Generic Services Model for the Automatic Improvement of Business Processes

Modelo de servicios genéricos para la mejora automática de los procesos empresariales

Valentina Vera Paz
Universidad del Cauca. Popayán (Colombia)
valentinav@unicauca.edu.co

Anyi Aracely Ramírez Obando
Universidad del Cauca. Popayán (Colombia)
anyiramirez@unicauca.edu.co

Carlos Alberto Cobos Lozada
Universidad del Cauca. Popayán (Colombia)
ccobos@unicauca.edu.co

Resumen

Introducción— Las organizaciones requieren día a día más productividad y eficiencia en sus procesos de negocio. En la actualidad se cuenta con diversas herramientas que brindan soporte para optimizar tiempos y recursos de acuerdo con la complejidad de las actividades de los procesos de negocio. Sin embargo, al automatizar procesos, pocas empresas logran definir un flujo de trabajo exitoso, por lo que no pueden prever las dificultades que surgen en un entorno de producción. En consecuencia, no es posible dar una solución anticipada a los problemas, lo que implica sobrecostos, pérdida de tiempo y en algunos casos afectación al talento humano de la organización.

Objetivo— En este artículo se presenta un modelo de servicios genérico para la mejora automática de procesos de negocio que permite identificar cuellos de botella, reprocesos, fallas y retrasos al analizar los logs de eventos de un proceso de negocio. También resume la implementación del servicio de gestión de cuellos de botella para soportar la toma de decisiones en un proceso simulado, aplicando modelos de regresión para predecir el rendimiento de las actividades manuales del proceso con base en los retrasos y la longitud de las colas. Mediante la predicción del rendimiento y la elaboración de sugerencias de asignación de recursos, se determinó el nivel de mejora del proceso.

Metodología— La investigación se realizó siguiendo el Patrón de Investigación Iterativa propuesto por Pratt. Primero se identificaron los principales problemas en la gestión de procesos, luego se realizó una revisión del estado del arte para conocer las propuestas de solución a estos problemas, después se propuso un modelo de solución independiente del software de gestión de procesos que se use y finalmente se realizaron dos evaluaciones, una a nivel conceptual con la técnica de grupo focal y la otra basada en la implementación de uno de los servicios propuestos y datos recolectados de un experimento en un simulador de procesos de negocio.

Resultados— La evaluación conceptual de los servicios propuestos en el modelo se realizó por parte de un grupo de expertos, con base en los lineamientos de diseño y contenido del estándar de la nomenclatura de modelado BPMN, otorgando una calificación de 4,8 sobre 5,0 para cada servicio. La experimentación con el simulador de procesos de negocio y las recomendaciones entregadas por el servicio implementado (gestión de cuellos de botella) permitió evaluar la reducción en tiempo del procesamiento de las instancias de un proceso en relación con los recursos adicionados.

Conclusiones— El modelo propuesto este compuesto de tres servicios principales, la gestión de cuellos de botella, la gestión de recursos y la gestión de entradas. El primer servicio ayuda a establecer las medidas correctivas para que el proceso fluya y las instancias de este no se encolen en tareas o actividades específicas, lo que ayuda a mejorar el tiempo de respuesta y la calidad del servicio. La gestión de recursos busca optimizar el tiempo de ejecución de las actividades manuales y la gestión de entradas busca asegurar que una instancia del proceso cuente con los datos y documentos requeridos para ser procesado de inicio a fin, evitando reprocesos y mejorando la calidad de los datos que se reciben y procesan.

Palabras clave— Procesos de Negocio; Minería de Procesos; Minería de Datos; Cuellos de Botella; Servicios

Abstract

Introduction— Organizations require more productivity and efficiency in their business processes every day. Currently, various tools provide support to optimize time and resources according to the complexity of the activities of the business processes. However, by automating processes, few companies can define a successful workflow, thus failing to anticipate the difficulties in a production environment. Consequently, it is impossible to provide an early solution to problems, which implies cost overruns, loss of time, and in some cases, affectation of the organization's human talent.

Objective— This article presents a generic service model for the automatic improvement of business processes that allows identifying bottlenecks, reprocesses, failures, and delays when analyzing the event logs of a business process. It also summarizes the implementation of the bottleneck management service to support decision-making in a simulated process, applying regression models to predict the performance of manual process activities based on delays and queue lengths. By predicting performance and making resource allocation suggestions, the level of process improvement was determined.

Methodology— The research was conducted following the Iterative Research Pattern proposed by Pratt. First, the main problems in process management were identified, then a review of the state of the art was carried out to find out the proposed solutions to these problems. A solution model independent of the process management software was used was proposed, and finally, two evaluations were carried out, one at a conceptual level with the focus group technique and the other based on the implementation of one of the proposed services and data collected from an experiment in a business process simulator.

Results— The conceptual evaluation of the services proposed in the model was conducted by a group of experts, based on the design and content guidelines of the BPMN modeling nomenclature standard, giving a rating of 4.8 out of 5.0 for each service. Experimentation with the business process simulator and the recommendations provided by the implemented service (bottleneck management) made it possible to evaluate the reduction in the processing time of the instances of a process in relation to the added resources.

Conclusions— The proposed model is composed of three main services, bottleneck management, resource management, and input management. The first service helps to establish the corrective measures so that the process flows and the instances of this do not get stuck in specific tasks or activities, which helps to improve the response time and the quality of the service. Resource management seeks to optimize the execution time of manual activities and input management seeks to ensure that an instance of the process has the data and documents required to be processed from start to finish, avoiding reprocessing and improving the quality of the data received and processed.

Keywords— Business Process; Process Mining; Data Mining; Bottleneck; Services

I. Introduction

Process mining is a discipline where the analysis of business processes developed by organizations is carried out, to generate greater profits and profitability. It consists of three main phases: Discovery, Compliance, and Improvement, which provide, respectively, the possibility to observe, monitor and control the behavior of the processes, as well as to optimize their performance, by solving the problems identified during the analysis of their performance [1].

Within organizations it is important to consolidate, in a clear and detailed way, the information of the processes, obtaining step by step the tasks that are carried out in each of the work areas. However, one of the risks that arise in the construction of the Business Model of a process, is that it is not able to predict the difficulties that arise in a real environment (or its use in production) and, consequently, cannot give an advance solution to such difficulties or problems, which implies over costs, demands, loss of time and in some cases the impact of both human and technical resources of the organization. The phase within process mining, which solves these problems, is the Improvement phase, which identifies bottlenecks, reprocesses, failures, and delays when analyzing the event logs of a process. These event logs enable you to evaluate process behavior and are the primary source of data that facilitates the integration of data analytics techniques for optimizing the tasks that run in each process. To achieve this, it is necessary to define a model that generates generic services, which allow the automation of processes at the enterprise level, supporting the problems that may arise in the daily life, and thereby achieve that the disadvantages are predictable and can be corrected in time, achieving the greatest benefit in the organization by minimizing risks.

Today, organizations have chosen to employ modeling standards that help maintain order in business processes. Therefore, it is increasingly common to use business process tools that allow visualizing the different tasks, resources and constraints that are part of the workflow. However, the methodologies used in Modeling and Business Process Notation (BPMN) have been oriented to identify the specific actions of a process. Some modeling tools allow simulation of a process, which counts the time each process instance takes to determine bottlenecks.

In the literature there are several proposals based on process mining which identify the main research topics and the most used algorithms. In addition, there are several investigations that make proposals for process discovery algorithms, and others specialized in compliance verification, architectural improvements, and tools [2]. Although this type of initiatives helps define the central objectives of process mining, there is still no general approach that provides the necessary support for decision-making and is directed at continuous improvement of processes.

Motivated by the above, this article presents a model that integrates a set of generic services for the automatic improvement of business processes. The services were modeled in BPMN nomenclature and represented as processes. The services defined in this research are: 1) Bottleneck management, 2) Resource management and, 3) Ticket management. It is expected that these services can be embedded transversally on any tool that executes and monitors business processes of any type in organizations.

These services were reviewed by a group of experts in a first Focus Group where comments, comments and suggestions were received and then in a second Focus Group evaluation. After that, a bottleneck management service implementation was performed and a business process simulation was used in Process Simulator, from which the execution log was obtained. Then, a regression model was applied to determine the delays generated in the manual activities of the process, a recommendation was made and then the result of the suggested improvement was evaluated. Also, a user interface was designed to show the results obtained in the execution of the service.

II. Theoretical Context

Business process mining is an area of Research, Development, and Innovation (R&D&I) that aims to discover, monitor, and improve business processes, obtaining information from the event logs that are extracted from the execution of each process that is monitored in an information system. In process mining, there are three fundamental phases to analyze a business process

[4]: 1) Process discovery: Consists in defining the model of a process based on the analysis of the current execution of a real process, such analysis is performed on the data of the event log (logs) of the process. This enables you to determine the actual way the process and business-related resources are executed [1], [4]. 2) Compliance Check: Seeks to evaluate whether the behavior of a running process matches the previously defined behavior. And 3) Improvement or extension of processes: This phase seeks to analyze the information extracted from the execution records of a process. For example, using data about the execution time of a business process's tasks/activities, stored in the event log (execution log), you can analyze the performance of the process from a time perspective or discover an activity that delays the execution of the process (bottleneck). In addition, the effect of certain elements on a decision-making point in an activity can be analyzed from the analysis of previously processed cases [4].

