2. Artefact and Desing of Artefact for Discover the Process Model

In the case study, the artifact built in DSR shows a methodology from obtaining data, and types of documents that make up the knowledge base to proceed with the application of the alpha mining algorithm that built the real billing process and collection of the case study, in which reprocessing points and activities susceptible to improvement are evident.

In the process model creation or discovery phase, data preparation is performed, data collection, identification of data sources in SAP S4Hanna, selection of necessary

variables and parameters, data extraction, and data transformation and cleaning, as shown in Figure 2.

Figure 2. Process for data preparation. Source: Self-made.

In Figure 2, the sequence for data preparation is illustrated.

Data Collection: The process is initiated by collecting financial transaction data relevant for invoicing analysis, specifically in transaction FB03-View Document in SAP S4Hanna, which provides access to accounting document records.

- Identification of data sources in SAP S4Hanna: Identifies specific data sources in SAP that contain the information required for the analysis, including the Finance (FI) and Sales and Distribution (SD) modules. Transaction FB03-View Document is identified as the main data source for the study. This transaction allows you to view existing accounting documents in SAP S4Hanna, including invoices, cancellations, cash receipts, value credits, and inventory notes.
- Choice of variables and parameters: The variables and parameters relevant for the analysis are defined. These include accounting document number, document date, posting date, recording time, document type, and accounting sequence, among others.
- Data extraction: Transaction FB03-Display Document is used to extract data from SAP accounting documents. When entering the transaction, it is important to ensure that the transaction has the appropriate layout (data organization) parameterized for consulting the information, in order to extract the information correctly with the necessary variables and parameters, as shown in Figure 3.
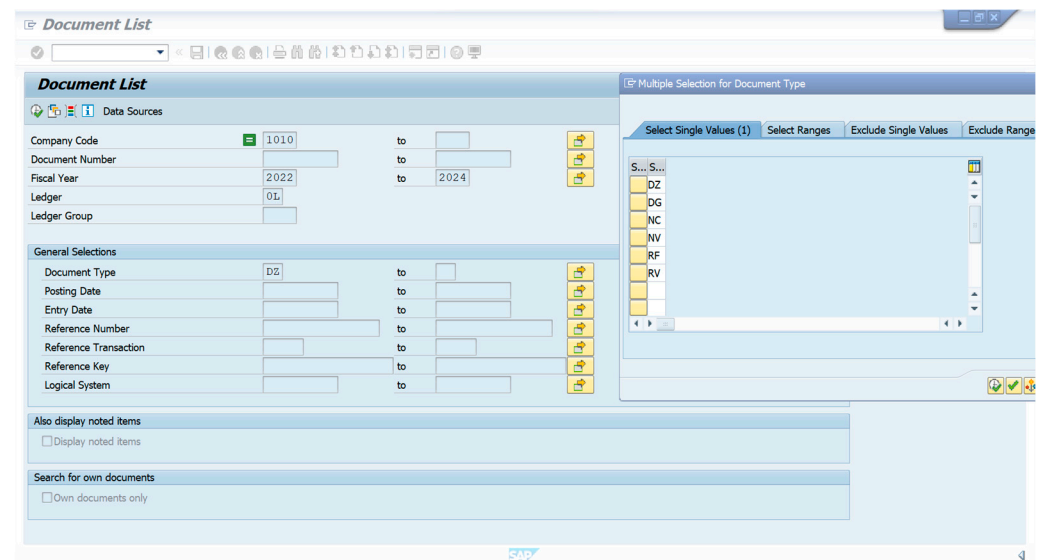
Figure 3. Transaction FB03-view document to extract data. Source: SAP S4Hanna information system.

Figure 3 shows transaction FB03 in the SAP S4Hanna Information System, which is a basic tool for viewing and analyzing accounting documents recorded in the Finance (FI) module. This transaction allows users to access a wide range of financial information, including:

- Accounting document details: Document number, accounting date, company, transaction type, accounting account, amount, descriptive text, etc.

- Document lists: Filter and sort accounting documents according to various criteria, such as company code, fiscal year, date range, transaction type, accounting account, amount, etc.
- Totals by accounting account: View the current balance and historical transactions for each accounting account within a specified date range.
- Line-item analysis: Break down accounting document amounts into their components, such as taxes, discounts, withholdings, etc.
- Transactional functions are
- Document Selection: Allows you to enter the specific document number you want to view or use selection criteria to search for documents that meet certain conditions.
- Detail Display: Displays detailed information for each accounting document, including all fields relevant to the transaction.
- Export Data: Allows you to export search results to an Excel file or other format for later analysis.
- Search Functions: Provides various search options to quickly locate the desired documents, including free text search, date range search, accounting account search, etc.

If you continue with the process, after selecting the appropriate variables, you will see a list of documents as shown in Figure 4 below.

Figure 4. Data listing. Source: SAP S4Hanna information system.

The list of documents in Figure 4 shows the information of the documents stored in the SAP S4Hanna system, this information is used to design the invoicing process, within the figure, there are criteria:

- Company Code: The company code for which the document was created.
- Document Number: The unique identifier for the document.
- Fiscal Year: Fiscal year for which the document was created.
- Document Name: A short description of the document.
- User: The user who created the document.
- Document Type: The type of document, such as invoice, credit memo, or payment.
- Document Date: The date on which the document was created.
- Accounting Date: The date the document was published in the accounting system.
- Reference: A reference number for the document.
- Reference Key: A unique identifier for the reference.
- Canceled With: Accounting document of the corresponding reversal, if applicable.
- Time: The time at which the document was registered in the module (FI).

Data transformation and cleansing: After the data has been extracted, the data transformation and cleansing process is performed. This may include removing duplicate or irrelevant data, standardizing data formats and correcting errors or inconsistencies, and storing the data as shown in Figure 5 below.

Figure 5. Data extracted from SAP S4Hanna system to Excel. Source: SAP S4Hanna information system.

Considering the transformation and cleaning of the data, Figure 5 shows the data extracted from the SAP S4Hanna system to Excel. The information extracted from the transaction is separated by sheets, in this case the data of cash receipts, credit notes affecting finished goods inventory, value credit notes and invoices are displayed along with their cancellations, if any. The data kept in the data are: Document Number, Entry Time, User Name, Document Type, Document Date, Posting Certificate, Reference Key, and Canceled With.

In the process, the validation of artifacts generated by the DSR methodology is carried out through measurement criteria, it is the most common artifact validation technique, the validation of compliance of the first artifact to the appropriate methodology for the construction of the collection and billing process in a case study of the pharmaceutical sector of Colombia through process mining and business intelligence, taking into account that the structure of the event log meets the criteria established by the SAP regulations in the ERP system.

3.3. Work Related and Conclusion of Artifact

In the modeling phase, it is important to keep in mind that process mining tools have their specific functionalities, among the most popular tools are CELONIS v 3.4, ProM v 6.9.129, and DISCO v 4.0.8, among others. For this paper, the process was constructed using DISCO software because of its more user-friendly interface, which is particularly suitable for personnel without technical training in open-source software. On the other hand, CELONIS software was not used due to its high cost and the limitations of its trial version, which limits the number of activities. In summary, the DISCO software is more accessible to non-experts in process mining and offers a more affordable membership for project creation.

