



UNIVERSIDAD NACIONAL DE LA PLATA
Facultad de Informática

**Minería de Procesos y Minería de Datos
en la Ingeniería de Software**

Análisis del proceso de distribución de piezas postales

Trabajo Final presentado para obtener el grado de
Magister en Ingeniería de Software

Lic. Victor Martinez

Julio, 2023



UNIVERSIDAD NACIONAL DE LA PLATA
Facultad de Informática

**Minería de Procesos y Minería de Datos
en la Ingeniería de Software**

Análisis del proceso de distribución de piezas postales

**Trabajo Final presentado para obtener el grado de
Magister en Ingeniería de Software**

Autor: Lic. Victor Martinez

Director: Dra. Laura Lanzarini

CoDirector: Dr. Franco Ronchetti

Julio, 2023

TRIBUNAL:

Presidente: _____

Vocal: _____

Secretario(a): _____

FECHA DE DEFENSA: _____

CALIFICACIÓN: _____

PRESIDENTE

VOCAL

SECRETARIO(A)

Fdo.:

Fdo.:

Fdo.:

Completar A mis familia
Por acompañarme en cada paso

Resumen

Completar

En una página como máximo, el resumen explicará de modo breve la problemática que trata de resolver el trabajo (*el ‘qué’*), la metodología para abordar su solución (*el ‘cómo’*) y los resultados obtenidos. En los trabajos cuyo idioma principal sea el inglés, el orden de Resumen y Abstract se invertirá.

Aunque la plantilla se ajusta a las necesidades y reglamentación de la ESI-UCLM, su adaptación es sencilla a otras titulaciones, instituciones y otros documentos de carácter académico. Esta plantilla permite de modo automático la elaboración de la memoria del documento en idioma inglés y puede ser utilizada en cualquier SO (Windows, Linux, Mac OSX, etc.).

Abstract

Completar *English version for the abstract.*

Agradecimientos

Completar Aunque es un apartado opcional, haremos bueno el refrán «*es de bien nacidos, ser agradecidos*» si empleamos este espacio como un medio para agradecer a todos los que, de un modo u otro, han hecho posible que el trabajo realizado *llegue a buen puerto*. Esta sección es ideal para agradecer a directores, profesores, mentores, familiares, compañeros, amigos, etc.

Estos agradecimientos pueden ser tan personales como se deseé e incluir anécdotas y chascarrillos, pero *nunca deberían ocupar más de una página*.

Lic. Victor Martinez
La Plata, 2023

Acrónimos

BI:	Business Intelligence - Inteligencia Empresarial
BPM:	Business Process Management - Gestión de Procesos de Negocio
BPMN:	Business Process Model and Notation - Modelado y Notación de Procesos de Negocio
CRM:	Customer Relationship Management - Gestión de Relaciones con los Clientes
ERP:	Enterprise Resource Planning - Planificación de Recursos Empresariales
GIT:	Sistema de control de versiones distribuido
IDE:	Integrated Development Environment - Entorno de Desarrollo Integrado
IEEE:	Institute of Electrical and Electronics Engineers - Instituto de Ingenieros Eléctricos y electrónicos
TFPM:	Task Force Process Mining - Grupo de Trabajo de Minería de Procesos
TMS:	Transportation Management System - Sistema de Gestión de Transporte
WMS:	Warehouse Management System - Sistema de Gestión de Almacenes
XES:	eXtensible Event Stream - Flujo de Eventos Extensible
XML:	eXtensible Markup Language - Lenguaje de Marcado Extensible

Glosario

Análisis de Datos:	Data Analysis. Es la ciencia de analizar datos con el objetivo de sacar conclusiones que permitan tomar decisiones o expandir el conocimiento.
Business Intelligence:	Inteligencia Empresarial. Es la transformación de datos en información y conocimiento para optimizar la toma de decisiones en procesos de negocio, generalmente se muestran los resultados en tiempo real.
Business Process Management	Gestión de Procesos de Negocio. Es una metodología para diseñar, ejecutar, analizar y mejorar procesos de negocio. Permite definir claramente todos los procesos de distintos modelos de negocio.
Business Process Model and Notation	Modelado y Notación de Procesos de Negocio. Es una metodología estándar para documentar procesos de negocio de forma gráfica. Actualmente se encuentra la especificación 2.0
Customer Relationship Management	Gestión de Relaciones con los Clientes. Es un conjunto de herramientas para administrar las relaciones con los clientes (registro, ventas, marketing, reclamos y todos los puntos de contacto), generalmente se agrupan bajo un conjunto de aplicaciones denominados CRM.
Data Mining:	Minería de Datos. Técnicas para descubrir y analizar información en grandes volúmenes de datos
Enterprise Resource Planning	Planificación de Recursos Empresariales. Es un conjunto de herramientas que centraliza todos los procesos de negocio necesarios para operar una empresa (finanzas, RRHH, facturación, manufactura, cadena de suministros, compras y otros). Generalmente se integran en una única aplicación.
Data Science:	Ciencia de Datos. Disciplina que transforma los datos en conocimiento, analizando grandes volúmenes de información.
Ingeniería de Software:	Es la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento del software; en otras palabras, es la aplicación de la ingeniería al software.
Institute of Electrical and Electronics Engineers	Instituto de Ingenieros Eléctricos y electrónicos. Es una asociación profesional internacional sin fines de lucro dedicada a generar estándares, entre otras cosas, para las nuevas tecnologías.
Integrated Development Environment	Entorno de Desarrollo Integrado. Es una herramienta que permite desarrollar y probar aplicaciones de software.
Log:	Registro o historial de log. Es una grabación en forma secuencial en un archivo de texto o base de datos que registra los eventos o acciones de un proceso en particular.

Machine Learning:	Aprendizaje Automático. Es una técnica de análisis de datos que enseña a las computadoras a aprender de la experiencia, de forma análoga a como aprenden las personas o los animales.
Process Mining:	Minería de Procesos. Técnicas para descubrir y analizar un modelo de negocio a partir de logs de eventos.
Sistema de Información:	Es un conjunto de proceso o componentes con un fin común, en informática se utilizan para recopilar, recuperar, administrar y analizar información importante para cada organización según sus procesos o características.
Transportation Management System:	Sistema de Gestión de Transporte. Es un sistema de gestión de la cadena de suministro y entrega de una empresa. Generalmente abarca todos los aspectos relevantes de la logística.
Task Force Process Mining:	Es un grupo de trabajo promovido por la IEEE con el objetivo de promover, investigar y desarrollar la Minería de Procesos.
Warehouse Management System:	Sistema de Gestión de Almacenes. Es un sistema que brinda soporte a las operaciones diarias de un almacén y brinda una visibilidad amplia de los recursos que dispone la empresa.
eXtensible Event Stream:	Es un formato de archivos de texto para registrar eventos, está basado en XML. Actualmente se considera el formato estándar para Minería de Procesos.
eXtensible Markup Language:	Es un formato de archivo de texto que utiliza una serie de etiquetas personalizadas para describir tanto la estructura, datos y otras características del documento.

Índice general

Resumen	v
Abstract	vii
Agradecimientos	ix
Acrónimos	xI
Glosario	xIV
Índice de figuras	xIX
Índice de tablas	xxI
Índice de listados	xxIII
1. Introducción	1
1.1. Motivación	4
1.2. Objetivos	5
1.3. Trabajos Publicados	5
1.4. Estructura del documento	6
2. Contexto General	7
2.1. Proceso de Distribución Postal	7
2.2. Business Process Management (BPM)	10
2.3. BPMN	13
2.4. Minería de Procesos	15
2.5. Material Relacionado	21
3. Arquitectura de la propuesta	23
3.1. Introducción	23
3.2. Producto elegido	24
3.3. Definición de la prueba	24
3.4. Herramientas de software utilizadas	25
3.5. Hardware utilizado	25
4. Minería de Procesos en la distribución Postal	27
4.1. Introducción	27
4.2. Extracción de los datos	28
4.3. Conversión de datos	29
4.4. Filtrado	30
4.5. Descubrimiento del proceso	31

4.6. Verificación de Conformidad	34
4.7. Problemas encontrados	39
5. Solución Propuesta	47
5.1. Introducción	47
5.2. Metodología aplicada	47
5.3. Arquitectura de la solución	48
5.4. Análisis de la necesidad	50
5.5. Diseño del prototipo	51
5.6. Desarrollo	54
5.7. Testeo	61
5.8. Implementación	61
6. Conclusiones	63
6.1. Justificación de competencias adquiridas	63
7. Resultados	65
Bibliografía	69
A. Publicación en CACIC 2021	73
B. Publicación en Springer 2022	85
C. Anexo Herramientas utilizadas	99

Índice de figuras

2.1.	Registro de entrega de OCA y Correo Argentino	9
2.2.	Proceso de Distribución Postal para el caso de estudio	11
2.3.	Ciclo BPM	12
2.4.	Principales componentes de BPMN	16
2.5.	Ejemplo de un proceso modelado en BPMN	16
2.6.	Ejemplo de registro de eventos en un archivo log	18
2.7.	Play-In, Play-Out, Replay	19
2.8.	Diferentes tipos de Minería de Procesos	21
3.1.	Minería de Procesos aplicada a la Distribución Postal	24
4.1.	Arquitectura de la propuesta	28
4.2.	Datos Originales	29
4.3.	Datos en formato XES	31
4.4.	Filtrado de archivos XES	32
4.5.	Proceso de Distribución Postal modelado en BPMN	33
4.6.	Proceso descubierto con Alpha+ que deberán cumplir las trazas	35
4.7.	Recorrido de las trazas: eventos más comunes en oscuro y recorridos más frecuentes en línea más gruesa	36
4.8.	Visualización de ProM de trazas con exceso de movimientos y ajuste muy bajo . . .	37
4.9.	Visualización de trazas según análisis de Performance	38
4.10.	Trazas de dos movimientos con inactividad	38
4.11.	Trazas con desvíos leves	40
4.12.	Trazas con desvíos graves	41
4.13.	Trazas exceso de movimientos	43
4.14.	Trazas con excesivo tiempo de finalización	44
4.15.	Trazas con tiempo extremadamente alto de finalización	44
5.1.	Arquitectura del prototipo a desarrollar	49
5.2.	Tablero Kanban con funcionalidades a desarrollar	52
5.3.	Épica API de Gestión de Incidencias	52
5.4.	DER del Prototipo de Minería de Procesos aplicado a la distribución postal	53
5.5.	Prueba de concepto visualización del entorno de desarrollo Spider	54
5.6.	Tabla donde se almacenan las trazas encontradas con ajuste bajo	57

Índice de tablas

4.1. Detalle de trazas y eventos utilizados	31
4.2. Detalle de irregularidades encontradas	39
4.3. Irregularidades encontradas	39

Índice de listados

5.1.	Código fuente del proceso de extracción de datos de Oracle	55
5.2.	Código SQL tabla trazasajustebajo	56
5.3.	Código en Python del componente de Minería de Procesos	57
7.1.	Código fuente en Java	66
7.2.	Ejemplo de código C	66
7.3.	Ejemplo escrito en Matlab	67

CAPÍTULO 1

Introducción

En todo proceso de negocio hay una búsqueda constante de reducir costos y mejorar el servicio a los clientes. Estas acciones contribuyen a hacer más competitivas las empresas y posicionarse mejor en el mercado. Para ello es indispensable conocer en detalle el funcionamiento del proceso de negocio, encontrar sus problemas, debilidades, cuellos de botella, posibles mejoras o nuevas oportunidades que posibiliten una mejora de servicio. Con el advenimiento de las nuevas tecnologías y la *era de la información* fueron apareciendo infinidad de nuevas técnicas, herramientas y metodologías que brindan un gran espectro de posibilidades para mejorar procesos de negocio y experiencia de los clientes.

Actualmente existe una amplia variedad de sistemas de información que dan soporte a las diferentes áreas de cada empresa ya sean propias del negocio o administrativas. Entre estas herramientas pueden encontrarse desarrollos ad hoc para cada compañía o herramientas de gestión como CRM, ERP, WMS, TMS, BI, etc.

Dada la facilidad y el bajo costo para almacenar datos en estos días, la mayoría de estos sistemas deja registro de las actividades realizadas con información para auditoría o resolución de errores en archivos de texto o bases de datos. Entre otras cosas, se registra qué tarea se realizó, cuándo se realizó, quién la hizo, si hubo algún error, etc. Analizando estos datos puede obtenerse información acerca de errores o problemas ocurridos, cual es el proceso de negocio que se realiza, como proponer mejoras o encontrar soluciones a diversos inconvenientes operativos, cuál es el comportamiento de los clientes y sus preferencias.

Esta gran disponibilidad de información contribuye al desarrollo de nuevas metodologías que permiten analizar datos en sus diferentes formas y proporcionar valor agregado en casi cualquier área de negocio.

El análisis y estudio de grandes volúmenes de información es conocido como *Data Science* o *Ciencia de Datos*.

En [1] se define como "La Ciencia de Datos es un campo interdisciplinario que tiene como objetivo convertir datos en información con valor real. Los datos pueden ser estructurados o no estructurados, grandes o pequeños, estáticos o dinámicos. El valor obtenido puede ser en forma de predicciones, decisiones automatizadas, modelos aprendidos de los datos o cualquier forma de visualización que nos brinde información."

También se expresa que "La ciencia de datos incluye la extracción, preparación, exploración y

transformación de datos. Su almacenamiento y recuperación, infraestructuras, diversos tipos de minería y aprendizaje, presentación de explicaciones y predicciones. Además en la explotación de resultados se debe tener en cuenta la ética y los aspectos sociales, legales y comerciales"^[1]

La información obtenida permite tomar decisiones o incrementar el conocimiento de algún aspecto en particular. Para ello se utilizan funciones matemáticas y estadísticas que permitan manipular y evaluar los datos.

Si bien el análisis de datos existe desde hace mucho tiempo, la denominación *Ciencia de Datos* tiene unos pocos años, en los cuales algunas metodologías han crecido mucho impulsadas por el avance tecnológico.

En general el análisis de datos se compone de varias etapas entre las que se encuentra la recopilación de requisitos, aquí es donde se define que es lo que se quiere obtener, la recopilación de datos de una o varias fuentes, la limpieza o pre procesamiento donde se sacan los datos incompletos o duplicados, y finalmente el Análisis y presentación de los resultados.

Dentro del análisis de datos podemos encontrar la **Minería de Datos (Data Mining)** y la **Minería de Procesos (Process Mining)** entre otras disciplinas.

La Minería de Datos es un proceso que combina estadística, técnicas de aprendizaje automático y tecnología para producir modelos que permitan analizar grandes volúmenes de datos.

En [2] se define la Minería de Datos como el "*Proceso de descubrimiento de patrones interesantes y conocimiento a partir de grandes volúmenes de datos*". También se expresa que la Minería de Datos satisface la necesidad actual de datos efectivos, flexibles y escalables.

El ciclo de vida de un proceso de Data Mining se define en el libro "*Data Mining Practical Machine Learning Tools and Techniques*"^[3] como una secuencia de tareas que pueden realizarse en una o varias iteraciones y que se compone de las siguientes etapas:

1. **Entendimiento del negocio:** En esta etapa se definen los objetivos de negocio y los requerimientos. Se decide qué tipo de análisis se puede aplicar para cumplirlos y se determina qué tipo de información se recolectará.
2. **Entendimiento de los datos:** En esta etapa se define y se estudia un set inicial de datos para evaluar si es posible realizar el análisis. Si la calidad de los datos es baja puede ser necesario recolectar nuevos datos con un criterio más riguroso. Con la comprensión de los datos es probable que haya que reconsiderar los objetivos de negocio y replantearse los requerimientos.
3. **Preparación de los datos:** Involucra el pre procesado de los datos iniciales para que los algoritmos puedan generar el modelo.
4. **Modelado:** Es donde se realiza el análisis. Está fuertemente relacionado con la preparación de los datos, generalmente los resultados obtenidos por el modelado brindan nuevas perspectivas que se relacionan con la preparación de los datos generando una iteración desde esa etapa con nuevas técnicas de pre procesamiento.
5. **Evaluación:** Analiza si las descripciones inferidas de los datos tienen algún valor predictivo o simplemente reflejan irregularidades espurias. Si la etapa de evaluación muestra que el

¹Traducido por el autor

modelo es pobre puede ser probable reconsiderar el proyecto entero y volver a la etapa de Entendimiento del negocio e identificar nuevos objetivos. Por otro lado si la precisión del modelo es lo suficientemente alta se procede con el despliegue.

6. **Despliegue:** Generalmente se integra el modelo en un entorno productivo en grandes sistemas de información por lo tanto el proyecto debe trasladarse a los ingenieros de software para su análisis y desarrollo (es posible que se deba realizar en un nuevo lenguaje de programación)²

La Minería de Procesos es una disciplina que permite descubrir, analizar y mejorar procesos de negocio a través de los archivos de registro dejados por los sistemas de información. Para ello, utiliza técnicas capaces de modelizar dichos procesos para luego extraer información útil relacionada con su funcionamiento que ayude a la toma de decisiones. La modelización del proceso no sólo permite comprender el funcionamiento real (basado en las acciones que se ejecutan día a día), sino que facilita la detección de desvíos operativos y cuellos de botella; aspectos que redundan en una mejora del proceso agilizando su funcionamiento y reduciendo las congestiones.

En el *Manifiesto de la minería de procesos*[4] se define la Minería de Procesos como "Una disciplina de investigación relativamente joven que se ubica entre Machine Learning y el Data Mining, por una parte, y la modelización y análisis de procesos, por otra". El registro generado por los sistemas de información puede utilizarse para obtener una mejor perspectiva de los procesos actuales, por ejemplo los desvíos pueden analizarse en detalle lo que lleva a una mejora del modelo.

La minería de procesos provee un puente importante entre la minería de datos y el modelado y análisis de procesos de negocio[4]

Al trabajar directamente con datos productivos se obtiene el comportamiento real que se realiza para el negocio y el proceso completo que se está realizando, el cual en algunos casos puede diferir del que se diseñó originalmente debido a desvíos operativos que no están registrados.

La minería de procesos puede responder interrogantes de tipo: ¿qué ocurrió realmente?, ¿por qué ocurrió?, ¿qué podría suceder en el futuro?, ¿cómo se puede mejorar el control del proceso?, ¿cómo se puede mejorar el proceso para incrementar la performance? [1].

En [4] se definen tres tipos de minería de procesos a partir de los archivos de eventos, estos son:

1. **Descubrimiento:** Se trata de técnicas capaces de modelizar el proceso de negocio que se está realizando únicamente a partir de la secuencia de logs correspondientes. La calidad del modelo generado está relacionada en forma directamente proporcional a la calidad de los datos de entrada. Esta es una característica común a cualquier técnica de extracción de conocimiento inductiva.
2. **Conformidad:** Se compara un modelo de proceso existente con un registro de eventos del mismo proceso. La verificación de conformidad puede ser usada para verificar si la realidad, tal como está almacenada en el registro de eventos, es equivalente al modelo y viceversa.
3. **Mejoramiento:** la idea es extender o mejorar un modelo de proceso existente usando la información acerca del proceso real almacenada en algún registro de eventos. Mientras la verificación de conformidad mide el alineamiento entre el modelo y la realidad, este tercer tipo de minería de procesos busca cambiar o extender el modelo.[4]

²Traducido por el autor

En este trabajo se aplicará *Ciencia de Datos* a la distribución de productos postales en la República Argentina. Puntualmente se aplicarán técnicas de *Minería de Datos* y *Minería de Procesos* para analizar los archivos de registro generados por los sistemas de información postales con el objetivo de analizar y proponer mejoras al proceso de distribución.

En el proceso de distribución postal se lleva un registro de todas las actividades, desde el ingreso del producto hasta su entrega al cliente, estas deben realizarse en un orden y en un intervalo de tiempo específicos.

Se trata de un proceso en el que se registran distintos movimientos conformados por eventos tales como la recepción del producto en sucursal, el ingreso a un centro de distribución, los movimientos internos y los diversos intentos de entrega con sus resultados (pudiendo ser una entrega efectiva o no). Todas las tareas están claramente definidas y deben cumplirse para asegurar la calidad del servicio a los clientes. Ocasionalmente se producen desvíos, estos pueden por ser redundancia o inconsistencia de tareas (se repiten o no se realizan en el orden correspondiente), excesivo tiempo para su finalización u otros motivos.

Este trabajo continúa la investigación publicada en [5] donde se aplican de técnicas de Minería de Procesos a la distribución Postal en la República Argentina y en [6] donde se incorpora un análisis de performance por medio del cual es posible identificar un nuevo tipo desvíos operativos. En ambos casos se pudieron detectar diferentes tipos de desvíos operativos con gran precisión.

Los datos analizados corresponden a la distribución postal efectuada en la República Argentina en el período comprendido entre los años 2017 y 2020.

A la fecha de generación de este documento los autores no conocen la existencia de trabajos con estas características para el negocio postal en la República Argentina.

1.1. MOTIVACIÓN

Durante mi desempeño de más de 12 años en desarrollo de Sistemas de Información para el sector postal tuve la oportunidad de conocer detalladamente el proceso de entrega de varios productos con diferentes características (entrega bajo puerta, con firma, producto de alta seguridad, etc)³ En los diversos proyectos que participé tuve la oportunidad de trabajar en conjunto con integrantes de otras áreas de negocio como Marketing, Ingeniería de procesos y Operaciones encontrando situaciones o inconvenientes operativos transversales a los diversos sectores y que afectan la calidad de servicio. Al no encontrar una herramienta capaz de solucionar estos inconvenientes o al menos disminuir su impacto decidí realizar un análisis detallado utilizando Ingeniería de Software y Ciencia de Datos.

Según la definición de la IEEE[7], la *Ingeniería de Software* es la aplicación de un enfoque sistemático, disciplinado y cuantificable al desarrollo, operación y mantenimiento del software; es decir, la aplicación de la ingeniería al software.

La ingeniería de software se ha desarrollado desde hace aproximadamente 50 años buscando diseñar, desarrollar y mantener herramientas de software que se ajusten a las necesidades de las personas y de las organizaciones.

³Distintas modalidades de entrega en el negocio postal. Bajo Puerta: se deja en domicilio sin verificar si hay alguien. Con firma: la persona que recibe debe firmar una planilla y acreditar identidad. Producto de alta seguridad: se realiza un seguimiento detallado de cada movimiento sumado a los requisitos de entrega con firma.

Basado en mi experiencia profesional en el desarrollo de software, mi conocimiento de la *Ingeniería de Software* y el proceso postal decidí utilizar técnicas de Ciencia de Datos para analizar el proceso de distribución postal en la República Argentina.

1.2. OBJETIVOS

1.2.1. Objetivo principal

Analizar el proceso de distribución postal en la República Argentina utilizando Ciencia de Datos para encontrar desvíos operativos de forma temprana y generar un activo de información que permita tomar decisiones.

1.2.2. Objetivos secundarios

- Aplicar Minería de Procesos y Minería de Datos a la distribución Postal
- Identificar rápidamente desvíos en la operación
- Realizar el análisis de la solución requerida para cubrir la problemática
- Generar entregables de fácil interpretación para personas no técnicas
- Documentar los pasos realizados en el análisis
- Analizar el modelo obtenido para evaluar la incorporación de alarmas

1.3. TRABAJOS PUBLICADOS

Los resultados de las investigaciones realizadas en el marco de esta tesis fueron publicados a través de los siguientes artículos:

1.3.1. Minería de Procesos aplicada a la Distribución Postal

Presentado en el *XXVII Congreso Argentino de Ciencias de la Computación (CACIC 2021)*. [5]

En este trabajo se realizó el descubrimiento del proceso postal a partir de una muestra del log de distribución utilizando técnicas de *Minería de Procesos*.

Luego se comparó el proceso descubierto contra otro archivo de log más extenso (aproximadamente 16.000 trazas con alrededor de 44.000 eventos de distribución) para verificar el cumplimiento del mismo. Se comprobó que la gran mayoría de los envíos son entregados y los que no se pueden entregar se distribuyen uniformemente entre los diversos motivos de no entrega. El proceso se cumple en tiempo y forma para el 85 % de las trazas. También se encontraron desvíos operativos y se realizó un análisis sobre los mismos.

Estos desvíos estaban dados en dos grandes grupos, por un lado los que tenían eventos repetidos, es decir el mismo evento cargado en más de una oportunidad generalmente en días diferentes y por otro lado inconsistencias donde los eventos secuenciales no se ejecutaban en el orden correcto incluso se repetían en sucesivos días. Estos desvíos estaban en el orden del 10 % del total. El 5 % restante eran trazas cuyos desvíos eran menores para tener en cuenta en el análisis.

1.3.2. Distribution Analysis of Postal Mail in Argentina Using Process Mining

Publicado en *Communications in Computer and Information Science. Springer.* [6]

En este trabajo se continúa con el análisis sobre la muestra de 16.000 trazas que se realizó en [5] adicionando un control de performance que permite identificar aquellos envíos que cumplen correctamente el proceso pero en un intervalo de tiempo muy alto, ocasionando demoras en la entrega. Con este nuevo análisis se encontró que alrededor de un 4 % de las trazas tenían este problema. En ambos trabajos publicados se comprobó que la utilización de técnicas de *Minería de Procesos* ayuda a mejorar un proceso de negocio permitiendo identificar rápidamente desvíos operativos.

1.4. ESTRUCTURA DEL DOCUMENTO

El presente documento está estructurado de la siguiente forma:

1. **Introducción:** Descripción general del trabajo a realizar
2. **Material Relacionado** Descripción de las diferentes metodologías de Ciencia de Datos que se utilizarán en la investigación y el material adicional.
3. **Arquitectura propuesta** Detalle de las técnicas que se aplicarán para el análisis.
4. **Implementación** Aplicación de *Minería de Procesos* a la distribución postal con datos de la República Argentina entre los años 2017 al 2020.
5. **Mejoras** Diseño de un prototipo para incluir las técnicas de *Minería de Procesos* a la distribución postal y alertar de forma temprana posibles desvíos.
6. **Conclusiones:** Se expresarán en este capítulo los resultados obtenidos y futuras líneas de investigación.

CAPÍTULO 2

Contexto General

2.1. PROCESO DE DISTRIBUCIÓN POSTAL

La distribución postal consiste en llevar un producto desde un lugar de origen a un destino. En la República Argentina esta distribución o entrega puede ser nacional donde el origen y el destino se encuentran dentro del país (se puede llevar de cualquier punto a cualquier otro dentro del país) o internacional donde o la persona que envía o la persona que recibe están fuera del país (puede ser internacional saliente o internacional entrante según sea el caso).

El cliente que solicita enviar un producto se conoce como *Remitente*, la persona que recibe el envío se conoce como *Destinatario*.

El envío puede ser de varios productos con distintas características, entre ellos se encuentran los diferentes tipos de cartas (certificada, expreso, simple), cartas documento o telegramas, estos son los productos *Tradicionales* del negocio postal. También se incluyen productos de *Paquetería* que pueden ser encomiendas o entregas de tipo *Puerta a Puerta* o productos de *Alta Seguridad* como tarjetas de crédito, débito o medicamentos.

Cada producto cuenta con características diferentes en cuanto a estándares de entrega, forma de recepción, cantidad de visitas, registro de entrega, seguimiento o incluso adicionales como *Aviso de recibo*, *Valor declarado* o *telegrama colacionado* que cambian las condiciones de servicio.

La descripción de cada producto y su implementación dependen de cada empresa, por ejemplo en el caso particular del Correo Oficial de la República Argentina estos se detallan en el *Manual de Productos y Servicios (PS-OO-001)* un documento interno que especifica cada producto que brinda la empresa a sus clientes.

Los productos postales se ofrecen en tres segmentos Corporativo, PyME y minorista, con condiciones diferentes en cada uno. Esta diversidad de productos, segmentos y tipos de envío genera una gran variedad de formas y procesos de entrega. Esto genera una actividad de alta complejidad donde intervienen varias personas de diferentes lugares en distintas etapas de cada proceso.

En la distribución postal se lleva un registro de todas las actividades, desde el ingreso del producto hasta su entrega al cliente. Entre estas actividades se encuentra la recepción del envío, ingreso a un almacén, traslados internos o intentos de entrega, entre otras.

Todas las tareas deben realizarse en un orden y en un intervalo de tiempo específicos. Dada la complejidad del proceso es frecuente que se produzcan errores o desvíos, estos pueden ser redundancia

o inconsistencia de tareas (se repiten o no se realizan en el orden correspondiente), excesivo tiempo para su finalización u otras. Bajo estas circunstancias es de vital importancia monitorear y evaluar el o los procesos de distribución para proporcionar un servicio de calidad a los clientes.

Las empresas de correo proporcionan información a los clientes para realizar el seguimiento del envío y observar los principales movimientos, esta información es parte del producto y permite brindar transparencia y una mejor calidad de servicio a los clientes. En general se dispone de un portal donde se pueden visualizar los movimientos previamente registrados del envío.

En la figura 2.1 se muestra el resultado de entrega y la historia del envío en los sitios web de Correo Argentino¹, Ocasa², Andreani³ y de OCA⁴.

En todos los casos, el proceso postal consta al menos de las siguientes etapas:

1. **Ingreso a Correo:** El producto a entregar ingresa a través de una sucursal, un punto de retiro o un sistema informático según sea el caso dependiendo de las características del producto y del cliente.
2. **Asignación de número de seguimiento:** A cada producto se le asigna un identificador único sobre el cual se registrarán todos los eventos relacionados con la distribución.
3. **Envío Sede Central Operativa:** Los productos se envían a un centro de clasificación y distribución donde agrupan de forma manual o automatizada según lugar de destino (por localidad y código postal) y características.
4. **Traslado a Centro de Distribución:** Los productos se envían al centro de distribución más cercano al destino para la etapa final de entrega.
5. **Entrega en domicilio:** Los distribuidores (carteros) realizan el reparto de los productos, dependiendo de las características se puede intentar la entrega en más de una oportunidad.
6. **Rendición:** Se brinda al cliente (Remitente) un informe con el resultado de la entrega y los diferentes estados del proceso.

En la República Argentina el servicio postal está regulado por el Ente Nacional de Comunicaciones (ENACOM)⁵ y la principal empresa del rubro es el *Correo Oficial de la República Argentina S.A.* con más de 1800 sucursales y llegada a todos los puntos del país.

Para la realización de este trabajo se eligió el producto *Carta Documento* ya que es un producto de alta seguridad, con registro y seguimiento de todas las etapas del proceso, entrega bajo firma del destinatario, con dos intentos de entrega dejando aviso de visita en ambos casos. Si el envío no se pudo entregar se espera en sucursal un período de tiempo a que el destinatario lo venga a retirar y una eventual devolución al cliente en caso de no entrega.