For organizations it is important to have a well-defined business model that helps establish workflows that determine the proper functioning of a process within the company. In addition to being able to establish a process model adjusted to reality, it is also essential to detect what faults are committed when the process moves from a manual version to an automated or semi-automated computer. It is currently evident that more companies integrate mining workflow to study the behavior of events and to find the most appropriate model to manage the business process.

III. State of the Art

In the literature there are several proposals based on process mining which identify the main research topics and the algorithms most used [2].

In 2009, ASML, the world leader in chip manufacturing and a key supplier to the industry, conducted a case study from the event logs of one of its scanner equipment, with the aim of improving the testing process of these machines [4]. The test process consisted of three phases: 1) the calibration phase; 2) the test phase (the actual test); and 3) the final qualification phase. This research sought to minimize testing time, as this generated cost overruns and sales losses, without leaving aside the quality of the hardware. ProM was used to perform the respective analysis by converting the data to an MXML4 format, where data fields, timestamps, and transactional information (i.e., whether the activity was scheduled, started, or completed) were tracked. After finishing the process, it was obtained as a result that the model discovered is "spaghetti type", that is, it is huge and unstructured. These spaghetti models were not caused by limitations of the algorithm but by the inherent complexity of the testing process.

In 2012, an introduction to process mining is made explaining each phase of it with their respective examples [5]. Regarding the improvement phase in process mining, the article reports that it is possible to extend or correct a process model by making use of the event log. A non-eligible process model can be corrected using the diagnostics provided by the model alignment and logging, thus representing the event log in the model, and analyzing the additional attributes of the model. For example, you can analyze wait times between activities, measuring the time difference between causally related events, and calculating basic statistics such as averages, variations, and confidence intervals. In this way, the main bottlenecks can be identified. Similarly, resource data can be used to discover roles, that is, groups of people frequently performing related activities, using standard clustering techniques. It is also possible to build social networks based on workflow and analyzing resource performance, for example, the relationship between workload and service times. Finally, standard classification techniques that analyze the decision points in the process model are used to build a decision tree that explains the observed behavior.

In 2013, an approach was developed to find delays in a purchasing process based on information provided from the process event log [6]. The approach is based on a logical representation of the event log and the application of induction with a decision tree to separate the process instances according to their duration. This research focused on finding the reason why some instances of the process are delayed. To identify bottlenecks, several experiments were performed with several induced hypotheses that are represented as decision trees. In each experiment, there is a different cause of delay, and the approach manages to find that cause.

A popular algorithm was used to induce logical decision trees, TILDE (Top-Down Induction of Logical Decision Trees), which is based on Inductive Logical Programming (ILP). The results showed that the decision tree was able to capture the delay condition, although in some experiments it was necessary to include additional logical predicates to better express the cause of the delay. They also conducted experiments to study the scalability and noise tolerance of the proposed approach. From the experiments it was possible to extract many rules, according to the users involved in the task and the estimated impact. All these reasons are induced by the TILDE algorithm to identify the impact of the delay.

In 2019, a systematic review was conducted to identify process mining techniques and applications between 2002 and 2018 [2]. The main objective was to map the most active research topics in process mining. It was possible to observe that the topics are associated with process discovery algorithms, followed by verification of compliance and improvements in architecture and tools. Although process discovery was identified as the primary focus, it was also clear that several documents named process improvement, indicating that researchers aim to improve actual business processes. They also showed research related to using extended processes to make predictions with statistics and machine learning on the fly, as they are expected to provide better operational support [7]. This year, a study was conducted on an academic information system called "Helpdesk I-Thank" at the University of Telkom [8]. The goal was to discover the activities that took the longest to be done by students through a process. After obtaining a process model, an audit was conducted to find bottlenecks for each activity. Bottlenecks show that business processes have problems resolving complaints through the academic information system. To discover the bottleneck that caused the loss of time in the request made by a student, it was used a mining algorithm based on genetic algorithms, event log and support methods (Excel, Disco, and ProM) to provide support to "Helpdesk". The results showed that application bottlenecks are in two activities and that both exceed the average waiting time. Finding this type of delay opens the door to make an improvement process for each activity. It was recommended that the institution develop a real-time progress function of the status of the request, so that the user knows the progress of their complaint and notifies the user when the process is complete so that they do not perform repetitive activities.

Also, in 2019, was sought to predict the temporal performance of any business process [9], for this a method was proposed consisting of the following steps: 1) construction of the process model, 2) generation of a process representation matrix and 3) construction of the prediction model. For the development of the first step, a state transition system was used that allows having different forms of process models. To conduct the second step, a temporal aggregation function is defined, which is used to measure the temporal performance. This function generates two different types of matrices that represent the temporal evolution of the process, which will be the training set used in step three. In the next step, three Deep Neural Networks (DNNs) are implemented: Convolutional NN, Long-Short-Term Memory NN, and Long-Term Re-Current Convolutional Networks. Three case studies with true records are used to evaluate the proposed method: healthcare service, BPI Challenge 2012, and helpdesk. In each experiment, cross-validation was performed to measure and compare the prediction accuracy using two metrics: Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The article has the repositories with the source code of the proposed method and detailed results (https://github.com/gyunamister/performance_prediction). Finally, the results indicate that the method successfully predicts yields at the level of the process model.

In 2020, a novel approach was developed, aimed at systematic modeling and error detection and correction of data flow in business processes [10]. This approach was based on a petri dish (WFIO-net) to model business processes. They introduce an algorithm to detect errors in the data flow called ADIM (Activity-Data Incidence Matrix), which employs an incidence matrix on the data flow. Additionally, a case study is conducted to evaluate this strategy. The case study is about a business process of approving a real estate loan. The results show that the proposed approach is very practical and can be applied to real life environments.

In 2021, a new prediction approach is proposed to analyze the performance of activities in manufacturing systems [11]. In this study, a framework involving process mining techniques is proposed to support the development of a probabilistic model based on Bayesian networks

and a set of predictive models. This is done to estimate the future behavior of the process and the completion time of an assigned production order, perform simulations or even estimate the probability that the system will fail. The validation of the proposed framework was performed with the following steps: 1) To perform a structured event log, taking into account the following columns: series, Instance ID, activity, start time, end time and duration of the activity; 2) To generate the structure of the Bayesian Network; 3) To structure the case file to obtain the conditional probabilities; 4) To carry out the learning of the Bayesian Network and; 5) To perform the simulation of the scenarios to validate the model. The information obtained in response to applying the probabilistic model, helps to analyze the probability of potential failures to occur, which allows guiding efforts to reduce unplanned stops and maintenance inspections of restructuring. Two case studies were performed, the first with data of a manufacturing process consisting of a single equipment and the second with multiple equipment. For both cases, three approaches were considered: 1) performance analysis and simulations, 2) prediction of completion time, and 3) time prediction for incomplete instances. Finally, production and maintenance managers can make decisions regarding the estimated time of completion of a single activity or group of activities, the probability of a single event occurring in an incomplete follow-up, the probability of a group of activities occurring until the end of an instance, have the estimated time of completion of multiple instances, know the future behavior of an activity or group of activities and, know the estimated variation of production cycle times.

Considering the previous theoretical context and state of the art, this research focused on proposing a model that integrates a set of generic services for the automatic improvement of business processes within organizations. The proposed solution is designed to work with any type of business process, not a specific business process, and is expected to be deployed over existing tools running business processes in organizations.

IV. Proposed Generic Service Model

A. *Development Methodology*

The Iterative Research Pattern (IIP) proposed by TAMU was used in this research [12]. PII consists of four main stages, namely: field observation, problem identification, solution development and field testing. In the field observation stage, the study of process modeling in BPMN nomenclature was carried out, considering the elements that compose it and the significant importance that process diagramming has acquired in organizations. This allowed to identify the way to represent the integration of services to help in the improvement of processes. Subsequently, in the identification stage of the problem, a search was made for the diverse ways in which bottlenecks, reprocesses, failures or delays have been determined in a business process, to support preventive decision-making to optimize resources. Likewise, we considered the case studies found in the state of the art, which focused on event logs to analyze the behavior of the processes. Once the most appropriate elements were identified to represent the services and the contributions and shortcomings of the previous works, the development stage of the solution started, where the Bizagi modeling tool (www.bizagi.com) was used to design each of the service models whose objective was to help the automatic improvement of business processes. Finally, in the field testing stage, two evaluations were performed, in the first one it was evaluated that the model met the representation defined in BPMN and allowed to achieve the objectives defined for this. This evaluation was called conceptual evaluation of the models and for this the focus group technique was used, conducted as follows: A first session where the opinion of five experts was known regarding the proposed models. This allowed to amend the proposed based on the feedback provided by the participants, to perform a conceptual evaluation of the updated version of the services with the participation of three (3) experts. In addition, an evaluation of the implementation of the bottleneck management service was performed based on experimental data obtained from the simulation of a business process. This evaluation allowed to analyze the reduction in processing time of the instances of a process in relation to the resources added in manual activities and showed the practical validity of the service.