Arias, M., and Rojas, E. (2016) [61] mention that the CELONIS tool allows monitoring and has associated audit functions to facilitate the analysis of information, it also has business intelligence functions, statistics, management of performance indicators and visualization of results through dashboards. Likewise, Van der Aalst, W. V. D. B., G. C. R. A., V. E., and W. T. (2009) [62] mention that DISCO is an open-source tool that brings together the main techniques developed in process mining, it also supports a wide variety of control flow models, Petri nets, and develops plug-ins to support the prediction, detection and recommendation of activities and processes to improve the ability to perform process mining analysis.

Similarly, Van der Aalst, W. M. P. (2011) [63] mentions that DISCO offers a user-friendly interface, presents functionalities to make filters, obtain statistics and information on cases and variants, in addition to visualize the results, which allows an analysis. process mining. Taking into account these tools, the construction of the invoicing process, a case study of the pharmaceutical sector in Colombia, was carried out with the DISCO tool with the license in evaluation version for the import phase of the event record, the discovery of the process and performance analysis in general with performance and frequency metrics, likewise the DISCO tool was used to apply process mining techniques and organizational analysis. The application is made with data extracted from the SAP S4Hanna system shown in Figure 6 below.
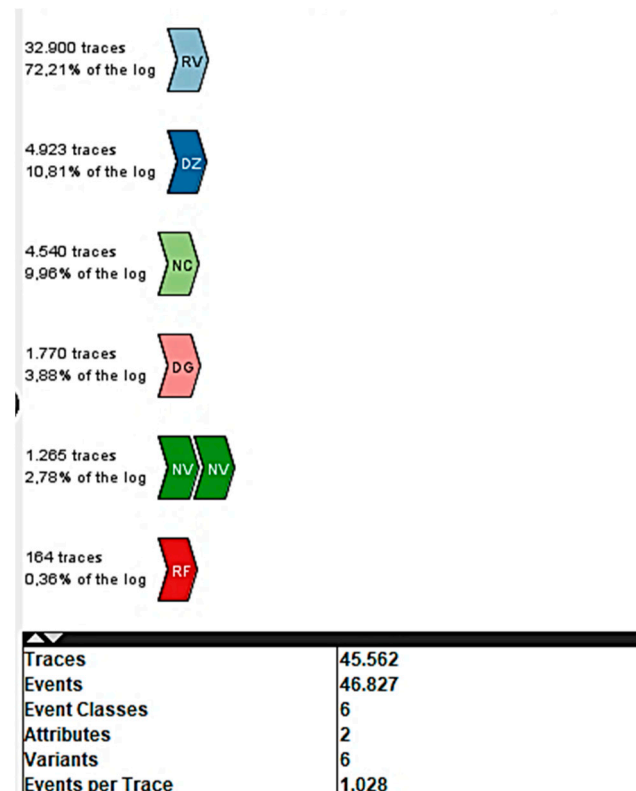

Traces	45.562
Events	46.827
Event Classes	6
Attributes	2
Variants	6
Events per Trace	1,028

Figure 6. General process data. Source: Own elaboration of the DISCO program.

Figure 6 shows the specific data for each type of document. In the case of (RV) invoices, a total of 32.900 documents were registered with a participation percentage of 72.21% of the total data processed, (DZ-DG) cash receipts with a total of 6.693 records and a participation percentage of 14,69%, the cash receipt documents are grouped taking into account that it is the same information but differentiated by the type of customer, for example, which has a different type of document, the (NC) value credit notes that do not move the inventory have a total of 4540 records and a participation percentage of 9,96%, (NV) credit notes that affect the inventory have a total of 1265 records and a participation of 2,78%, (RF) which corresponds to invoice cancellations have a total of 164 records and a participation of 0,36%, with a favorable percentage for this process and the error level is below 1%. In summary, this figure clearly shows the descriptive statistics of the event log from the case study.

This visualization provides a graphical representation of the variability in the process by highlighting the frequency of different execution paths.

- X axis: Represents the number of events per case, varying from the minimum to the maximum values observed in the data.
- Y-axis: Represents the number of cases, which indicates the frequency with which each number of events occurs per case.

- Bars: The height of each bar represents the number of cases that have the corresponding number of events.

According to the information in the figure, the event log contains 6 cases with a total of 46,827 events. The average number of events per case is 7805, with a minimum of 164 events and a maximum of 32,900 events. This suggests that the analyzed process has a significant degree of variability in its execution, with some cases requiring more steps or activities than others.

Considering the case study, the proposal of a methodology to implement process mining in pharmaceutical sector organizations in Colombia supported by the Design Science Research (DSR) methodology from the acquisition of data and types of documents that make up the knowledge base, supported by process mining in the construction of the real process of the case study, also combined with Business Intelligence can allow a new level of process modeling and support to address specific billing problems that have previously been managed evaluate the alignment between the execution of real-world activities and the pre-defined billing process.

3.4. Validation of Artifact

According to Brocke, J. vom, Hevner, A., and Maedche, A. (2020) [64], the validation of the artifacts generated by the DSR methodology goes through the techniques of consistent validation, criterion validation, measurement criteria validation, input data validation, internal design validation, linguistic validation, relative improvement validation, representational validation, requirements validation and theoretical validation. In the same research, validation by measurement criteria is the most common artifact validation technique, given that a form of validation developed in the proposed artifact is the case study of the pharmaceutical sector in Colombia.

Validation by measurement criteria is applied to the case study and allows the validation of compliance with the first artifact on the appropriate methodology for the construction of the invoicing process in a case study of the Colombian pharmaceutical sector, Through process mining and business intelligence, taking into account that the structure of the event log meets the criteria established by the SAP rules in the ERP system, and also guarantees that the input data are consistent with the information generated in the ERP system, and proceeds to data cleaning, which guarantees that the artifact does not suffer from errors resulting from a lack of consistent structure in the structure of the event log and the quality of the data [64].

The validation of the artifacts produced by the DSR methodology is carried out through measurement criteria, it is the most common artifact validation technique, the validation of the compliance of the first artifact to the appropriate methodology for the construction of the collection and billing process in a case study of the pharmaceutical sector of Colombia through process mining and business intelligence, taking into account that the structure of the event log complies with the criteria established by the SAP regulations in the ERP system and likewise it is guaranteed that the input data are consistent with the information generated in the ERP system. Likewise, the evaluation of the knowledge results, the effectiveness, the usefulness of the artifacts produced in the research, the validity in the quality of scientific research and the reliability of the research results is of utmost importance, which constitutes an integral aspect of research in design sciences.

4. Results

By taking the data related to the invoicing processes, these data are imported into the DISCO program and the process map of the invoicing process in the pharmaceutical sector of Colombia is obtained through process mining, which is observed below. continued in Figure 7.