El proceso del producto elegido se detalla en la figura 2.2 y se describe de la siguiente manera:

El producto ingresa a una sucursal de correo donde se recepciona por un operador. El operador verifica que el envío contenga todos los datos necesarios para su correcta distribución (que los datos

¹<https://www.correoargentino.com.ar/formularios/ondnc>

²<https://www.ocasa.com/>

³<https://www.andreani.com/personas>

⁴<https://www.oca.com.ar/Busquedas/Envios>

⁵<https://www.enacom.gob.ar/>

The figure displays four screenshots of tracking interfaces:

- OCA Tracking (Left):** Shows a history of movements for tracking number CD-192870025-AR. Movements include Intento de Entrega at CDD Monte Grande, Poder del Cartero, and Domicilio Cerrado/Visita; Llegada al Centro de Procesamiento at CTP Buenos Aires; and Ingreso al Correo at CTP-BUE Grandes Clientes.
- OCA Tracking (Top Right):** A simplified timeline showing the status of the shipment from Retirado to En proceso de Entrega.
- Correo Argentino Tracking (Bottom Left):** An interface showing the movement of tracking number 360000197263480 to CORDOBA, CORDOBA. It includes a table of events and a bell icon for notifications.
- Correo Argentino Tracking (Bottom Right):** A detailed timeline of events for tracking number CAA40740552, starting with Receipt at Sucursal OCA - LOMAS DE ZAMORA and ending with Delivery at LOMAS DE ZAMORA.

Figura 2.1: Registro de entrega de OCA y Correo Argentino

para la entrega estén correctamente cargados, tanto del remitente como del destinatario) y que el producto cumpla con las características requeridas de tamaño y peso. Si todo es correcto se procede con la recepción, en caso contrario se rechaza el envío.

Al ingresar el producto se le asigna un número de seguimiento único y diferente al de cualquier otro envío, este número se utilizará para registrar todos los eventos asociados al envío, generando una historia completa del proceso, comenzando por su recepción en sucursal.

Luego se trasladan los envíos al centro de distribución donde se registra su ingreso y se realiza la clasificación a destino. A continuación se traslada el producto a sucursal de destino registrando su ingreso en la misma. Luego se procede con los intentos de entrega, si el envío se entrega se correctamente se registra el evento y finaliza el proceso. Si el envío no puede entregarse debido a un motivo definitivo (por ejemplo problemas con la dirección, no existe el número, dirección incompleta, se mudó o falleció el destinatario, etc) se registra el motivo y finaliza el proceso.

Si el motivo de no entrega no es definitivo (el destinatario no estaba) se deja un aviso de visita y se procede con un segundo intento de entrega al próximo día hábil. En caso de entrega en el segundo intento finaliza el proceso, si no se pudo entregar nuevamente se deja un aviso de visita y se guarda el envío en sucursal durante una semana esperando que el destinatario haga el retiro. En caso de que el destinatario no pase a retirar el envío luego del tiempo especificado se procede a la devolución del mismo al remitente.

Dada la complejidad del proceso y la cantidad de actores intervenientes se producen habitualmente demoras y problemas con la entrega, entre los más comunes están omitir algún paso en el proceso, no cumplimiento con los estándares de entrega.

2.2. BUSINESS PROCESS MANAGEMENT (BPM)

La Administración de Procesos de Negocio (BPM por sus siglas en inglés) es un enfoque disciplinado para identificar, diseñar, ejecutar y monitorear procesos de negocio. Estos pueden ser automatizados o no automatizados.

BPM Se enfoca en encontrar las causas principales de los problemas de performance y su mejora. Busca descubrir las debilidades del proceso y no de la gente que lo realiza, esto puede ser extremadamente complejo debido a la cantidad de procesos que puede tener una organización, incluso los procesos pueden funcionar correctamente y aún no alcanzar los niveles de performance requeridos por la organización o los clientes.

Utilizando BPM se busca obtener resultados que estén alineados con los objetivos de la organización. En [8] se define el proceso de BPM como un ciclo que comienza con la creación de un proceso formal, este debe estar claramente definido y documentado. Este primer paso es muy importante ya que muchas veces las organizaciones carecen de un proceso definido de extremo a extremo (por ejemplo desde la adquisición de materia prima hasta la entrega al cliente), en ocasiones se resuelven temas con improvisación y sin que quede registro de ello. Luego de que el proceso está en marcha debe monitorearse de forma continua, y se debe evaluar su rendimiento con métricas que se relacionan con las necesidades del cliente y de la empresa. Estas métricas se comparan con los objetivos del proceso, los mismos pueden basarse en las expectativas de los clientes, los puntos de referencia de

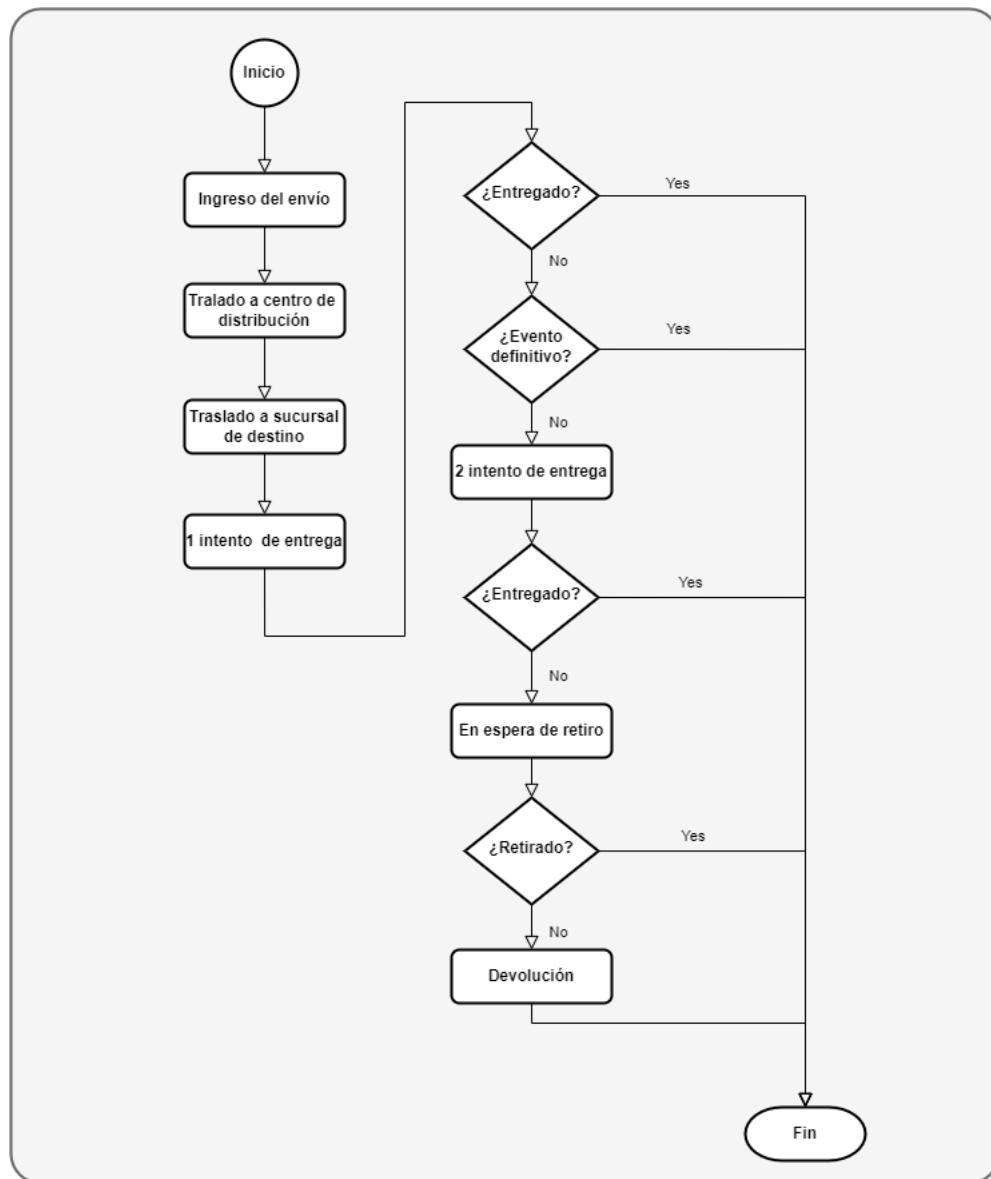


Figura 2.2: Proceso de Distribución Postal para el caso de estudio

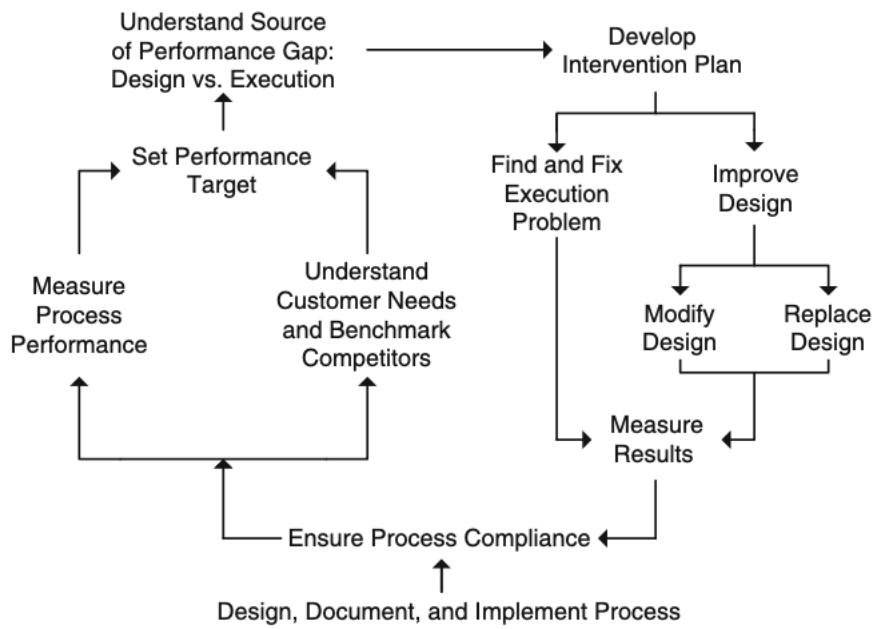


Figura 2.3: Ciclo BPM

la competencia, las necesidades de la empresa y otras fuentes. Si el rendimiento no cumple con los objetivos, se debe determinar el motivo y rediseñar el proceso.

En general los procesos no cumplen con los requisitos de rendimiento, ya sea por un diseño defectuoso o una ejecución parcial o deficiente. Si la falla está en la ejecución, se debe determinar la causa raíz (como entrenamiento inadecuado, o recursos insuficientes, o equipo defectuoso, u otras). Hacerlo es una tarea difícil, debido al gran número de posibles causas fundamentales, sin embargo en general, una vez que se ha encontrada la causa raíz, es fácil de solucionar. Por el contrario si el problema está en el diseño del proceso, en general es fácil de encontrar (debido al rendimiento inadecuado) pero mucho más difíciles de solucionar ya que pueden requerir un rediseño completo del proceso. Una vez que se ha elegido e implementado la modificación adecuada, se evalúan los resultados y todo el ciclo comienza de nuevo.

En la figura 2.3 se muestra el ciclo de BPM propuesto en [8]

BPM se basa en que el rendimiento de una organización no se enfoca en prueba y error, o empujando a la gente más fuerte, a través de la manipulación financiera, sino a través de la gestión deliberada de los procesos comerciales de extremo a extremo a través de los cuales se crea todo el valor para el cliente. De hecho, BPM es un enfoque centrado en el cliente para la gestión organizativa. [8] Los clientes evalúan los resultados. Tales resultados provienen de los procesos comerciales, de las secuencias de actividades que trabajan juntas. Optimizando los procesos se obtienen mejores resultados que derivan en satisfacción de los clientes. De todas maneras, tener un buen diseño para un proceso no garantiza buenos resultados debido a la complejidad del mundo real, las máquinas se rompen, las personas se van, los datos se corrompen, etc. Ocasionalmente que un proceso bien diseñado no logre el rendimiento que puede dar.

Las empresas que utilizan la gestión de procesos para supervisar y verificar el rendimiento de cada proceso pueden reconocer y corregir problemas de rendimiento de forma mucho más dinámica.

También descubren más fácilmente las oportunidades de mejorar el diseño del proceso

En [8] se define que BPM establece cinco "Facilitadores" para diseñar un proceso de alto rendimiento, estos son:

1. **Diseño del Proceso:** Es el aspecto más importante del proceso. Aquí se detalla que tareas se deben realizar, por quién, en qué momento y en qué lugar bajo qué circunstancias, el grado de precisión, con qué información y similares. El diseño es la especificación del proceso; sin un diseño, solo hay una actividad individual descoordinada y un caos organizativo.
2. **Métricas del Proceso:** Los procesos necesitan métricas de extremo a extremo que se derivan de las necesidades del cliente y los objetivos de la empresa. Es necesario establecer objetivos en términos de estas métricas y el rendimiento supervisado en relación con ellas. Se debe implementar un conjunto equilibrado de métricas de proceso (como costo, velocidad y calidad), para que las mejoras en un área no oculten las caídas en otra.
3. **Ejecutantes del Proceso:** Las personas que trabajan en procesos necesitan un conjunto diferente de habilidades y comportamientos de los que trabajan en funciones y departamentos convencionales. Necesitan una comprensión del proceso general y sus objetivos, la capacidad de trabajar en equipo y la capacidad de gestionarse a sí mismos. Sin estas características, no podrán realizar el potencial del trabajo de extremo a extremo.
4. **Infraestructura del Proceso:** Los ejecutantes deben contar con el apoyo de los sistemas de información y recursos humanos para cumplir con las responsabilidades del proceso. Los sistemas de información funcionalmente fragmentados no admiten procesos integrados, y los sistemas convencionales de recursos humanos (formación, compensación y carrera, etc.) refuerzan las perspectivas de trabajo fragmentadas. Se necesitan sistemas integrados (como sistemas ERP y sistemas de compensación basados en resultados) para los procesos integrados.
5. **Dueño del Proceso:** En una organización convencional, nadie es responsable de un proceso de extremo a extremo, por lo que nadie estará en condiciones de gestionarlo de extremo a extremo (es decir, llevar a cabo el ciclo de gestión de procesos). Una organización que se toma en serio sus procesos debe tener propietarios de procesos: altos directivos con autoridad y responsabilidad por un proceso en toda la organización en su conjunto. Ellos son los que realizan el trabajo ilustrado en la figura 2.3.

2.3. BPMN

Business Process Model and Notation (BPMN) es una notación gráfica estándar diseñada para representar y documentar las actividades que conforman los procesos de negocio de una organización y los mensajes que se intercambian entre los participantes de las diferentes actividades. BPMN es mantenido por el Object Management Group (OMG) una organización internacional sin fines de lucro con el objetivo de generar estándares tecnológicos que provean valor agregado cuantificable para la industria⁶.

Actualmente hay una gran variedad de herramientas de gestión de procesos empresariales (BPM), varias basadas en XML proporcionando mecanismos formales para la definición de procesos de

⁶<https://www.omg.org/>

negocio. En la mayoría de los casos estas herramientas están diseñadas para interactuar con sistemas de BPM y están orientados a usuarios avanzados o con grandes conocimientos técnicos o matemáticos. Por sus características un proceso de negocio complejo puede organizarse en un formato complejo y poco intuitivo, que se ajusta muy bien a las herramientas de software (o un programador) pero de difícil comprensión para analistas o gerentes de una organización que son los que generalmente se encargan de desarrollar, administrar y monitorear el proceso.

BPMN pretende proporcionar una notación fácilmente legible y entendible por todos los actores que participan en el proceso. En general, estos actores (gerentes, analistas) se sienten muy cómodos visualizando los procesos de negocio en un formato de diagrama de flujo.

En [9] se expresa que hay una gran cantidad de analistas de negocio con diagramas de flujo simples. Esto crea una brecha técnica entre el formato en que los procesos de negocio son diseñados y el formato de los lenguajes, que ejecutarán estos procesos de negocio. Esta brecha debe cerrarse con un mecanismo formal que mapee la visualización adecuada de los procesos de negocio (una notación) al formato de ejecución apropiado (un lenguaje de ejecución BPM) para estos procesos de negocio. Esta brecha se puede resolver con la estandarización del Modelo y Notación de Procesos de Negocio (BPMN)

BPMN proporciona múltiples diagramas, que están diseñados para su uso por las personas que diseñan y gestionan los procesos de negocio, además proporciona un mapeo a un lenguaje de ejecución. Por lo tanto, BPMN permite tener un mecanismo de visualización estándar para los procesos de negocio definidos en un lenguaje de procesos de negocio optimizado para la ejecución.

BPMN proporciona a las empresas la capacidad de comprender sus procedimientos comerciales internos en una notación gráfica y brinda a las organizaciones la capacidad de comunicar estos procedimientos de manera estándar. Actualmente, hay decenas de herramientas y metodologías de modelado de procesos. Dado que las personas se moverán de una empresa a otra y que las empresas se fusionarán y divergirán, es probable que los analistas de negocios necesiten comprender múltiples representaciones de los procesos de negocio, representaciones diferentes del mismo proceso a medida que se mueve a través de su ciclo de vida de desarrollo, implementación, ejecución, monitoreo y análisis. Por lo tanto, una notación gráfica estándar facilita la comprensión de las colaboraciones de rendimiento y las transacciones comerciales dentro de las organizaciones y entre ellas.[9]

La versión actual es BPMN 2.0. Esta versión permite el modelado de procesos de negocio extremo a extremo.

Hay tres tipos básicos de submodelos dentro de un modelo BPMN de extremo a extremo:

1. **Procesos:**Estos pueden ser

- a) Procesos privados: Son los internos de una organización específica. Se conocen como flujos de trabajo y pueden ser de dos tipos : ejecutables y no ejecutables. Un proceso ejecutable fue modelado para realizar una acción específica, un proceso no ejecutable se realiza para documentar un proceso existente en forma detallada, es decir tiene toda la información necesaria para su ejecución
- b) Procesos públicos: Un proceso público representa la interacción de un proceso privado y algún otro participante externo a la organización

2. **Coreografías:** Una coreografía es una definición del comportamiento esperado, básicamente un contrato de procedimiento, entre los participantes que interactúan. Mientras que existe un proceso normal dentro de un grupo, existe una coreografía entre grupos (o participantes). La coreografía se parece a un proceso de negocio privado, sin embargo, una coreografía es diferente en el hecho de que las actividades son interacciones que representan un conjunto (1 o más) de intercambios de mensajes, que involucra a dos o más participantes. Además, a diferencia de un Proceso normal, no hay un controlador central, una entidad responsable o un observador del Proceso.
3. **Colaboraciones:** Una colaboración describe las interacciones entre dos o más entidades comerciales. También se pueden mostrar los mensajes asociados con los flujos de mensajes. La colaboración se puede mostrar como dos o más procesos públicos que se comunican entre sí.

Fuente: [9]

Todos los procesos tienen un evento de inicio y un evento de fin. BPMN proporciona una gran cantidad de componentes para documentar cada etapa del proceso, los principales se muestran en la figura 2.4

Hay 5 categorías donde se agrupan los componentes de BPMN, estas son:

1. **Objetos de flujo:** son los principales elementos gráficos para definir el comportamiento de un proceso de negocio. Hay tres objetos de flujo: eventos, actividades y puertas de enlace
2. **Información:** La información se representa con cuatro elementos: entradas de datos, salidas de datos, objetos de datos y almacenes de datos.
3. **Conexión de objetos:** La conexión puede ser de los objetos de flujo entre sí o con otra información y puede ser a través de: flujos de secuencia, flujos de mensaje y asociaciones
4. **Grupos:** permiten agrupar los elementos, puede ser con grupos o carriles
5. **Artefactos:** se utilizan para proporcionar información adicional sobre el proceso. Hay dos artefactos estandarizados agrupación y anotación de texto.

Fuente:[9]

En la figura 2.5 se muestra un ejemplo de un proceso modelado y documentado con BPMN⁷. En el presente trabajo se utilizará BPMN para modelar el proceso de distribución postal.

2.4. MINERÍA DE PROCESOS

Como se mencionó anteriormente la *Minería de Procesos* es parte de la *Ciencia de Datos*.

En la Ciencia de Datos hay algunas preguntas genéricas que se pueden realizar en cualquier situación. Por ejemplo la pregunta ¿Qué pasó?, si hay desvíos o un cuello de botella y estamos registrando datos sobre esos eventos podemos ver que qué cosas pasaron.

La siguiente pregunta es ¿Por qué pasó?, es decir ¿por qué hubo tal demora o desvío?. Nuevamente si registramos datos de los eventos podemos obtener esta información. Estas preguntas son sobre el pasado pero también con la *Ciencia de Datos* podemos responder preguntas sobre el futuro, por lo tanto podemos preguntarnos ¿Qué pasará ante determinada situación?, ¿Qué podemos aprender de

⁷<https://www.bpmn.org/>

Elemento	Descripción	Notación
Evento	Un evento es algo que "sucede" durante el curso de un proceso. Generalmente tienen una causa (disparación) o un impacto (resultado)	○
Actividad	Una actividad es un término genérico para el trabajo que la empresa realiza. Una actividad puede ser atómica o no atómica (compuesta). Los tipos de actividades que forman parte de un modelo de proceso son: Subproceso y Tarea	□
Gateway	Una puerta de enlace se utiliza para controlar la divergencia y convergencia de los flujos de secuencia en un proceso y en una coreografía. Por lo tanto, determinará la ramificación, la bifurcación, la fusión y la unión de caminos.	◇
Flujo de secuencia	Un flujo de secuencia se utiliza para mostrar el orden en que las actividades se realizarán en un proceso y en una coreografía.	→
Flujo de mensajes	Un flujo de mensajes se utiliza para mostrar el flujo de mensajes entre dos participantes que están preparados para enviarlos y recibirlos.	—→
Asociación	Una asociación se utiliza para vincular información y artefactos con elementos gráficos de BPMN. Una punta de flecha en la Asociación indica una dirección de flujo (por ejemplo, datos), cuando corresponda→
Grupo	Un grupo es la representación gráfica de un participante en una colaboración. Es un contenedor gráfico para particionar un conjunto de actividades de otros grupos. Puede no tener detalles internos, es decir, puede ser una "caja negra".	[Name]
Carril	Un carril es una subpartición dentro de un proceso, a veces dentro de un grupo, y se extenderá toda la longitud del proceso, ya sea vertical u horizontalmente. Los carriles se utilizan para organizar y categorizar las actividades.	[Name] [Name]
Datos	Proporcionan información sobre qué actividades deben realizarse y/o qué producen. Pueden representar un objeto singular o una colección de objetos. La entrada y salida de datos proporcionan la misma información para los procesos	File
Mensaje	Un mensaje se utiliza para describir el contenido de una comunicación entre dos participantes o una entidad asociada empresarial	✉
Inicio	Indica el comienzo de un proceso	○
Fin	Indica el fin de un proceso	●

Figura 2.4: Principales componentes de BPMN

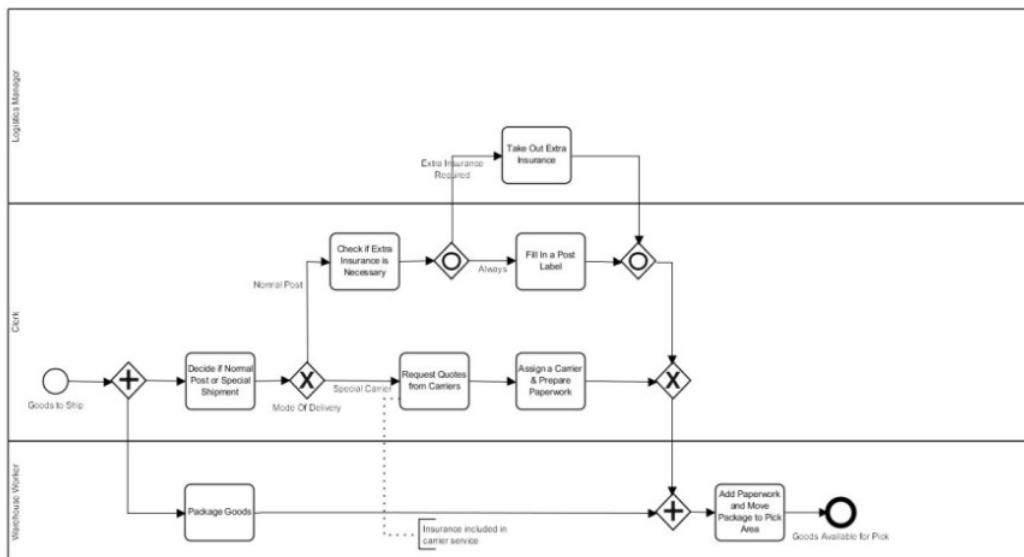


Figura 2.5: Ejemplo de un proceso modelado en BPMN

la información histórica para hacer predicciones sobre lo que sucederá?, o ¿Cómo podemos mejorar esta demora o desvío?

Por lo tanto con la *Ciencia de Datos* podemos usar analítica para mejorar un comportamiento, para ello debemos tener en cuenta que deberemos trabajar con grandes volúmenes de datos, que se obtendrán de una gran variedad de fuentes, que están en constante crecimiento, además de considerar factores éticos y sociales que nos permitan extraer conocimiento de esos datos. El desafío no es guardar datos, en los últimos tiempos se incrementó notablemente el volumen de datos que se registra, el desafío es verificar la veracidad de los datos almacenados y su interpretación.

Podemos usar estadística y Minería de Datos para obtener información muy valiosa, pero estas técnicas están principalmente enfocadas en los datos y no en los procesos que registran esos datos.

La *Minería de Procesos* pretende responder estas preguntas y obtener esta información con el foco en análisis de procesos basados en datos.

No se desea mejorar los datos, se desea mejorar el proceso que genera esos datos. Se busca una mejora en el proceso de extremo a extremo. Este enfoque nos permite generarnos nuevas preguntas como ¿Cuál es el proceso que la gente realmente sigue? ¿Qué hace realmente?, ¿Cuáles son los cuellos de botella?, ¿Qué los está causando?, ¿Dónde y por qué la gente o máquinas se desvían de un proceso esperado?

La Minería de Procesos busca aprender los procesos de negocio y mejorarlo.

En [1] se pone a la *Minería de Procesos* como una metodología que cierra la brecha entre el modelado clásico de procesos (BPM) donde no se tienen en cuenta los datos a la hora de diseñar el proceso y la *Minería de Datos* o la *Estadística* que se centran en los datos pero no analizan el proceso que los genera.

Se realiza *Minería de Procesos* para responder preguntas relacionadas con el rendimiento y con el cumplimiento con el objetivo de tener mejores resultados.

El punto de partida para la minería de procesos son los datos de eventos, estos son almacenados por los sistemas de información en archivos de log o en bases de datos en varios formatos. Un evento es algo que ocurrió y que tiene propiedades, generalmente se registra que es lo que pasó (descripción del evento), quién lo realizó, la fecha en que se hizo, además de un identificador del evento y un identificador del caso (o traza) que permita agrupar todos los eventos que le pertenezcan. Los formatos de los archivos de eventos pueden diferir dependiendo de los sistemas de información que los generan o del propósito para el que fueron generados, pero casi siempre contienen al menos los datos anteriormente mencionados.

El punto más importante en los archivos de registro es la veracidad de la información, es decir la calidad del archivo. Debido a que son el punto de entrada de la *Minería de Procesos* se debe contar con archivos de gran calidad ya que los resultados obtenidos dependen en gran cantidad de ellos, por eso al momento de diseñar que datos se guardarán se deberá evaluar cuales son los más descriptivos de las tareas que se están registrando.

El formato elegido por la Task Force on Process Mining⁸ de IEEE para los archivos de log es el CES⁹

⁸<https://www.tf-pm.org/>

⁹<https://xes-standard.org/>

Identificador de traza	Identificador de evento		Descripción del evento		Fecha evento
	123 trazalD	123 EvelD	abc eveDescrip		
31	481,053	0	INGRESADO	2017-08-16 10:45:22	
32	481,053	2	1 INTENTO DE ENTREGA	2017-08-18 11:15:00	
33	481,053	9	DEVOLUCION	2017-08-18 13:00:00	
34	481,054	0	INGRESADO	2017-08-16 10:45:28	
35	481,054	2	1 INTENTO DE ENTREGA	2017-08-22 12:05:00	
36	481,054	9	DEVOLUCION	2017-08-22 17:25:00	
37	481,055	0	INGRESADO	2017-08-16 10:45:27	
38	481,055	1	ENTREGADO	2017-08-22 15:13:00	

Figura 2.6: Ejemplo de registro de eventos en un archivo log

el cual es soportado por las herramientas principales herramientas de *Process mining* como ProM¹⁰ y Disco¹¹

En la figura 2.6 se muestra un ejemplo de registro de eventos en un archivo de log con el formato necesario para aplicar *Minería de Procesos*.

Por ejemplo en el caso de una biblioteca se puede tener el caso de un préstamo de un libro donde posiblemente se registre la reserva del libro por el socio, su retiro y devolución en tres eventos distintos. Estos tres eventos componen la historia de ese préstamo y deben ejecutarse en ese orden. No sería lógico tener una devolución del préstamo antes que la entrega, aunque estas situaciones pueden darse por errores en el registro de los datos o en el proceso.

Como se mencionó la *Minería de Procesos* opera sobre los archivos de registro de uno o varios sistemas de información. Se comienza con una extracción de datos que contenga todos los eventos de una determinada actividad, luego de filtrar estos datos se puede realizar un análisis automático donde se determina cual es el proceso de negocio, que actividades lo componen y que secuencia se debe cumplir. De esta forma es posible hacer un análisis y responder preguntas como ¿Qué fue lo que pasó?, ¿Por qué pasó?, ¿Qué es lo que podría pasar en el futuro?, ¿Cómo se puede mejorar el control?, ¿Cómo se puede mejorar la performance? [1] Con la información obtenida luego de la interpretación del modelo se podrán realizar diferentes tipos de análisis como por ejemplo: quién realiza la actividad, si la actividad está completa y si se realiza en un intervalo de tiempo coherente.

Si nos enfocamos en la relación entre los modelos de proceso y los datos de eventos encontramos en la *Minería de Procesos* tres tipos de relaciones, dependiendo de en que sentido se relacionen llamadas Play-out, Play-in y Replay.

- *Play-out*: Se comienza desde el diseño del modelo de forma manual y a partir de ahí se genera comportamiento, es decir se generan registros con información que verifica el proceso. Esta información cumple con el proceso *ideal*, generalmente se usa para verificar manualmente el proceso generado.

Play-out ejemplifica el uso clásico de BPM para generar un modelo sin interactuar con datos de eventos.