B. *Service Model*

For organizations in general, it is essential to identify when a process begins to lower its service quality indexes and thus generate improvement actions to prevent this situation becomes a problem that affects business production. For this reason, this research defined three generic services that can be applied transversally to the manual activities (performed by human talent) of the business processes that an organization develops. Services along with their processes and threads are listed in the Fig. 1 represented by a component and package diagram.

The first service that is illustrated is the *Bottleneck Management* service, which consists of four main processes and five threads, which relate to each other, according to the objectives set for each. The start of each process is presented in sequence, as indicated by the horizontal blue arrows, but they are executed independently since each of them has its own activity flow.

In the proposed services there are four (4) roles, namely: 1) *Support Software*: Corresponds to the system or software design and execution of business processes in the company. The tools considered are those that incorporate BPMN modeling and allow the integration of services such as BonitaSoft, BizagiStudio, Camunda, among others. 2) *Process Manager*: The process manager is responsible for managing the business processes of the company. It works hand in hand with the Support Software and is responsible for configuring, designing, validating, and evaluating the activities and human resources of the processes. 3) *Analyst*: It oversees building and making improvements, analyzing the statistics that report the current functioning of the business process. And 4) *User*: Represents a natural or legal person who makes a request through the Company Support Software.
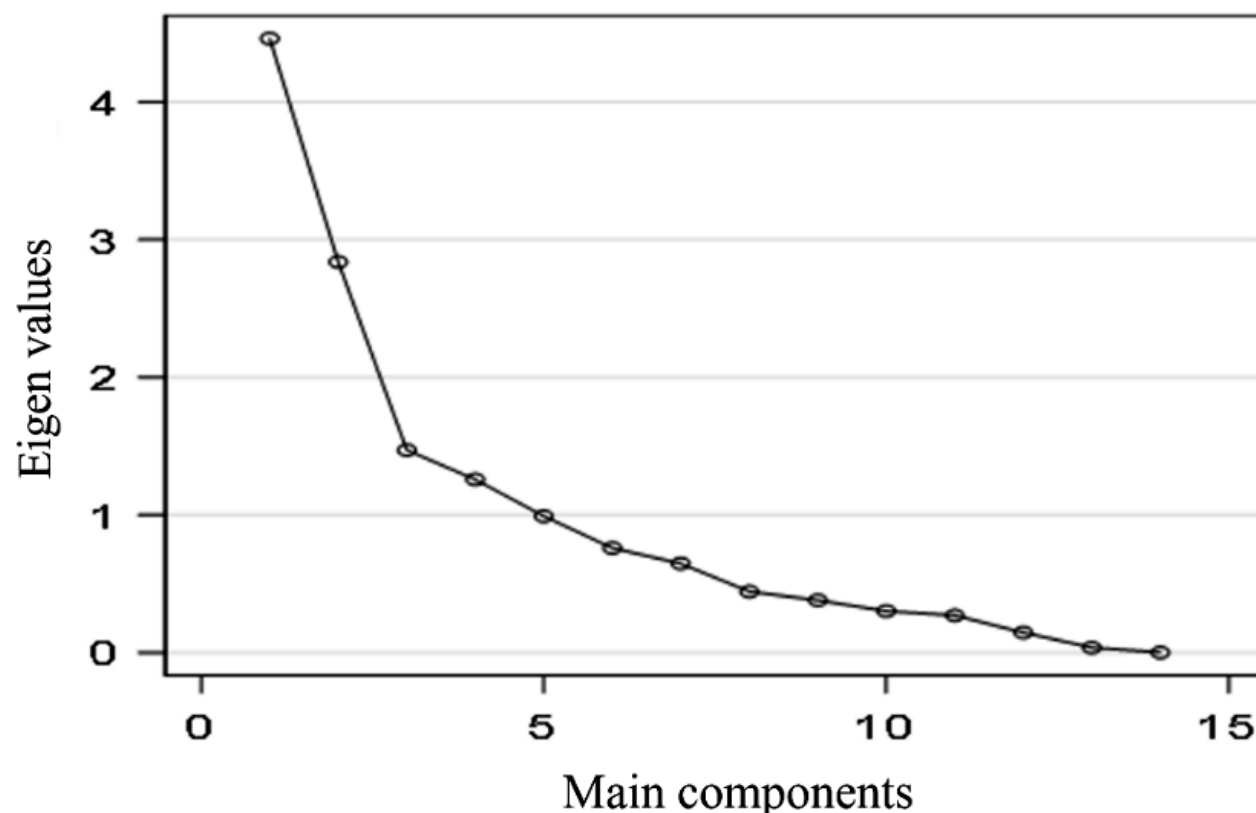

Fig. 1. Generic Services Model.
Source: Authors.

C. *Service: Bottleneck Management*

This service analyzes the transaction log of the different business processes to forecast significant delays in the manual activities that make up the process. To achieve this, a prediction model is generated that, when applied, informs the user of delays that may arise in manual activities and how they can be prevented, through a prediction report. The suggestions addressed in this report seek to make better use of resources and better execute activities. Based on these tips, the user can apply changes to the process to improve the productivity of the process. The service is defined in four processes and then represented with their proper BPMN diagram. The diagrams and their detailed explanation are available at https://drive.google.com/drive/folders/12yFIpoTLbG2A7qzmRSwS76RAKydZrxwE?usp=sharing.

1) *Prediction model training*: This process ([Fig. 2](#)) generates prediction models by executing various artificial intelligence techniques and selects the model with the best performance, stores it and uses it to improve the business process that presents bottlenecks in its activities. This process takes as input the minable view generated by the *Extract Business Process Data* thread and the *Configuration File*, which outputs the *Modify Model Settings* process. It then runs several regression models (decision trees, vector support machines, neural networks, among others) with different configuration parameters and transforming the data with different time windows (windowing). As a result, the different models are obtained, with their parameters, window sizes and performance, and then select the model with the best performance and this is returned as a result of the process.

2) *Predicting bottlenecks in activities*: This process ([Fig. 3](#)) generates a prediction report by applying a regression model to propose recommendations on the human resources of the activities of the process that help in the improvement of this.

3) *Performance measurement of the prediction model*: This process ([Fig. 4](#)) determines whether the differences in the data obtained in the prediction and the actual data of the running process are below the defined threshold, in order to measure the success of the prediction model and thus determine whether or not an update of the current prediction model should be made. This process measures, from the value of a threshold defined by the *Process Manager* in the Configuration File, whether the differences between the prediction data versus the actual data of the running business process are significant to update or not the current prediction model. To do this, it uses the *Calculate Model Performance* in Production thread, which is responsible for finding differences between the prediction data and the actual process data. Once you have calculated differences from the data, they are compared to the defined threshold. Finally, if it is established that the data significantly exceeds the threshold, the *Update* Prediction *Model* thread must run to generate a new predictive model.

4) *Evaluation of the improvement of the prediction model*: This process ([Fig. 5](#)) evaluates the quality of the prediction model by changes made to the process according to the suggestions provided from the *Prediction* process *bottlenecks in activities* to determine if the changes were favorable to the business process.

5) *Modify Model Settings*: This process ([Fig. 6](#)) modifies the initial *configuration file* defined in the Support Software, to establish the execution parameters of the techniques/algorithms that allow the creation of regression models or classification models to be used in the different services that require it. In addition, for the proper functioning of the service, the above processes are complemented by a set of threads listed below (these processes are not explained in detail to be able to comply with the restriction on the number of pages set by the event/journal): Extract data from the business process; Define the prediction model to use; Assign resources to business process activities; Calculate the performance of the model in production; Update the prediction model and Modify model configuration.
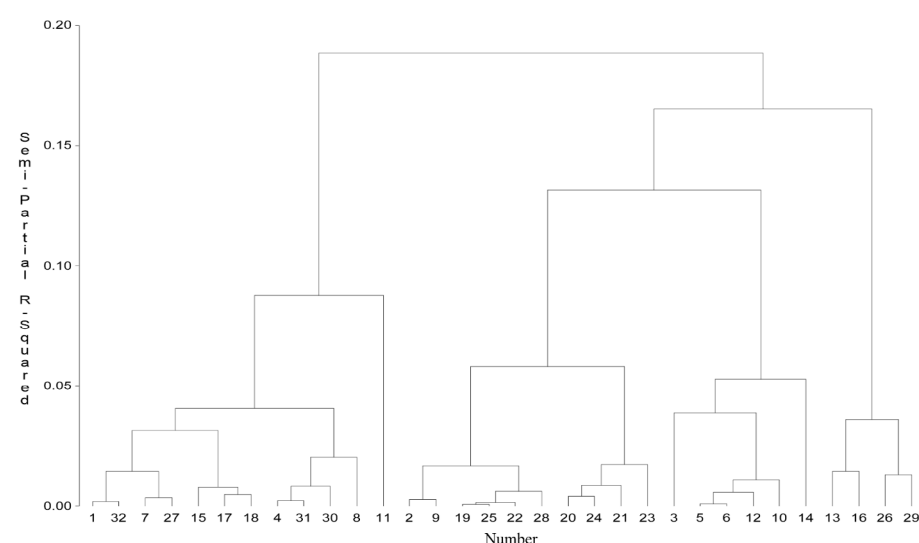