Figure 7 illustrates the map of the invoicing processes in the pharmaceutical sector of Colombia, which allows us to visualize the real flow of the invoicing processes, taking into account the activities that are invoiced (RV), cancellations (RF), cash receipts (DZ-DG),

value credit notes (NC) and inventory credit notes (NV). In addition, the absolute frequency of the activities is shown, the frequency of RF was 164, RV 32.900, NV 2530, NC 4540, DZ 4923 and DG 1770. Taking into account the absolute frequency, the 3 main processes in the case study that occur most frequently according to the process map are Invoices (IR), Credit Notes (NC), and Cash Receipts (CR). Similarly, Figure 8 illustrates the relational model used to analyze the activities of the invoicing process under study.
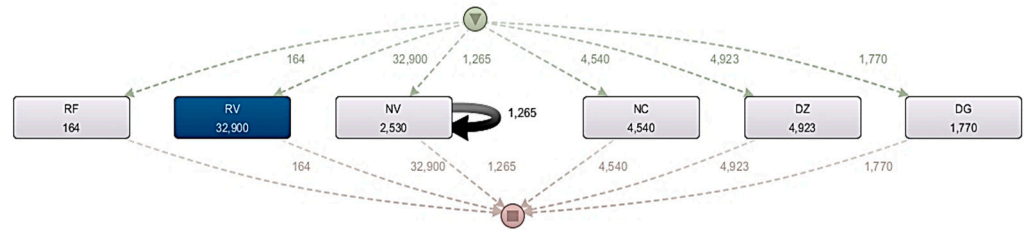

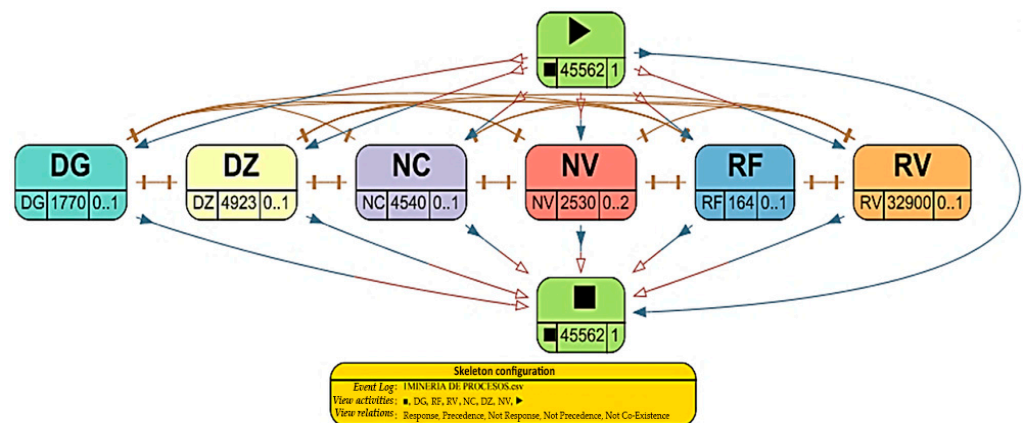**Figure 7.** Process map. Source: Own elaboration of the DISCO program.

Figure 8. Relational model of collection and billing processes. Source: Own elaboration of the DISCO program.

The relational model presented in Figure 8 allows us to analyze in more detail the workflow and the behavior of the activities related to the invoicing processes, these activities are Cash Receipts (DG-DZ), Value Credits (NC), Inventory Credits (NV), Cancellations (RF) and Invoices (RV), which according to the edges it is evident that they are all related to each other, It is also possible to observe the activities that lead to other events, it is possible to see where loops and repetitions occur, which are important to take into account because they can create bottlenecks, in the case study, in the invoicing process, the relational model does not allow to clearly show inefficiencies and bottlenecks, but it does provide important information about the workflow and the behavior of activities. The relational model represents the first version of the actual activity flowchart for the billing process in the case study.

Next, Figure 9 shows the frequency of the cases in the invoicing processes of the case study.

The frequency of cases shown in Figure 9 shows the frequency for each activity of the invoicing process; the frequency of RF was 0.36%, of RV was 72.2%, of NV was 2.78%, of NC was 9.96%, of DZ was 10.8%, and of DG was 3.88%. The frequency of cases provides valuable information about the efficiency of the processes, understanding the general distribution of cases, and identifying behavioral patterns in the processes. In the case study, a higher frequency is reflected in the activities RV invoices of 72.2%, followed by DZ cash receipts of 10.8% and NC value credit notes of 9.96%, likewise, the least frequent activities are RF cancellations of 0.36%, NV credit notes for inventory of 2.78% and DG cash receipts of 3.88%. When analyzing that the highest frequency of cases is reflected in the RV invoices of 72.2%, this indicates the efficiency and high performance of this activity in relation to

others, since invoices constitute a crucial activity in the billing process, also when observing the high frequency in DZ cash receipts of 10.8% represents efficient management in the collection process, likewise, the lower frequency of RF cancellations of 0. They do not reflect an inefficiency or a bottleneck in the invoicing process, since having a lower number of cancellations and credit notes for inventory reflects the efficiency in the invoicing process and does not generate risks in the invoicing process, which is the main activity of the process. In this paper, business intelligence facilitates the monitoring of the process derived from process mining through visual dashboards. In summary, the above figure helps to identify activities that could potentially become bottlenecks or an additional burden in their normal execution within the process. The frequency data allows the decision makers in this case study to determine which billing activities may be at risk of saturation and rework.
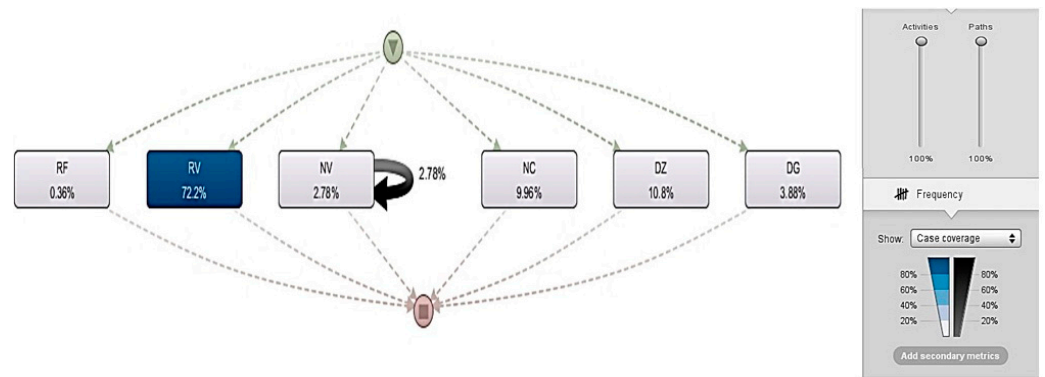

Figure 9. Frequency of cases. Source: Own elaboration of the DISCO program.