- *Play-in*: Por el contrario, aquí se parte de los registros de eventos y se genera el modelo

¹⁰<https://www.promtools.org/doku.php>

¹¹<https://fluxicon.com/disco/>

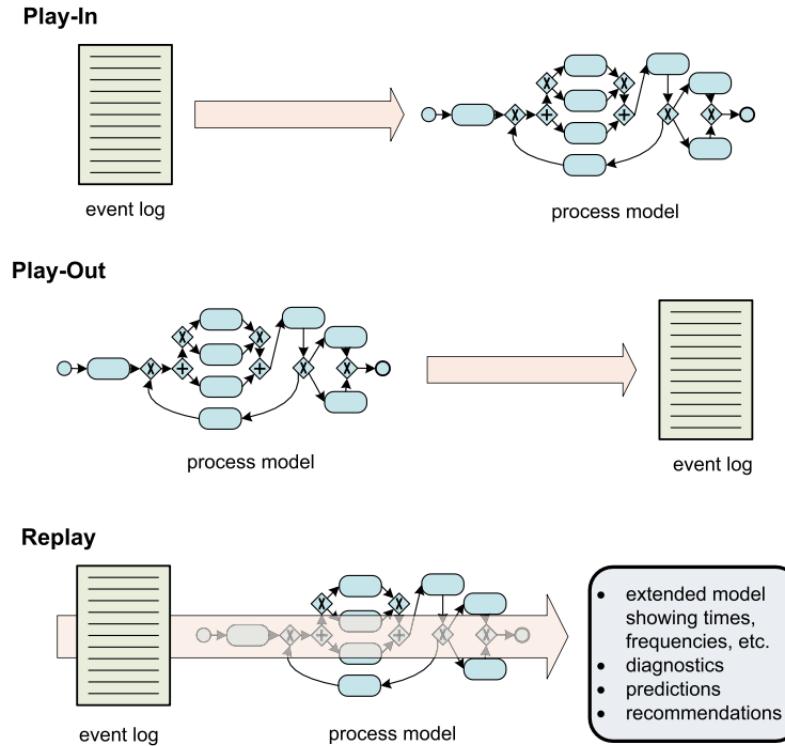


Figura 2.7: Play-In, Play-Out, Replay

correspondiente, en este caso el proceso es *descubierto* desde los archivos de log. La idea básica es mirar el número de comportamientos que se encuentran registrados. Y automáticamente deducir un modelo. No se genera ningún modelo a mano. Se infiere automáticamente modelos de proceso a partir de datos de eventos sin procesar.

Play-in se relaciona con el descubrimiento, aprender automáticamente un modelo de negocio sin ninguna información adicional que los registros de datos.

- *Replay:* Se toman datos reales y se verifican contra el proceso generado. Se realiza una conformidad de los datos sobre el modelo, y se evalúa cuan cercano es este a la realidad. Replay permite además analizar el rendimiento del modelo. Si se cuenta con el instante en que se generó el evento se puede analizar cuanto tiempo toman las actividades, cuales son sus retrasos, con que frecuencia se ejecuta una actividad o se sigue un camino en particular.

Replay permite verificar conformidad, investigar problemas de rendimiento y proponer mejoras al proceso.

La figura 2.7 extraída de [1] muestra las diferentes relaciones de los modelos y los archivos de datos.

La minería de procesos puede clasificarse en tres tipos [4]:

- *Descubrimiento:* Se trata de técnicas capaces de modelar un proceso de negocio que se está ejecutando únicamente a partir de la secuencia de archivos de log correspondientes, estos logs pueden extraerse de archivos texto, bases de datos o algún otro medio. Para ello se utilizan algoritmos que de forma automatizada analizan los archivos y establecen el orden en que se ejecutan las tareas. Como resultado se obtiene un proceso de negocio que representa la realidad de lo que se está realizando en la operación, la cual puede diferir del proceso diseñado originalmente. Esta información es de vital importancia para los analistas de proceso

y proporciona un activo que se puede utilizar para mejoras al proceso.

Possiblemente el algoritmo más popular para realizar esta tarea es el algoritmo *Alpha* [10] que genera un red de Petri con el proceso descubierto. Este algoritmo escanea los eventos en busca de patrones y con ellos se arma el proceso. Es un algoritmo simple y eficiente pero no se ajusta a todos los tipos de proceso, además es muy sensible al ruido (comportamiento anormal) en los datos.

Otros algoritmos son *Euristics Mining*[11], *Fuzzy Mining*[12] e *Inductive Mining*[13]. Algunos son muy buenos para manejar ruido y trazas incompletas. En general se basan la frecuencia de actividades y la cantidad de veces que una actividad es seguida de otra para generar el proceso de salida.

La calidad del modelo generado está relacionada en forma directamente proporcional a la calidad de los datos de entrada. Esta es una característica común a cualquier técnica de extracción de conocimiento inductiva. Es fundamental que las secuencias de eventos relevadas representen completamente al proceso a modelizar. De lo contrario, quedarán aspectos sin descubrir.

- *Verificación de Conformidad*: Permite contrastar una secuencia de eventos contra un modelo de proceso (puede ser el descubierto anteriormente u otro) para determinar cuanto se cumple y cuáles son las ocurrencias que se desvían. Las técnicas de verificación de conformidad consideran cada uno de los eventos del archivo registro como actividades del modelo. De esta manera, los comportamientos observados en el registro de eventos se pueden comparar con el comportamiento que hay en el modelo. Hay varias aplicaciones de verificación de conformidad. Por ejemplo, identificar desvío, evaluar la calidad de un modelo de proceso descubierto o mejoras al proceso y auditoría.

Pueden utilizarse aplicaciones que representan gráficamente los resultados y realizan animaciones para una mejor visualización. El análisis resultante de esta operatoria puede ser de una gran precisión ya que se trabaja con datos reales, no es una simulación.

- *Mejoramiento*: Se trata de mejorar el proceso existente a partir del análisis realizado. En el mejoramiento del procesos se toma como entrada un archivo de eventos y el modelo que se desea mejorar. Es decir se puede mejorar el proceso analizando lo que ocurrió en el pasado. Permite analizar tiempos medios de ejecución de trazas, cuales son los desvíos y los cuellos de botella.

Se diferencia de la Verificación de Conformidad porque aquí el foco está puesto en el proceso y no en la ejecución de las trazas.

La figura 2.8 extraída de [1] muestra los distintos tipos de *Minería de Procesos* y su interacción con los procesos reales y sus datos.

Existen adicionalmente cuatro aspectos a tener en cuenta para obtener resultados de calidad en la Minería de Procesos. Estos son: precisión, ajuste, generalización y simplicidad [1]. Debe mantenerse una relación balanceada entre estas cuatro fuerzas para que el modelo descubierto sea representativo del proceso y fácil de comprender.

Por último pero, no menos importante, cabe destacar que generalmente se encuentra una gran cantidad de ruido en los datos de entrada que puede deberse a duplicidad de datos o trazas incompletas que pueden distorsionar el resultado del análisis [14]. Como en otras técnicas de minería se suele realizar

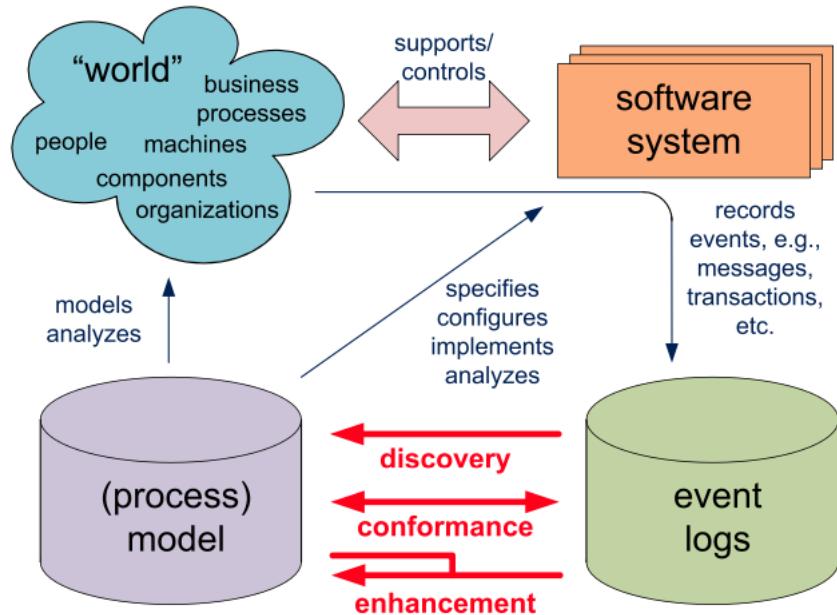


Figura 2.8: Diferentes tipos de Minería de Procesos

un procesamiento previo para mejorar la calidad de los datos de entrada con el fin de eliminar datos incompletos o corruptos que puedan llevar a una modelización errónea.

2.5. MATERIAL RELACIONADO

Existen algunas publicaciones que relacionan el negocio postal, la logística, la inteligencia de datos y la minería de procesos, por ejemplo [15] que utiliza *Data Mining* en un entorno de Big Data para el correo de China. Debido a la problemática del negocio postal, en dicho artículo se utilizaron técnicas de Clustering para agrupar a los clientes según el comportamiento, los hábitos de consumo y el foco de interés logrando generar una estrategia de Marketing postal mas precisa y efectiva con resultados muy satisfactorios.

Otro caso es [16] donde se aplica *Minería de Procesos* en logística buscando similitudes y diferencias entre distintos procesos de entrega en un contexto cambiante de manufactura y logística. En dicho trabajo se comparan diferentes procesos utilizando técnicas de agrupamiento en busca de lograr una documentación automatizada de procesos en un contexto cambiante.

En [17] se vincula la *Ingeniería de Software* con la *Minería de Procesos* para mostrar la utilidad de esta última en el proceso de desarrollo, presentando varios casos de estudio y mostrando los desafíos que enfrentan ambas disciplinas al trabajar en conjunto.

Más recientemente encontramos [18] donde se propone utilizar la *Minería de Procesos* para modelar, analizar y mejorar la cadena de suministro de productos. Por ejemplo con la identificación de procesos, la mejora del rendimiento y la detección de problemas. Además el artículo analiza cómo la minería de procesos puede integrarse con otras técnicas analíticas, como el aprendizaje automático, para mejorar aún más la gestión de la logística lo que puede conducir a beneficios para la empresa.

Una visión interesante se plantea en [19] donde se pone foco en el grado de abstracción que se debe tomar para proyectos de *Minería de Procesos* en logística. El nivel de abstracción se refiere a la

granularidad de los datos que se utilizan para modelar y analizar los procesos. En el artículo describe un enfoque para seleccionar el nivel de abstracción adecuado. El enfoque se basa en la identificación de los objetivos de análisis y la selección de los datos que sean relevantes para esos objetivos. También incluye un caso de estudio en el que se aplicó este enfoque a los datos de seguimiento de los envíos en una empresa de logística. Ahí se demostró que la selección del nivel adecuado de abstracción es crítica para obtener resultados precisos y útiles de la *Minería de Procesos*.

En [20] se presenta una metodología que sirve de guía para la ejecución de proyectos de *Process Mining* y que describe las diferentes etapas. Además se muestra como caso de estudio su aplicación práctica en el proceso de compras de IBM.

CAPÍTULO 3

Arquitectura de la propuesta

3.1. INTRODUCCIÓN

En esta tesis de maestría se utilizará un proceso existente para la distribución postal de productos trazables como por ejemplo Carta Documento. Se tomarán algunos archivos productivos de distribución para realizar el descubrimiento con las técnicas de *Minería de Procesos* anteriormente mencionadas.

Luego se realizará la verificación de conformidad sobre el modelo descubierto con otro set de datos productivos para verificar consistencia. Finalmente se procederá al mejoramiento del proceso productivo de distribución postal para el producto elegido y se utilizará Adaptive Case Management para detectar en forma temprana desvíos en el proceso operativo.

El detalle del flujo de trabajo se muestra en la figura 3.1 y se compone de la siguiente manera:

Se inicia con la definición del proceso de distribución postal con BPM y su modelado en BPMN. Luego de la definición del proceso se lanza el producto al mercado y comienza su comercialización, los eventos relacionados con la distribución se registran y se almacenan en uno o más sistemas de información. El diseño original del proceso, su lanzamiento y registro de eventos se encuentran en funcionamiento desde hace algunos años lo que nos permite obtener información suficiente para aplicar las técnicas de *Minería de Procesos*.

A continuación se procede a realizar el descubrimiento del proceso *real*, es decir lo que sucede en la operación diaria y en el cual se visualizarán las diferencias con el diseñado originalmente. Para ello se utilizarán algoritmos de descubrimiento automático de procesos como el algoritmo *Alpha*[10] o el algoritmo *Inductive Visual Miner*[13]

Luego del descubrimiento se realizará una verificación del proceso descubierto, esto se realiza contrastando una muestra de datos contra el proceso descubierto. La muestra de datos que se utilizará para la verificación debe ser diferente a la que se utilizó para descubrir el proceso.

Finalmente se realizará un análisis de los resultados obtenidos en la verificación, este análisis puede proporcionar mejoras al proceso las cuales se reflejarán en la documentación BPMN y se incluirán en la distribución iniciando nuevamente el proceso. Adicionalmente se podrá ejecutar la verificación del proceso con los envíos que se encuentran *activos* (aún no finalizó su distribución) para encontrar desvíos operativos rápidamente y proceder a su ajuste mejorando la calidad del servicio brindado a los clientes.

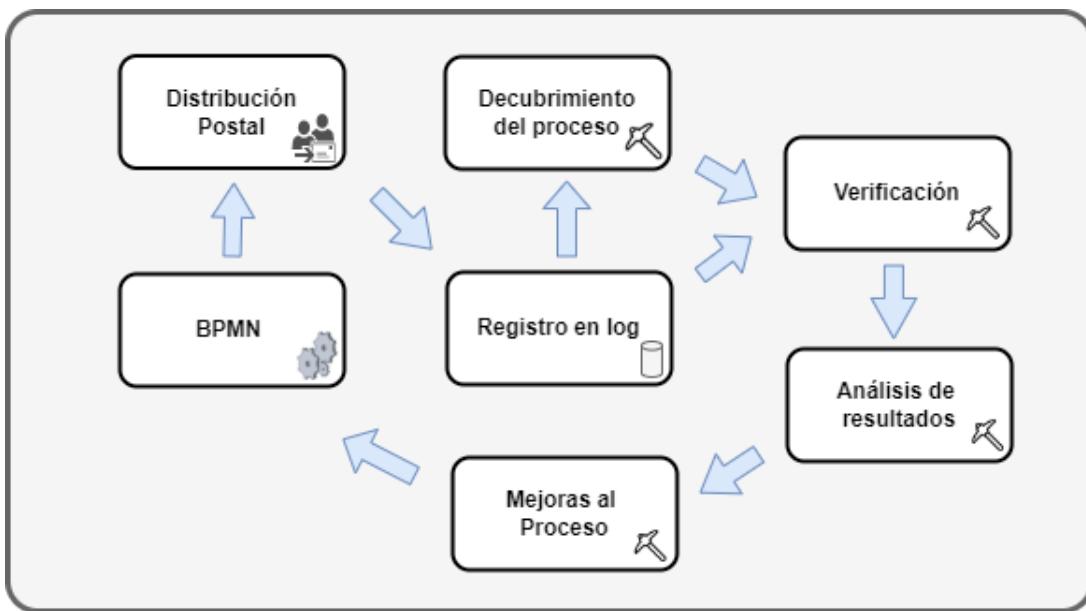


Figura 3.1: Minería de Procesos aplicada a la Distribución Postal

3.2. PRODUCTO ELEGIDO

Para el presente trabajo se eligió trabajar con el producto Carta Documento ya que es uno de los productos postales más representativos en Argentina y posee proceso complejo para su distribución.

3.3. DEFINICIÓN DE LA PRUEBA

El trabajo estará compuesto de las siguientes etapas:

1. Recopilación de datos: Este componente implica la recopilación de datos relevantes del proceso postal, como registros de eventos, logs de procesos, datos de seguimiento de envíos, datos de tiempo y fecha, datos de volumen de procesamiento, y otros datos pertinentes al proceso postal.
2. Preparación de datos: En esta etapa, los datos recopilados se limpian, transforman y preparan para su posterior análisis. Esto puede incluir la eliminación de datos duplicados, la normalización de los datos, la corrección de errores, la agregación de datos, y otras tareas de preparación de datos necesarias.
3. Análisis de procesos: En esta etapa, se aplican técnicas de minería de procesos para analizar los datos recopilados y descubrir patrones, tendencias, desviaciones y oportunidades de mejora en el proceso postal. Esto puede incluir la identificación de actividades, flujos de trabajo, roles de los participantes, reglas de negocio, tiempos de ejecución, y otras características del proceso postal.
4. Verificación de conformidad: Como se mencionó anteriormente, la verificación de conformidad es una parte importante de la minería de procesos. En esta etapa, se comparan los datos reales del proceso con el modelo de proceso teórico o de referencia para identificar desviaciones y discrepancias.
5. Análisis de rendimiento: En esta etapa, se evalúa el rendimiento del proceso postal en términos de eficiencia, tiempos de ejecución, cuellos de botella, recursos utilizados, y otros indicadores de

rendimiento relevantes. Esto puede ayudar a identificar oportunidades de mejora y optimización del proceso.

6. Visualización y presentación de resultados: Los resultados del análisis de procesos y del análisis de rendimiento se visualizan y presentan de forma clara y comprensible para los stakeholders relevantes en el proyecto de minería de procesos. Esto puede incluir informes, gráficos, dashboards u otras representaciones visuales de los hallazgos del proyecto.
7. Mejora y optimización del proceso: Basado en los hallazgos del análisis de procesos y del análisis de rendimiento, se pueden identificar oportunidades de mejora y optimización del proceso postal. Estas oportunidades pueden ser implementadas en el proceso real para obtener beneficios y mejoras en la eficiencia y calidad del proceso.

3.4. HERRAMIENTAS DE SOFTWARE UTILIZADAS

A continuación se listan las principales herramientas de software utilizadas para la elaboración de esta tesis de maestría. El listado completo se detalla en el Anexo C *Herramientas Utilizadas*

- Repositorio GIT: se genera un repositorio GIT¹ en GitHub² para administrar el versionado de los fuentes, gráficos y documentación realizados. El repositorio es <https://github.com/vimartinez/UNLP-Tesis>
- DBeaver Community: Herramienta de Base de Datos universal y gratuita para conectarse a diversos motores de Base de Datos³
- RapidMiner: Plataforma de análisis y minería de datos de código abierto que permite a los usuarios realizar tareas de análisis de datos, minería de datos y minería de procesos de manera visual y fácil de usar.⁴
- ProM Tools: Conjunto de herramientas de software de código abierto para el análisis y descubrimiento de procesos en el campo de la minería de procesos. Puede realizar análisis y validaciones sobre grandes volúmenes de datos.⁵

3.5. HARDWARE UTILIZADO

A continuación se detalla el hardware utilizado para el procesamiento de los datos y la generación del presente trabajo.

- Servidor de Procesamiento de datos:
 - Procesador: Intel(R) Core(TM) i7 2.90 GHz
 - Memoria RAM: 32 GB
 - Disco Rígido: 2x 1 TB
- Estación de trabajo:
 - Marca: Hewlett Packard

¹<https://git-scm.com/>

²<https://github.com/>

³<https://dbeaver.io/>

⁴<https://rapidminer.com/>

⁵<https://promtools.org/>

- Procesador: Intel(R) Core(TM) i3 2.40 GHz
- Memoria RAM: 16 GB
- Disco Rígido: SSD 240 GB

CAPÍTULO 4

Minería de Procesos en la distribución Postal

4.1. INTRODUCCIÓN

En el capítulo anterior se definieron las técnicas y metodologías que se utilizarán en este documento para aplicar *Minería de Procesos* a la Distribución Postal. En este capítulo se aplicarán estas técnicas y se detallarán los desafíos encontrados durante el proceso.

Se obtendrán los datos de entrada de archivos de texto y una base de datos Oracle, se utilizará Rapid-Miner para realizar una visualización de los datos y de ser necesario eliminar algunas inconsistencias. Se empleará PROM para eliminar trazas incompletas y aplicar las técnicas de *Minería de Procesos* requeridas según se definió anteriormente.

La arquitectura detallada de la solución se muestra en la Figura 4.1.

El prototipo que se realizará consta de las siguientes partes:

1. Extracción de los datos: se extrae la muestra de los archivos de origen. Para este caso de estudio se utilizarán datos productivos de distribución de los años 2019 al 2021.
2. Conversión: Se transforman los datos a un formato de fácil interpretación para las herramientas de *Minería de Procesos*.
3. Filtrado: se realiza un procesamiento previo de los datos, se hace un muestreo aleatorio con una revisión visual para verificar consistencia. Luego se eliminan datos inconsistentes y se divide la muestra en dos tandas, una para descubrimiento y otra para utilizar en la verificación.
4. Descubrimiento: se aplican los algoritmos de *Minería de Procesos* para realizar el descubrimiento del modelo actual utilizando una parte del set de datos.
5. Verificación: se realizará la verificación de conformidad del modelo descubierto utilizando la otra parte del set de datos.
6. análisis: se realizará un análisis de los desvíos y problemas encontrados en la distribución
7. presentación: se presentarán los resultados obtenidos
8. mejoras: se desarrollarán las mejoras sugeridas al proceso y la propuesta de ACM.

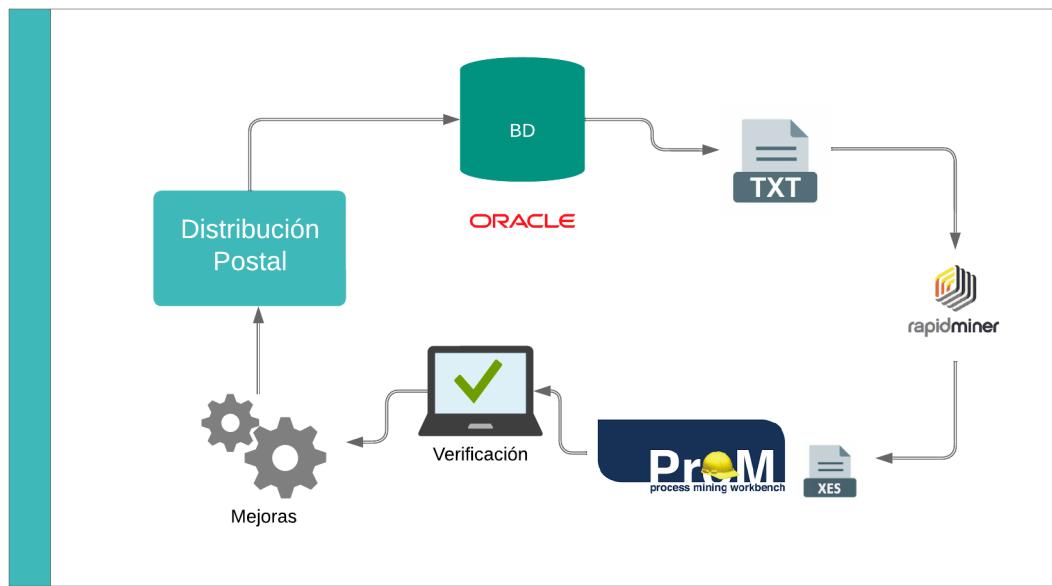


Figura 4.1: Arquitectura de la propuesta

Con el objetivo de comprender las técnicas y herramientas de *Minería de Procesos* y poder aplicarlas de forma efectiva, previo a la realización del presente trabajo y la implementación de la propuesta los siguientes cursos:

- Process Mining: Data Science in Action de la Eindhoven University of Technology ¹.
- Introduction to process mining with ProM ²

Luego se realizaron prácticas y ejercitaciones con sets de datos públicos implementando los conocimientos adquiridos, hasta conocer mínimamente el funcionamiento de las herramientas y como interpretar los resultados obtenidos. Finalmente se publicaron los papers [5] y [6] abordando la temática y realizando un análisis preliminar sobre una muestra menor de datos.

4.2. EXTRACCIÓN DE LOS DATOS

Los datos de distribución postal se encuentran almacenados en una base de datos Oracle, se realiza una exportación del período entre enero de 2017 hasta diciembre de 2020.

Obteniendo más de **9.000.000** de registros que corresponden aproximadamente a **3.750.000** piezas postales distribuidas. Cada uno de esos registros representa un evento de un envío.

Se considera cada envío como una traza y cada movimiento como un evento. La traza se considerará completa cuando el envío tiene registrado el ingreso, y el fin, pudiendo ser la entrega exitosa o no. Se exporta para cada traza los siguientes datos: identificador de traza e identificador de evento, descripción del evento, fecha del evento. El resto de los campos disponibles no aportan valor para el procesamiento que se va a realizar así que no se tienen en cuenta. Para considerar una trazas como válida cada traza deberá tener al menos dos eventos, uno de inicio y otro de fin. Se descartaran los registros que no cumplan con estas condiciones en un procesamiento posterior.

¹<https://www.coursera.org/learn/process-mining>

²<https://www.promtools.org/doku.php?id=mooc:prom>

123 LOGTTID	123 NOTTID	123 PIED	LOGTTFECHA	LOGTTOBSEVAC
99,889,052	9	41,230,484	2018-05-07 12:00:00	PLAZO VENCIDO NO RECL.
99,889,053	1	41,230,487	2018-05-07 13:20:00	ENTREGADO
99,889,054	2	41,230,292	2018-05-04 12:36:00	CERRADO/AUSENTE/1-VISIT
99,889,057	9	41,230,275	2018-05-07 11:50:00	PLAZO VENCIDO NO RECL.
99,889,058	2	41,230,260	2018-05-07 16:00:00	CERRADO/AUSENTE/1-VISIT
99,889,060	2	41,230,025	2018-05-07 11:30:00	CERRADO/AUSENTE/1-VISIT
99,889,061	6	41,230,026	2018-05-07 13:30:00	SE MUDO
99,889,062	9	41,230,490	2018-05-07 10:00:00	PLAZO VENCIDO NO RECL.
99,889,064	1	41,218,737	2018-05-04 10:05:00	ENTREGADO
99,889,065	9	41,218,677	2018-05-07 08:30:00	PLAZO VENCIDO NO RECL.
99,889,066	9	41,218,676	2018-05-07 08:30:00	PLAZO VENCIDO NO RECL.
99,889,067	1	41,219,809	2018-05-03 12:10:00	ENTREGADO
99,889,069	2	41,229,739	2018-05-07 12:30:00	CERRADO/AUSENTE/1-VISIT
99,889,070	1	41,229,736	2018-05-07 10:45:00	ENTREGADO
99,889,071	9	41,229,718	2018-05-07 12:00:00	PLAZO VENCIDO NO RECL.
99,889,072	9	41,229,713	2018-05-07 11:00:00	PLAZO VENCIDO NO RECL.
99,889,073	4	41,229,711	2018-05-07 12:48:00	CERRADO/AUSENTE/2-VISIT
99,889,075	9	41,229,702	2018-05-07 08:53:00	PLAZO VENCIDO NO RECL.
99,889,076	9	41,229,700	2018-05-07 08:53:00	PLAZO VENCIDO NO RECL.
99,889,077	2	41,229,699	2018-05-07 13:33:00	CERRADO/AUSENTE/1-VISIT
99,889,079	1	41,229,698	2018-05-07 13:00:00	ENTREGADO
99,889,081	9	41,222,293	2018-05-07 09:10:00	PLAZO VENCIDO NO RECL.
99,889,082	1	41,222,301	2018-05-04 12:11:00	ENTREGADO
99,889,083	2	41,222,567	2018-05-02 10:00:00	CERRADO/AUSENTE/1-VISIT
99,889,084	4	41,222,317	2018-05-03 14:22:00	CERRADO/AUSENTE/2-VISIT
99,889,085	2	41,222,324	2018-05-04 11:35:00	CERRADO/AUSENTE/1-VISIT

Figura 4.2: Datos Originales

Originalmente los datos se almacenan según se muestra en la figura 4.2.

Los datos originales se exportan de la base de datos como archivos de texto con formato csv para su posterior utilización, dado el gran volumen de datos a procesar se pone un límite de 150 MB para cada archivo ya que se comprobó con operaciones previas que archivos de un tamaño mayor se tornan muy difíciles de trabajar en las etapas posteriores de la *Minería de Procesos* con los recursos de hardware disponibles para el presente trabajo.

Se importan uno a uno los archivos en RapidMiner³ para realizar una comprobación visual, no se detecta nada extraño y se verifica que la exportación fue exitosa.

Finalmente los datos originales quedan en 9 archivos de 150 MB, que totalizan 3.753.495 envíos con 9.105.396 eventos.

A continuación se comienza con el procesamiento de los datos.

4.3. CONVERSIÓN DE DATOS

Luego se procesan estos archivos en Prom⁴, se ingresan uno a uno y se convierten a formato formato XES [21].

XES (eXtensible Event Stream) es un formato de archivo utilizado basado en etiquetas, similar a XML⁵ se utiliza para almacenar y compartir datos de eventos. Es el principal formato de archivo utilizado en el ámbito de la minería de procesos y la gestión de procesos empresariales. Algunas de las ventajas de usar el formato XES son:

³<https://rapidminer.com/>

⁴<https://promtools.org/>

⁵<https://www.w3.org/XML/>

1. Flexibilidad: El formato XES es muy flexible y permite a los usuarios definir sus propios esquemas de datos, lo que significa que puede adaptarse a diferentes tipos de procesos y flujos de trabajo.
2. Interoperabilidad: El formato XES es compatible con una amplia gama de herramientas de software de minería de procesos, lo que facilita la interoperabilidad entre diferentes sistemas y la colaboración entre diferentes equipos y organizaciones.
3. Capacidad de registro detallado: El formato XES puede registrar una gran cantidad de información detallada sobre cada evento, lo que puede ayudar a los usuarios a identificar patrones y tendencias en los procesos y a tomar decisiones informadas basadas en los datos.
4. Facilidad de uso: El formato XES es fácil de usar y leer, lo que facilita la comprensión de los datos por parte de los usuarios, incluso si no tienen experiencia en minería de procesos.
5. Aplicabilidad: El formato XES se puede utilizar en una amplia gama de aplicaciones, como la gestión de procesos empresariales, la monitorización de sistemas y la gestión de proyectos, lo que lo convierte en una herramienta valiosa para muchas organizaciones.

Durante la conversión los datos se organizan por traza y con sus eventos en orden cronológico esto facilita su procesamiento en etapas posteriores.

Luego de la conversión se exportan nuevamente los 9 archivos originales.

Los archivos resultantes de la conversión tienen un tamaño aproximado de 260 MB debido al agregado de las etiquetas propias del formato XES pero están optimizados para su utilización con las herramientas de *Minería de Procesos*.

El formato del archivo XES se muestra en la figura 4.3 donde se visualiza que las trazas están claramente delimitadas incluyendo cada uno de los eventos que la componen. Además cada traza y cada uno de sus eventos están anotados como tales.

Luego de la conversión los datos originales quedan en 9 archivos de aproximadamente 260 MB, manteniendo los totales de 3.753.495 envíos con 9.105.396 eventos.

4.4. FILTRADO

Para tener una mayor consistencia en los datos y facilitar el análisis se eliminaron todas las trazas incompletas, ya sea porque no tienen un estado inicial (ingreso) o un estado final (entregado, devuelto, no existe dirección, falleció, o se mudó). Esto puede ocurrir por algún error en la carga de los datos o porque son trazas que todavía no completaron el proceso.

Esta tarea se realiza en ProM importando cada uno de los archivos de datos en formato XES que se obtuvieron en la etapa anterior y aplicando un filtro de reglas heurísticas simples para quitar las trazas incompletas. Para ello el algoritmo escanea el archivo y nos permite seleccionar cuales son los eventos iniciales válidos de todos los eventos iniciales que se detectaron en las trazas, luego debemos realizar la misma selección pero con los eventos finales. Basándose en estos datos de entrada se procede a descartar todas las trazas que no cumplen estas condiciones, como resultado se obtiene un nuevo archivo en formato XES que contiene únicamente las trazas que se utilizarán en este caso de estudio reduciendo considerablemente el ruido en los datos de entrada.