Fig. 2. Prediction Model Training.
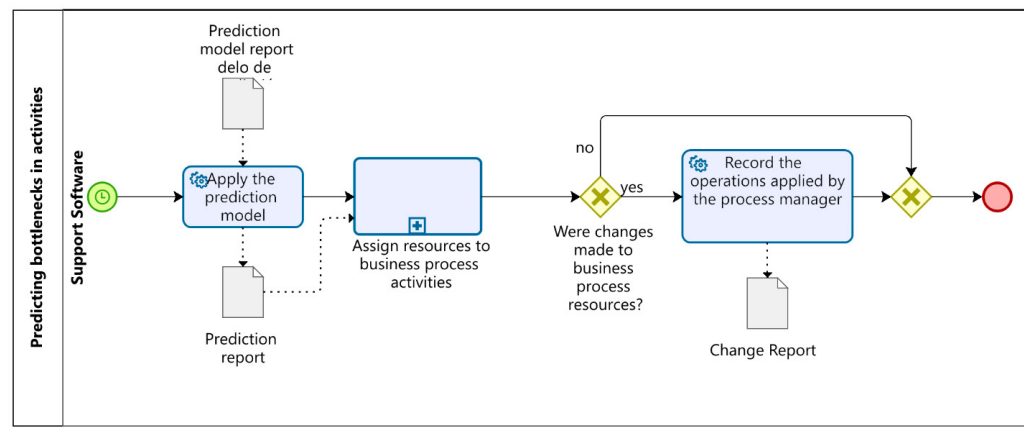Source: Authors [15].

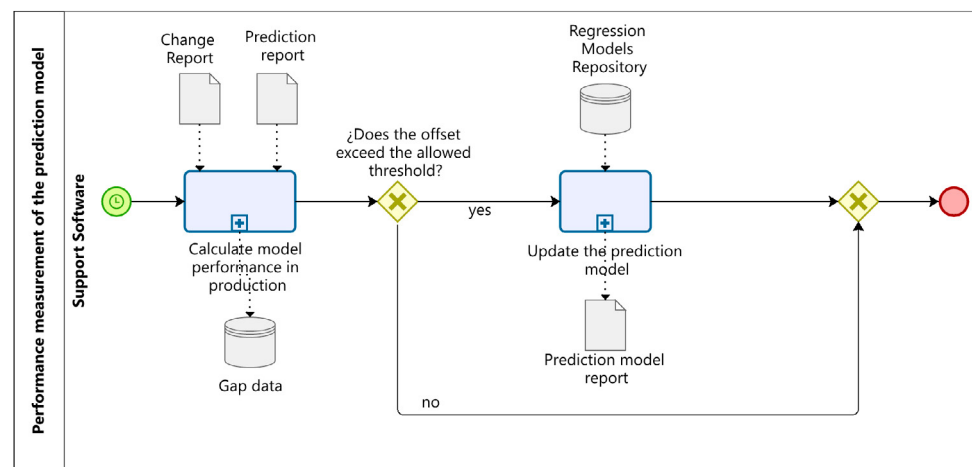Fig. 3. Predicting bottlenecks in activities.
Source: Authors [15].

Fig. 4. Performance measurement of the prediction model.
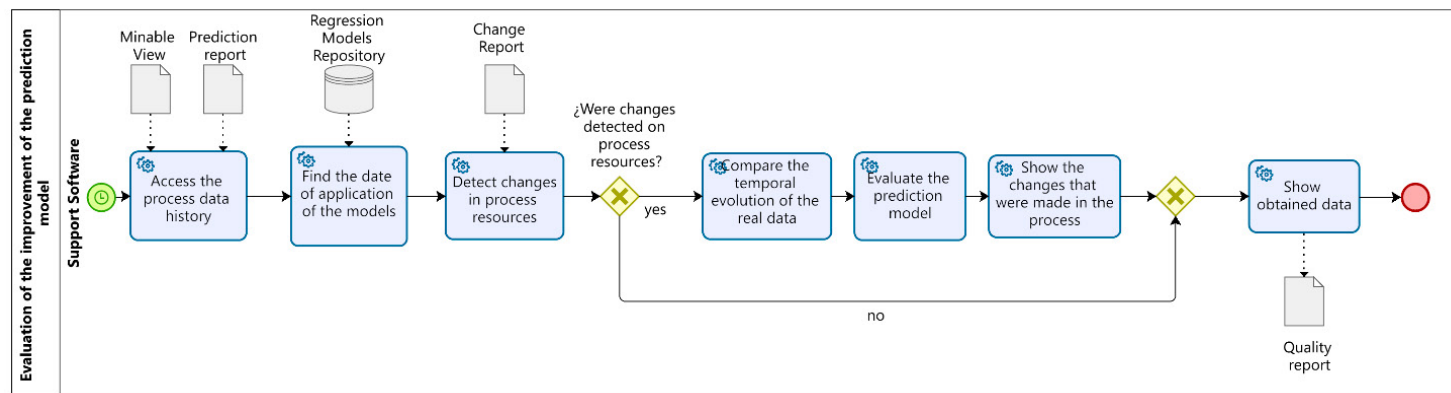Source: Authors [15].

Fig. 5. Evaluation of the improvement of the prediction model.
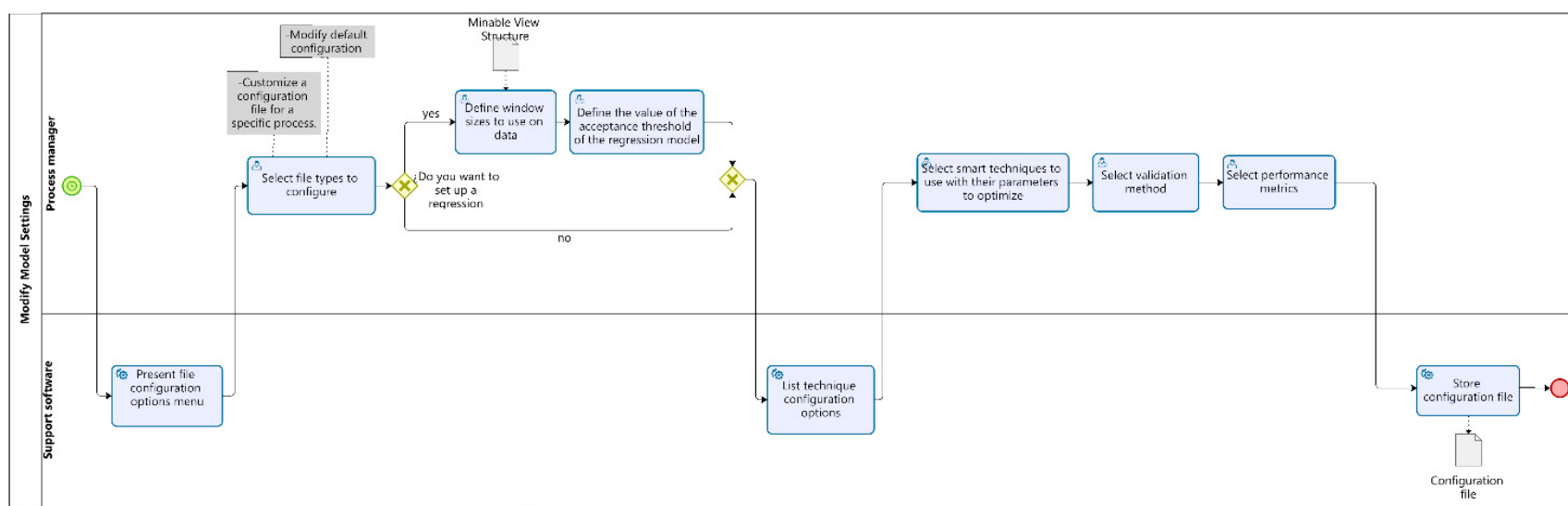Source: Authors [15].

Fig. 6. Modify Model Settings.
Source: Authors [15].