Figure 10 shows the business intelligence results developed in Oracle.
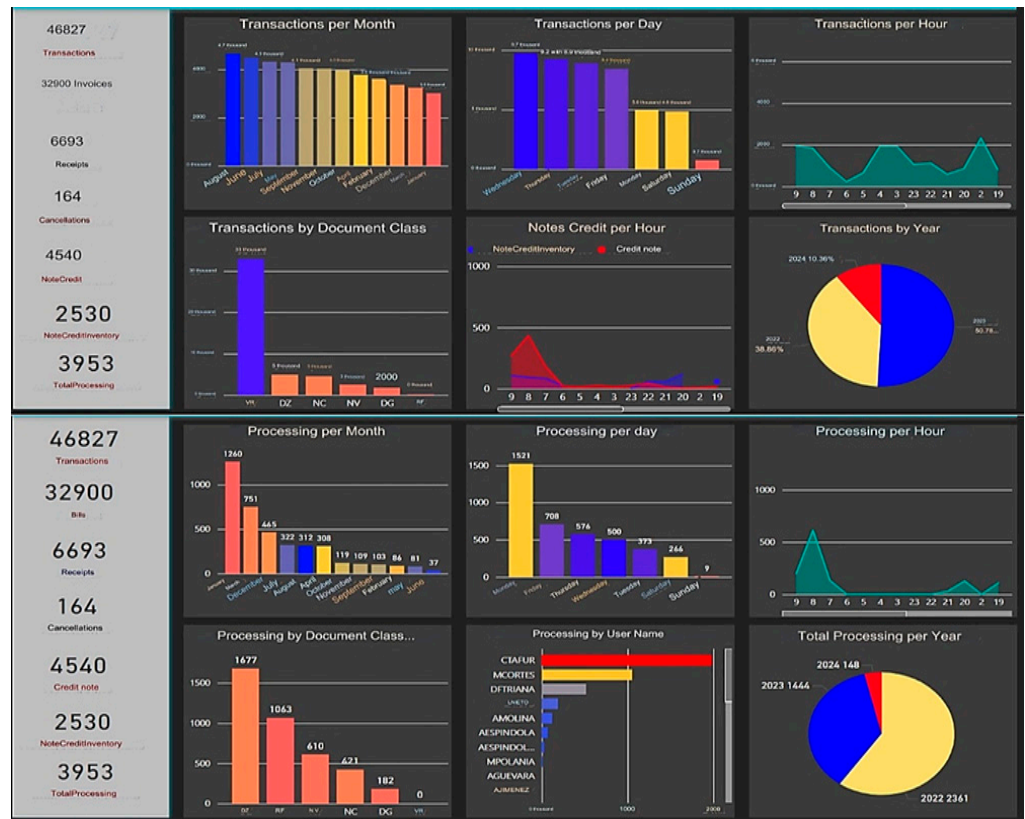

Figure 10. Frequency by items in the case study. Source: Own elaboration in Oracle dashboard.

Figure 10 shows the analysis of the results presented in the dashboard, highlighting several key aspects of transaction processing. A total of 46,827 transactions were recorded, divided into 32,900 invoices, 6693 receipts, 164 cancellations, 4540 credits, and 2530 inventory credits, for a total of 3953 processes.

The processing graph by month shows that the period with the highest number of transactions is January with 1260, followed by December with 751 and November with 465. The periods with the least activity were June and May, with 37 and 81 transactions, respectively. In terms of days of the week, Friday stands out with 1521 transactions, while Sunday is the least active day with only 9 transactions. The analysis by type of document shows that, in terms of the duration of the transactions, those of type DZ were the most frequent with 1677, followed by RF with 1063 and NV with 610. The least frequent were those of type RV, which did not record a longer duration in the transaction. The graph of processing times by user name shows that the user CTAFUR, followed by MCORTES and DFTRI-ANA have the lowest frequency of transaction registration times.

In the graph of processing time per hour, a significant peak in transactions is observed at 9 a.m., with sporadic activity at other times of the day. Finally, the graph of total processing by year shows that 2022 had the highest number of transactions with 2361, followed by 2023 with 1444, and 2024 with 148 at the time of the report. In summary, the results show clear patterns in the temporal and user distribution of transactions, highlighting certain months, days, and hours of higher activity, as well as certain users with higher workloads. This provides a solid foundation for making informed decisions about resource management and financial planning in the case study's invoicing process. Figure 11 shows the dashboards related to billing and the number of receipts.

Figure 11. Frequency of billing and receipts in the case study. Source: Own elaboration in Oracle dashboard.

Figure 11 shows that the results indicate that January and Monday are the periods with the highest activity in terms of transaction processing and billing, while June, July, and weekends show a significant decrease. The peak hours of activity are concentrated in the morning and early afternoon, with a significant decrease in the evening. The distribution by user shows that some users handle a significant volume of transactions, suggesting different roles within the organization. In addition, certain types of documents are processed more frequently, with the DZ class being highlighted. A comparison between the invoice and receipt levels shows a significant disparity, with a greater focus on invoices. These consistent temporal patterns at both levels indicate opportunities for transactional process optimization and a better understanding of organizational activity patterns.

5. Discussion

The construction of the invoicing process, a case study of the Colombian pharmaceutical sector, through process mining and business intelligence, presents unique challenges due to the complexity of regulations and the need for precision in the management of large volumes of data. In this context, process mining emerges as a critical tool to analyze and improve workflows by discovering patterns and bottlenecks in administrative activities. The application of process mining techniques can identify inefficiencies in invoicing management, optimizing the revenue cycle and improving process transparency and traceability. This approach not only facilitates compliance with strict regulations, but also improves customer satisfaction and company profitability by ensuring more agile and accurate processes.

In the specific case of the Colombian pharmaceutical sector, where the integration of technology and regulatory compliance is essential, Process Mining is positioned as an innovative strategy for invoicing management in pharmaceutical industries. In this section, the discussion of the construction of the invoicing process in the pharmaceutical sector of Colombia is addressed, using the statistics and graphs obtained in the DISCO program. The different results obtained by DISCO will be analyzed. Below, Figure 12 shows the total duration of the activities related to the invoicing process.
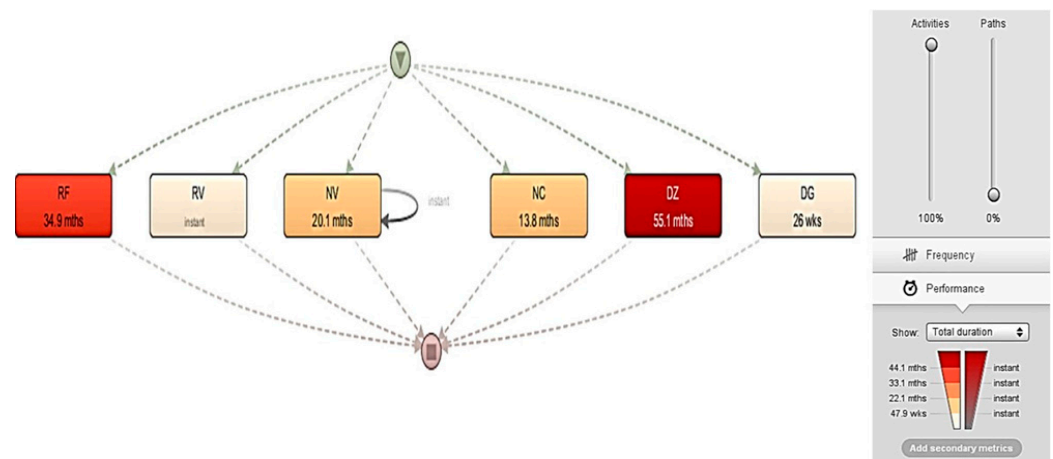

Figure 12. Total duration of activities. Source: Own elaboration of the DISCO program.

The total duration of the activities shown in Figure 12 indicates the duration of each activity, for RF the duration is 34.9 min, for RV the duration is instantaneous, for NV the duration is 20.1 min, for NC 13.8 min, DZ 55.1 min and DG 26. The total duration of the activities allows to identify the duration of the activities and reflects a higher duration in the DZ of 55.1 min, in the RF of 34.9 min and in the NV of 20.1 min, which reflects an inefficiency in these 3 activities that directly affects the invoicing process, creating a bottleneck in the organization. Likewise, activities with a shorter total duration such as RV, which is instantaneous, and DG, which is 26 times, reflect the efficiency and behavior in these activities. Figure 13 shows the frequency of activities in the invoicing processes.