La utilización de este filtro se visualiza en la figura 4.4

```

</event>
</trace>
<traces>
    <string key="concept:name" value="493094"/>
    <event>
        <string key="concept:name" value="INGRESADO"/>
        <string key="lifecycle:transition" value="complete"/>
        <string key="EveID" value="0"/>
        <date key="time:timestamp" value="2017-08-17T11:12:55.000-03:00"/>
    </event>
    <event>
        <string key="concept:name" value="1 INTENTO DE ENTREGA"/>
        <string key="lifecycle:transition" value="complete"/>
        <string key="EveID" value="2"/>
        <date key="time:timestamp" value="2017-08-23T17:00:00.000-03:00"/>
    </event>
    <event>
        <string key="concept:name" value="1 INTENTO DE ENTREGA"/>
        <string key="lifecycle:transition" value="complete"/>
        <string key="EveID" value="2"/>
        <date key="time:timestamp" value="2017-08-28T15:00:00.000-03:00"/>
    </event>
    <event>
        <string key="concept:name" value="1 INTENTO DE ENTREGA"/>
        <string key="lifecycle:transition" value="complete"/>
        <string key="EveID" value="2"/>
        <date key="time:timestamp" value="2017-08-29T14:22:00.000-03:00"/>
    </event>
    <event>
        <string key="concept:name" value="DEVOLUCION"/>
        <string key="lifecycle:transition" value="complete"/>
        <string key="EveID" value="9"/>
        <date key="time:timestamp" value="2017-09-13T10:28:00.000-03:00"/>
    </event>
</traces>
<trace>
    <string key="concept:name" value="493152"/>
    <event>
        <string key="concept:name" value="INGRESADO"/>

```

Figura 4.3: Datos en formato XES

El proceso se realiza con cada uno de los archivos obtenidos en el paso anterior y aunque no hay una reducción considerable en el tamaño de los archivos (ahora están en el orden de los 230 MB) los datos de entrada ya están listos para utilizar las técnicas de *Minería de Procesos*

Tabla 4.1: Detalle de trazas y eventos utilizados

Archivo	Trazas Originales	Eventos Originales	Trazas Filtradas	Eventos Filtrados
1	469.288	1.115.026	433.256	1.041.854
2	466.271	1.096.119	427.240	1.018.169
3	455.551	1.089.551	419.241	1.015.692
4	455.931	1.088.746	430.914	1.025.856
5	432.879	1.078.439	405.369	981.950
6	400.576	1.040.031	354.174	908.407
7	403.035	1.041.523	371.039	919.938
8	398.503	1.044.665	357.374	913.666
9	271.461	511.296	258.325	452.680
TOTAL	3.753.495	9.105.396	3.456.932	8.278.212

Luego de este filtrado los archivos resultantes tienen aproximadamente 230 Mb y totalizan 3.456.932 trazas con un total de 8.278.212 eventos.

El detalle de cada archivo se visualiza en la tabla 4.1

4.5. DESCUBRIMIENTO DEL PROCESO

El proceso de distribución postal que deben cumplir los envíos que se utilizaron en este trabajo se describió anteriormente y se modeló según BPMN como se muestra en la figura 4.5

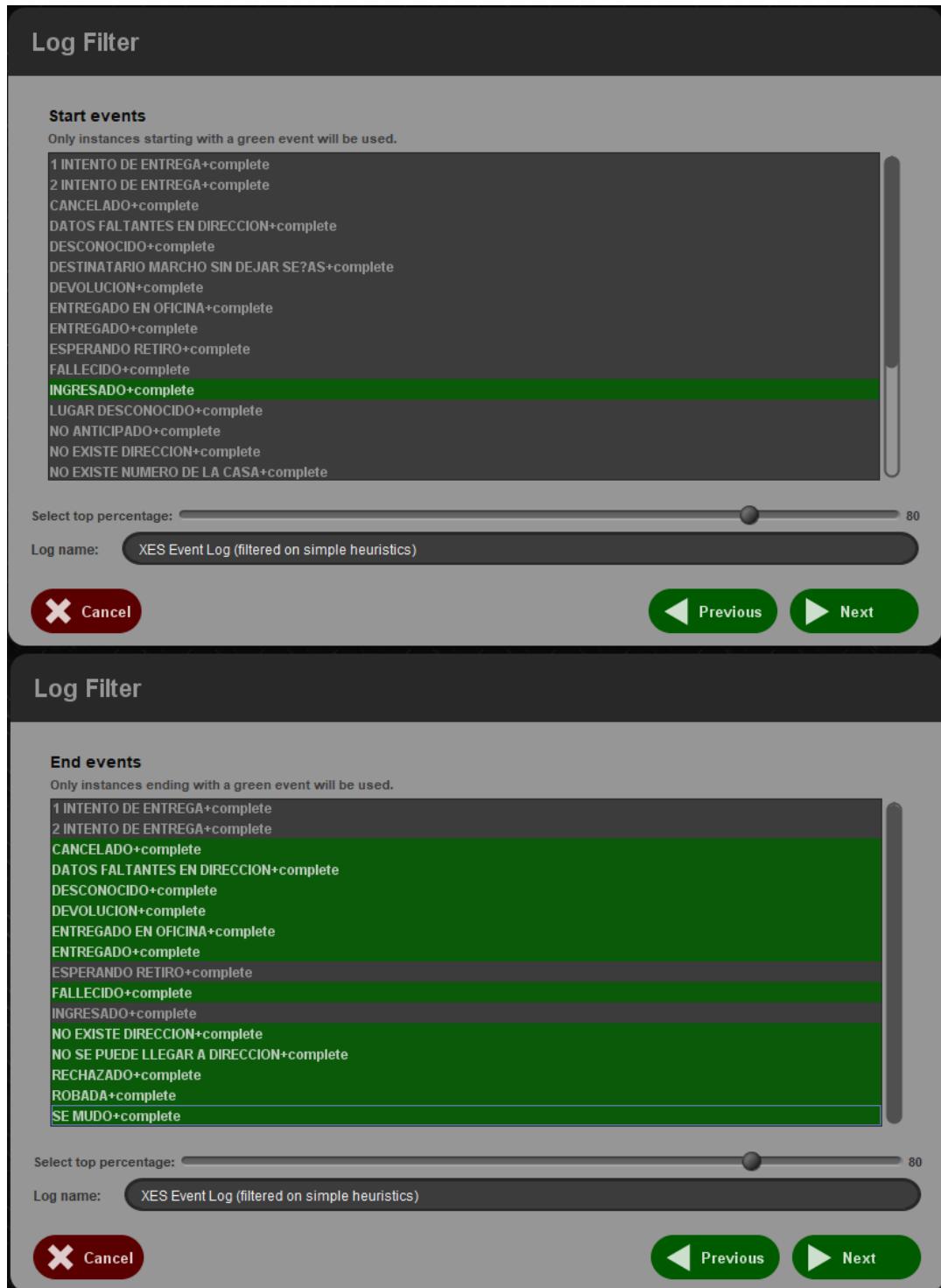


Figura 4.4: Filtrado de archivos XES

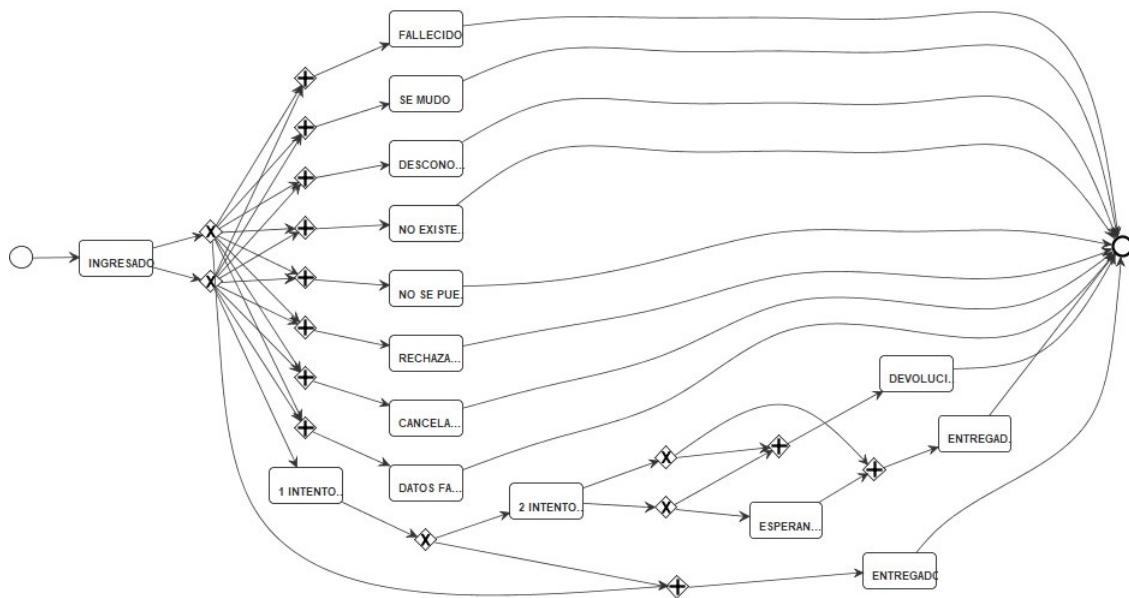


Figura 4.5: Proceso de Distribución Postal modelado en BPMN

Este proceso fue validado por personal de *Ingeniería de Procesos* y se verificó que cumple con las características del producto utilizado en este trabajo.

Para realizar el descubrimiento del proceso que cumplen las piezas se utilizó el algoritmo Alpha+ dado que se verificó en los trabajos que se realizaron previamente para dar soporte al presente trabajo su correcto funcionamiento y ajuste a este tipo de datos.

Tanto el algoritmo *Alpha*^[10] como Alpha+ son ampliamente usados en *Minería de Procesos* para descubrir procesos de negocio a partir de registros de eventos.

El algoritmo Alpha busca descubrir modelos de proceso basados en la estructura de flujo de eventos. El proceso de descubrimiento del modelo comienza con la creación de un grafo que representa la estructura de los eventos. Luego se eliminan las transiciones redundantes, y se agrupan las transiciones similares para formar lugares en el modelo. Finalmente, se fusionan lugares similares para obtener un modelo de proceso simplificado. Los pasos son los siguientes:

1. Creación del grafo de eventos: En este paso se crea un grafo dirigido que representa la estructura de flujo de eventos en el registro de eventos. Los nodos del grafo representan las actividades del proceso y los arcos representan la relación de orden entre las actividades.
2. Eliminación de transiciones redundantes: En este paso, se eliminan las transiciones redundantes del grafo. Una transición se considera redundante si su eliminación no afecta la estructura del grafo. Esto se hace para simplificar el modelo y reducir la complejidad.
3. Agrupación de transiciones similares: En este paso, se agrupan las transiciones que tienen un comportamiento similar para formar lugares en el modelo. Un lugar representa un conjunto de actividades que tienen un comportamiento similar. Esto se hace para simplificar aún más el modelo.
4. Fusión de lugares similares: En este paso, se fusionan los lugares que tienen un comportamiento similar para obtener un modelo de proceso simplificado. La fusión de lugares se basa en la similitud de las actividades y los arcos que los conectan.

Por otro lado, el algoritmo *Alpha+* mejora el algoritmo *Alpha* al tener en cuenta los tiempos de los eventos. Además de la estructura de flujo de eventos, también considera los tiempos de los eventos para descubrir un modelo de proceso que refleje las relaciones de orden entre los eventos. [22]

El descubrimiento es el primer tipo de *Minería de Procesos* de los que se describieron anteriormente. Se utiliza ProM para pasar la muestra de datos a través del algoritmo *Alpha+* y encontrar el proceso. El proceso descubierto se conforma de la siguiente manera:

- Siempre hay un evento de ingreso que deben cumplir todas las trazas luego del cual se realiza la distribución.
- Si se puede entregar se registra la entrega y finaliza la historia de la traza.
- Si no se puede entregar por un evento final (persona fallecida, no existe la dirección, no hay suficientes datos para encontrar la dirección, etc) se registra la no entrega con el motivo y finaliza el proceso.
- Si el motivo de no entrega no es un evento final (por ejemplo no había nadie para entregar) se registra un primer intento de entrega y se intenta nuevamente al día siguiente.
- Si al día siguiente tampoco se puede entregar se mantiene el envío en la oficina durante un tiempo esperando que el destinatario se presente a retirar (se deja un aviso en cada visita)
- Si el destinatario pasa a retirar el envío por la sucursal se registra la entrega y finaliza el proceso.
- si el destinatario no pasa a retirar el envío luego de un tiempo especificado (usualmente 2 semanas) se devuelve el envío al remitente y finaliza el proceso.

Como resultado el algoritmo nos devuelve el proceso representado en una *Red de Petri*. Una red de Petri es un modelo matemático y gráfico que permite representar sistemas distribuidos y concurrentes. Se compone de dos elementos principales: lugares y transiciones.

Un lugar representa una condición o estado del sistema. Los lugares se representan como círculos en el diagrama y pueden contener tokens, que representan la existencia o disponibilidad de un recurso o condición. Por ejemplo stock de un producto.

Una transición representa un evento o acción que cambia el estado del sistema. Las transiciones se representan como barras en el diagrama y se activan cuando hay suficientes tokens en los lugares de entrada correspondientes. Al activarse, una transición consume los tokens de los lugares de entrada y produce tokens en los lugares de salida correspondientes.

En la Figura 4.6 se muestra la red de Petri descubierta que representa el proceso que deberán cumplir las trazas. Se realiza una exportación desde ProM para utilizar el proceso descubierto en etapas posteriores.

4.6. VERIFICACIÓN DE CONFORMIDAD

Para realizar la verificación de conformidad se contrastará cada uno de los archivos de datos filtrados contra el modelo descubierto previamente.

Se busca determinar cuánto se ajustan las trazas al proceso, es decir que tanto cumplen las etapas requeridas. Cuanto más se acerque la traza (en cantidad de pasos y orden) al proceso esta tendrá un

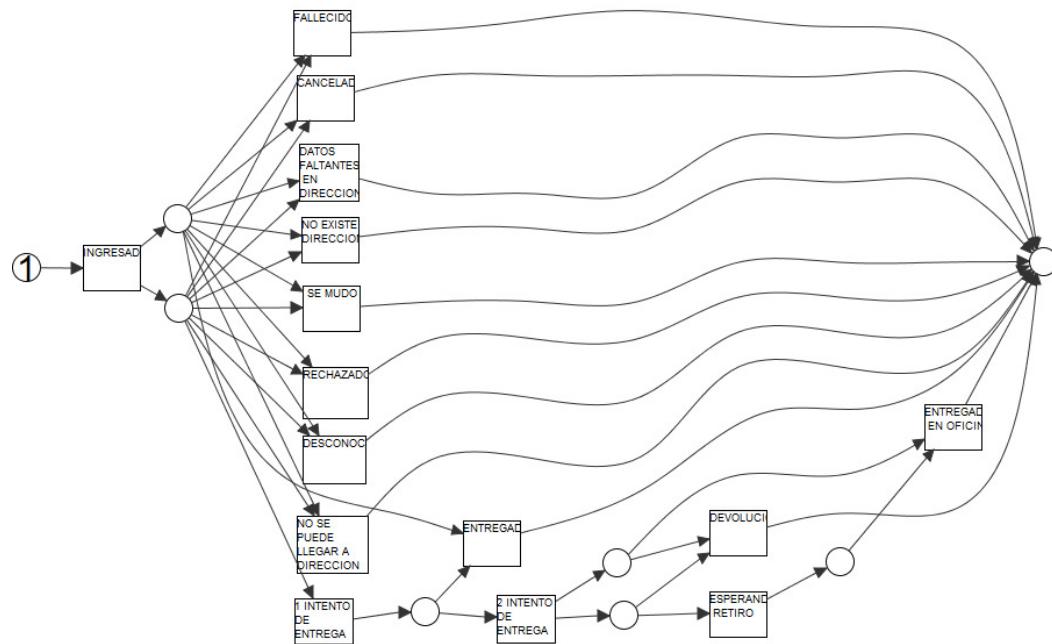


Figura 4.6: Proceso descubierto con Alpha+ que deberán cumplir las trazas

mayor ajuste (o fitness) con respecto al mismo. Por el contrario cuanto menos se acerque al proceso (pasos sobrantes u omitidos o que no se realizan en el orden correcto) la traza tendrá un ajuste menor. Por lo tanto para este análisis se tendrán en cuenta qué pasos se cumplen, cuáles no, si hay repetidos o adicionales que no estén reflejados en el proceso descubierto y su correcta sucesión.

A continuación se procede con la reproducción de cada uno de los archivos de datos sobre el proceso descubierto anteriormente, para ello se importa la Red de Petri del proceso y luego se importa cada uno de los archivos de datos y se ejecuta el comando “*Replay a Log on Petri Net for Conformance Analysis*”.

Un ejemplo se muestra en la Figura 4.7. Allí se resaltan los eventos más comunes en color oscuro y los recorridos más frecuentes con una línea más gruesa.

Luego de procesar todos los archivos por la Red de Petri se comienza con el análisis de los resultados verificando las estadísticas de los datos.

Inicialmente se observa que un 73 % de los casos el proceso finaliza con la pieza es entregada, ya sea en el primer o segundo intento y que el resto se distribuye uniformemente entre los diversos motivos de no entrega. Esto no implica un mal funcionamiento ya que la no entrega es parte del proceso, solo nos muestra una característica del producto Carta Documento.

Se observa además que en todo el conjunto de datos las trazas tienen una media de 2.37 eventos en cada traza y una ajuste con respecto al proceso de 0,72.

Estos valores son similares a los encontrados en los trabajos publicados [5] y [6] para dar soporte a esta tesis.

Se observa también que muchas trazas no verifican el proceso exactamente ya sea porque no cumplen los pasos en el orden correcto o porque repiten algunos. Un ejemplo de esto se muestra en la figura 4.8. tomado del visualizador de trazas de ProM, allí se pueden ver trazas con muchos movimientos lo que implica un ajuste muy bajo (entre 0,23 y 0,33) con respecto al proceso. Se establece un umbral

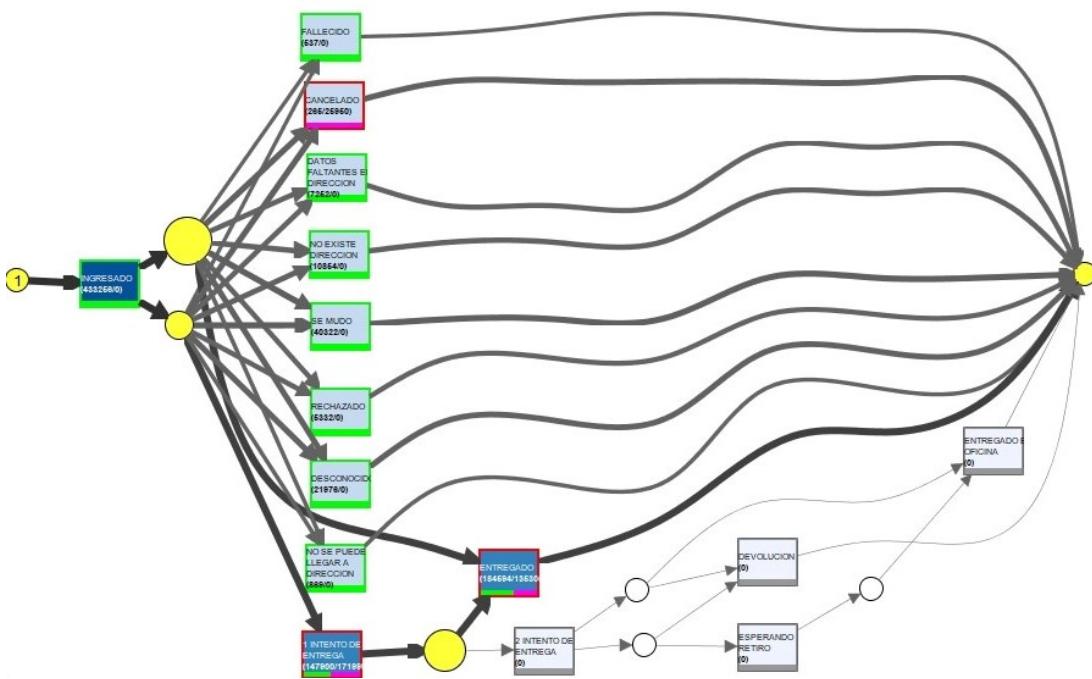


Figura 4.7: Recorrido de las trazas: eventos más comunes en oscuro y recorridos más frecuentes en línea más gruesa

donde se define que las trazas deberán tener al menos un ajuste de 0,5 para no considerarse casos sospechosos de incumplimiento del proceso.

Se decide analizar con mayor detalle las trazas que tienen un ajuste más bajo del umbral establecido, por lo tanto se exportan de cada archivo aquellas que tengan un ajuste menor al **50 %** ya que representan un gran desvío operativo y se consolidan en un único archivo para un análisis posterior.

Encontrando un total de **374.876** trazas en estas condiciones.

Se verifica que la cantidad de movimientos por traza sigue una distribución normal con una media **2,37** (como se mencionó anteriormente) y un desvío estándar de **0,95**. Por lo tanto se decide tomar un valor superior a la *media mas tres desvíos estándar* como un valor representativo de una cantidad excesiva de movimientos, esto implica tener en cuenta las trazas con más de **5,22** movimientos. En base a esto, se procedió a filtrar y exportar las trazas con más de **6** movimientos para realizar un análisis posterior.

Encontrando un total de **1.724** trazas en estas condiciones.

Adicionalmente se procede a realizar un análisis de tiempos de finalización sobre las trazas del modelo en busca de aquellas que, si bien pueden haber completado el proceso correctamente, el mismo se realizó en un intervalo de tiempo extremadamente superior a la media de resolución de cada traza. Para ello se vuelve a reproducir cada uno de los archivos de datos sobre el proceso descubierto anteriormente pero en este caso con una herramienta orientada al análisis de performance llamada "*Replay a Log on Petri Net for Performance Analysis*", esta se enfoca en los tiempos de resolución y no tanto en ver si las trazas se ajustan o no al proceso.

La figura 4.9 muestra una visualización obtenida luego de ejecutar esta herramienta y que representa en tonos más oscuros las tareas que demoran más tiempo en realizarse y en más claro las que finalizan



Figura 4.8: Visualización de ProM de trazas con exceso de movimientos y ajuste muy bajo

más rápidamente.

Analizando los resultados se encuentran trazas que se ajustan bien al proceso pero con tiempos extremadamente altos. En algunos casos las trazas tienen más de tres movimientos y un intervalo de hasta 3 días entre cada uno de ellos lo que justifica un tiempo total tan elevado.

Adicionalmente se encontraron muchos casos que cuentan solo con son de solamente dos movimientos, donde el envío ingresó y estuvo sin actividad, en algunos casos hasta **15 días** antes de tener un evento de salida a distribución. Luego el proceso se cumple correctamente con una entrega o devolución, por lo tanto tiene un ajuste muy bueno. Estos casos representan un desvío operativo importante ya que impactan directamente en la calidad del servicio a los clientes incumpliendo los estándares de entrega del producto y son extremadamente difíciles de detectar debido a que la operación finaliza correctamente. En la figura 4.10 se observan trazas que cumplen el proceso con solo dos movimientos pero con una diferencia de 10 días o más entre uno y otro. Con el análisis de performance se descubrieron nuevos desvíos operativos que no aparecían en el análisis anterior ya que se trata de trazas que se ajustan perfectamente al modelo pero que están con los tiempos muy por encima de la media.

Se observa un tiempo medio de resolución para cada traza de **5,79** días, con un desvío de **4,96** días. Utilizando un criterio similar al usado en la cantidad de movimientos por traza se decide enfocar el análisis en aquellos casos que tengan un tiempo de finalización mayor a la media más un desvío (**10,75** días) por considerar que un tiempo de resolución mayor a este valor se trata de un desvío operativo.

Se realiza la exportación de las trazas que cumplen con ese criterio y se obtiene un total de **595.880** casos.

El resultado de la verificación del proceso nos muestra una gran cantidad de posibles desvíos operativos a verificar. El detalle de cada las irregularidades encontradas en cada archivo visualiza en la tabla 4.2

A continuación se analizarán en detalle los diferentes desvíos encontrados.

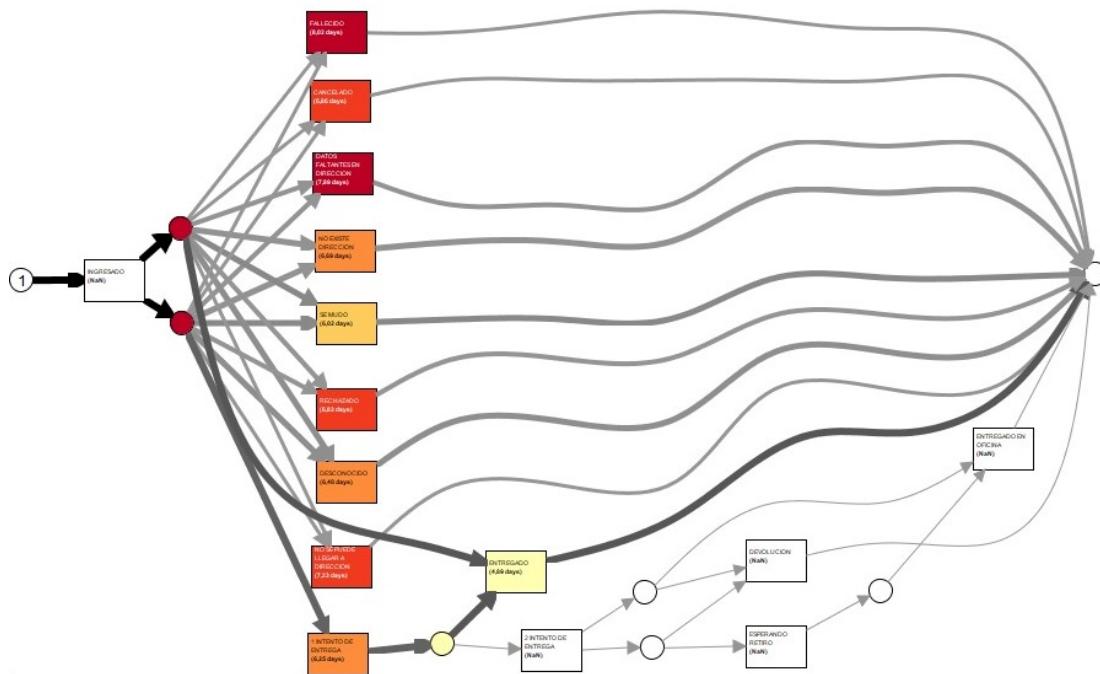


Figura 4.9: Visualización de trazas según análisis de Performance

2 events	
INGRESADO	#1 complete
	11.12.2017 08:43:35.000
ENTREGADO	
	#2 complete
	21.12.2017 19:14:00.000
3306727	
2 events	
INGRESADO	#1 complete
	19.07.2019 19:15:45.000
ENTREGADO	#2 complete
	29.07.2019 12:20:00.000
3451408	
2 events	
INGRESADO	#1 complete
	15.08.2019 08:34:55.000
ENTREGADO	#2 complete
	26.08.2019 10:23:00.000

Figura 4.10: Trazas de dos movimientos con inactividad

Tabla 4.2: Detalle de irregularidades encontradas

Archivo	Trazas	Ajuste menor a 0,5	Más de 6 movimientos	Más de 10 días
1	433.256	35.729	91	64.397
2	427.240	38.395	51	107.722
3	419.241	46.642	88	76.782
4	430.914	41.215	76	76.627
5	405.369	60.049	324	101.26
6	354.174	57.295	338	50.271
7	371.039	55.949	257	52.728
8	357.374	20.063	347	33.359
9	258.325	19.539	152	32.727
TOTAL	3.456.932	374.876 (10,84 %)	1.724 (0,05 %)	595.880 (17,24 %)

4.7. PROBLEMAS ENCONTRADOS

Los resultados de la verificación de conformidad realizada anteriormente proporcionan una evaluación objetiva de cómo el modelo de proceso descubierto se ajusta a los datos reales, es decir nos muestran que está pasando en la operación y nos permiten identificar posibles mejoras en el proceso real.

En este caso particular se encontró que para los envíos realizados entre enero de 2017 y Diciembre de 2020 el **28,13 %** son posibles desvíos operativos, lo que representa una gran cantidad en cuanto en términos de volumen y porcentaje, abriendo una amplia gama de posibilidades para realizar algunos ajustes que permitan mejorar la calidad del servicio a los clientes.

Estos inconvenientes pueden ser debido a desviaciones menores donde algunas trazas puntuales no cumplen con exactitud el proceso o también a desviaciones significativas de las trazas lo que implica incumplimientos graves en el proceso que requieren de inmediata atención y medidas predictivas para evitar que se repitan en el futuro.

Se decide analizar más en detalle estos casos realizando un muestreo de las trazas exportadas en cada caso para verificar si realmente son problemas en la entrega del producto y de ser así conocer el motivo. Se analizaron tres tipos de inconvenientes, trazas que se ajustan muy poco al proceso, trazas que tienen un ajuste mejor pero que tienen gran cantidad de movimientos y trazas que ajustan bien pero que finalizan el proceso en un intervalo de tiempo muy elevado.

El resumen de las irregularidades encontradas sobre el total de **3.456.932** trazas procesadas visualiza en la tabla **4.3**

Tabla 4.3: Irregularidades encontradas

Irregularidad	Cantidad	Porcentaje
Trazas con ajuste menor a 0,5	374.876	10,84 %
Trazas con más de 6 movimientos	1.724	0,05 %
Trazas que demoraron más de 10 días	595.880	17,24 %
TOTAL	972.480	28,13 %

4.7.1. Trazas con ajuste menor a 0,5

Como se mencionó anteriormente el ajuste dentro de la *Minería de Procesos* se refiere a una medida que cuantifica la calidad o adecuación de los datos reales al modelo de proceso descubierto. Es una



Figura 4.11: Trazas con desvíos leves

métrica que indica qué tan bien se ajustan entre sí los datos observados y el modelo descubierto. Cuanto mayor sea el ajuste, mayor será la conformidad entre el modelo y los datos reales, lo que indica que el modelo es una representación precisa del proceso observado.

Se realizó un muestreo estratificado de 300 elementos entre las 374.876 trazas que tienen un ajuste menor a 0.5 respecto del modelo, dividiendo las mismas en tres estratos y analizando puntualmente los casos seleccionados.

Los estratos se formaron de la siguiente manera:

- Estrato 1: Ajuste hasta 0,2.
- Estrato 2: Ajuste entre 0,2 y 0,4.
- Estrato 3: Ajuste mayor a 0,4.

Se encontraron algunos casos que representan desvíos menores por ejemplo repetición de un evento, pero el proceso continúa y finaliza correctamente. Los mismos se muestran en la figura 4.11 donde se observan tres trazas repiten el evento 1 intento de entrega pero el proceso continúa correctamente, posiblemente esto se deba a un error de carga (se cargó nuevamente primer intento en lugar del segundo) ya que las fechas de los eventos duplicados son posteriores al original. De todas maneras el producto se entrega o se devuelve respetando la esencia del proceso postal.

También se encontraron varios casos con desvíos graves y varias inconsistencias en el proceso, estos casos requieren una mayor atención ya que no se está cumpliendo correctamente el servicio.

La Figura 4.12 ilustra estas situaciones en las que se observan casos que hay eventos que no pueden estar en la misma traza.