D. *Service: Resource Management*

This service is responsible for verifying the availability of resources considering their level of experience and working hours. To do this, it looks for certain features that are requested from the *Prediction Report* issued by the *Bottleneck Management Service.* In addition, it consists of a thread called *Notify Changes Made*, which is used in other services to update resource availability status in real time, either because it was assigned to a new activity, changed with another resource, or became fully available. The service is defined in a single process called *Assign Resources to Activities.*

This service is responsible for verifying the availability of resources considering their level of experience and working hours. To do this, it looks for certain features that are requested from the *Prediction Report* issued by the *Bottleneck Management Service.* In addition, it consists of a thread called *Notify Changes Made*, which is used in other services to update resource availability status in real time, either because it was assigned to an original activity, changed with another resource, or became fully available. The service is defined in a single process called *Assign Resources to Activities.*

1) *Assign resources to activities*: Assign the appropriate resources to activities that show delays or queues in your requests, to make the right changes that benefit the business process.

E. *Service: Ticket Management*

This service is responsible for verifying that the data entered by *users* when initiating a request is correct in form and content. This is achieved by properly defining the business rules that must be considered during the development cycle of the established process. It requires that data entered through forms be validated in real time through the company's *Support Software* automatically to avoid human errors in terms of the types of data and content in the information being processed. In addition, it considers a classification model, which, when applied, helps to organize optimally the documents that are entered in support of an application, and thus determine that those documents have the correct format and content. The service seeks to establish the type of decisions that are made in accordance with a request. Finally, it analyzes whether the decisions taken in response to incoming requests present significant inconsistencies to structure the business process activity flow in a more appropriate way. The service is defined in five processes:

1) *Receipt of documents*: Verifies that the data and documents that are taken as input into the business processes are those requested in form and quantity, through validations performed by the *Support Software*, to have the complete and correct data and documents during the execution of the business process.

2) *Create classification minable view*: Organizes incoming documents from the business process to generate a classification minable view with structured and pre-classified data that will be used as input for training the document classifier.

3) *Define Classification Model*: Runs multiple classification models with different configuration parameters to select the best classification model.

4) *Verification of data extraction*: Validates if the information provided by users is correct and true, to process in an appropriate way your request, according to business rules defined in the process.

5) *Monitoring of results*: This process seeks to identify by reviewing the decisions previously made during the execution of the process if there are irregularities in the responses to requests to act and correct those anomalies in time.

According to the conceptual model proposed, in this research, the service of management of bottlenecks was implemented to validate the benefit it brings to organizations, supporting the flow of activities of a business process.

V. Implementing the Bottleneck Management Service

Prior to the service implementation, the business process to be simulated was defined, which required capturing data from its execution log. The selected business process was the *Realization of Clothing*, which is composed of 10 activities, and in its initial configuration 7 human resources were assigned. The main objective of the process is to carry out manual preparations, starting with the request of the order, followed by the design and patterning of the garments, the preparation of the raw material, the manual elaboration of the garments and finally, the delivery and distribution of the orders. The mapping of the process was done in BPMN nomenclature and was done in the Process Simulator tool, as shown in the Fig. 7.
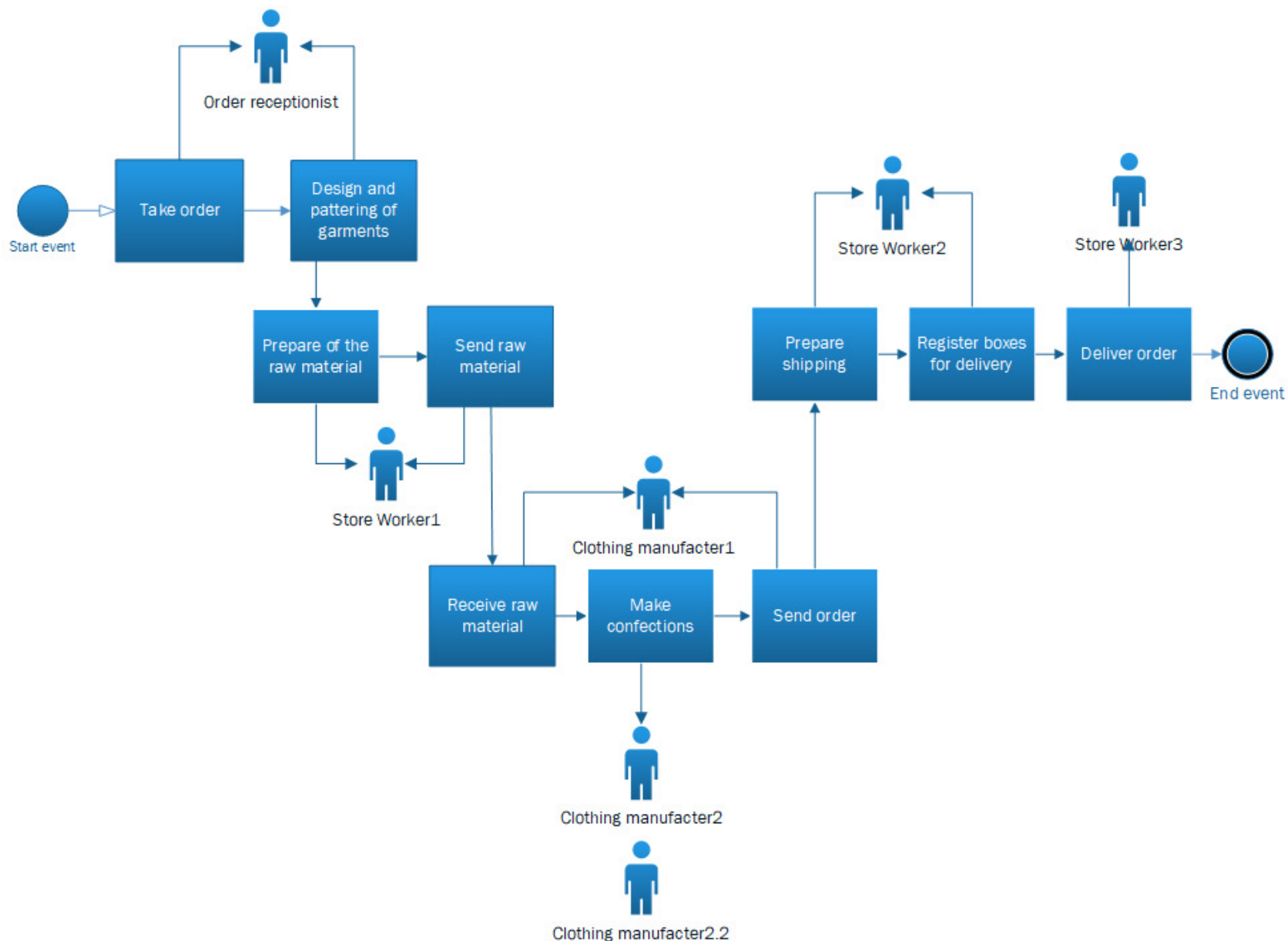

Fig. 7. Manufacturing process.
Source: Authors.

Once the model was made, a simulation scenario was created with working hours from Monday to Friday from 8:00 am to 6:00 pm, and Saturdays from 8:00 am to 01:00 pm. An order overhead was generated for Tuesday. In addition, a rest period was stipulated from 12:00 m. to 2:00 pm. Each resource has the same level of experience in the activity assigned to them. When executing the defined scenario, the first data was obtained to analyze the behavior of the simulated process. First, it was evident that during the period executed 532 order records were generated, of which 302 were successfully evacuated, indicating that 230 orders did not complete satisfactorily. However, it was essential to detect which activities were generating representation of requests, for this purpose a graph was generated showing the capacity of the states of each activity, measured in percentages shown in the Table 1 and Fig. 8.
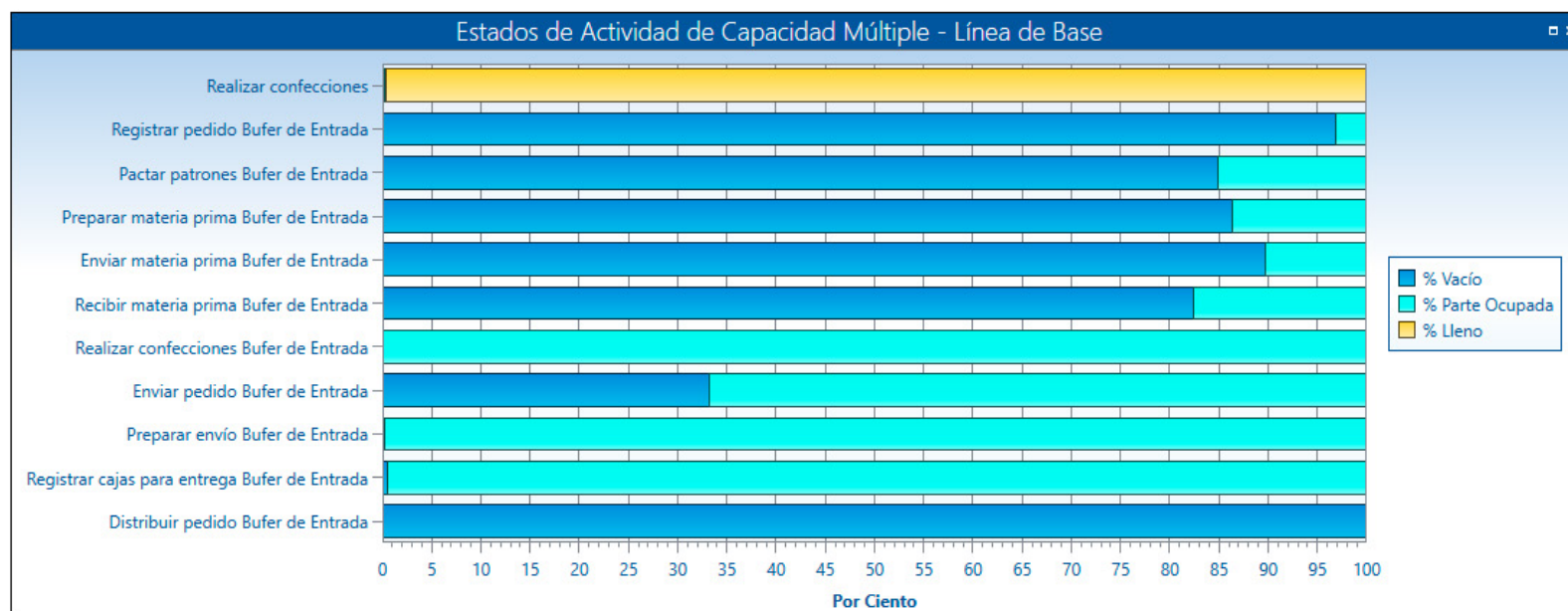
Fig. 8. "Capacity activity states" graph.
Source: Authors.