Figure 13 allows us to observe the frequency of the activities of the invoicing processes, which reflects a higher frequency in the RVs with 70.26%, which are the invoices and the central axis of the processes analyzed, and also the frequency of the DZ and DG, which are the cash receipts with 12.28% of the records each. The apparent discrepancy between the high percentage of invoice records and the lower percentage of cash receipts may be due to payment consolidation, as customers pay multiple invoices at once, with individual records being grouped into a single payment. The difference in percentages does not indicate an error. It reflects the nature of the payment process, where customers consolidate payments and cash receipts record the entire transaction, not each invoice.
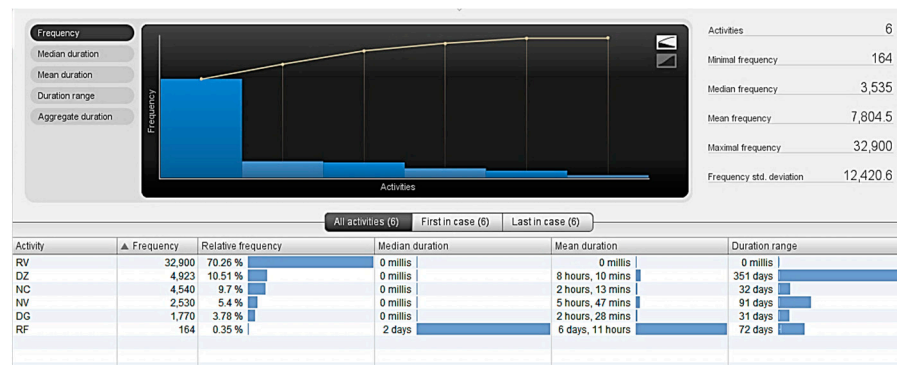
Figure 13. Frequency of activities. Source: Own elaboration of the DISCO program.

The Relative Frequency column shows the percentage of times each activity was performed out of the total number of activities. RV is the activity with the highest relative frequency at 70.26%. In the median duration of each activity, the column shows the RV activity, which has the shortest average duration, with 0 milliseconds, while the RF activity has the longest average duration, with 6 days and 11 h, RF corresponds to the cancellation of invoices, which is about avoiding errors in accounting facts due to the implications that this entails in the operation.

Frequency Standard Deviation: This column shows the standard deviation of the frequency of each activity. RV activity has the highest frequency standard deviation at 12,420.6, while RF activity has the lowest frequency at 164. It is also important to analyze the statistics by user, which is shown below in Figure 14.
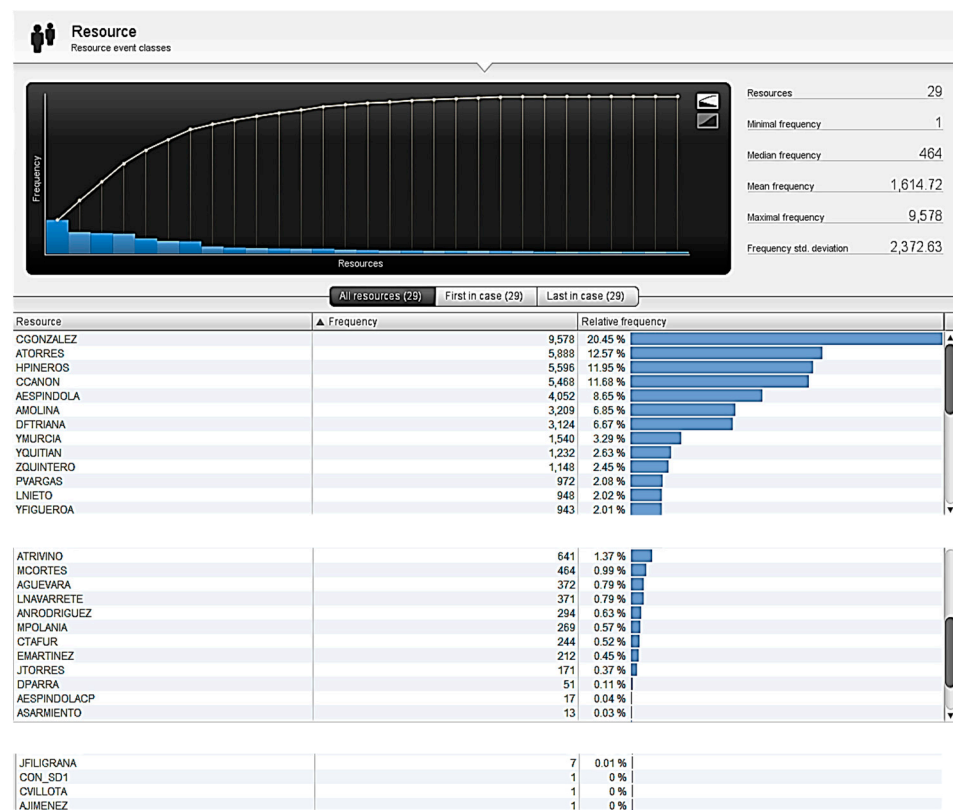

Figure 14. Statistics by users. Source: Own elaboration of the DISCO program.

The statistics by user are illustrated in Figure 14, which shows a dashboard containing statistics on the resource usage by users. The figure shows that the user CGONZALEZ has appeared in the largest number of events with a frequency of 9578 times. This represents

20.45% of the total number of occurrences of all resources. The next most frequent users are ATORRES, HPINEROS, CCANON, and AESPINDOLA, as well as the resources CGONZA-LEZ, ATORRES, HPINEROS, CCANON, and AESPINDOLA that have been used the most by users.

Likewise, users with lower percentages correspond to users who have been connected to the organization for a short time. It is important to mention that a bottleneck is observed in the statistics of users, in the user CGONZALEZ of 20.45% regarding the frequency of other users of the organization, this information is key to identify the users who intervene in the process and be able to track the responsibility of users in the process.

The table of resource event classes shows that the minimum frequency of a resource appearing in an event is 1 time. The median frequency is 464 times, the average frequency is 1614.72 times, and the maximum frequency is 9578 times. The standard deviation of the frequency is 2372.63.

The designed process revealed that the flowchart developed by the organization in the case study did not follow the suggested order of activities. In addition, the flowchart did not explicitly highlight control points and reported bottlenecks. Leveraging the process mining and business intelligence paradigm can provide several benefits, including the development of generic process models. The standardized representation of a process model allows the efficient use of a wide range of search algorithms with minimal effort, and process mining algorithms employ heuristics that can efficiently scale any process.

Based on these advantages, several recommendations for professional researchers are necessary: it is crucial to identify the conditions under which process mining is feasible, such as the availability of data. Furthermore, it is important to explore the integration of process mining and business intelligence in dynamic process environments that require adaptability. It is recommended that the event log be structured declaratively, either by extracting attributes from a structured database or by other means that can improve the execution time of mining algorithms. It is also important to investigate the implementation of more advanced mining algorithms that incorporate large language models (LLMs) and other techniques such as automated scheduling. The practical result of this research is the automation of an artifact that represents the model of the real process aligned with the activities executed in real time. In this way, changes in the order of activities or the timeline would allow the artifact to provide an updated and flexible process in the face of emerging changes. These recommendations aim to improve the application of process mining and business intelligence by exploiting flexibility and reasoning power in dynamic organizational environments.