Se detalla a continuación los casos que se muestran:

1. Ingresado, entregado en oficina y devuelto.
2. Ingresado, devuelto y entregado en oficina.
3. Ingresado, no se puede llegar a la dirección, segundo intento de entrega, primer intento de entrega y devolución.
4. Ingresado, segundo intento de entrega, devolución, otro segundo intento de entrega y una segunda devolución.

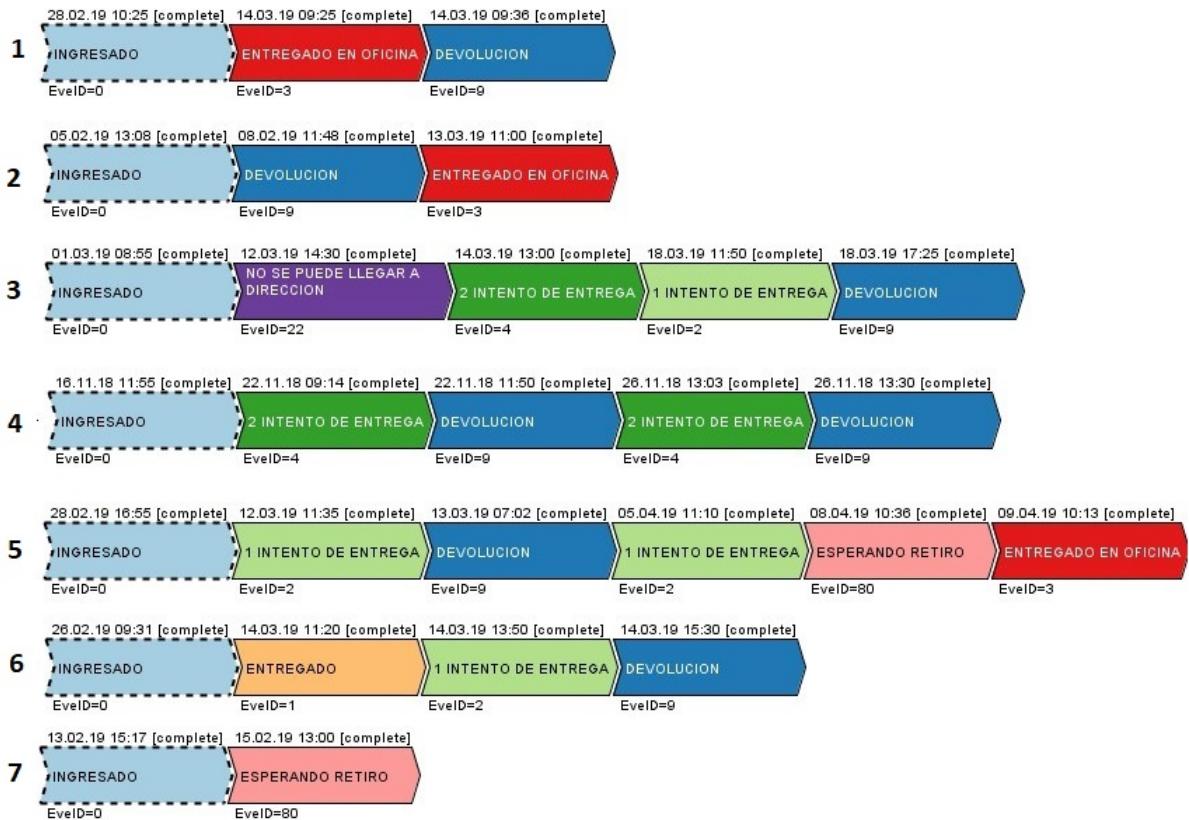


Figura 4.12: Trazas con desvíos graves

5. Ingresado, primer intento de entrega, devolución, otro primer intento de entrega, esperando retiro y entregado.
6. Ingresado, entregado, primer intento de entrega, devolución.
7. Ingresado, esperando retiro y no hay nuevos movimientos.

Claramente son problemas graves ya que por ejemplo entregado y devuelto es imposible que se cumplan para un envío porque no se puede entregar el producto al destinatario y devolverlo al remitente.

Todas Estas inconsistencias graves del proceso deben ser revisadas en tiempo y forma para tomar acciones correctivas ya que implican que no se estén cargando eventos realizados o que estos eventos nunca se hayan realizado.

4.7.2. Trazas con exceso de movimientos

Para realizar el análisis sobre los casos que tenían muchos movimientos se realizó un muestreo aleatorio simple debido a la baja cantidad de casos con estas características. A pesar de ser solo el 0,05 % se encontraron algunos casos con inconsistencias graves y movimientos contradictorios por ejemplos sucesivas entregas y devoluciones del mismo envío en diferentes fechas.

Algunos de los casos más representativos se muestran en la figura 4.13 donde se observan las siguientes irregularidades.

A continuación se explica en detalle los eventos de los casos mostrados:

1. Ingreso, seis intentos de primer entrega y devolución.
2. Ingreso, primer intento de entrega, devolución, primer intento de entrega, espera en sucursal, otra devolución, otro primer intento de entrega, espera en sucursal, otra devolución, y finalmente entregado. Es imposible devolver más de una vez un envío y luego entregarlo.
3. Ingreso, primer intento de entrega, segundo intento de entrega, devolución, luego dos veces segundo intento de entrega, dos devoluciones y finalmente entregado.
4. Ingreso, cuatro primer intento de entrega, un segundo intento de entrega y finalmente entregado.
5. Ingreso, no se pudo llegar a domicilio, esperando retiro, devolución, luego un primer y un segundo intento de entrega y finalmente otra devolución.
6. Ingreso, siete primer intento de entrega y luego entregado.
7. Ingreso, primer intento de entrega, devolución, luego tres primer intento de entrega, espera en sucursal, otra devolución y finalmente entregado.
8. Ingreso, primer intento de entrega, devolución, nuevamente primer intento de entrega dos veces, espera en sucursal, devolución y entregado
9. Ingreso, primer intento, devolución, datos faltantes en dirección, esperando retiro, otra devolución y entregado.
10. Ingreso, dos veces primer intento de entrega, devolución, otro primer intento de entrega, un segundo intento y luego entregado.
11. Ingreso, primer intento de entrega, devolución, nuevamente dos veces primer intento de entrega, esperando en sucursal y entregado.
12. Ingreso, no existe dirección, dos veces primer intento de entrega, esperando en sucursal, devolución, nuevamente un primer intento de entrega, otra vez esperando en sucursal, una segunda devolución y finalmente entregado.

En estos ejemplos se muestran graves inconsistencias en los datos que reflejan el incumplimiento del proceso y peor aún no tener información fidedigna de lo que pasó con el envío. En más de un ejemplo se ven sucesivas devoluciones y entregas de una misma pieza postal en sucesivos días, movimientos que no solo no representan el proceso sino que son contradictorios y no podrían ocurrir con el mismo envío. Nunca una carta puede ser entregada y devuelta a la vez, o devuelta varias veces en días sucesivos.

En el análisis de trazas con exceso de movimientos se encontraron graves incumplimientos del proceso.

4.7.3. Trazas que demoraron mucho tiempo

Finalmente se realizó el análisis en las trazas que tuvieron un excesivo tiempo para su finalización. Para ello se realizó también un muestreo estratificado entre las 595.880 trazas que tardaron más de 10 días en completar el proceso y dividiendo las mismas en tres estratos de la siguiente forma:

- Estrato 1: Hasta 20 días de duración (493.241 trazas).
- Estrato 2: Entre 20 y 40 días de duración (91.687 trazas).

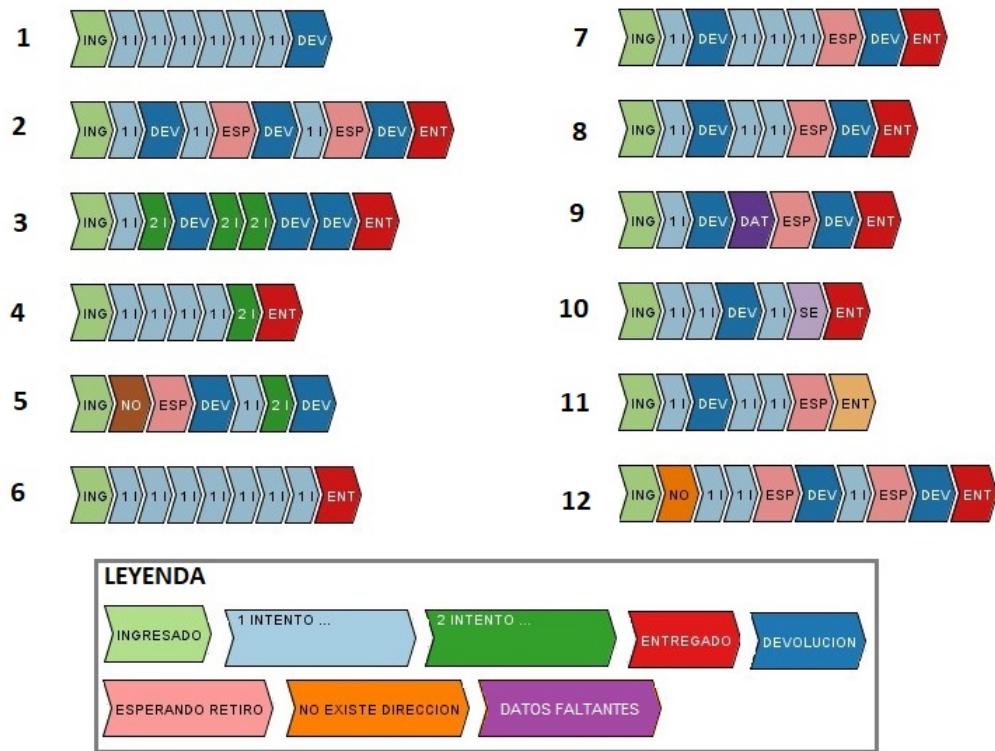


Figura 4.13: Trazas exceso de movimientos

- Estrato 3: Más de 40 días de duración (10.952 trazas)

La figura 4.14 muestra algunos ejemplos de los estratos 1 y 2, donde se observa que el proceso se cumple correctamente pero el envío permanece hasta más de 10 días sin actividad. Esto se repite en la mayoría de los casos analizados, aunque también se encontraron trazas con exceso de movimientos, y se verificó varias de las que se analizaron en la sección "trazas con exceso de movimientos" también estaban en este conjunto de datos.

La figura 4.15 muestra algunos ejemplos del estrato 3 donde se encontraron casos con tiempos extremadamente altos que de hasta 8 meses y que cumplen el proceso. Por ejemplo está el caso 3059379 que ingresa el 09/05/2019 y permanece sin movimientos hasta el 24/06/2019 donde sale por primera vez a distribución y no se pudo entregar registrando el primer intento de entrega. Luego no registra actividad hasta el 2/10/2019 donde se ingresa un segundo intento de entrega y una devolución al remitente el mismo día. Este caso particular registra una demora de casi 5 meses como el 3742021 que ingresa el 12/12/2019 y se ajusta correctamente al proceso registrando un primer intento de entrega, un segundo intento de entrega y la devolución pero finalizando el 06/05/2020.

4.7.4. Resultado del análisis

Con la suma del trabajo realizado en los trabajos previos [5] y [6] y el presente análisis aplicando *Minería de Procesos* a la distribución postal se demuestra que aplicando estas técnicas se puede hacer un aporte considerable al negocio postal en la República Argentina proporcionando herramientas que permitan evaluar y monitorear el cumplimiento del proceso con el objetivo de mejorar el servicio a los clientes. Entre otras cosas se encontraron desvíos considerables al proceso operativo, inconsistencias en la información cargada y entregada a los clientes y casos en los que a pesar de que el proceso se

2885916 2 events	2885943 2 events	2885934 2 events
INGRESADO #1 complete 01.03.2019 16:32:39.000	INGRESADO #1 complete 01.03.2019 16:25:17.000	INGRESADO #1 complete 01.03.2019 16:27:55.000
ENTREGADO #2 complete 21.03.2019 09:57:00.000	ENTREGADO #2 complete 18.03.2019 14:20:00.000	DATOS FALTANTES EN DIRECCION #2 complete 18.03.2019 14:00:00.000
2885918 2 events	3196771 4 events	3031122 4 events
INGRESADO #1 complete 01.03.2019 16:29:59.000	INGRESADO #1 complete 01.07.2019 13:52:08.000	INGRESADO #1 complete 09.05.2019 09:44:20.000
ENTREGADO #2 complete 14.03.2019 12:52:00.000	1 INTENTO DE ENTREGA #2 complete 12.07.2019 13:37:00.000	1 INTENTO DE ENTREGA #2 complete 27.05.2019 11:00:00.000
2885923 2 events	ESPERANDO RETIRO #3 complete 15.07.2019 08:03:31.000	ESPERANDO RETIRO #3 complete 30.05.2019 07:43:54.000
INGRESADO #1 complete 01.03.2019 16:29:54.000	DEVOLUCION #4 complete 30.07.2019 14:09:04.000	ENTREGADO EN OFICINA #4 complete 31.05.2019 13:15:01.000
SE MUDO #2 complete 14.03.2019 11:02:00.000		

Figura 4.14: Trazas con excesivo tiempo de finalización

3991130 3 events	3742021 4 events	4066783 5 events
INGRESADO #1 complete 30.03.2020 14:19:10.000	INGRESADO #1 complete 12.12.2019 18:56:33.000	INGRESADO #1 complete 24.04.2020 16:38:13.000
1 INTENTO DE ENTREGA #2 complete 02.06.2020 12:02:00.000	1 INTENTO DE ENTREGA #2 complete 07.02.2020 09:00:00.000	1 INTENTO DE ENTREGA #2 complete 06.05.2020 14:50:00.000
DESCONOCIDO #3 complete 30.07.2020 12:00:00.000	2 INTENTO DE ENTREGA #3 complete 18.02.2020 09:45:00.000	2 INTENTO DE ENTREGA #3 complete 07.05.2020 13:00:00.000
3059379 4 events	DEVOLUCION #4 complete 06.05.2020 11:00:00.000	DEVOLUCION #4 complete 08.07.2020 12:00:00.000
INGRESADO #1 complete 09.05.2019 18:09:28.000	3656878 3 events	DEVOLUCION #5 complete 08.07.2020 13:00:00.000
1 INTENTO DE ENTREGA #2 complete 24.06.2019 18:40:00.000	INGRESADO #1 complete 26.11.2019 18:43:55.000	
2 INTENTO DE ENTREGA #3 complete 02.10.2019 11:10:00.000	1 INTENTO DE ENTREGA #2 complete 19.12.2019 11:15:00.000	
DEVOLUCION #4 complete 02.10.2019 11:15:00.000	DEVOLUCION #3 complete 06.05.2020 09:40:00.000	

Figura 4.15: Trazas con tiempo extremadamente alto de finalización

cumplió correctamente los envíos se demoraron de forma excesiva sin ningún motivo aparente.

En el próximo capítulo se diseñará y se desarrollará un prototipo que permita incorporar estas herramientas en la operación diaria alertando al personal responsable de posibles desvíos operativos en tiempo y forma para que puedan tomar medidas predictivas o correctivas.

CAPÍTULO 5

Solución Propuesta

5.1. INTRODUCCIÓN

En este capítulo se diseñará y desarrollará un prototipo completamente funcional que aplique las técnicas de *Minería de Procesos* explicadas anteriormente y las introduzca dentro del circuito postal con el objetivo de proporcionar una nueva herramienta que permita evaluar posibles desvíos operativos en un intervalo de tiempo cercano a la operatoria real para poder corregir dichos desvíos lo más rápidamente posible y mitigar el impacto de los mismos en el servicio al cliente.

Luego de análisis realizado en los capítulos anteriores y en los trabajos previamente publicados se decide tomar solo dos de los tres tipos de incidencias analizadas. Anteriormente se consideraron como posibles desvíos las trazas que tengan un ajuste menor al 0,5 con respecto al modelo, las trazas que tengan más de 6 movimientos y las trazas que demoraron más de 10 días.

En la aplicación que se desarrollará se tomarán solo las trazas que tengan un ajuste menor a 0,5 y aquellas que tardaron más de 10 días dado que las trazas que tenían más de 6 movimientos representaban un porcentaje muy bajo con respecto al total y se encontraron algunas trazas con estas características en los otros conjuntos exportados.

5.2. METODOLOGÍA APLICADA

El prototipo de la aplicación se desarrollará utilizando metodologías ágiles ya que presentan algunas ventajas frente al ciclo de vida clásico, están diseñadas para ser flexibles, adaptarse a las necesidades de cada equipo o proyecto y son las más comunes en la actualidad. Se seguirán los lineamientos de Scrum¹ utilizando un Backlog para registrar el desarrollo junto con un tablero Kanban² para documentar los requisitos y tareas que se llevarán a cabo durante para este prototipo y para hacer un seguimiento de su avance.

El ciclo de vida clásico o cascada incluye al menos las siguientes etapas: recopilación de los requerimientos, análisis, diseño, desarrollo, testeo e implementación. En esta metodología, no se puede comenzar una fase hasta que no se haya completado la actual y luego, no se puede volver a ella, lo que significa que cualquier cambio en los requerimientos tiene un impacto muy grande y debe abordarse en una fase posterior a la finalización del desarrollo. Una de las principales ventajas de esta metodología es que es fácil de entender y aplicar, lo que la hace adecuada para proyectos bien

¹<https://www.scrum.org/>

²<https://kanban.university/>

definidos y estables. Sin embargo, su rigidez y falta de flexibilidad pueden llevar a problemas si los requisitos cambian durante el proceso de desarrollo, lo que puede resultar en retrasos y costos adicionales. Además el cliente no puede tener contacto con el producto hasta etapas avanzadas del desarrollo lo que puede ocasionar el descubrimiento de algún error de negocio en etapas muy avanzadas.

En contraste las metodologías ágiles poseen un enfoque iterativo e incremental para la gestión y desarrollo de software lo que permite adaptarse rápidamente al cambio y entregar un producto de mayor calidad.

Entre sus características están:

- Mayor flexibilidad y capacidad de respuesta al cambio: al enfocarse en pequeñas iteraciones, se adapta fácilmente a los cambios en los requisitos y necesidades del cliente.
- Mayor eficiencia en el desarrollo: Al enfocarse en entregas frecuentes, la metodología ágil permite identificar y corregir problemas de manera temprana, lo que aumenta la eficiencia del proceso de desarrollo.
- Mejora de la calidad del software: como la metodología ágil se enfoca en la entrega de software funcional en cada iteración, tiene una mayor eficacia en la identificación y corrección de problemas, reduciendo el tiempo y los costos asociados con la corrección de errores.

Si se verifica que la implementación del prototipo contribuye a la mejora del proceso y la calidad del servicio se propondrá el desarrollo de una solución definitiva.

5.3. ARQUITECTURA DE LA SOLUCIÓN

Para desarrollar el prototipo y posteriormente la aplicación definitiva se utilizará una arquitectura orientada a microservicios, la cual utiliza una estrategia de diseño que busca dividir una aplicación en pequeños componentes independientes y autónomos que se comunican entre sí mediante APIs. Esta arquitectura permite una mayor flexibilidad, escalabilidad y facilidad de mantenimiento, ya que cada microservicio se puede desarrollar, probar y desplegar de forma independiente. Los componentes se estructurarán en contenedores independientes usando Docker³ para una rápida integración con la nube y futura orquestación con Kubernetes⁴

Se desarrollarán de manera independiente los siguientes componentes:

1. Proceso de exportación de archivos: proceso automático que realizará una exportación de la base de datos Oracle productiva y tomará las trazas que hayan ingresado desde la última ejecución. Se cronará para ejecutarse en un intervalo de tiempo especificado (por defecto una vez por semana) y dejará las trazas en un archivo csv.
2. API de Minería de Procesos: aplicación que leerá el archivo exportado en el paso anterior y aplicará las técnicas de minería de procesos en la distribución postal como se describe en este trabajo, luego del análisis exportará los resultados a una base de datos PostgreSQL.
3. API de gestión: se desarrollará una API que permita gestionar las incidencias encontradas por la API de minería de procesos.

³<https://www.docker.com/>

⁴<https://kubernetes.io/es/>

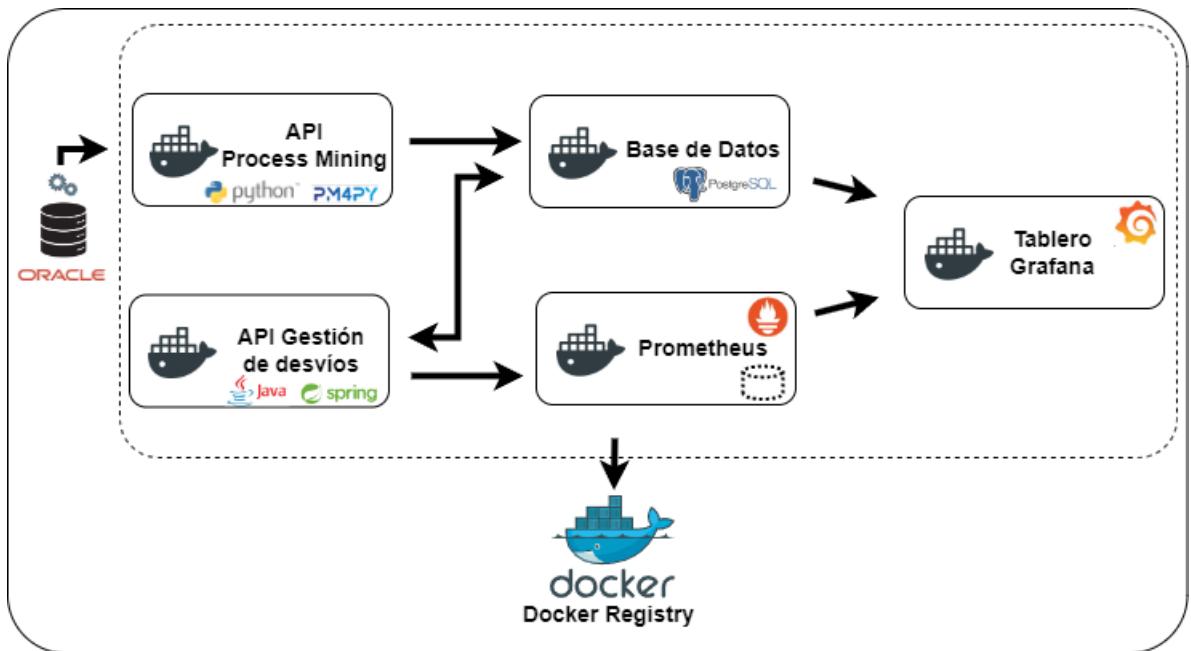


Figura 5.1: Arquitectura del prototipo a desarrollar

4. Base de Datos: Se utilizará una base de datos relacional para almacenar las incidencias encontradas y registrar su resolución.
5. Observabilidad: Se utilizará Prometheus para verificar el funcionamiento de los servicios y Grafana para mostrar estadísticas

La arquitectura de las solución se muestra en la figura 5.1

Para el desarrollo de este prototipo se seguirán las buenas prácticas de desarrollo de software actuales entre las que se encuentran los casos de prueba automatizados, el uso de estrategias de desarrollo colaborativo y estrategias enfocadas al desarrollo incremental como GitFlow⁵, el código limpio (Clean Code)[23] generando un código fácil de entender y mantener, la refactorización constante[24] que permite mejorar continuamente el diseño y la estructura del código y el uso de herramientas de evaluación de código como SonarQube⁶.

La implementación de estas prácticas permitirá acoplarse rápidamente en un ambiente de Integración continua y despliegue continuo (CI/CD) donde se automatiza la construcción, pruebas y despliegue de los microservicios de manera automática mediante la utilización de pipelines generados con herramientas de DevOps como Jenkins, GitLab CI/CD o GitHub Actions.

La API de gestión de incidencias se desarrollará con una arquitectura de tres capas como se estila actualmente en este tipo de aplicaciones desarrolladas con Spring o Spring Boot implementando la capa de controlador, servicio y repositorio. Adicionalmente se generará la documentación contenida en la aplicación y se generará un mecanismo que permita verificar su estado, es decir, si el servicio está activo o tiene algún inconveniente (health check)

⁵<https://aulasoftwarelibre.github.io/taller-de-git/gitflow/>

⁶<https://docs.sonarqube.org/latest/>

5.4. ANÁLISIS DE LA NECESIDAD

5.4.1. Requerimientos

Para el desarrollo del prototipo para aplicar *Minería de Procesos* a la distribución postal se recopilaron los siguientes requisitos:

- Generación de Prueba de Concepto: Desarrollo de una herramienta mínima que permita verificar la viabilidad de utilizar las técnicas de *Minería de procesos* desde el código, la misma deberá incluir:
 - Importación de un archivo de trazas en formato .xes
 - Importación de un proceso modelado como una Red de Petri
 - Realizar la verificación de conformidad del archivo de trazas contra la Red de Petri
 - Filtrar las trazas que no cumplan con los criterios establecidos (ajuste o duración)
 - Exportar las trazas resultantes en un archivo de formato .csv
- Verificación de Prueba de Concepto: Se deberán contrastar los resultados obtenidos por la prueba de concepto contra los que brinda una herramienta de *Minería de Procesos* como por ejemplo ProM tools⁷
- Generación de Proceso de exportación de datos: se generará un proceso batch que se ejecutará de manera automática en los servidores de la empresa y realizará una extracción de la base de datos productiva con todos los envíos que estén pendientes de análisis. Este proceso estará cronado en el servidor y se ejecutará una vez a la semana
- API de *Minería de Procesos*: Este componente se encargará de procesar los archivos que se exporten semanalmente, deberá realizar el análisis completo de las trazas realizando los siguientes pasos:
 - Importación del archivo resultante del proceso de exportación.
 - Conversión de los datoas a formato .xes
 - Filtrado de trazas inconsistentes.
 - Importación del proceso a verificar
 - Verificación de conformidad del archivo log
 - Filtrado de trazas con ajuste menor a 0.5
 - Filtrado de trazas que tengan más de diez días de duración.
 - Ingreso de datos resultantes del análisis a Base de Datos de gestión de incidencias.
 - Envío de alerta por mail con resumen de posibles desvíos encontrados.
- API de Gestión de incidencias: Permitirá listar las incidencias encontradas y brindará información adicional de las mismas para tomar acciones predictivas o correctivas con el desvío que representan, deberá contener las siguientes funcionalidades:
 - Ingreso con usuarios nominados con usuario y clave o token.

⁷<https://promtools.org/>

- Listado de posibles desvíos pendientes de tratamiento (por ajuste bajo).
 - Listado de posibles desvíos pendientes de tratamiento (por excesivo tiempo de ejecución).
 - Obtención de datos adicionales del desvío (Responsable, sucursal, etc)
 - Registro de resultado de tratamiento (resuelto, informado, no es desvío)
 - Listado consolidado de desvíos por sector/sucursal o responsable
- Base de Datos: Se generará una base de datos relacional que permita almacenar la información obtenida del análisis realizado.

5.4.2. Requerimientos no funcionales

- Observabilidad de la aplicación: Se implementarán herramientas que permitan visualizar el estado de los servicios en tableros dinámicos, así como monitorear su uso y su estado a través de algún mecanismo que permita realizar el health check de la aplicación.
- Rápida visualización de incidencias: Se implementará algún mecanismo de tableros que permita visualizar rápidamente las incidencias encontradas y su estado, pudiendo filtrar por diferentes estados y rangos de fecha.
- Contenerización: Los servicios deberán disponibilizarse en contenedores para un despliegue productivo rápido, de fácil mantenimiento y escalamiento vertical y horizontal adecuado..

Las funcionalidades se mapearon en un tablero Kanban siguiendo los lineamientos de Scrum organizando las principales tareas en épicas que se componen de tareas más pequeñas con el objetivo de atomizar al máximo las tareas para organizar y distribuir mejor la carga de trabajo. Se generó un ticket por cada tarea, el tablero se muestra en la figura 5.2

En la figura 5.3 se muestra el detalle de la épica *API Gestión de Desvíos* con la lista de tareas necesarias para completar el desarrollo y el estado actual de cada una.

5.5. DISEÑO DEL PROTOTIPO

5.5.1. Consideraciones generales

El prototipo estará compuesto por tres partes principales, el componente que analizará los datos de distribución postal y aplicará las técnicas de minería de procesos, el componente que permitirá gestionar las incidencias encontradas y la base de datos donde se persistirá la información. Adicionalmente se implementarán herramientas actuales de monitoreo y visualización de datos.

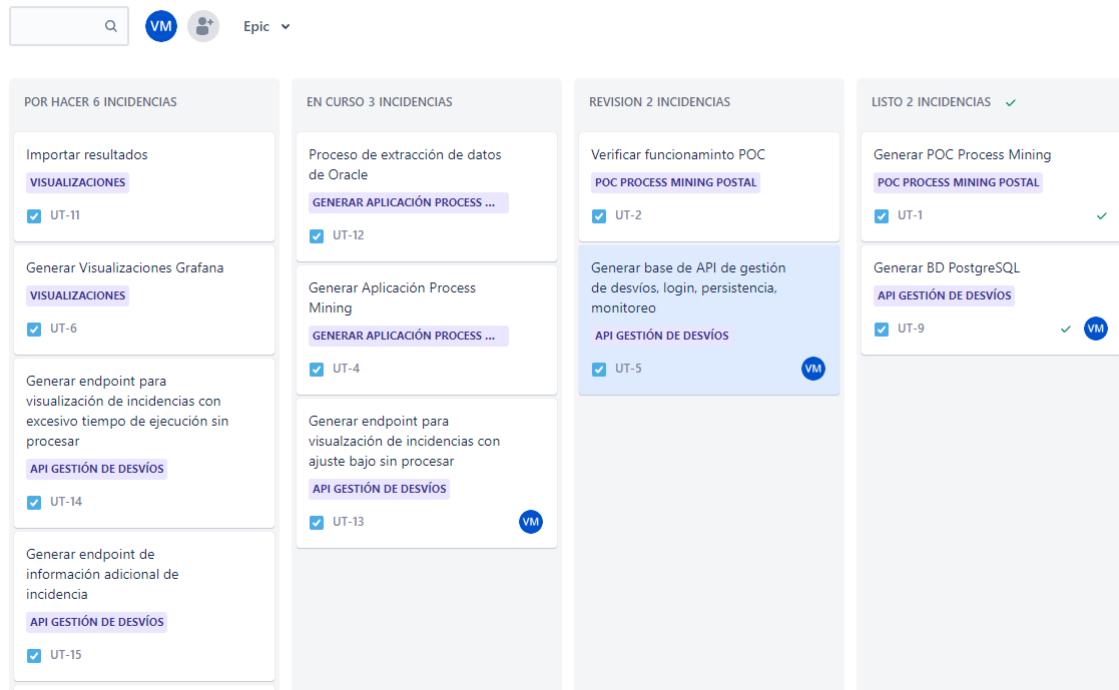
El proceso que tomará los datos originales de la Base de Datos Oracle será un shell script que ejecutará una consulta a la base de datos y guardará los resultados en un archivo csv.

Para desarrollar la aplicación de minería de procesos se utilizará Python como lenguaje de programación y la librería PM4PY⁸ para implementar las técnicas de minería de procesos.