TABLE 1.
ACTIVITIES DESCRIPTION.

Activities	Description
Take order.	This activity is carried out by the order receptionist who is in charge of taking the orders made during the process.
Design and pattering of garments.	This activity is carried out by the order receptionist who is in charge of carrying out the pattern making of the orders.
Prepare the raw material.	This activity is carried out by store worker 1 who is in charge of organizing the resources and supplies necessary to carry out the preparation.
Send raw material.	This activity is carried out by store worker 1 who is in charge of sending the resources and supplies to carry out the preparation.
Receive raw material.	This activity is carried out by clothing manufacturer 1 who is in charge of receiving and verifying the necessary material to make the garments.
Make confections.	This activity is carried out by clothing manufacturer 2 which is in charge of making the seams.
Send order.	This activity is carried out by clothing manufacturer 1 which is in charge of sending the orders.
Prepare Shipping.	This activity is carried out by store worker 2 who is in charge of preparing the shipment according to the order.
Register boxes for delivery.	This activity is carried out by store worker 2 who is in charge of filing the shipment boxes before distributing them to their destination.
Delivery order.	This activity is carried out by store worker 3 who is in charge of delivering the orders.

Source: Authors.

In the Fig. 8 it can be observed that the activities that show a high percentage in their capacity of capacity are, *Making Confections, Preparing Shipping and Registering boxes for delivery*. Once, it was identified that these activities that are causing problems in the productivity of the process, the detailed records of *Make Confections* and *Prepare Shipment* are extracted, to identify through these entries' records, how is their behavior and how many requests are arriving per day. Data used to apply regression models for each activity are date, day of the week, number of hours worked per day, and number of orders queued per day.

To properly implement the proposed service, the processes of 1 were prioritized. Prediction model training and 2. Prediction of bottlenecks in activities. To do this, a project was created based on the ASP.NET Core framework with Angular, available in the Visual Studio suite. Thus, a graphical user interface was obtained with Bootstrap and a REST API was developed, whose objective was to make the connection between the Backend and the Frontend designed to display the data when running the service.

The Backend was implemented in Python language and used LazyPredict, a free and open source library that allows the execution of several machine learning models. Considering that the service to be implemented includes among its objectives the prediction of the behavior of the simulated business process, the set of regressors that are available in the library was used. Once the training was completed, an ordered list of regression models with their performance metrics was obtained, from this, the best performance model for the data set was selected and the training and prediction process continued.

A. *Training Module*

The training data set of the activities "make preparations" and "prepare shipment" was structured to define the initial data sets to use. The dataset had 126 records and were structured with the fields of the day of the week, date, hours worked per day and the number of requests that arrive. In the Fig. 9 a sample dataset is presented. The required variables and parameters were configured throughout the training process, and in the case of available regression models, their default configuration parameters were left. Depending on [13] the circumstances, a time series is defined as a chronologically ordered set of data sampled at a constant frequency of time (hours, days, months, or others). In the series, the dependent variable is the value of the target series at the instant of time to predict, and the independent variables consist of data extracted from time windows of previous hours, days, or months. Therefore, to carry out the preprocessing of the data set, a window size was defined with a constant value of 3 (although this is a parameter that can be refined), in order to make the prediction of each value in the time series, for the case study, the dependent variable is the number of orders queued per day in the analyzed activity.

Once obtained the independent variables, the LazyRegressor procedure was executed in the library, where the models were trained and validated, to finally show the user the performance of the regressors used. From the results in the performance metrics, the model with the highest performance was selected and trained and evaluated by the technique of cross validation, since so the results obtained are more real or close to the production environment (neither very optimistic, nor very pessimistic), this was saved, with the use of pickle library, then be applied in the prediction of the activity of the business process.

date	day	hours_day	request
2/08/21	1	8	3
3/08/21	2	8	6
4/08/21	3	8	13
5/08/21	4	8	16
6/08/21	5	8	15

Fig. 9. Sample data set used.
Source: Authors.

B. *Prediction Module*

Continuing with the proof of concept within a simulated environment, to generate the prediction about the analyzed activities, the business process was executed for one week after the date of the last record in the training data set. Logs within this week were formatted and preprocessed according to the previous module. The saved model was loaded back into the service and the prediction was run for the new week obtained. At this point, it was possible to contrast the data of the execution of the business process with the data obtained from the prediction made.

There are several metrics to examine the quality of the prediction obtained by the model, in this case the available regression metrics have been taken to evaluate the prediction error in relation to the actual data of the execution of the business process. Specifically, two of these metrics were chosen, which clearly show the quality of the difference between predicted and actual data. The Mean Absolute Error (MAE), is the metric corresponding to the mean deviation of the expected value with respect to the real data, in the analysis of the activity "Make preparations", this gave a value of 4.74 and for the activity "Prepare shipment", a value of 2.03 was obtained.

The Root Mean Quadratic Error (RMSE) metric reported a value of 6.12 for the activity "Send" and a value of 2.42 for the activity "Prepare sent".

Considering the values obtained with the metrics, it was evident that there is a low difference between the predicted and actual data, so it was considered that the prediction of the queue length of requests for the activities analyzed was appropriate. Based on the above, a set of suggestions were made regarding the amount of additional resources, with which it would be possible to free up the accumulation of requests for each of the analyzed activities, this allows the business process administrator, determine how many additional resources you want to allocate, depending on how much queue length is considered acceptable for the defined business process. Consequently, for this research, a support resource was added for each of the analyzed activities.

In the Fig. 10 the list of regression models, organized from highest to lowest performance metrics, is displayed according to the Make Readings activity dataset. Similarly, in the Fig. 11 the list of models for the Prepare Shipment activity is displayed. It should be noted that as expected, the best classifier is not the same for the two activities and in general to be able to define a single model for all the activities of the different business processes, other variables that complement the data set and a greater amount of data will be required. The Fig. 12 presents the regression model selected for the Make Confections activity with its corresponding performance metrics and the Fig. 13 illustrates one-week predictions for this activity in its queue length according to the level of orders entering per day.

Model	Adjusted R-Squared	R-Squared	RMSE
Lasso	0.997912607	0.9983474806	1.4797427128
LarsCV	0.9977938248	0.9982534446	1.5212622977
OrthogonalMatchingPursuitCV	0.9977216771	0.9981963277	1.54593682
LassoCV	0.9976963687	0.9981762919	1.5544994892
LassoLarsCV	0.9976684897	0.998154221	1.5638776422
ARDRegression	0.997629622	0.9981234507	1.5768591695
BayesianRidge	0.997610685	0.998108459	1.5831453973
TransformedTargetRegressor	0.9975751355	0.9980803156	1.5948793704

Fig. 10. Show regression models of the activity Make confections.
Source: Authors.

Model	Adjusted R-Squared	R-Squared	RMSE
TransformedTargetRegressor	0.999964801	0.9999713193	0.1540754728
RANSACRegressor	0.999964801	0.9999713193	0.1540754728
LinearRegression	0.999964801	0.9999713193	0.1540754728
LassoLarsIC	0.999964801	0.9999713193	0.1540754728
OrthogonalMatchingPursuitCV	0.999964801	0.9999713193	0.1540754728
LassoLarsCV	0.999964801	0.9999713193	0.1540754728
LarsCV	0.999964801	0.9999713193	0.1540754728
Lars	0.999964801	0.9999713193	0.1540754728

Fig. 11. Show regression models for Prepare Shipment activity.
Source: Authors.