6. Conclusions and Future Works

The results obtained in this research work through the different process mining tools allow us to identify the inefficiencies in the invoicing processes that can be generated in the activities related to the invoicing processes.

The construction of the invoicing process, a case study of the pharmaceutical sector in Colombia, through process mining and business intelligence based on its application and results, allows a real description of the invoicing processes and the activities involved. in these processes and identify their inefficiencies.

An artifact related to the method for constructing the real invoicing process in the pharmaceutical sector industries in Colombia is obtained, which allowed it to be validated in a case study. The contribution of this research focuses on the development of a new approach to build models of invoicing processes supported by process mining, which leads to filling the gap in modeling these accounting processes in organizations from an alternative methodology and technique to the flowcharts pre-established by biases and regulations.

Validation in design science is a process that ensures the quality, effectiveness and relevance of the artifacts, these artifacts for the case study are what is the appropriate methodology and how to validate the methodology for the construction of the invoic-

ing process in a case study of the pharmaceutical sector in Colombia, Through process mining and business intelligence, this validation will be carried out with validation by measurement criteria, taking into account the structure of the event log and compliance with SAP regulations in the systems ERP and the consistency of the input data with the information generated by the ERP, which guarantees the consistency of the event log and the quality of the data. Likewise, it is important to evaluate the results of the knowledge, the effectiveness, the usefulness of the artifacts produced in the research, the validity of the quality of scientific research and the reliability of the research results, which is an integral aspect of design science research.

Process Mining is an important tool for organizations because it allows to discover the real execution model of the processes, to determine the compliance in the processes with the established guidelines, to discover the existing bottlenecks, to monitor the productivity of the personnel who execute the different processes, to predict the execution times in the processes and to determine the relationship between different variables related to the processes.

From a business intelligence perspective, it complements the results of process mining as a prelude to considering process design. The modeling of processing and billing data reveals significant temporal patterns in the organization's transactional activity. January and Mondays stand out as the busiest periods, while June, July, and weekends show a significant decrease in transaction volume. Peak hours are concentrated in the morning and early afternoon, with a decrease in the evening. In addition, the uneven distribution of work across users and document types suggests potential areas for load balancing and process optimization. The comparison between invoice and receipt levels shows a greater focus on invoices, suggesting a possible difference in operational priorities. These findings provide a solid foundation for implementing process management optimization and improvement strategies within the organization, thereby improving the efficiency and effectiveness of day-to-day operations.

The limitation of this research is related to the application of a single case study for invoicing processes in organizations in the pharmaceutical sector in Colombia, for which future work must carry out more case studies in the pharmaceutical sector and compare with cross-sectional studies with countries in the same region to evaluate the modeling of the billing process and real collection with references to the particularities of culture and regulations, which allows complementing the process modeling approach by process mining.

It should be noted that the application of Process Mining requires information available in organizations, for which the existence of ERP information systems that function as data repositories that can be extracted by the methodology proposed in this article is required, which implies that for the size of the Colombian context, the processes that can be modeled by Process Mining are in medium and large organizations.

It is also important to highlight that, in many cases, the time in which the data are recorded in the information systems does not correspond to the actual execution time of the processes, which reduces the reliability of the data and information available.

For future lines of research, the design of all processes developed in an organization can be carried out and other process mining tools or algorithms can be used. Carry out studies on the integration of process mining in organizations as a management and optimization tool for processes and activities, as well as allowing the redesign of processes and implementation of required changes. This research is part of the Accounting Theory research line of the University of Manizales.

References

1. Besson, P.; Rowe, F. Strategizing information systems-enabled organizational transformation: A transdisciplinary review and new directions. *J. Strateg. Inf. Syst.* **2012**, *21*, 103–124. [CrossRef]
2. Rivas, M.H.; Bayona-Oré, S. Algoritmos de Minería de Proceso para el Descubrimiento Automático de Procesos. *Rev. Ibérica Sist. Tecnol. Informação* **2019**, *31*, 33–49. [CrossRef]
3. Van der Aalst, W.M.P. *Process Mining: Discovery, Conformance and Enhancement of Business Process*; Springer Science & Business Media: Berlin/Heidelberg, Germany; Dordrecht, The Netherlands; London, UK, 2011.
4. SAP Spain Centro de Noticias. Impacto de un Software de Gestión Empresarial en la Transformación Digital-SAP España News Centerfrom. 2021. Available online: https://news.sap.com/spain/2021/10/impacto-de-un-software-de-gestion-empresarial-en-la-transformacion-digital/ (accessed on 7 June 2024).
5. STEL Order. ¿Qué es un ERP? Definición, Tipos, Ventajas E Inconvenientes. 2023. Available online: https://www.stelorder.com/blog/que-es-erp/ (accessed on 7 June 2024).
6. Castro Game, J.K.; Marila Isabel, C.; Castro, G.; Geovanna, P.; Córdova, G. Software informático como alternativa para la gestión contable. *RECIMUNDO* **2023**, *7*, 186–196. [CrossRef]
7. Hugo, A.; Rincón, N. Minería de procesos: Desarrollo, aplicaciones y factores críticos. *Cuad. Adm.* **2015**, *28*, 137–157.
8. Lapa, J.; Bernardino, J.; Figueiredo, A. A comparative analysis of open source business intelligence platforms. In *ACM International Conference Proceeding Series*; ACM: Lisbon, Portugal, 2014. [CrossRef]
9. Monsalve, B.J.F. Principales Desafíos de la Industria Farmacéutica 2023 | LinkedIn. Principales Desafíos de la Industria Farmacéutica. 2022. Available online: https://www.linkedin.com/pulse/principales-desaf%C3%ADos-de-la-industria-farmac%C3%A9utica-2023-monsalve/ (accessed on 2 January 2023).
10. Velasco, C. Industria Farmacéutica en Colombia 2023: Tendencias. 2023. Available online: https://cercal.co/envinculo/industria-farmaceutica-en-colombia-2023/ (accessed on 2 May 2024).
11. Svieta, L.; Gallegos, L. Aplicación de Una Técnica de Minería de Procesos para el Modelamiento del Proceso de Ventas de Una Empresa Textil en Arequipa Tesis Presentada por la Bachiller. Universidad Nacional de San Agustín de Arequipa Escuela de Posgrado Unidad de Posgrado Facultad de Ingeniería de Producción y Servicios. 2018. Available online: https://repositorio.unsa.edu.pe/items/3e5b9399-b514-44b4-80f6-6ed46a6df675/ (accessed on 2 May 2024).
12. Inverbis Analytics. Cómo Usar Process Mining en Salud: Entrevista con Datasalus. 2022. Available online: https://web.inverbisanalytics.com/es/entrevista-datasalus-process-mining-salud/ (accessed on 10 March 2024).
13. Van der Aalst, W.M.P.; Van Dongen, B.F.; Herbst, J.; Maruster, L.; Schimm, G.; Weijters, A.J.M.M. Workflow mining: A survey of issues and approaches. *Data Knowl. Eng.* **2003**, *47*, 237–267. [CrossRef]
14. Meneses, A. Generalidades de las Redes de Petri. 2002. Available online: https://computacion.cs.cinvestav.mx/~ameneses/pub/tesis/mtesis/node6.html (accessed on 15 August 2024).
15. Caballero, F. Introducción a BPM | Mindware SRL. 2015. Available online: https://mindwaresrl.com/2015/08/05/introduccion-a-bpm/ (accessed on 15 August 2024).
16. Sydle. BPM, BPMN y BPMS: ¿Qué significan? Comprende Cada Término | Blog SYDLE. 2023. Available online: https://www.sydle.com/es/blog/bpm-bpmn-bpms-60ba98c3a5c829237349b32f (accessed on 15 August 2024).
17. Amblard-Ladurantie, C. ¿Qué es la Minería de Procesos? Aquí Tiene Una Guía | MEGA International. 2023. Available online: https://www.mega.com/es/blog/que-es-la-mineria-de-procesos-guia (accessed on 2 February 2024).
18. Gregorio, T. La Minería de Procesos Como Herramienta Para Generar Ventajas Competitivas-Stratesys | Consultoría Tecnológica | Consultoría Estratégica | SAP. 2023. Available online: https://www.stratesys-ts.com/es/la-mineria-procesos-herramienta-generar-ventajas-competitivas/ (accessed on 5 June 2024).
19. Rodríguez, A.; Romero Castro, V.F.; Rodríguez Gonzalez, A.D.C.; Cabezas Baque, N.A.; Pino Tarragó, J.C. Aplicaciones de la Inteligencia Artificial en técnicas de minería de procesos. *Ser. Científica Univ. Cienc. Informáticas* **2021**, *14*, 136–155. Available online: https://dialnet.unirioja.es/servlet/articulo?codigo=8590663&info=resumen&idioma=SPA (accessed on 30 March 2024).
20. Rouhiainen, L. Inteligencia Artificial 101 Cosas Que Debes Saber Hoy Sobre Nuestro Futuro Inteligencia Artificial. 2018. Available online: https://www.amazon.com/-/es/Lasse-Rouhiainen/dp/8417568085 (accessed on 12 May 2024).
21. Camargo-Vega, J.J.; Camargo-Ortega, J.F.; Joyanes-Aguilar, L.; Camargo-Vega, J.J.; Felipe, J.; Joyanes-Aguilar, C.-O.-L. Conociendo Big Data. *Rev. Fac. Ing.* **2015**, *24*, 63–77. Available online: https://www.researchgate.net/publication/304208360_Conociendo_Big_Data (accessed on 7 July 2024). [CrossRef]