Para desarrollar la API de gestión de incidencias se utilizará java con Spring boot, JPA, se documentará

⁸<https://pm4py.fit.fraunhofer.de/>

Proyectos / UNLP-Tesis

Tablero UNLP-Tesis**Figura 5.2:** Tablero Kanban con funcionalidades a desarrollar

API Gestión de desvíos

Detalles:

- Responsable: Sin asignar
- Asignarme a mí
- Etiquetas: Ninguno
- Start date: Ninguno
- Fecha de vencimiento: Ninguno
- Informador: Victor M

Campos anclados: Haz clic en el situado junto a la etiqueta de un campo para anclarlo.

Tareas por hacer:

Tarea	Estado	Opciones
UT-9 Generar BD PostgreSQL	FINALIZADA	VM
UT-5 Generar base de API de gestión de desvíos, login, persistencia, monitoreo	REVISIÓN	VM
UT-13 Generar endpoint para visualización de incidencias con ajuste bajo sin procesar	EN CURSO	VM
UT-14 Generar endpoint para visualización de incidencias con excesivo tiempo de ejecución sin procesar	TAREAS POR HACER	VM
UT-15 Generar endpoint de información adicional de incidencia	TAREAS POR HACER	VM
UT-16 Generar endpoint de registro de incidencia procesada	TAREAS POR HACER	VM
UT-17 Generar casos de prueba	TAREAS POR HACER	VM

Añadir un comentario... Consejo de expertos: pulsa para comentar

Figura 5.3: Épica API de Gestión de Incidencias

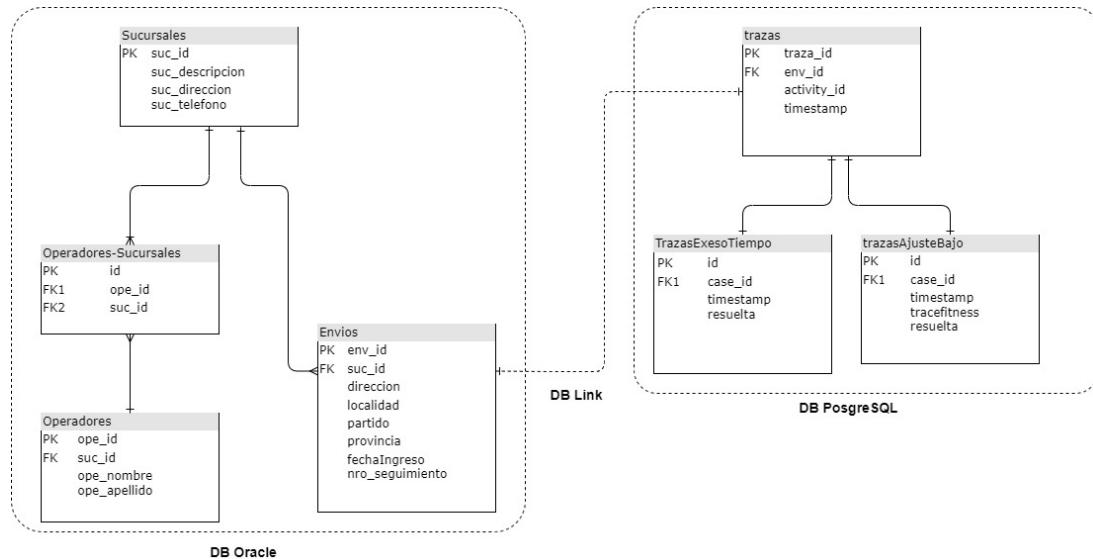


Figura 5.4: DER del Prototipo de Minería de Procesos aplicado a la distribución postal

con con Open API⁹ a través de Swagger, y se utilizará el componente Actuator¹⁰ de Spring¹¹ para exponer las estadísticas de uso, realizar el monitoreo y la comprobación del estado de la aplicación (health check)

Para la persistencia de los datos se consideró necesario elegir una herramienta de almacenamiento de datos confiable y escalable, por lo que se optó por un motor de base de datos relacional. Para ello, se seleccionó PostgreSQL, una herramienta open source ampliamente reconocida en el mercado por su robustez y estabilidad.

La elección de PostgreSQL como herramienta de almacenamiento se debió a su capacidad para manejar grandes volúmenes de datos, su arquitectura orientada a objetos y su compatibilidad con diversos sistemas operativos. Además, se podrá utilizar un dblink para la integración de datos provenientes de distintas fuentes y su integración en un solo repositorio.

Para realizar el monitoreo de los servicios se usará Prometheus lo que permitirá recopilar las estadísticas de uso y funcionamiento. Adicionalmente se utilizará grafana para su visualización. En grafana también se crearán tableros que permitan observar y diagnosticar rápidamente las incidencias encontradas.

5.5.2. Diagrama de entidad relación

La base de datos que de este prototipo contendrá las tablas necesarias para almacenar la información de las incidencias encontradas y su gestión. Adicionalmente se conectará a la base de datos Oracle productiva a través de un dblink para traer información correspondiente a sucursales y envíos necesaria para la gestión.

La figura 5.4 muestra el Diagrama de Entidad Relación con las tablas necesarias para el prototipo y las principales tablas que se consultarán de la Base de datos Externa.

⁹<https://www.openapis.org/>

¹⁰<https://spring.io/guides/gs/actuator-service/>

¹¹<https://spring.io/>

```

Spyder (Python 3.10)
File Edit Search Source Run Debug Consoles Projects Tools View Help
D:\Facultad\UNLP - Tests\UNLP-Tests\App\PythonPrototype\POC-ProcessMining.py
ProcessMiningPostI.py  POC-ProcessMining.py  ProcessDiscover.py

1 # -*- coding: utf-8 -*-
2 ...
3 # Created on Tue Apr 25 22:34:36 2023
4 #
5 # @author: Vlc
6 #
7 #
8 import pm4py
9 import pandas as pd
10
11 POSTAL_PROCESS_PETRI_NET = "PostalProcess.pnml"
12 INPUT_LOG_FILE = "TransactionsJustifiedjo.csv"
13 OUTPUT_LOG_FILE = "TracesConBajaPerformance.csv"
14 OUTPUT_LOG_FILE2 = "TracesConBajaPerformance.xes"
15 OUTPUT_LOG_FILE_XES = "TracesConBajaPerformance.xes"
16 OUTPUT_LOG_FILE_XES2 = "TracesConBajaPerformance.xes"
17
18 def mine_postal_data():
19
20     # Import log from process
21     event_log = pm4py.read_xes(INPUT_LOG_FILE)
22     petri_net, initial_marking, final_marking = pm4py.read_pnml(POSTAL_PROCESS_PETRI_NET)
23
24     # Verificación de conformidad
25     conformance_token_based_replay = pm4py.conformance_diagnostics_token_based_replay(event_log, petri_net, initial_marking, fit)
26     conformance_alignments = pm4py.conformance_diagnostics_alignments(event_log, petri_net, initial_marking, fi)
27
28     # Filtrado por ajuste menor a 0.5
29     filtered_traces = event_log.filter(lambda x: x['trace_fitness'] < 0.5, conformance_token_based_replay)
30
31     # Export filtrados ajuste menor a 0.5
32     data_frame1 = pd.DataFrame(filtered_traces)
33     data_frame1.to_csv(OUTPUT_LOG_FILE)
34     pm4py.write_xes(filtered_traces, OUTPUT_LOG_FILE_XES)
35     pm4py.write_xes(filtered_traces, OUTPUT_LOG_FILE_XES2)
36
37     # Filtrado por duración
38     filtered_traces_performance = pm4py.filter_case_performance(event_log, 8000000.0, 19000000.0, timestamp_key="time:timestamp")
39
40     # Export filtrados duración
41     data_frame2 = pd.DataFrame(filtered_traces_performance)
42     data_frame2.to_csv(OUTPUT_LOG_FILE)
43     pm4py.write_xes(filtered_traces_performance, OUTPUT_LOG_FILE_XES2)
44     print(data_frame2.count())
45
46     # print("Conformance Token Based Replay")
47     # data_frame1 = pd.DataFrame(conformance_token_based_replay)
48     # print("Conformance Alignments")
49     # data_frame2 = pd.DataFrame(conformance_alignments)
50     # print(data_frame2.to_string(index=False))
51
52 if __name__ == "__main__":
53     mine_postal_data()

```

Figura 5.5: Prueba de concepto visualización del entorno de desarrollo Spyder

5.6. DESARROLLO

5.6.1. Prueba de Concepto

Como se mencionó anteriormente la prueba de concepto se desarrolló con el lenguaje de programación Python, utilizando la librería PM4PY para aplicar las técnicas de *Minería de Procesos* y verificar viabilidad del desarrollo. Se eligió Anaconda Navigator¹² una interfaz gráfica que permite administrar extensiones y lanzar paquetes de Anaconda y entornos de desarrollo. Puntualmente se utilizó el IDE Spyder¹³ para hacer el desarrollo por ser uno de los más utilizados en Python para hacer desarrollos e investigación científica.

Se observa el desarrollo de la prueba de concepto en el IDE Spyder en la figura 5.5

El prototipo desarrollado importa un archivo de trazas en formato .xes y el proceso postal descubierto anteriormente en formato de Red de Petri, luego reproduce una a una las trazas sobre el modelo del proceso Postal y filtra las trazas que tengan un ajuste menor a 0,5 o más de 10 días de duración, siguiendo los lineamientos establecidos en este trabajo. Finalmente exporta los resultados en archivos .xes y .csv para su posterior utilización.

5.6.2. Verificación de la Prueba de Concepto

Para comprobar el funcionamiento de la prueba de concepto se procesaron nuevamente los archivos de datos que se utilizaron en este trabajo y luego se realizó una comprobación visual de cada uno de los archivos de datos resultantes en ProM Tools, la misma herramienta que se utilizó originalmente. Se comparó la cantidad de trazas con ajuste menor a 05 y con más de 10 días de duración encontrando resultados similares.

Con estos resultados se da por válida la implementación de *Minería de Procesos* en la prueba de concepto desarrollada en Python y se procede con el desarrollo de la aplicación.

¹²<https://docs.anaconda.com/free/navigator/>

¹³<https://www.spyder-ide.org/>

5.6.3. Proceso de exportación de Datos desde BD Oracle

Para el proceso de extracción de datos de la base de datos Oracle, se decidió utilizar un shell script que estará programado para ser ejecutado periódicamente en el servidor de Base de Datos. La tarea principal de este script será ejecutar una consulta SQL para obtener todos los envíos que hayan sido ingresados desde la última ejecución del script.

Una vez obtenidos los resultados de la consulta, el script los exportará a un archivo .csv. Este archivo se generará en una ubicación específica del servidor, y contendrá los datos de los envíos ordenados por traza. Este ordenamiento se ha realizado con el fin de optimizar el proceso de procesamiento posterior de los datos.

Luego de la correcta generación del archivo .csv, se procederá a marcar los envíos como ".exportados". Esta tarea se realizará para asegurar que no se vuelvan a exportar los mismos envíos en la próxima ejecución del script. El script se ejecutará una vez a la semana, siguiendo un cronograma específico en el servidor Unix. Esto permitirá una actualización regular de los datos y un flujo de información más constante y confiable.

De esta manera se automatizará el proceso de extracción de datos de la base de datos Oracle, reduciendo el tiempo y la complejidad de esta tarea. Además, gracias a la marcación de envíos exportados, se evita la duplicidad de datos y se asegura la integridad y calidad de la información.

El código fuente del script de extracción de datos de Oracle se muestra en el listado 5.1

Listado 5.1: Código fuente del proceso de extracción de datos de Oracle

```
1  #!/bin/bash
2
3  # Configuración de la conexión a la base de datos Oracle
4  ORACLE_USER="userName"
5  ORACLE_PASSWORD="userPassword"
6  ORACLE_HOST="192.168.0.54"
7  ORACLE_PORT="5001"
8  ORACLE_SID="bdProd"
9
10 # Consulta SQL a ejecutar
11 SQL_QUERY="SELECT env_id AS case_id , eve_descrip AS activity , ↵
12   ↵ eve_fecha AS timestamp , eve_id AS activityid
13   FROM envios
14 WHERE informado IS NULL
15 ORDER BY env_id"
16
17 # Archivo de salida CSV
18 OUTPUT_FILE="TrazasEntrada.csv"
19
20 # Conectarse a la base de datos Oracle y ejecutar la consulta
21 sqlplus -s ${ORACLE_USER}/${ORACLE_PASSWORD}@${ORACLE_HOST}: ↵
22   ↵ ${ORACLE_PORT}/${ORACLE_SID} << EOF
23 set pagesize 0
24 set colsep ,
25 set feedback off
26 set echo off
```

```

25 set heading off
26 spool ${OUTPUT_FILE}
27 ${SQL_QUERY};
28 spool off
29 exit
30 EOF

```

5.6.4. Base de Datos

La base de datos utilizada en este prototipo cuenta con las tablas necesarias para almacenar la información de las incidencias encontradas y su respectiva gestión. Para el ambiente productivo se establecerá una conexión entre la base de datos PostgreSQL y la base de datos Oracle mediante el uso de dblink, con el objetivo de acceder a la información correspondiente a sucursales y envíos necesarios para la gestión. En el ámbito de este prototipo se replicarán los datos provenientes de Oracle dado que no se cuenta con acceso a los servidores productivos.

Se generaron las tablas según lo establecido en el diseño de la Base de Datos, las principales son:

- trazas: Contiene el listado de trazas exportado desde la base de datos Oracle. Guarda un registro por cada movimiento de la traza para poder reconstruir su historia.
- trazasajustebajo: Contiene las trazas que tengan un ajuste por debajo del umbral establecido como posible desvío. Actualmente fijado en 0,5
- trazasexcesotiempo: Contiene las trazas que tengan una duración mayor a la establecida como umbral de posibles desvío. Para este caso 10 días corridos.

La figura 5.6 como se almacenan las trazas que tienen un ajuste bajo para su posterior tratamiento. El componente que aplica las técnicas de *Minería de Procesos* sobre los archivos de entrada es quien inserta los datos en la BD PostgreSQL.

El listado 5.2 muestra el código en SQL utilizado para generar la tabla *trazasajustebajo* donde se almacenan las trazas que tengan un ajuste menor a 0,5 con respecto al modelo.

Listado 5.2: Código en SQL par generar la tabla trazasajustebajo

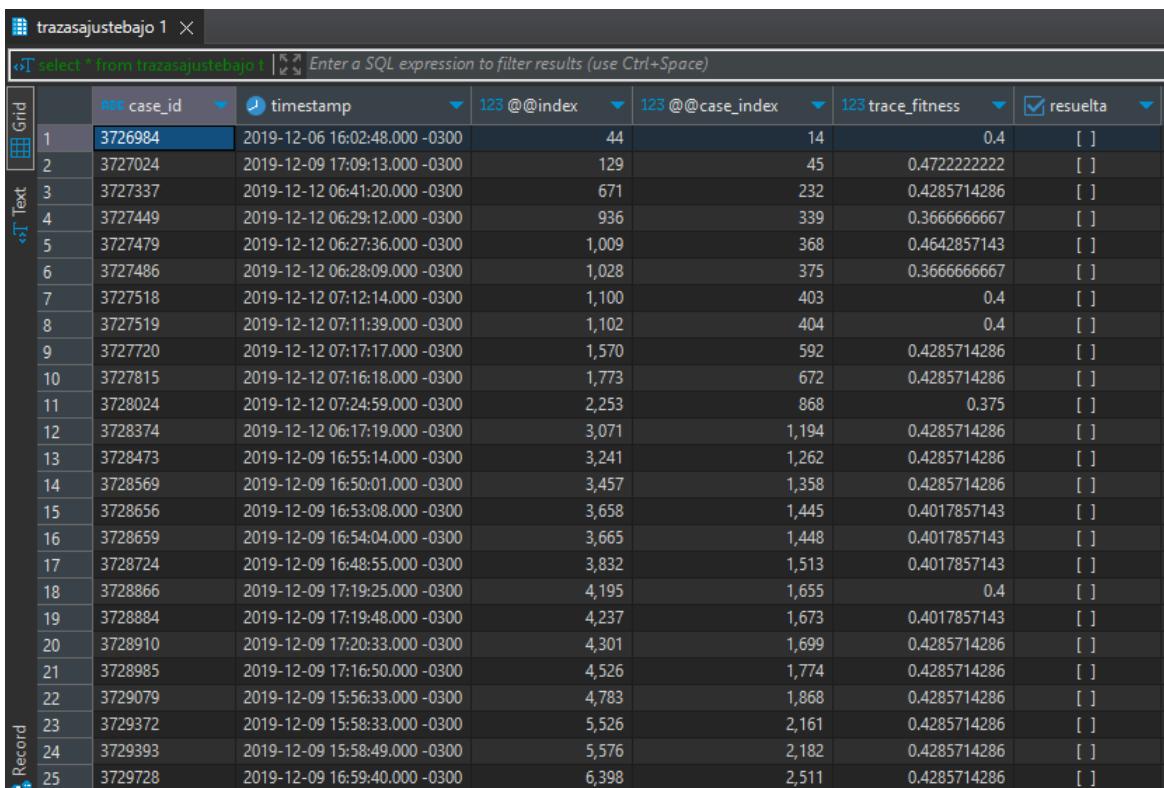
```

1 CREATE TABLE public.trazasajustebajo (
2     case_id text NOT NULL,
3     "timestamp" timestampz NULL,
4     "@@index" int8 NULL,
5     "@@case_index" int8 NULL,
6     trace_fitness float8 NULL,
7     resuelta bool NULL DEFAULT false,
8     CONSTRAINT trazasajustebajo_pk PRIMARY KEY (case_id),
9     CONSTRAINT trazasajustebajo_fk FOREIGN KEY (case_id) ←
10      REFERENCES public.trazas(case_id)
);

```

5.6.5. Componente de Minería de Procesos

Luego de la verificación de la prueba de concepto se desarrolló el componente que aplica las técnicas de *Minería de Procesos* sobre los archivos resultantes del proceso de exportación de datos. Este



The screenshot shows a database query results window titled "trazasajustebajo 1". The query is "select * from trazasajustebajo t". A search bar at the top right says "Enter a SQL expression to filter results (use Ctrl+Space)". The table has columns: Grid, case_id, timestamp, @@index, @@case_index, trace_fitness, and resuelta. There are 25 rows of data, each with a unique case_id and timestamp, and varying fitness values ranging from 0.4 to 0.472222222.

Grid	case_id	timestamp	@@index	@@case_index	trace_fitness	resuelta
1	3726984	2019-12-06 16:02:48.000 -0300	44	14	0.4	[]
2	3727024	2019-12-09 17:09:13.000 -0300	129	45	0.472222222	[]
3	3727337	2019-12-12 06:41:20.000 -0300	671	232	0.4285714286	[]
4	3727449	2019-12-12 06:29:12.000 -0300	936	339	0.3666666667	[]
5	3727479	2019-12-12 06:27:36.000 -0300	1,009	368	0.4642857143	[]
6	3727486	2019-12-12 06:28:09.000 -0300	1,028	375	0.3666666667	[]
7	3727518	2019-12-12 07:12:14.000 -0300	1,100	403	0.4	[]
8	3727519	2019-12-12 07:11:39.000 -0300	1,102	404	0.4	[]
9	3727720	2019-12-12 07:17:17.000 -0300	1,570	592	0.4285714286	[]
10	3727815	2019-12-12 07:16:18.000 -0300	1,773	672	0.4285714286	[]
11	3728024	2019-12-12 07:24:59.000 -0300	2,253	868	0.375	[]
12	3728374	2019-12-12 06:17:19.000 -0300	3,071	1,194	0.4285714286	[]
13	3728473	2019-12-09 16:55:14.000 -0300	3,241	1,262	0.4285714286	[]
14	3728569	2019-12-09 16:50:01.000 -0300	3,457	1,358	0.4285714286	[]
15	3728656	2019-12-09 16:53:08.000 -0300	3,658	1,445	0.4017857143	[]
16	3728659	2019-12-09 16:54:04.000 -0300	3,665	1,448	0.4017857143	[]
17	3728724	2019-12-09 16:48:55.000 -0300	3,832	1,513	0.4017857143	[]
18	3728866	2019-12-09 17:19:25.000 -0300	4,195	1,655	0.4	[]
19	3728884	2019-12-09 17:19:48.000 -0300	4,237	1,673	0.4017857143	[]
20	3728910	2019-12-09 17:20:33.000 -0300	4,301	1,699	0.4285714286	[]
21	3728985	2019-12-09 17:16:50.000 -0300	4,526	1,774	0.4285714286	[]
22	3729079	2019-12-09 15:56:33.000 -0300	4,783	1,868	0.4285714286	[]
23	3729372	2019-12-09 15:58:33.000 -0300	5,526	2,161	0.4285714286	[]
24	3729393	2019-12-09 15:58:49.000 -0300	5,576	2,182	0.4285714286	[]
25	3729728	2019-12-09 16:59:40.000 -0300	6,398	2,511	0.4285714286	[]

Figura 5.6: Tabla donde se almacenan las trazas encontradas con ajuste bajo

componente está diseñado para procesar y analizar trazas de los envíos postales, realiza una serie de tareas predefinidas en función de los parámetros de entrada establecidos.

En primer lugar, se realiza una verificación de la existencia del archivo CSV que contiene las trazas a procesar. Si este archivo existe, se importa para poder procesar los datos contenidos en él. A continuación, se importa el modelo de red de petri en el que se define el proceso postal y se procede a realizar la verificación de conformidad de las trazas contra este modelo.

Una vez verificada la conformidad, se remueven las columnas adicionales que se hayan generado durante el análisis. A continuación, se filtran las trazas en función de los resultados obtenidos en el análisis. Se separan aquellas que tengan un ajuste menor a 0,5 y las que tardaron más de diez días en finalizar. Finalmente, se insertan los datos resultantes en la base de datos PostgreSQL en las tablas diseñadas para tal fin: trazas, trazasajustebajo y trazasexcesotiempo, agregando los nuevos valores a los existentes en cada tabla.

Después de realizar la inserción en la base de datos, se envía una notificación por correo electrónico a la dirección de correo electrónico establecida en uno de los parámetros de entrada.

Finalmente, se renombra el archivo CSV original concatenando la fecha actual al nombre y se mueve a la carpeta "procesados" para guardarla como copia de seguridad de las trazas procesadas.

De esta manera este componente es idempotente y altamente configurable por lo tanto puede ser adaptado a distintas necesidades específicas.

El código del componente de minería de procesos se muestra en el listado 5.3

Listado 5.3: Componente de Minería de Procesos en distribución Postal

```
1 # -*- coding: utf-8 -*-
2 """
3 Created on Mon Apr 26 22:38:59 2023
4
5 @author: Vic
6 """
7
8 import pm4py
9 import pandas as pd
10 import sqlalchemy
11 import smtplib
12 import shutil
13 import os
14 import shutil
15 from datetime import datetime
16
17 POSTAL_PROCESS_PETRI_NET = "PostalProcessAlpha.pnml"
18 INPUT_LOG_FILE = "TrazasEntrada-01.csv"
19 DB_HOST = "localhost"
20 DB_NAME = "UNLP-Tesis"
21 DB_USER = "postgres"
22 DB_PASSWORD = "1234"
23 EXPORT_FILES = False
24
25 def mine_postal_data():
26
27     if os.path.exists(INPUT_LOG_FILE):
28
29         #importar csv
30         event_log = pd.read_csv(INPUT_LOG_FILE, sep=',')
31         event_log = event_log.applymap(str)
32         event_log = pm4py.format_dataframe(event_log, ↪
33                                         case_id='case_id', activity_key='activity', ↪
34                                         timestamp_key='timestamp')
35
36         #importar proceso
37         petri_net, initial_marking, final_marking = ↪
38             pm4py.read_pnml(POSTAL_PROCESS_PETRI_NET)
39
40         #verificación de conformidad
41         conformance_token_based_replay = ↪
42             pm4py.conformance_diagnostics_token_based_replay ↪
43             (event_log, petri_net, initial_marking, ↪
44             final_marking, activity_key='activity', ↪
45             case_id_key='case_id', ↪
46             timestamp_key='time:timestamp')
47
48         data_frame = pd.DataFrame(conformance_token_based_replay)
49         event_log_traces = ↪
50             event_log.drop_duplicates(subset=['case_id'])
51         event_log_traces.drop('activityid', inplace=True, axis=1)
52         event_log_traces.drop('activity', inplace=True, axis=1)
53         merged_df = pd.merge(event_log_traces, data_frame, ↪
```

```
    ↪ left_index=True, right_index=True, how='inner')
```

44
45 #Remover columnas adicionales
46 merged_df.drop('case:concept:name', inplace=True, axis=1)
47 merged_df.drop('concept:name', inplace=True, axis=1)
48 merged_df.drop('time:timestamp', inplace=True, axis=1)
49 merged_df.drop('activated_transitions', inplace=True, ↪
 ↪ axis=1)
50 merged_df.drop('reached_marking', inplace=True, axis=1)
51 merged_df.drop('enabled_transitions_in_marking', ↪
 ↪ inplace=True, axis=1)
52 merged_df.drop('transitions_with_problems', ↪
 ↪ inplace=True, axis=1)
53 merged_df.drop('missing_tokens', inplace=True, axis=1)
54 merged_df.drop('consumed_tokens', inplace=True, axis=1)
55 merged_df.drop('remaining_tokens', inplace=True, axis=1)
56 merged_df.drop('produced_tokens', inplace=True, axis=1)
57 merged_df.drop('trace_is_fit', inplace=True, axis=1)
58
59 #Filtrado por ajuste menor a 0.5
60 traces_with_low_alignemt = ↪
 ↪ merged_df.query("trace_fitness<0.5")
61
62 #Filtrado por tiempo excesivo de resolución
63 filtered_traces_performance = ↪
 ↪ pm4py.filter_case_performance(event_log, ↪
 ↪ 800000.0, 19000000.0, ↪
 ↪ timestamp_key='time:timestamp', ↪
 ↪ case_id_key='case:concept:name')
64
65 #Remover columnas adicionales
66 traces_with_excessive_time_detail = ↪
 ↪ pd.DataFrame(filtered_traces_performance)
67 traces_with_excessive_time_detail.drop('case:concept:name', ↪
 ↪ inplace=True, axis=1)
68 traces_with_excessive_time_detail.drop('concept:name', ↪
 ↪ inplace=True, axis=1)
69 traces_with_excessive_time_detail.drop('time:timestamp', ↪
 ↪ inplace=True, axis=1)
70 traces_with_excessive_time = ↪
 ↪ traces_with_excessive_time_detail.drop_duplicates ↪
 ↪ (subset=['case_id'])
71 traces_with_excessive_time.drop('activityid', ↪
 ↪ inplace=True, axis=1)
72 traces_with_excessive_time.drop('activity', ↪
 ↪ inplace=True, axis=1)
73
74 event_log.drop('case:concept:name', inplace=True, axis=1)
75 event_log.drop('concept:name', inplace=True, axis=1)
76 event_log.drop('time:timestamp', inplace=True, axis=1)
77
78 #Export a BD Postgre

```

79         engine = ↵
80             ↪ sqlalchemy.create_engine("postgresql://" +DB_USER+ ":" ↵
81             ↪ +DB_PASSWORD+"@"+DB_HOST+"/"+DB_NAME)
82     event_log.to_sql('trazas', engine, if_exists='append', ↵
83             ↪ index=False)
84     traces_with_low_alignemt.to_sql('trazasajustebajo', ↵
85             ↪ engine, if_exists='append', index=False)
86     traces_with_excessive_time.to_sql('trazasexcesotiempo', ↵
87             ↪ engine, if_exists='append', index=False)
88     send_notification()
89     move_file_proceced()
90
91
92
93     if EXPORT_FILES == True:
94
95
96         OUTPUT_LOG_FILE = "TrazasConAjusteBajoAPP.csv"
97         OUTPUT_LOG_FILE2 = "TrazasConBajaPerformanceAPP2.csv"
98         OUTPUT_LOG_FILE_XES = "TrazasConAjusteBajoApp.xes"
99         OUTPUT_LOG_FILE_XES2 = ↵
100             ↪ "TrazasConBajaPerformanceAPP.xes"
101
102
103         #Export a csv las trazas con ajuste bajo
104         data_frame = pd.DataFrame(traces_with_low_alignemt)
105         data_frame.to_csv(OUTPUT_LOG_FILE, index=False, ↵
106             ↪ header=True)
107         #Export a csv las trazas con más de diez días
108         traces_with_excessive_time_detail.to_csv ↵
109             ↪ (OUTPUT_LOG_FILE2, index=False, header=True)
110         #Export a xes las trazas con ajuste bajo
111         pm4py.write_xes(data_frame, OUTPUT_LOG_FILE_XES)
112         #Export a xes las trazas con más de diez días
113         pm4py.write_xes(filtered_traces_performance, ↵
114             ↪ OUTPUT_LOG_FILE_XES2)
115
116
117     else:
118         print("No hay archivos de trazas para procesar")
119
120
121
122     def send_notification():
123         remitente = "Notificaciones<notificaciones@mail.com.ar>"
124         destinatario = "Responsable de envíos<↵
125             ↪ <responsable@mail.com.ar>""
126         asunto = "Posibles desvíos operativos en envíos de trazas"
127         mensaje = """Atención:<br/> <br/>
128         Se encontraron posibles desvíos operativos en el ↵
129             ↪ procesamiento de varios envíos. <br>
130         Ingresar en la aplicación de seguimiento para su gestión. ↵
131             ↪ <br><br>
132         Este mensaje se envió de forma automática, no responder al ↵
133             ↪ mismo.
134         """
135
136
137         email = """From: %s

```

```

118     To: %s
119     MIME-Version: 1.0
120     Content-type: text/html
121     Subject: %s
122
123     %s
124     """ % (remitente, destinatario, asunto, mensaje)
125     try:
126         smtp = smtplib.SMTP('localhost')
127         smtp.sendmail(remitente, destinatario, email)
128         print ("Correo enviado")
129     except:
130         print ("Error: no se pudo enviar el mensaje")
131
132     def move_file_proceded():
133         fecha_actual = datetime.now().strftime("%Y-%m-%d")
134         nuevo_nombre = ↵
135             f"{INPUT_LOG_FILE.split('.')[0]}_{fecha_actual}.csv"
136         ruta_procesados = "procesados"
137         os.rename(INPUT_LOG_FILE, nuevo_nombre)
138         shutil.move(nuevo_nombre, os.path.join(ruta_procesados, ↵
139             nuevo_nombre))
140
141     if __name__ == "__main__":
142         mine_postal_data()

```

5.6.6. API de administración de desvíos

API de backend de resultados

Front end (no se desarrolla)

5.6.7. Visualizaciones

Tablero de grafana

5.7. TESTEO

unit test

integration test (no en prototipo)

regression test (no en prototipo)

end to end test (no en prototipo)

stress test (no en prototipo)

5.8. IMPLEMENTACIÓN

La implementación en el ambiente productivo se realizará luego del desarrollo del aplicativo final.

Este trabajo ha culminado con la creación de un prototipo completamente funcional que se enfoca en la aplicación de la minería de procesos en la distribución postal. Este prototipo se presenta como

una sólida base sobre la cual se podrá desarrollar la aplicación final para su implementación en el ambiente productivo. El desarrollo del aplicativo final se llevará a cabo siguiendo estrictas buenas prácticas de programación, tales como el uso de Clean Code y la metodología Git Flow, que han sido implementadas y seguidas durante la creación del prototipo.