Modelo	Lasso
R Squared	0.9975271575102891
Adjusted R Squared	1.3016867837447315
RMSE	1.9350395682285955

Fig. 12. Train Prediction Model to Make Confections.
Source: Authors.

Los resultados de la predicción obtenida, junto a sus fechas correspondientes, día de la semana (1: Lunes - 7: Domingo) y horas laboradas por día, se muestran a continuación.

Fecha	Día	Horas laborales	Predicción - Longitud de cola
6/12/2021	1	8	151.9932775479
7/12/2021	2	8	153.7516177291
8/12/2021	3	8	152.582666776
9/12/2021	4	8	154.7467001648
10/12/2021	5	8	156.7368650361
11/12/2021	6	5	157.4591651026
12/12/2021	7	0	163.8283221323

Fig. 13. One-week prediction for Make-up.
Source: Authors.

VI. EVALUATION RESULTS

A. *Conceptual evaluation*

The evaluation process performed to the proposed Service Model was performed using the Focus Group technique. This technique was applied through the qualitative research method to have quick feedback with reliable information. The planning conducted to evaluate each of the proposed services focused on determining the feasibility of integrating the services into the business processes of companies. Likewise, to establish if the representation made in BPMN nomenclature was clear and if it corresponded with the description raised in the characterization of the models. For the definition of discussion groups (expert group), we sought to have professionals in the areas of Software Engineering, Process and Product Quality and Artificial Intelligence. In addition, with experience in business processes.

A first session was held to obtain the opinion of five experts on the proposed models. The initial objective was to obtain relevant information to improve modeling services, which was collected using different techniques (Notes, audio, and survey with open questions), from which comments, observations and suggestions made by participants were identified.

Then in the Table 2 it groups together the areas that the experts commented should be improved and the Table 3 here are the other comments that participants made about the proposed models.

TABLE 2.
AREAS FOR IMPROVEMENT IDENTIFIED AT THE FOCUS GROUP SESSION.

Areas for improvement	Indicated by participant				
	P1	P2	P3	P4	P5
Better specify the inputs and outputs tliat are used in each process that makes up the Service.	x	x	x		x
See granularity of Services in their respective models.		x	x		
Explain in more detail what each activity does so that it is understood in the model. Put more meaningful ñames.	x	x	x		
Properly define the sequence, order, and description of diagrams, as numbering confiases.	x	x	x		x
Simplify activities that have the same objective to perform.		x			x
Specify the model referenced in the diagrams: business process or prediction model.		x			x
Processes must not be modeled with a single Thread. Put the activities that make it up.			x		x
Use the Parallel Activities representaron when you indicate that múltiple things are running at the same time			x		
When an activity represents two actions within the intent of the diagram, divide into two activities.			x		
Avoid using words that are too large and too broad. Like all techniques or the best model. These are things that you can't measure, and for some people, it can be different.					x

Source: Authors.

TABLE 3.
COMMENTS IDENTIFIED AT THE FOCAL GROUP SESSION.

Remarks	Indicated by participant				
	P1	P2	P3	P4	P5
Make a diagram explaining the relationship and iteration of the proposed Services.		x			
The diagrams are described at High Level, it is important to consider that a feedback cycle is allowed.	x				
Consider the Identify, Choose, or Select verbs in the Activities that yon define.	x	x	x		x
Make it clear in the User Activities that they are performed by People. That is, there is human intervention.	x				
Consider whether to put an error flow into the diagrams that represent the Services.			x		

Source: Authors.

Based on the feedback obtained, improvements were made for each of the services presented. The diagrams were adjusted, especially including the input and output artifacts of each activity and defining more appropriate names for the activities. These improvements were based on the modeling guidelines stipulated by *Montes de Oca*.

Subsequently, a second review was conducted with the availability of three of the experts attending the first session of the Focal Group. For this evaluation session, a complete document was sent with the characterization of each of the models corresponding to each service, to receive a rigorous evaluation in aspects concerning its structure, content, and clarity in relation to the concepts discussed in this research. To this end, the evaluators were provided with quality guidelines to be considered during the evaluation of the proposed service models. Based on the conceptual evaluation conducted by the participants during the second session, it was established that the proposed Generic Services Model meets the objectives and guidelines in terms of design, concept, and structure, concluding that the processes created in each service solve the problem addressed in this research. The results of the evaluation indicate that the Bottleneck Management service was approved by the experts with a rating of 4.81 out of 5.0. The Resource Management service was subsequently approved with a rating of 4.81, and the Entry Management service was approved with a rating of 4.83. In addition, future improvement observations were indicated for the integration of the service in the organizations, considering limitations and rules depending on the business process where it is applied.

B. *Experimentation with the simulator*

The experimentation developed in this research was based on determining the improvements regarding the execution time of the business process, which can be generated by adjusting the human resources involved in the business activities.

From the addition of human resources proposed in this research, the defined business process was left with a total of 9 resources allocated (This is presented in the Fig. 14). When applying the corresponding adjustments and executing the scenario, it could be determined that the prediction of the queue length with an additional resource for the activities to make and prepare the shipment, have an average absolute error of 4.06 and 3.59 respectively; with respect to the results obtained when executing the scenario with the changes made. Similarly, it was evident that during the simulated week 34 requests were completed, a notable difference with respect to the simulated process without the suggested support resources, where for the same week, a total of 17 requests were completed, obtaining as a result that there was indeed a decrease in the queue length when applying the suggested changes. Thus, it was determined that the requests were completed in less time, reducing the average time of execution of the simulated business process.
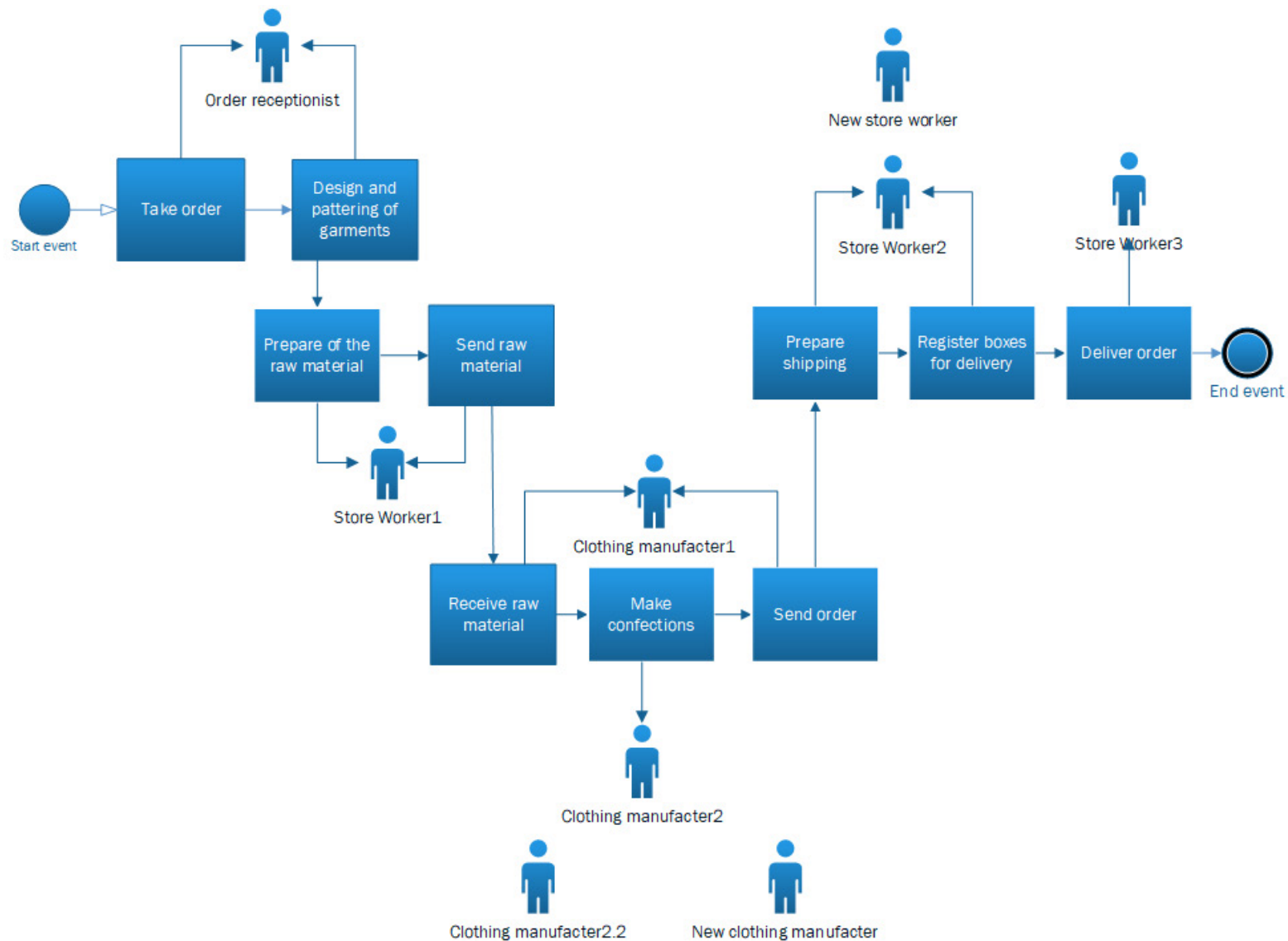
Fig. 14. Process of making preparations with addition of resources.
Source: Authors.