22. Arreaga Serpa, J.D. Evolución de las Tecnologías de la Inteligencia de Negocios y la Toma de Decisiones Administrativas. Universidad Católica de Cuenca, 2024. Available online: https://dspace.ucacue.edu.ec/handle/ucacue/17535 (accessed on 2 April 2024).

23. Agrawal, R.; Gunopulos, D.; Leymann, F. Mining Process Models from Workflow Logs. In *International Conference on Extending Database Technology*; Springer: Berlin/Heidelberg, Germany, 1998.

24. Cook, J.E.; Wolf, A.L. Discovering models of software processes from event-based data. ACM Transactions on Software Engineering and Methodology. *ACM Trans. Softw. Eng. Methodol.* **1998**, *7*, 215–249. [CrossRef]

25. Van Der Aalst, W.M.; Reijers, H.A.; Weijters, A.J.; van Dongen, B.F.; De Medeiros, A.A.; Song, M.; Verbeek, H.M.W. Business process mining: An industrial application. *Inf. Syst.* **2007**, *32*, 713–732. [CrossRef]

26. Mans, R.S.; Schonenberg, M.H.; Song, M.; van der Aalst, W.M.; Bakker, P.J. Application of process mining in healthcare. A case study in a Dutch Hospital. In *Biomedical Engineering Systems and Technologies: International Joint Conference, BIOSTEC 2008 Funchal, Madeira, Portugal, 28–31 January 2008*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 425–438.

27. Mǎruşter, L.; Van Beest, N.R.T.P. Redesigning business processes: A methodology based on simulation and process mining techniques. *Knowl. Inf. Syst.* **2009**, *21*, 267–297. [CrossRef]

28. Rozinat, A.; de Jong, I.S.; Günther, C.W.; van der Aalst, W.M. Process mining applied to the test process of wafer steppers in ASML. *IEEE Trans. Syst. Man Cybern.* **2010**, *39*, 474–479. [CrossRef]

29. Jans, M.; Van Der Werf, J.M.; Lybaert, N.; Vanhoof, K. A business process mining application for internal transaction fraud mitigation. *Expert Syst. Appl.* **2011**, *10*, 13351–13359. [CrossRef]

30. van der Aalst, W.M.P.; Schonenberg, M.H.; Song, M. Time prediction based on process mining. *Inf. Syst.* **2011**, *36*, 450–475. [CrossRef]

31. Rebuge, Á.; Ferreira, D.R. Business process analysis in healthcare environments: A methodology based on process mining. *Inf. Syst.* **2012**, *37*, 99–116. [CrossRef]

32. Aguirre, S.; Parra, C.; Alvarado, J. Combination of process mining and simulation techniques for business process redesign: A methodology approach. In *Data-Driven Process Discovery and Analysis: Second IFIP WG 2.6, 2.12 International Symposium, SIMPDA 2012, Campione d'Italia, Italy, 18–20 June 2012*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 24–43.

33. De Weerdt, J.; Schupp, A.; Vanderloock, A.; Baesens, B. Process Mining for the multi-faceted analysis of business processes-A case study in a financial services organization. *Comput. Ind.* **2013**, *64*, 57–67. [CrossRef]

34. Carlos Andrés Villacreses Parrales. Aplicación de Minería de Procesos Para la Verificación de la Conformidad del Workflow de Producción en la Compañía EuroFish. 2023. Available online: https://repositorio.unesum.edu.ec/handle/53000/5502 (accessed on 10 April 2024).

35. Van der Aalst, W.M.; Weijters, A.J. Process mining: A research agenda. *Comput. Ind.* **2004**, *53*, 231–244. Available online: https://www.researchgate.net/publication/222696704_Process_mining_A_research_agenda (accessed on 20 February 2024). [CrossRef]

36. Yaksilik, D.; Sakipova, T. Algoritmos y técnicas de descubrimiento de procesos poco estructurados: Estado del arte Algorithms and discovering techniques for unstructured processes: State of art. *Rev. Cuba. Cienc. Informáticas* **2014**, *8*, 121–148.

37. Van der Aalst WM, P.; Mylopoulos, J.; Ram, S.; Rosemann, M.; Szyperski, C. *Process Mining Handbook*; Springer: Berlin/Heidelberg, Germany, 2022.

38. Weijters, A.J.M.M.; Van der Aalst, W.M.P. Process Mining with the Heuristics Miner-Algorithm. 2006. Available online: https://www.researchgate.net/publication/229124308_Process_Mining_with_the_Heuristics_Miner-algorithm (accessed on 20 February 2024).