Asimismo, el uso de contenedores ha sido considerado como una herramienta clave en la integración de los diferentes componentes del sistema, permitiendo su fácil despliegue en un ambiente de Integración y Despliegue Continuo (CI/CD) productivo.

El presente prototipo ha sido desarrollado con el objetivo de poner en práctica los conocimientos y herramientas de la minería de procesos, para optimizar los procesos de distribución postal. Una vez finalizado el aplicativo final, se espera que este pueda ser integrado en el ambiente productivo, brindando una mejora significativa en la eficiencia y eficacia en el proceso de distribución postal.

CAPÍTULO 6

Conclusiones

Conclusiones y futuras líneas de investigación, ver si se unifica con capítulo resultados)

En este capítulo se realizará un juicio crítico y discusión sobre los resultados obtenidos. *Cuidado, esta discusión no debe confundirse con una valoración del enriquecimiento personal que supone la realización del trabajo como culminación de una etapa académica.* Aunque de gran importancia, esta última valoración debe quedar fuera de la memoria del trabajo y solo debe ahondarse en ella ante requerimiento explícito del comité en el acto de defensa.

Si es pertinente deberá incluir información sobre trabajos derivados como publicaciones o ponencias en preparación, así como trabajos futuros (*solo si estos están iniciados o planificados en el momento que se redacta el texto*). Evitar hacer una lista de posibles mejoras. Contrariamente a lo que alguno pueda pensar generalmente aportan impresión de trabajo incompleto o inacabado.¹

6.1. JUSTIFICACIÓN DE COMPETENCIAS ADQUIRIDAS

Es muy importante recordar que según la normativa vigente en la ESI, el capítulo de conclusiones debe incluir *obligatoriamente* un apartado destinado a justificar la aplicación en el TFG de competencias específicas (una o más) adquiridas en la tecnología específica cursada.

En el TFG se han aplicado las competencias correspondientes a la Tecnología Específica de [*poner lo que corresponda*]:

Código de la competencia 1: [*Texto de la competencia 1*]. Explicación de cómo se ha aplicado en el TFG.

... otras más si las hubiera.

¹Puede reflexionarse en ello por si en la defensa del trabajo se pregunta sobre estas posibles mejoras.

CAPÍTULO 7

Resultados

Revisar si da para que sea un capítulo a parte o si se fusiona con conclusiones)

En esta sección se describirá la aplicación del método de trabajo presentado en el capítulo

Cuando se tiene en cuenta la puesta en marcha de un proyecto de ingeniería, la planificación y presupuesto que se realizan de modo previo a su ejecución son críticos para gestionar los recursos que permitan alcanzar los objetivos de calidad, temporales y económicos previstos para el proyecto.

Es muy importante que todas las justificaciones aportadas se sustenten no solo en juicios de valor sino en evidencias tangibles como: historiales de actividad, repositorios de código y documentación, porciones de código, trazas de ejecución, capturas de pantalla, demos, etc.

$$a^2 = b^2 + c^2 \quad (7.1)$$

Algoritmos y listados de código fuente

En los textos científicos relacionados con las TIC¹ (Tecnologías de la Información y Comunicaciones) suelen aparecer porciones de código en los que se explica alguna función o característica relevante del trabajo que se expone. Muchas veces lo que se quiere ilustrar es un algoritmo o método en que se ha resuelto un problema abstrayéndose del lenguaje de programación concreto en que se realiza la implementación. El paquete `algorithm2e` proporciona un entorno `algorithm` para la impresión apropiada de algoritmos tratándolos como objetos flotantes y con mucha flexibilidad de personalización.

Algoritmo 7.1: Cómo escribir algoritmos

```

Datos : este texto
Resultado: como escribir algoritmos con LATEX2e
1 inicialización;
2 while no es el fin del documento do
3   leer actual;
4   if comprendido then
5     ir a la siguiente sección;
6     la sección actual es esta;
7   else
8     ir al principio de la sección actual;
9   end
10 end
```

La inclusión de porciones de código fuente se puede formatear de modo sencillo en L^AT_EX mediante el uso del paquete ‘listings’. A continuación, se muestran varios ejemplos.

Listado 7.1: Ejemplo de código fuente en lenguaje Java

```

1 // @author www.javadb.com
2 public class Main {
3 // Este método convierte un String a un vector de bytes
4
5 public void convertStringToByteArray() {
6
7 String stringToConvert = "This\u00a0String\u00a0is\u00a015";
8 byte[] theByteArray = stringToConvert.getBytes();
9 System.out.println(theByteArray.length);
10 }
11
12 public static void main(String[] args) {
13 new Main().convertStringToByteArray();
14 }
15 }
```

Listado 7.2: Ejemplo de código C

```

1 // Este código se ha incluido tal cual está en el fichero LATEX
2 #include <stdio.h>
```

¹Por supuesto en un TFG (Trabajo Fin de Grado) o tesis de un centro superior de informática.

```
3
4 int main(int argc, char* argv[]) {
5     puts("¡Hola mundo!");
6 }
```

Listado 7.3: Ejemplo escrito en Matlab

```
1 function f = fibonacci(n)
2 % FIBONACCI Fibonacci sequence
3 % f = FIBONACCI(n) generates the first n Fibonacci numbers.
4 % Copyright 2014 Cleve Moler
5 % Copyright 2014 The MathWorks, Inc.
6 f = zeros(n,1);
7 f(1) = 1;
8 f(2) = 2;
9 for k = 3:n
10 f(k) = f(k-1) + f(k-2);
11 end
```


Bibliografía

- [1] Wil van der Aalst. *Process Mining: Data Science in Action*. Springer, 2nd. edition, 2016.
- [2] Jiawei Han; Micheline Kamber; Jian Pei. *Data Mining Concepts and Techniques*. Morgan Kaufmann, 3rd edition, 2012.
- [3] Ian H. Witten; Eibe Frank; Mark A. Hall; Christopher J. Pal. *Data Mining Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 4th edition, 2017.
- [4] Wil van der Aalst. The process mining manifesto by the ieee task force. *IEEE Task Force*, 2012.
- [5] Victor Martinez; Laura Lanzarini; Franco Ronchetti. Minería de procesos aplicada a la distribución postal. *XXVII Congreso Argentino de Ciencias de la Computación (CACIC)*, pages 271–280, 2021.
- [6] Victor Martinez; Laura Lanzarini; Franco Ronchetti. Distribution analysis of postal mail in argentina using process mining. *Communications in Computer and Information Science*. Springer, 1584:159–169, 2022.
- [7] IEEE Computer Society. *Guide to the Software Engineering*. IEEE, v3 edition, 2014.
- [8] Michael Rosemann Jan vom Brocke. *Handbook on Business Process Management 1*. Springer, 1st edition, 2015.
- [9] Inc Object Management Group. Business process model and notation. *Object Management Group*, 2010.
- [10] W. van der Aalst; T. Weijters; L. Maruster. Workflow mining: discovering process models from event logs. *IEEE*, 2004.
- [11] Wil Van der Aalst A. J. M. M. Weijters. Process mining with the heuristicsminer algorithm. *ResearchGate*, 2006.
- [12] Wil M. P. van der Aalst Christian W. Günther. Fuzzy mining – adaptive process simplification based on multi-perspective metrics. *Springer*, 2006.
- [13] Sander J.J. Leemans. inductive visual miner. *leemans*, 2017.
- [14] Christian Walter Gunther. Process mining in flexible environments. *Technische Universiteit Eindhoven*, 2004.
- [15] Xia Hu1; Yanfeng Jin1; Fan Wang. Research of postal data mining system based on big data. *3rd International Conference on Mechatronics, Robotics and Automation*, 2015.
- [16] Mitchell M. Tseng; Hung-Yin Tsai; Yue Wang. Context aware process mining in logistics. *The 50th CIRP Conference on Manufacturing Systems*, 2017.
- [17] Vianca Vega Zepeda Brian Keith Norambuena. Process mining applications in software engineering. *risti - Revista Ibérica de Sistemas y Tecnologías de Información*, 2017.
- [18] Patrick SIEGFRIED Teresa REIL, Erich GROHER. Process mining in supply chain management. *Supply Chain Management Journal*, 2022.

- [19] Dennis Maneschijn. Finding the appropriate level of abstraction for process mining in logistics. *University of Twente*, 2021.
- [20] Maikel L. van Eck; Xixi Lu; Sander J.J. Leemans; Wil M.P. van der Aalst. Pm2: a process mining project methodology. *Eindhoven University of Technology, The Netherlands*, 2017.
- [21] IEEE Std 1849-2016. Ieee standard for extensible event stream (xes) for achieving interoperability in event logs and event streams. *IEEE*, 2016.
- [22] Wil van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 1st edition, 2011.
- [23] Robert C. Martin. *Clean Code: A Handbook of Agile Software*. Pearson Educatio, 1st edition, 2008.
- [24] Kent Beck Martin Fowler. *Refactoring: Improving the Design of Existing Code*. Addison Wesley, 1st edition, 2018.

ANEXOS

ANEXO A

Publicación en CACIC 2021

Se incluye a continuación el Paper *Minería de Procesos aplicada a la Distribución Postal* presentado en CACIC 2021 sobre el tema.

Minería de Procesos aplicada a la distribución postal

Victor Martinez¹, Laura Lanzarini^{1,2}, and Franco Ronchetti^{1,2,3}

¹ Facultad de Informática, Universidad Nacional de La Plata

² Instituto de Investigación en Informática LIDI (UNLP-CIC)

³ Comisión de Investigaciones Científicas de la provincia de Buenos Aires
(CIC-PBA)

martinezvictor@hotmail.com, {laural, fronchetti}@lidi.info.unlp.edu.ar

Resumen La minería de procesos es una técnica que permite analizar procesos de negocio a través de registros de eventos. En este artículo se utilizan distintas técnicas de Minería de Procesos para analizar datos basados en la distribución postal de productos en la República Argentina durante los años 2017 a 2019. Los resultados obtenidos permiten afirmar que el 85 % de los envíos realizados se ajustan exactamente al modelo. El análisis de las situaciones con bajo nivel de ajuste al proceso descubierto constituyó una herramienta de identificación rápida de algunos problemas recurrentes en la distribución facilitando el análisis de los desvíos ocurridos. A futuro se espera poder incorporar estas técnicas en la construcción de alertas tempranas que adviertan de la existencia de desvíos excesivos del proceso.

Keywords: Process Mining, Data Mining, Postal Distribution, Postal Processes, Business Process Management

1. Introduction

En la actualidad la mayoría de los sistemas de información dejan registro de las actividades realizadas, ya sean sistemas de gestión empresarial (ERP, CRM, WMS, BI, etc) como también desarrollos propios de cada empresa. Por lo general se registra qué se realizó, quién lo realizó y cuándo se realizó, entre otras cosas. A partir de estos datos almacenados puede obtenerse información que describa el proceso y buscar mejoras o solucionar problemas.

Las técnicas de Minería de Procesos se utilizan para analizar estos datos y encontrar patrones de comportamiento. Se puede encontrar con que tareas comienza un proceso, cual es la secuencia que se cumple, y con que tareas termina. De esta forma se puede “descubrir” el proceso que se está realizando para una determinada actividad.

Con la minería de procesos se pueden responder preguntas como: ¿qué fue lo que pasó realmente?, ¿por qué pasó?, ¿qué podría pasar en el futuro?, ¿cómo se puede mejorar el control del proceso?, ¿cómo se puede rediseñar el proceso para mejorar la performance? [1].

Una de sus principales ventajas es que permite trabajar directamente sobre datos reales y obtener el verdadero comportamiento del proceso, el cual, en algunos casos, no es el diseñado originalmente.

En este trabajo se aplicarán las técnicas de minería de procesos a la distribución postal en la República Argentina para analizar su funcionamiento con el objetivo de encontrar desvíos operativos, cuellos de botella y otros problemas que impactan negativamente en la calidad del servicio.

En la distribución postal se lleva un registro de todas las actividades, desde el ingreso del producto hasta su entrega al cliente. Entre estas actividades se encuentra la recepción del envío, ingreso a un almacén, traslados internos o intentos de entrega, entre otras. Todas deben realizarse en un orden y en un intervalo de tiempo específicos. Ocasionalmente se producen desvíos, estos pueden ser redundancia o inconsistencia de tareas (se repiten o no se realizan en el orden correspondiente), excesivo tiempo para su finalización u otras.

En la literatura se pueden encontrar trabajos que relacionan el negocio postal, la inteligencia de datos y la minería de procesos. Tal es el caso de [2] que utiliza data mining en un entorno de Big Data en el correo de China. Debido a la problemática del negocio postal, en dicho artículo se utilizaron técnicas de clustering para agrupar a los clientes según el comportamiento, los hábitos de consumo y el foco de interés logrando generar una estrategia de Marketing postal más precisa y efectiva con resultados muy satisfactorios. Otro caso es [3] donde se aplica process mining en logística buscando similitudes y diferencias entre distintos procesos de entrega en un contexto cambiante de manufactura y logística. En dicho trabajo se comparan diferentes procesos utilizando técnicas de agrupamiento en busca de lograr una documentación automatizada de procesos en un contexto cambiante. En [4] se presenta una metodología que sirve de guía para la ejecución de proyectos de process mining que describe las diferentes etapas. Además se muestra como caso de estudio su aplicación práctica en el proceso de compras de IBM.

En este artículo se utilizarán distintas técnicas de Minería de Procesos para analizar datos basados en la distribución postal de productos en la República Argentina durante los años 2017 a 2019. Los autores de este trabajo desean manifestar que, a la fecha de generación del presente documento, no tienen conocimiento de la existencia de trabajos similares que implementen la minería de procesos en la distribución postal en la República Argentina.

2. Minería de procesos (Process Mining)

El punto de partida de la Minería de Procesos es el registro de eventos. Se asume que el proceso a analizar requiere del registro de una serie de eventos secuenciales pertenecientes a la actividad y que se relacionan con un caso particular. Si bien se pueden almacenar datos adicionales de cada evento deben registrarse obligatoriamente la fecha del evento (día y hora), el identificador del caso y el tipo de evento.

Los tres tipos básicos de minería de procesos [5] son:

- *Descubrimiento:* las técnicas de descubrimiento toman como punto de entrada un log de procesos y generan un modelo sin ninguna información adicional. Un ejemplo es el algoritmo Alpha [6] que toma los datos del log y genera una red de Petri que explica el comportamiento reflejado en el log. Como toda técnica que extrae conocimiento a partir de los datos, la calidad del proceso descubierto dependerá del grado de representación de los eventos relevados con respecto al funcionamiento de dicho proceso. No se pueden descubrirse partes del proceso que no se encuentren representadas por los eventos.
- *Verificación de Conformidad:* Consiste en realizar la comparación de un modelo existente (puede ser el descubierto anteriormente u otro) con la secuencia de eventos real para encontrar desvíos y verificar el funcionamiento. Suelen utilizarse aplicaciones capaces de generar representaciones gráficas y animaciones que posibilitan observar el comportamiento real y ver cuánto se ajusta la realidad al proceso definido originalmente. Su principal ventaja es que muestra la realidad, no es una simulación, por lo que permite realizar un análisis mucho más preciso.
- *Mejoramiento:* Se busca extender o mejorar el proceso existente a través de la información subyacente en la secuencia de eventos. A diferencia de la Verificación de Conformidad, donde se comparan los datos con el modelo, aquí se busca modificar el proceso

Se debe regular el grado de abstracción a utilizar durante el análisis considerando los siguientes aspectos: ajuste, precisión, generalización y simplicidad [1]. Existe una relación de compromiso entre ellos que debe ser tenida en cuenta si se espera obtener buenos resultados. El ajuste se refiere a la habilidad del modelo para explicar el comportamiento observado. La precisión se refiere a la exactitud con la que se ejecuta el proceso; aquí es importante no sobreajustar el modelo a los datos de entrada porque se perderá generalización impidiendo lograr el nivel de abstracción deseado. La simplicidad también se ve comprometida por el sobreajuste ya que se logra agregando mayor detalle a la descripción del proceso.

Finalmente es importante destacar que el resultado obtenido del análisis con minería de procesos está altamente vinculado a la calidad de los datos de entrada. De hecho se sabe que siempre hay una cantidad de ruido en los datos el cual puede deberse a trazas incompletas, intervalos que no se hayan registrado correctamente o duplicidad de datos. Esta información puede distorsionar o falsificar el resultado del análisis [7].

Generalmente es necesario verificar los datos de entrada y realizar un procesamiento previo para eliminar la mayor cantidad de ruido posible.

3. Descubrimiento del proceso de Distribución Postal

En esta sección se llevará a cabo el descubrimiento del proceso a partir de datos basados en la distribución postal de productos en la República Argentina durante los años 2017 a 2019.

El proceso de distribución postal abarca diferentes tipos de productos; en todos los casos el proceso consiste de una recepción del producto del remitente (ya sea por un medio físico o un canal digital) y una entrega al destinatario, pudiendo esta ser exitosa o no. La no entrega del producto no implica un error de proceso ya que puede haber motivos que lo justifiquen.

El proceso registra al menos las siguientes actividades: recepción e identificación, envío para su distribución, uno o más intentos de entrega, espera en centro de distribución y devolución (si no fue posible la entrega). En todos los casos se debe dejar registro de cada uno de los pasos realizados asociados a un identificador único de envío, lo que permite conocer el estado actual del mismo e informarlo al cliente.

Extracción de los datos

El primer paso en el descubrimiento del modelo consiste en la recolección y preprocesamiento de los datos a utilizar. En particular, para este caso de estudio se utilizaron envíos de productos con dos intentos de entrega en el domicilio. El procedimiento actual establece que aquellos productos que no pueden ser entregados se guardan un tiempo a la espera de que el destinatario venga a retirarlo; transcurrido ese tiempo se realiza la devolución al remitente.

Se define una traza como un envío. Cada movimiento que se registre de ese envío es un evento. Se considera una traza completa cuando el envío tiene registrado su ingreso y su fin, con entrega exitosa o no.

Como resultado de la recolección de datos se generó una tabla o muestra de alrededor de 33.000 trazas y un total de 78.000 eventos. Para cada evento se registraron los campos mínimos necesarios para el análisis; estos son: identificador de traza, identificador de evento, fecha de evento, descripción del evento (ver Figura 1). Cada traza puede tener uno o más eventos asociados, se necesita identificador de traza e identificador de cada uno de los eventos para armar la historia de la traza.

De esta forma se dispone de los distintos pasos o eventos que se produjeron a lo largo del envío de cada producto.

Luego se transformaron los datos utilizando el formato XES [8]. XES es una gramática para un lenguaje basado en etiquetas cuyo objetivo es proporcionar a los diseñadores de sistemas de información una metodología unificada y extensible para capturar los comportamientos de los sistemas mediante registros y flujos de eventos [8]. De esta forma se agiliza el manejo de los datos pudiendo ser procesados por distintas herramientas de manera más eficiente.

A fin de facilitar el descubrimiento del proceso correcto, se eliminaron todas las trazas incompletas entendiendo por tales a aquellas trazas que no tengan o bien un evento de inicio (Recepción) por algún error de carga o bien un evento final (entrega, devolución, destinatario inexistente, fallecido, rechaza el envío o se mudó). Esto último podría ocurrir por error de carga o porque todavía no transcurrió el tiempo estipulado para finalizar el proceso.

Identificador de traza	Identificador de evento		Descripción del evento		Fecha evento
	123 trazalD	123 EveID	abc eveDescrip	eveFecha	
31	481,053	0	INGRESADO	2017-08-16 10:45:22	
32	481,053	2	1 INTENTO DE ENTREGA	2017-08-18 11:15:00	
33	481,053	9	DEVOLUCION	2017-08-18 13:00:00	
34	481,054	0	INGRESADO	2017-08-16 10:45:28	
35	481,054	2	1 INTENTO DE ENTREGA	2017-08-22 12:05:00	
36	481,054	9	DEVOLUCION	2017-08-22 17:25:00	
37	481,055	0	INGRESADO	2017-08-16 10:45:27	
38	481,055	1	ENTREGADO	2017-08-22 15:13:00	

Figura 1. Ejemplo de eventos extraídos para el análisis

Utilizando un filtro de reglas heurísticas simples se identificaron las trazas completas dejando sólo aquellas que tienen su estado inicial y los estados finales válidos.

Como resultado de este proceso de filtrado se obtuvieron aproximadamente 16.000 trazas con 43.000 eventos.

3.1. Modelo del proceso

Para generar el modelo del proceso se utilizó un algoritmo clásico de minería de procesos, el algoritmo Alpha, propuesto por primera vez por van der Aalst, Weijters y Mărușter[6]. El objetivo de este algoritmo es reconstruir la causalidad a partir de un conjunto de secuencias de eventos. Construye Redes de Petri con propiedades especiales (redes de flujo de trabajo) a partir de registros de eventos (como los que podría recoger un sistema ERP). Cada transición en la red corresponde a una tarea observada.

En lo que se refiere al caso de estudio de la distribución postal, si bien es posible construir el modelo directamente a partir de las 16000 trazas mencionadas previamente, se seleccionaron manualmente algunos casos buscando los más representativos para encontrar el proceso ideal que deben cumplir las trazas, simplificando de esta manera esta etapa de descubrimiento del proceso.

La Figura 2 ilustra la red de petri correspondiente al proceso descubierto que deberán cumplir las trazas.

El proceso encontrado es el siguiente: Todas las trazas deben comenzar con un evento de ingreso y luego sale a distribución. Si se puede entregar, se registra el evento y finaliza el proceso. Si no se puede entregar, en caso de ser un evento final (fallecido, datos faltantes en dirección, no existe dirección, desconocido, etc.), se registra el motivo y finaliza el proceso. Si no es un evento final (por ejemplo se pudo llegar a la dirección pero no había nadie) se registra un primer intento de entrega. Posteriormente se realiza una nueva visita. Si se pudo entregar, se registra el evento y finaliza el proceso. Si en la segunda visita no

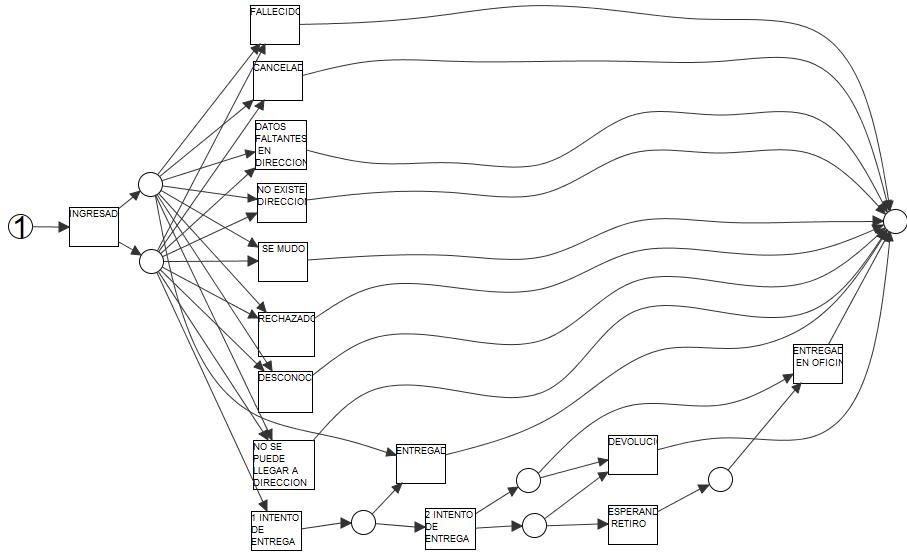


Figura 2. Proceso descubierto que deberán cumplir las trazas

se puede entregar, se guarda el envío en oficina por un tiempo esperando que el destinatario venga a retirarlo y luego se registra la entrega en oficina. Transcurrido el tiempo de espera, si el destinatario no vino a retirar el envío, se registra la devolución. Esos son los posibles escenarios que contempla el proceso.

3.2. Verificación del modelo

El modelo de proceso descubierto en la etapa anterior fue generado a partir de un subconjunto de trazas seleccionadas previamente. En esta sección se llevará a cabo la verificación de conformidad. Tal como se explicó previamente, se buscará determinar cuánto se ajustan las trazas de la muestra al proceso. Se tendrán en cuenta qué pasos se cumplen, cuáles no y si hay algunos adicionales que no están reflejados en el proceso descubierto. Para ello, se comparará cada una de las trazas obtenidas contra el proceso encontrado en el punto anterior. Se utilizará la muestra de datos obtenida en la sección anterior, la cual está conformada por 16.811 trazas con un total de 43.888 eventos.

El resultado se muestra en la Figura 3. Allí se resaltan los eventos más comunes en color oscuro y los recorridos más frecuentes con una línea más gruesa. Se observa que la mayoría de las trazas siguen el camino de la entrega, ya sea en el primer intento o en el segundo.

Las estadísticas de la muestra indican que en un 80 % de los casos la pieza es entregada, ya sea en el primer o segundo intento y que el resto se distribuye uniformemente. En base a estas observaciones se puede afirmar que el 85 % de las trazas se ajusta perfectamente al proceso descubierto con una cantidad promedio de 2.6 eventos en cada traza.

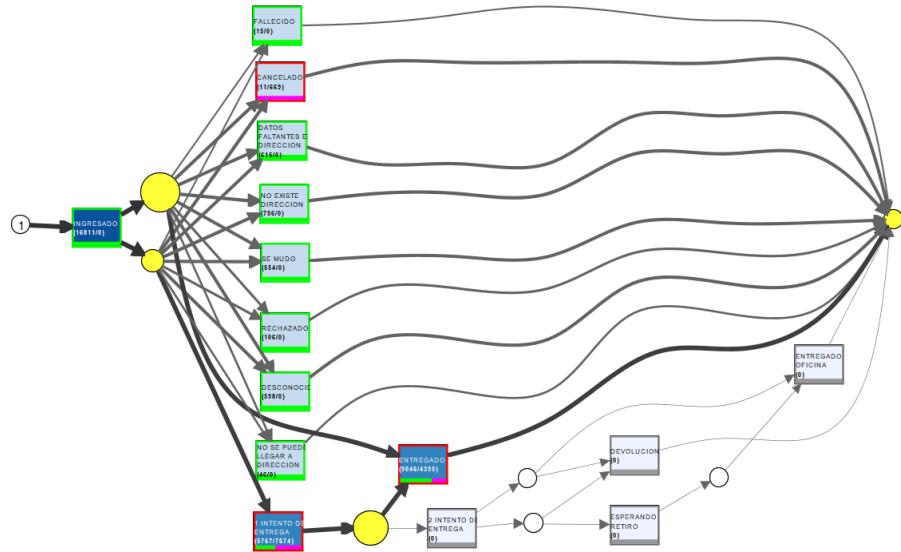


Figura 3. Recorrido de las trazas: eventos más comunes en oscuro y recorridos más frecuentes en línea más gruesa

Sin embargo, algunas trazas no verifican exactamente el proceso descubierto ya sea porque se saltean pasos, o no los hacen en el orden correspondiente o hacen pasos por fuera del modelo. Estas trazas poseerán un valor de ajuste que será más bajo cuanto más se diferencien del modelo. Aquellas que tengan un ajuste menor al 50 % son el objeto de análisis de esta sección ya que representan un gran desvío operativo.

A partir de estos resultados se enfoca el análisis en dos caminos, por un lado las trazas que tienen una excesiva cantidad de movimientos y por otro aquellas que no se ajustan al modelo, ya sea porque hay eventos faltantes o porque no cumplen la secuencia correspondiente.

En el primer caso, dado que la cantidad de movimientos por traza sigue una distribución normal con media 2,6 y desvío 0,95 se consideró adecuado tomar un valor superior a la media mas tres desvíos estándar como un valor representativo de una cantidad excesiva de movimientos.

En base a esto, se procedió a filtrar y posteriormente a analizar las trazas con más de 6 movimientos encontrando 108 casos.

El segundo caso analizado fue el de las trazas que no se ajustaban exactamente al proceso descubierto. En esta oportunidad se fijó el valor del umbral de ajuste en 0.5 por considerar que las trazas que contienen al menos el 50 % de sus movimientos desalineados (no se corresponden) con respecto a lo indicado por el modelo representan una fuerte distorsión del procedimiento y deben ser inspeccionadas. Como se dijo previamente, esto puede ocurrir o bien porque hay

3760173.0 8 events	3780088.0 6 events
INGRESADO #1 02.01.2020 16:29:45.000	INGRESADO #1 20.01.2020 21:00:14.000
1 INTENTO DE ENTREGA #2 13.01.2020 12:25:00.000	1 INTENTO DE ENTREGA #2 21.01.2020 11:12:00.000
2 INTENTO DE ENTREGA #3 14.01.2020 11:20:00.000	1 INTENTO DE ENTREGA #3 22.01.2020 11:16:00.000
2 INTENTO DE ENTREGA #4 15.01.2020 11:20:00.000	ESPERANDO RETIRO #4 22.01.2020 16:04:31.000
1 INTENTO DE ENTREGA #5 16.01.2020 11:20:00.000	DEVOLUCION #5 29.01.2020 15:02:07.000
2 INTENTO DE ENTREGA #6 17.01.2020 09:39:00.000	ENTREGADO #6 31.01.2020 11:15:00.000
2 INTENTO DE ENTREGA #7 20.01.2020 10:10:00.000	
DEVOLUCION #8 24.01.2020 10:10:00.000	

Figura 4. Eventos repetidos e inconsistencias

eventos en la trazas que no aparecen en el modelo o porque los eventos de la traza no cumplen con la secuencia correspondiente.

Como resultado de esto se observó que se trataban de repeticiones de eventos en diferentes días o de registros inconsistentes. La Figura 4 ilustra ambas situaciones. En el cuadro de la izquierda se puede observar un caso en el que aparece registrada una primera visita luego de haber ocurrido previamente una segunda; esta situación sólo puede deberse a un error en el registro. En el cuadro de la derecha de la misma figura se observa una inconsistencia ya que no es posible entregar un producto que previamente fue devuelto.

También es posible utilizar herramientas visuales para generar animaciones que faciliten la comprensión de estas situaciones. En este caso particular, con el conjunto de trazas que tienen un ajuste inferior al 50 % se utilizó Inductive Visual Miner[9] para generar una animación que permite ver cómo se suceden los eventos de cada traza en una línea de tiempo. La Figura 5 ilustra esta animación. En dicha figura cada traza está simbolizada con un token (o círculo) que va recorriendo las diferentes etapas del proceso. Esta representación visual permite observar que unas trazas demoran más que otras y que en ocasiones se presentan algunos comportamientos erróneos. Por ejemplo en la figura 5 se observan movimientos de retroceso en las trazas; en particular, se han marcado con círculos que 25 trazas vuelven del segundo intento al primero y 20 vuelven sobre sí mismas.



Figura 5. Trazas que no cumplen con el modelo, en círculo la cantidad que retroceden en de un estado en lugar de avanzar (ver sentido de la flecha)

La exportación de las trazas que no se ajustan al modelo permite un análisis detallado de los motivos que ocasionan los desvíos en busca de mejoras al proceso de distribución.

4. Conclusión y futuras líneas de investigación

En este artículo se han utilizado distintas técnicas de Minería de Procesos para analizar datos reales de distribución postal de productos en la República Argentina durante los años 2017 a 2019. No se encontraron a la fecha de generación del presente documento trabajos similares que implementen la minería de procesos en la distribución postal en la República Argentina.