To measure the average amount of time that was achieved in processing a request, the time obtained in the initial simulation without suggested resource changes was analyzed, taking the average time in the system, given by the simulation software during the week of prediction, which corresponds to a value of 676.5 hours, divided by the number of completed requests, obtaining a result of 39.8 hours, which correspond to the average processing time of a request throughout the business process. However, the same procedure was performed after adding the support resources, where an average system time of 656.29 hours was obtained that when divided among the 34 completed requests results in 19.30 hours, which is the average time of each request from the beginning to the end of the business process. This shows an improvement, as it ensures the customer a considerable decrease in the processing of their request.

VII. Conclusions

A model of generic services was developed for the automatic improvement of business processes. Visual representation was performed according to BPMN nomenclature modeling standards with the Bizagi Modeler tool. This allowed to identify and properly define the components and artifacts of each service through processes and threads, which evidenced the correct flow of the activities that make them up. Thus, the Bottleneck Management Service was composed of the processes of training prediction model, prediction of bottlenecks in activities, performance measurement of the prediction model and evaluation of the improvement of the model, having as complements the threads of extracting data from the business process, define the prediction model to use, allocate resources to the activities of the business process, calculate the performance of the model in production, update the prediction model, and modify model configuration, which jointly seek to detect significant delays in the manual activities of a business process, based on the analysis of the execution log of the process to improve the productivity of this.

The Resource Management Service consists of the process of Assigning Resources to Activities, which is related to the first service defined, which seeks to assign appropriate resources to activities based on the suggestions in the Prediction Report generated by the Bottleneck Management Service to optimize the execution time of manual activities.

Finally, the Entry Management service was incorporated, which was formed of the processes of receiving documents, create minable view of classification, define classification model, verification of data extraction, and monitoring of results, which jointly, is responsible for ensuring that the entries of business processes correspond to the requirements defined in the rules established by the company to give reliability of the decisions that are made during the execution of the process when a request is received.

The Bottleneck Management service was implemented, where several functionalities were structured to obtain a prediction according to the data sets obtained from the business process executed by the simulation tool used. The modules created were: running the available regression models, selecting the best performing regression model, and predicting the queue length per day in a week. Once the prediction was obtained, the respective suggestions on the resources that can be involved in the process were made, especially the activities analyzed that generated bottleneck. This was achieved by developing the backend of the service with Python libraries and the respective frontend and the global connection with the .Net Core Framework.

It was found that, by adding the suggested resources to each activity analyzed according to the prediction of the obtained queue length, the average execution time of the simulated business process was optimized, establishing an overall improvement of 100% in the efficiency and productivity of the same (adding 2 human resources), corroborating that it is also noticeable the agility with which a request is completely met, thus achieving an improvement in customer satisfaction by showing the reduction of the average execution time of a part.

The Focal Group technique validated concepts and provided useful information that provided the possibility of making appropriate adjustments based on the comments and views of five experts. Likewise, from the evaluation questionnaire applied in the conceptual evaluation session, with the participation of three experts, it was validated that each of the proposed services were approved considering the BPMN modeling design guidelines, according to *Montes de Oca* [14], obtaining results of *Totally agree and Agree*, by the evaluators.

Using Process Simulator made it easier to run scenarios by allowing you to view iterations for assigned activities and resources in the simulated business process. It also provided the data required to analyze in depth how the process behaved according to the distribution of requests and first and foremost production times, allowing adjustments to evidence improvements in production and average times of care.

As a future work it is considered important to integrate in a tool of business processes of the market (Bonita BPM, Bizagi, jBPM, jSonic, ProcessMaker, RunaWFE, Camunda or GFile) the model of generic services proposed in this research, including techniques and strategies in data mining proposed for the management of business processes in organizations.

References

[1] S. Park & Y. Kang, "A Study of Process Mining-based Business Process Innovation," *Procedia Comput. Sci.*, vol. 91, pp. 734–743, Dec. 2016. https://doi.org/10.1016/j.procs.2016.07.066

[2] C. dos Santos, A. Meincheim, E. Junior, M. Dallagassa, D. Sato, D. Carvalho, E. Santos & E. Scalabrin, "Process mining techniques and applications – A systematic mapping study," *Expert Syst. Appl.*, vol. 133, pp. 260–295, Nov. 2019. https://doi.org/10.1016/j.eswa.2019.05.003

[3] P. Harmon, "Chapter 15 - Software tools for business process work," in *Business Process Change*, 4th ed. BTV, USA: Morgan Kaufmann Publishers, 2019, ch 15, pp. 369–392. https://doi.org/10.1016/B978-0-12-815847-0.00015-7

[4] A. Rozinat, I. de Jong, C. Günther & W. van der Aalst, "Process mining applied to the test process of wafer scanners in ASML," *IEEE Trans. Syst. Man Cybern. Part C: Applications and Reviews*, vol. 39, no. 4, pp. 474–479, Mar. 2009. https://doi.org/10.1109/TSMCC.2009.2014169

[5]   W. van der Aalst, "Process Mining: Overview and Opportunities," *ACM Trans. Manage. Inf. Syst.*, vol. 3, no. 2, pp. 1–17, Jul. 2012. https://doi.org/10.1145/2229156.2229157

[6]   E. Vasilyev, D. Ferreira & J. Iijima, "Using Inductive Reasoning to Find the Cause of Process Delays," presented at *IEEE 15th Conference on Business Informatics*, WIE, VIE, AT, 15-18 Dec. 2013. https://doi.org/10.1109/CBI.2013.41

[7]   W. van der Aalst, *Process Mining: Data Science in Action*, 2 ed. LDN: UK, Springer, 2016. Available: https://doi.org/10.1007/978-3-662-49851-4

[8]   A. Dzihni, R. Andreswari & M. Hasibuan, "Business Process Analysis and Academic Information System Audit of Helpdesk Application using Genetic Algorithms a Process Mining Approach," *Procedia Comput. Sci*, vol. 161, pp. 903–909, Jan. 2019. https://doi.org/10.1016/j.procs.2019.11.198

[9]   G. Park & M. Song, "Predicting performances in business processes using deep neural networks," *Decis. Support Syst.*, vol. 129, pp. 1–34, Feb. 2020. https://doi.org/10.1016/j.dss.2019.113191

[10] C. Liu, Q. Zeng, H. Duan, L. Wang, J. Tan, C. Ren & W. Yu, "Petri Net Based Data-Flow Error Detection and Correction Strategy for Business Processes," *IEEE Access*, vol. 8, pp. 43265–43276, Feb. 2020. https://doi.org/10.1109/ACCESS.2020.2976124

[11] E. Ruschel, E. Loures, & E. A. Santos, "Performance analysis and time prediction in manufacturing systems," *CAIE*, vol. 151, pp. 1–17, Jan. 2021. https://doi.org/10.1016/j.cie.2020.106972

[12] K. Pratt, "Design Patterns for Research Methods: Iterative Field Research," presented at *2009 AAAI Spring Symposium: Experimental Design for Real*, STFD, CA, USA, 23-25 Mar. 2009. Available from http://kpratt.net/wp-content/uploads/2009/01/research_methods.pdf

[13] S. Beeram & S. Kuchibhotla, "Time Series Analysis on Univariate and Multivariate Variables: A Comprehensive Survey", In S. C. Satapathy, V. Bhateja, M. R. Murty, N. Gia, J. Kotti (eds) *Communication Software and Networks*, LDN, UK: Springer, 2021, pp. 119–126. https://doi.org/10.1007/978-981-15-5397-4_13

[14] I. Moreno, "Patrón y clasificación taxonómica para directrices prácticas en modelos de procesos de negocio," *Tesis doctorado*, UCLV, ST. CLRA, CU, 2015. Disponible en http://dspace.uclv.edu.cu:8089/handle/123456789/7212

[15] *Bizagi Modeler*, (version 11). Bizagi Studio. Available: https://www.bizagi.com/es/plataforma/modeler

Valentina Vera Paz. Universidad del Cauca (Popayán, Colombia). https://orcid.org/0000-0003-0199-5997

Anyi Aracely Ramírez Obando. Universidad del Cauca (Popayán, Colombia). https://orcid.org/0000-0002-2378-7841

Carlos Alberto Cobos Lozada. Universidad del Cauca (Popayán, Colombia). https://orcid.org/0000-0002-6263-1911