39. Castillo, E.C.E. *Desarrollo de Aplicación Web Para Gestionar las Tareas Asignadas a Equipos de Desarrollo de Software en la Empresa Ptesa Implementando un Algoritmo Heurístico*; Universidad Piloto de Colombia: Bogota, Colombia, 2020.

40. De Medeiros, A.K.A.; Weijters, A.J.M.M.; Van Der Aalst, W.M.P. Genetic process mining: An experimental evaluation. *Data Min. Knowl. Discov.* **2007**, *14*, 245–304. [CrossRef]

41. De Weerdt, J.; De Backer, M.; Vanthienen, J.; Baesens, B. A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Inf. Syst.* **2012**, *37*, 654–676. [CrossRef]

42. Completo, J.; Cruz, R.S.; Coheur, L.; Delgado, M. Design and Implementation of a Data Warehouse for Benchmarking in Clinical Rehabilitation. *Procedia Technol.* **2012**, *5*, 885–894. [CrossRef]

43. Gehrke, N. Basic Principles of Financial Process Mining a Journey through Financial Data in Accounting Information Systems. 2010. Available online: https://aisel.aisnet.org/cgi/viewcontent.cgi?referer=&httpsredir=1&article=1281&context=amcis2010 (accessed on 20 February 2024).

44. Butt, N.A.; Mahmood, Z.; Sana, M.U.; Díez, I.D.L.T.; Galán, J.C.; Brie, S.; Ashraf, I. Behavioral and Performance Analysis of a Real-Time Case Study Event Log: A Process Mining Approach. *Appl. Sci.* **2023**, *13*, 4145. [CrossRef]

45. Garcia, C.L. Casos de Uso de Minería de Procesos en El Sector Bancario. 2023. Available online: https://www.linkedin.com/pulse/desbloqueando-la-innovaci%C3%B3n-casos-de-uso-miner%C3%ADa-en-el-garc%C3%ADa-cabria/ (accessed on 10 February 2024).

46. Harmon, P. Business Process Change: A Business Process Management Guide for Managers and Process Professionals. Morgan Kaufmann. Available online: https://books.google.com.co/books?hl=es&lr=&id=S-WKDwAAQBAJ&oi=fnd&pg=PP1&dq=+Implementation+of+Business+Process+Management+Systems+for+Improving+Billing+and+Collection+Processes:+A+Case+Study+in+the+Services+Sector.+International+Journal+of+Process+Management+and+Benchmarking,+9(4),+412%E2%80%93429.+&ots=pH1CXZY2Ef&sig=wYFScFPPqCH30LOeRXVET3rN7Ng&redir_esc=y#v=onepage&q&f=false (accessed on 20 February 2024).
47. Zuhaira, B.; Ahmad, N. Business process modeling, implementation, analysis, and management: The case of business process management tools. *Bus. Process Manag. J.* **2021**, *27*, 145–183. [CrossRef]
48. Urrea-Contreras, S.J.; Flores-Rios, B.L.; Astorga-Vargas, M.A.; Ibarra-Esquer, J.E.; Burtseva, L. Adopción de Herramientas de Minería de Procesos en la Ingeniería de Software. *Abstr. Appl.* **2018**, *20*, 24–37.
49. Pavón-González, Y.; Ortega-González, Y.C.; Infante-Abreu, M.B.; Souchay-Alzugaray, S.; Cobiellas-Herrera, L.M. Conceptual modeling method of business processes at ontological and situational levels with enterprise architecture scope. *DYNA* **2021**, *88*, 227–236. [CrossRef]
50. Sanchis, R.; Poler, R.; Ortiz, Á. Técnicas para el Modelado de Procesos de Negocio en Cadenas de Suministro. *Inf. Tecnológica* **2009**, *20*, 29–40. [CrossRef]
51. Numminen, L. Minería de Procesos Algoritmos Explicados de Forma Sencilla-Workfellow. 2023. Available online: https://www.workfellow.ai/es/learn/process-mining-algorithms-simply-explained (accessed on 30 April 2024).
52. Yzquierdo Herrera, C. Ciencias de la Información. 2013, p. 44. Available online: http://www.redalyc.org/articulo.oa?id=181430077003 (accessed on 25 March 2024).
53. Chapela-Campa, D.; Dumas, M.; Mucientes, M.; Lama, M. Efficient edge filtering of directly-follows graphs for process mining. *Inf. Sci.* **2022**, *610*, 830–846. [CrossRef]
54. van der Aalst, W.M. Object-centric process mining: Unraveling the fabric of real processes. *Mathematics* **2023**, *11*, 2691. [CrossRef]
55. Fantinato, M.; Peres, S.M.; Reijers, H.A. X-Processes: Process model discovery with the best balance among fitness, precision, simplicity, and generalization through a genetic algorithm. *Inf. Syst.* **2023**, *119*, 102247. [CrossRef]
56. Rebmann, A.; van der Aa, H. Enabling semantics-aware process mining through the automatic annotation of event logs. *Inf. Syst.* **2022**, *110*, 102111. [CrossRef]
57. Vom Brocke, J.; Weber, M.; Grisold, T. Design Science Research of High Practical Relevance. In *Engineering the Transformation of the Enterprise: A Design Science Research Perspective*; 2021. Available online: https://link.springer.com/chapter/10.1007/978-3-030-84655-8_8 (accessed on 30 April 2024).
58. Dresch, A.; Lacerda, D.P.; Antunes, J.A.V. Design science research: A method for science and technology advancement. In Design Science Research: A Method for Science and Technology Advancement. 2015. Available online: https://www.researchgate.net/publication/265467276_Design_Science_Research_A_Method_for_Science_and_Technology_Advancement (accessed on 30 April 2024).
59. Zhao, W.; Zhao, X. Process mining from the organizational perspective. In *Foundations of Intelligent Systems: Proceedings of the Eighth International Conference on Intelligent Systems and Knowledge Engineering, Shenzhen, China, Nov 2013 (ISKE 2013)*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 701–708.
60. Salazar Hernández, R.; Peralta Miranda, P. Una mirada a la industria farmacéutica en Colombia. *Rev. Fac. Cienc. Contab. Económicas Adm.-FACCEA* **2014**, *4*, 107–115. [CrossRef]
61. Arias, M.; Rojas, E. Guía para gestionar procesos de negocio a través de minería de procesos. *InterSedes* **2016**, *17*, 36. [CrossRef]
62. Van der Aalst, W.M.; van Dongen, B.F.; Günther, C.W.; Rozinat, A.; Verbeek, H.M.W.; Weijters, A.J.M.M. Prom: The process mining toolkit. In Proceedings of the BPM 2009 Demonstration Track (BPMDemos 2009), Ulm, Germany, 8 September 2009; pp. 1–4.
63. Van der Aalst, W.M.P. Process mining: Discovering and improving Spaghetti and Lasagna processes. In Proceedings of the 2011 IEEE Symposium on Computational Intelligence and Data Mining CIDM, Paris, France, 11–15 April 2011; pp. 1–7.
64. Vom Brocke, J.; Winter, R.; Hevner, A.; Maedche, A. Special issue editorial–accumulation and evolution of design knowledge in design science research: A journey through time and space. *J. Assoc. Inf. Syst.* **2020**, *21*, 9. [CrossRef]