Por medio del modelo generado a partir de trazas seleccionadas se descubrió el proceso que realmente se lleva a cabo. Es importante remarcar el impacto que el criterio de selección de dichas trazas tiene en el modelo obtenido. Los primeros modelos, generados a partir de la totalidad de las trazas, tenían un detalle excesivo de la situación lo cual dificultaba su representación e interpretación posteriores. Actualmente se está trabajando para poder identificar las trazas más frecuentes a partir del modelo inicial para luego filtrar en forma automática las más representativas.

Por su parte, la verificación de conformidad ha permitido determinar situaciones anómalas de interés. Se identificaron tanto los casos que no cumplieron con el modelo como también aquellos que lo hicieron pero fuera de los tiempos esperados o con redundancia de tareas. Se deberá analizar si estos casos representan errores manuales en la carga y buscar una solución.

Se utilizó un umbral de ajuste para establecer el grado de distorsión mínimo que una traza debe cumplir para no afectar el desarrollo del proceso. Este factor debe ser analizado con más detalle para determinar su valor en base al caso de estudio del cual se trate.

Como línea de trabajo futura se propone continuar trabajando con las técnicas de minería de procesos no sólo para modelar situaciones que ya han ocurrido

sino para poder insertar en el sistema alertas tempranas que adviertan de la existencia de desvíos excesivos del modelo.

Referencias

1. Process Mining: Data Science in Action, Wil van der Aalst,978-3-662-49850-7,2016,Springer
2. Research of Postal Data mining system based on big data, Xia Hu1; Yanfeng Jin1; Fan Wang, 3rd International Conference on Mechatronics, Robotics and Automation, 2015, 10.2991/icmra-15.2015.124, https://www.researchgate.net/publication/300483008_Research_of_Postal_Data_mining_system_based_on_big_data
3. Context Aware Process Mining in Logistics, Mitchell M. Tseng; Hung-Yin Tsai; Yue Wang, 2017,The 50th CIRP Conference on Manufacturing Systems, <https://www.sciencedirect.com/science/article/pii/S2212827117303311>
4. PM2: a Process Mining Project Methodology,Maikel L. van Eck; Xixi Lu; Sander J.J. Leemans; Wil M.P. van der Aalst,Eindhoven University of Technology, The Netherlands, http://www.processmining.org/_media/blogs/pub2015/pm2_processminingprojectmethodology.pdf
5. Wil van der Aalst, The Process Mining Manifesto by the IEEE Task Force, 2012, <https://www.tf-pm.org/resources/manifesto>
6. Workflow mining: discovering process models from event logs, W. van der Aalst; T. Weijters; L. Maruster, 1041-4347, 2004, IEEE, <https://ieeexplore.ieee.org/document/1316839>
7. Process mining in flexible environments, Christian Walter Gunther,978-90-386-1964-4,2009,Technische Universiteit Eindhoven, <https://research.tue.nl/en/publications/process-mining-in-flexible-environments>
8. IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams, IEEE Std 1849-2016, 2016, DOI 10.1109/IEEEESTD.2016.7740858
9. inductive visual miner, Sander J.J. Leemans, 2017, <http://leemans.ch/leemansCH/publications/ivm.pdf>
10. Reinventing the Postal Sector in an Electronic Age, Michael A. Crew; Paul R. Kleindorfer,978-1849803601,2011,Edward Elgar Publishing

ANEXO B

Publicación en Springer 2022

Se incluye a continuación el Paper *Distribution Analysis of Postal Mail in Argentina Using Process Mining* publicado en Springer Computer Science – CACIC 2021 en mayo de 2022 sobre el tema.

Análisis de la distribución postal en Argentina utilizando Minería de Procesos

Victor Martinez¹[0000-0003-3581-5704], Laura Lanzarini²[0000-0001-7027-7564],
and Franco Ronchetti^{2,3}[0000-0003-3173-1327]

¹ Facultad de Informática, Universidad Nacional de La Plata

² Instituto de Investigación en Informática LIDI (UNLP-CIC)

³ Comisión de Investigaciones Científicas de la provincia de Buenos Aires (CIC-PBA)

Resumen La minería de procesos combina una serie de técnicas que permiten analizar procesos de negocio únicamente con registros de eventos. Este artículo continúa la investigación realizada en [1] para analizar datos basados en la distribución postal de productos en la República Argentina entre los años 2017 a 2020. Los resultados obtenidos inicialmente dejaron en evidencia que el 85 % de los envíos realizados cumplen el proceso correctamente. También se identificaron rápidamente los casos que no cumplían el modelo y se encontraron problemas recurrentes que facilitan el análisis para una mejora del proceso. Los problemas más comunes fueron trazas que no respetan el orden de tareas, exceso de movimientos o movimientos faltantes y trazas que cumplen el proceso pero demoran demasiado tiempo. En este artículo se incorporó un análisis de performance por medio del cual se descubrieron trazas que, si bien se ajustan correctamente al proceso, tienen desvíos operativos debido al excesivo tiempo de finalización. Se espera poder incorporar estas técnicas en el proceso mediante la construcción de alertas tempranas que adviertan de la existencia de este tipo de situaciones para mejorar la calidad del servicio.

Keywords: Process Mining · Data Mining · Postal Distribution · Postal Processes · Business Process Management

1. Introduction

Actualmente existe una amplia variedad de sistemas de información que dan soporte a las diferentes áreas de cada empresa ya sean propias del negocio o administrativas. Entre estas herramientas pueden encontrarse desarrollos ad hoc para cada compañía o herramientas de gestión como CRM, ERP, WMS, TMS, BI, etc.

Dada la facilidad y el bajo costo para almacenar datos en estos días, gran parte de estos sistemas dejan información para auditoría o resolución de errores en archivos de texto o bases de datos. Entre otras cosas, se registra qué se realizó, cuándo se realizó, quién lo hizo, si hubo algún error, etc. Analizando estos datos puede obtenerse información acerca de errores o problemas ocurridos, cual es el

proceso de negocio que se realiza y como proponer mejoras o encontrar soluciones a diversos inconvenientes operativos.

La minería de procesos proporciona técnicas que permiten analizar estos registros y obtener información.

La minería de procesos puede responder interrogantes cómo: ¿cómo se puede mejorar el control del proceso?, ¿qué ocurrió realmente?, ¿por qué ocurrió?, ¿qué podría suceder en el futuro?, ¿cómo se puede mejorar el proceso para incrementar la performance? [2].

Al trabajar directamente con datos productivos se obtiene el comportamiento real que se realiza para el negocio y el proceso completo que se está realizando, el cual en algunos casos puede diferir del que se diseñó originalmente.

El caso de estudio en este trabajo es la distribución de productos postales. Se trata de un proceso en el que se registran distintos movimientos conformados por eventos tales como la recepción del producto en sucursal, el ingreso a un centro de distribución, los movimientos internos y los diversos intentos de entrega. Todas las tareas están claramente definidas y deben realizarse en un orden e intervalo de tiempo determinado. En la operación diaria se producen desvíos por retrasos en la ejecución de tareas o inconsistencia (tareas repetidas o realizadas en un orden incorrecto)

Existen algunos trabajos que vinculan el negocio postal, la minería de procesos y la inteligencia de datos. Por ejemplo [3] donde se usa data mining para el correo de China en un entorno de Big Data. Dada la complejidad del negocio postal se utilizaron técnicas de clustering que agrupaban a los clientes según los hábitos de consumo, los principales intereses y el comportamiento para lograr una estrategia de Marketing más efectiva y precisa. Los resultados fueron muy satisfactorios. Otro caso es [4] aquí se aplica Process Mining en una cadena de logística y manufactura para buscar similitudes y diferencias de los diferentes procesos de entrega en un entorno cambiante. Para ello se comparan diferentes procesos utilizando técnicas de agrupamiento para automatizar la documentación de procesos. Finalmente en [5] se muestra una metodología que se puede utilizar como guía en proyectos de process mining y como ejemplo se da el caso de estudio de su aplicación en IBM.

Este trabajo continúa la investigación publicada en [1] referido a la aplicación de técnicas de Minería de Procesos a la distribución Postal en la República Argentina con el objetivo de analizar su funcionamiento, encontrar desvíos o inconvenientes en la distribución y proponer mejoras que mejoren la calidad del servicio. A lo allí reportado, en el presente trabajo se ha incorporado un análisis de performance por medio del cual es posible identificar un nuevo tipo desvíos operativos. Los datos analizados corresponden a la distribución postal efectuada en la República Argentina en el período comprendido entre 2017 y 2020. Cabe destacar que, a la fecha de publicación de este artículo los autores no conocen la existencia de trabajos con estas características para el negocio postal.

2. Minería de procesos (Process Mining)

La Minería de Procesos opera sobre los archivos de log del sistema de información a analizar. Comienza con la extracción de la información de todos los eventos de una determinada actividad y luego, mediante un análisis automático, se determina cual es el proceso de negocio, que actividades lo componen y que secuencia se debe cumplir. De esta forma es posible hacer un análisis y responder preguntas como ¿Qué fue lo que pasó?, ¿Por qué pasó?, ¿Qué es lo que podría pasar en el futuro?, ¿Cómo se puede mejorar el control?, ¿Cómo se puede mejorar la performance? [2] Con la información obtenida luego de la interpretación del modelo se podrán realizar diferentes tipos de análisis como por ejemplo: quién realiza la actividad, si la actividad está completa y si se realiza en un intervalo de tiempo coherente.

La minería de procesos puede clasificarse en tres tipos [6]:

- *Descubrimiento:* Se trata de técnicas capaces de modelizar el proceso de negocio que se está realizando únicamente a partir de la secuencia de logs correspondientes, estos logs pueden extraerse de archivos, bases de datos o algún otro medio. Posiblemente el algoritmo más popular para realizar esta tarea es el algoritmo Alpha [7] que genera un red de Petri con el proceso descubierto.

La calidad del modelo generado está relacionada en forma directamente proporcional a la calidad de los datos de entrada. Esta es una característica común a cualquier técnica de extracción de conocimiento inductiva. Es fundamental que las secuencias de eventos relevadas representen completamente al proceso a modelizar. De lo contrario, quedarán aspectos sin descubrir.

- *Verificación de Conformidad:* Permite contrastar una secuencia de eventos real contra un modelo de proceso (puede ser el descubierto anteriormente u otro) para determinar cuanto se cumple y cuáles son las ocurrencias que se desvían.

Pueden utilizarse aplicaciones que representan gráficamente los resultados y realizan animaciones para amenizar la presentación de los resultados. El análisis resultante de esta operatoria puede ser de una gran precisión ya que se trabaja con datos reales, no es una simulación.

- *Mejoramiento:* Se trata de mejorar el proceso existente a partir del análisis realizado. Se diferencia de la Verificación de Conformidad porque aquí el foco está puesto en el proceso y no en la ejecución de las trazas.

Existen adicionalmente cuatro aspectos a tener en cuenta para obtener resultados de calidad en la Minería de Procesos. Estos son: precisión, ajuste, generalización y simplicidad [2]. Debe mantenerse una relación balanceada entre estas cuatro fuerzas para que el modelo descubierto sea representativo del proceso y fácil de comprender.

Por último pero, no menos importante, cabe destacar que generalmente se encuentra una gran cantidad de ruido en los datos de entrada que puede deberse a duplicidad de datos o trazas incompletas que pueden distorsionar el resultado del análisis [8]. Como en otras técnicas de minería se suele realizar un procesamiento

previo para mejorar la calidad de los datos de entrada con el fin de eliminar datos incompletos o corruptos que puedan llevar a una modelización errónea.

3. Minería de Procesos en la Distribución Postal

Se comenzará con el descubrimiento del proceso a partir de datos de distribución postal en la República Argentina entre los años 2017 a 2020.

La distribución postal puede aplicarse a varios productos tales como cartas, telegramas o encomiendas o algo menos "tradicional" por ejemplo e-commerce.

En todos los casos durante el proceso se cumplen al menos los siguientes pasos: registro y recepción del producto a distribuir, encaminamiento interno por una o más plantas o centros de distribución, uno o más intentos de entrega, entrega efectiva o devolución si no se pudo entregar. Siempre queda un registro de cada paso realizado, dejando constancia de quién lo realizó, que realizó y cuando lo hizo. Este registro está asociado a un identificador único de envío que permite informar estos eventos al cliente.

3.1. Extracción de los datos

Para descubrir el modelo y luego analizar los eventos se realizará una extracción y procesamiento previo de los datos a utilizar. Para este caso de estudio en particular se utilizaron envíos de productos que requirieron dos intentos de entrega. El procedimiento en este caso es el siguiente: el producto sale a distribución, si no se pudo entregar por algún motivo que no sea definitivo se intenta una nueva entrega al día siguiente. Si tampoco se pudo entregar se aguarda en oficina un tiempo a que el destinatario venga a retirar (se deja un aviso en cada visita). Transcurrido el tiempo establecido si el cliente no retiró se devuelve el envío al remitente.

Se considera cada envío como una traza y cada movimiento como un evento. La traza se considerará completa cuando el envío tiene registrado el ingreso, y el fin, pudiendo ser la entrega exitosa o no.

Como resultado de la recolección de datos se generó una muestra que contiene alrededor de 33.000 trazas con más de 77.000 eventos (ver Figura 1). Cada traza tiene al menos dos eventos asociados: identificador de traza e identificador de cada evento. En particular, en todos los eventos del caso de estudio se registraron los campos necesarios para el análisis: identificador de traza, identificador de evento dentro de la traza, descripción, fecha de evento

Con esta información es posible reconstruir cada traza y armar el modelo correspondiente.

Luego, los datos se convierten a formato XES [9]. XES Propone un lenguaje basado en etiquetas que proporciona una metodología unificada y extensible para registrar comportamiento en los sistemas de información [9]. Los datos estructurados de esta forma se pueden procesar en una amplia variedad de herramientas de una forma eficiente.

Identificador de traza	Identificador de evento		Descripción del evento		Fecha evento
	123 trazalD	123 EveID	abc eveDescrip	eveFecha	
31	481,053	0	INGRESADO	2017-08-16 10:45:22	
32	481,053	2	1 INTENTO DE ENTREGA	2017-08-18 11:15:00	
33	481,053	9	DEVOLUCION	2017-08-18 13:00:00	
34	481,054	0	INGRESADO	2017-08-16 10:45:28	
35	481,054	2	1 INTENTO DE ENTREGA	2017-08-22 12:05:00	
36	481,054	9	DEVOLUCION	2017-08-22 17:25:00	
37	481,055	0	INGRESADO	2017-08-16 10:45:27	
38	481,055	1	ENTREGADO	2017-08-22 15:13:00	

Figura 1. Ejemplo de eventos extraídos para el análisis

Para tener una mayor consistencia en los datos y facilitar el análisis se eliminaron todas las trazas incompletas, ya sea porque no tienen un estado inicial (ingreso) o un estado final (entregado, devuelto, no existe dirección, falleció, o se mudó). Esto puede ocurrir por algún error en la carga de los datos o porque son trazas que todavía no completaron el proceso.

Para quitar las trazas incompletas se utilizó un filtro de reglas heurísticas simples donde se especificó cual es el estado inicial y cuales son los estados finales válidos. El filtro descarta todas las trazas que no cumplan con estos requisitos. Luego de este filtrado se obtiene una muestra de aproximadamente 16.000 trazas con 43.000 eventos.

3.2. Descubrimiento del proceso

Se eligió el algoritmo Alpha para realizar el descubrimiento del proceso. El algoritmo Alpha fue propuesto por van der Aalst, Weijters y Mărușter[7] y es ampliamente utilizado en Minería de Procesos. El algoritmo reconstruye causalidad a partir de una secuencia de eventos y como resultado devuelve una Red de Petri donde se refleja el proceso de negocio utilizado. Cada transición en la red representa una tarea.

Para este caso de estudio en particular se extrae una pequeña muestra representativa que se utilizará para descubrir el proceso, luego se contrastará este proceso descubierto con el resto de las trazas para observar el cumplimiento.

En la Figura 2 se muestra la red de petri descubierta que representa el proceso que deberán cumplir las trazas.

El proceso descubierto se conforma de la siguiente manera: siempre hay un evento de ingreso que deben cumplir todas las trazas luego del cual se realiza la distribución. Si se puede entregar se registra la entrega y finaliza la historia de la traza. Si no se puede entregar por un evento final (persona fallecida, no existe la dirección, no hay suficientes datos para encontrar la dirección, etc) se registra la no entrega con el motivo y finaliza el proceso. Si el motivo de no entrega no es

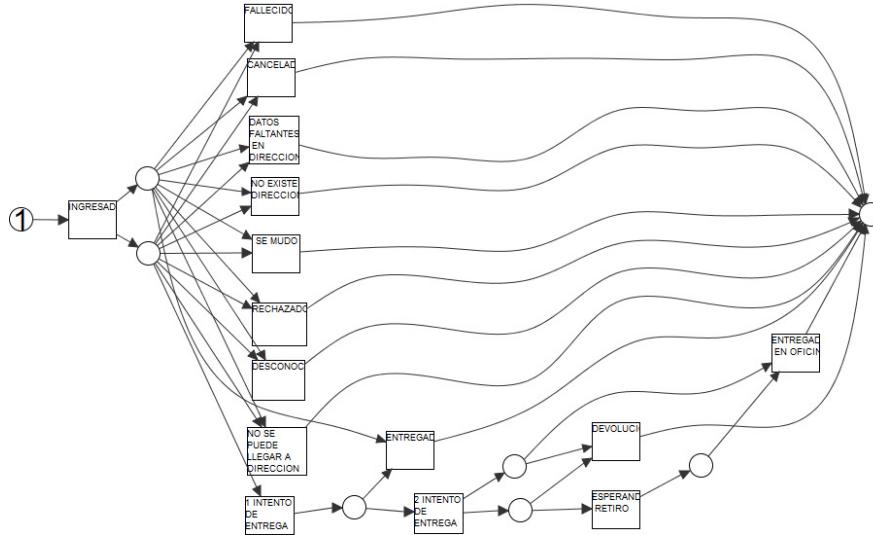


Figura 2. Proceso descubierto que deberán cumplir las trazas

un evento final (por ejemplo no había nadie para entregar) se registra un primer intento de entrega y se intenta nuevamente al día siguiente. Si al día siguiente tampoco se puede entregar se mantiene el envío en la oficina durante un tiempo esperando que el destinatario se presente a retirar (se deja un aviso en cada visita). Finalmente se registra el evento de entrega en la oficina o se devuelve el envío al remitente. La Figura 2 ilustra los escenarios que componen el proceso descubierto.

3.3. Verificación del modelo

Para llevar a cabo la verificación de conformidad se toman todas las trazas de la muestra (en total 16.811 trazas con 43.888 eventos) y se contrastan con el modelo descubierto.

Se busca determinar cuánto se ajustan las trazas al modelo y analizar aquellas que más se desvíen teniendo en cuenta qué pasos se cumplen, cuáles no y si hay pasos que no están reflejados en el proceso descubierto. El resultado se muestra en la Figura 3. Allí se muestran los eventos más comunes en color oscuro y los caminos más frecuentes con una línea más gruesa. Se puede ver que la mayoría de las trazas finalizan con una entrega, ya sea en el primer intento o en el segundo.

Analizando las estadísticas de la muestra se observa que en un 80 % de los casos la pieza es entregada, ya sea en el primer o segundo intento y que el resto se distribuye uniformemente. Según estas observaciones se puede afirmar que el 85 % de las trazas cumple por completo el proceso descubierto (ya sea entregado o no) con una media de 2.6 eventos en cada traza.

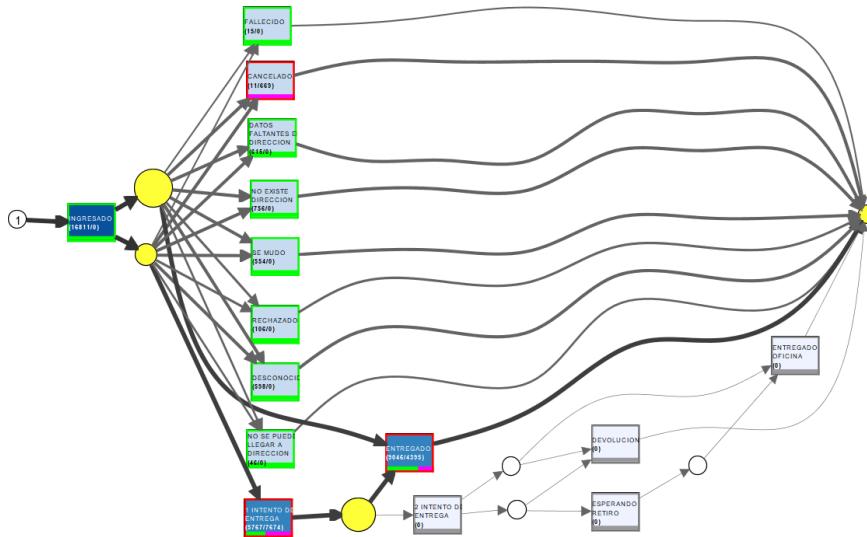


Figura 3. Recorrido de las trazas: eventos más comunes en oscuro y recorridos más frecuentes en línea más gruesa

Se observa también que muchas trazas no verifican el proceso exactamente ya sea porque no cumplen los pasos en el orden correcto o porque se saltean alguno. Estas trazas tienen un ajuste más bajo cuanto más se diferencien del modelo por lo tanto, se analizarán aquellas que tengan un ajuste menor al 50% ya que representan un gran desvío operativo.

Con estos resultados se abren dos caminos diferentes para analizar, por un lado las trazas que tienen demasiados movimientos y por otro aquellas que no realizan las tareas en el orden correspondiente o tienen eventos faltantes.

Adicionalmente se analizarán aquellas trazas que cumplen el proceso correctamente pero en un intervalo de tiempo excesivamente alto.

Se verifica que la cantidad de movimientos por traza sigue una distribución normal con media 2,6 y desvío 0,95, por lo tanto se decide tomar un valor superior a la media mas tres desvíos estándar como un valor representativo de una cantidad excesiva de movimientos.

En base a esto, para el primer caso se procedió a filtrar y exportar las trazas con más de 6 movimientos para realizar un análisis posterior. Se encontraron 108 casos con estas características.

En el segundo caso se analizaron las trazas que no se ajustan exactamente al proceso descubierto. Para ello se estableció el valor de ajuste en 0.5 al considerar que las trazas tienen al menos el 50% de sus movimientos fuera del modelo representando un gran desvío al procedimiento. Como se mencionó previamente esto puede ocurrir porque hay eventos de más en las trazas (no existen en el modelo) o no cumplen correctamente la secuencia.

<p>3760173.0 8 events</p> <p>INGRESADO #1 02.01.2020 16:29:45.000</p> <p>1 INTENTO DE ENTREGA #2 13.01.2020 12:25:00.000</p> <p>2 INTENTO DE ENTREGA #3 14.01.2020 11:20:00.000</p> <p>2 INTENTO DE ENTREGA #4 15.01.2020 11:20:00.000</p> <p>1 INTENTO DE ENTREGA #5 16.01.2020 11:20:00.000</p> <p>2 INTENTO DE ENTREGA #6 17.01.2020 09:39:00.000</p> <p>2 INTENTO DE ENTREGA #7 20.01.2020 10:10:00.000</p> <p>DEVOLUCION #8 21.01.2020 10:10:00.000</p>	<p>3780088.0 6 events</p> <p>INGRESADO #1 20.01.2020 21:00:14.000</p> <p>1 INTENTO DE ENTREGA #2 21.01.2020 11:12:00.000</p> <p>1 INTENTO DE ENTREGA #3 22.01.2020 11:16:00.000</p> <p>ESPERANDO RETIRO #4 22.01.2020 16:04:31.000</p> <p>DEVOLUCION #5 29.01.2020 15:02:07.000</p> <p>ENTREGADO #6 31.01.2020 11:15:00.000</p>
---	--

Figura 4. Eventos repetidos e inconsistencias

En este caso se observó que el mismo evento se repite en sucesivos días o existen registros inconsistentes La Figura 4 muestra estas situaciones. En la parte izquierda se observa un caso que aparece una primera visita luego de la segunda; esto no puede ocurrir y seguramente es debido a un error en la carga (ejemplo se cargó mal la fecha en uno de los movimientos). Del lado derecho, en la misma figura, se muestra una inconsistencia ya que aparece registrado un movimiento de entrega luego de que el producto fue devuelto.

Adicionalmente se utilizó la herramienta Inductive Visual Miner para generar una animación visual y representar de forma más amena estos casos. Para ello se utilizó el conjunto de trazas que tiene un ajuste menor al 50 % obteniendo la visualización de las trazas que más se desvían. La Figura 5 ilustra esta animación. En dicha figura cada traza se representa con un círculo que recorre las diferentes etapas del proceso. Esta representación visual muestra que unas trazas tardan mucho más que otras. También muestra que en ocasiones se presentan algunos comportamientos erróneos. Por ejemplo en la figura 5 se observan retrocesos en las trazas, es decir luego de una tarea se ejecuta otra que debería ser anterior; se resaltan con círculos que 25 trazas van del segundo intento al primero y 20 vuelven sobre sí mismas.

Esta exportación de trazas que no se ajustan al modelo muestran los motivos que ocasionan los desvíos y permiten hacer un análisis más detallado con el fin de encontrar mejoras al proceso.



Figura 5. Trazas que no cumplen con el modelo, en círculo la cantidad que retroceden en de un estado en lugar de avanzar (ver sentido de la flecha)

3.4. Análisis de Performance

Finalmente se procede a realizar un análisis de performance sobre las trazas del modelo en busca de aquellas que, si bien pueden haber completado el proceso correctamente, el mismo se realizó en un intervalo de tiempo extremadamente superior a la media de resolución de cada traza. Para ello se vuelve a reproducir la muestra de datos sobre la red de Petri pero en este caso con una herramienta orientada al análisis de performance que se enfoca en los tiempos de resolución y no tanto en ver si las trazas se ajustan o no al proceso.

Como resultado se obtiene una visualización que representa en tonos más oscuros las tareas que demoran más tiempo en realizarse.

Se observa un tiempo medio de resolución para cada traza de 3,75 días, con un desvío de 2,81 días. Utilizando el mismo criterio que para la cantidad de movimientos por traza se decide enfocar el análisis en aquellos casos que tengan un tiempo de finalización mayor a la media más dos desvíos (9,37 días) por considerar que un tiempo de resolución mayor a este valor se trata de un desvío operativo.

Se realiza la exportación de las trazas que cumplen con ese criterio y se obtiene un total de 687 casos.

Observando los casos filtrados se encuentran trazas que se ajustan perfectamente al proceso pero con tiempos extremadamente altos. En algunos casos las trazas tienen más de tres movimientos y un intervalo de hasta 3 días entre cada uno de ellos lo que justifica un tiempo total tan elevado. Si bien 3 días entre movimientos es un tiempo mayor a la media, podría justificarse con algunas situaciones que pueden ocurrir en el día a día, como por ejemplo la falta de personal debido a enfermedad o vacaciones. Por otro lado, la mayoría de los casos que se obtuvieron en esta muestra son de solamente dos movimientos, donde el envío ingresó y estuvo detenido, en algunos casos hasta 14 días antes de tener un evento de salida a distribución. Luego el proceso se cumple correctamente con una entrega o devolución. Estos casos representan un desvío operativo importante ya que impactan directamente en la calidad del servicio a los clientes incumpliendo los estándares de entrega del producto.

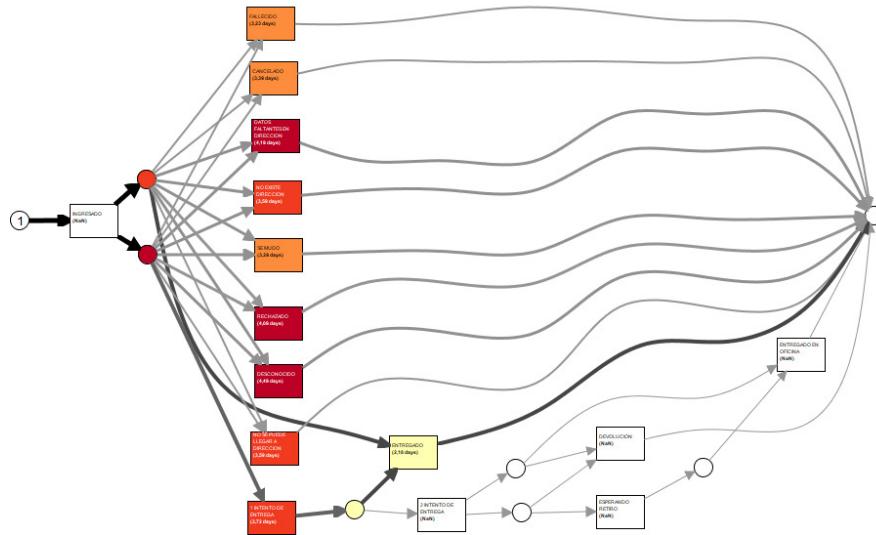


Figura 6. Análisis de Performance de las trazas de la muestra

En la figura 7 se observan trazas que cumplen el proceso con solo dos movimientos pero con una diferencia de más de 10 días entre uno y otro.

Con el análisis de performance se descubrieron nuevos desvíos operativos que no aparecían en el análisis anterior ya que se trata de trazas que se ajustan perfectamente al modelo pero que están con los tiempos muy por encima de la media, llegando en casos extremos a 24 días para una entrega. En base a este estudio se decidió tomar una muestra de los 10 casos más extensos y consultar a los responsables de cada sector / sucursal al que pertenezca para obtener información acerca del motivo de este retraso.

Como resultado del análisis de performance se encuentra la necesidad de generar una alarma que permita identificar estos casos luego de un intervalo de tiempo determinado para mejorar la calidad del servicio a los clientes.

4. Conclusión y futuras líneas de investigación

En este trabajo se utilizaron diferentes técnicas de Minería de Procesos para analizar datos reales de distribución postal en la República Argentina durante los años 2017 a 2020. No se encontraron a la fecha de generación de este documento investigaciones similares que analicen este mismo caso de estudio.

Luego de generar el modelo a partir de una muestra representativa de trazas se descubrió el proceso que realmente se lleva a cabo. Posteriormente se contrastó este modelo con la totalidad de trazas. Por su parte, la verificación de conformidad permitió identificar situaciones fuera de lo común que se traduje-

3765390.0 2 events	3769366.0 2 events	3763891.0 2 events	3763937.0 2 events
INGRESADO #1 07.01.2020 10:58:28.000	INGRESADO #1 10.01.2020 16:30:40.000	INGRESADO #1 03.01.2020 20:23:55.000	INGRESADO #1 03.01.2020 20:40:43.000
ENTREGADO #2 22.01.2020 09:55:00.000	ENTREGADO #2 28.01.2020 10:32:00.000	ENTREGADO #2 17.01.2020 13:00:00.000	DESCONOCIDO #2 16.01.2020 16:37:00.000
3764216.0 2 events	3769606.0 2 events	3767176.0 2 events	3775013.0 2 events
INGRESADO #1 03.01.2020 20:45:15.000	INGRESADO #1 06.01.2020 20:10:58.000	INGRESADO #1 07.01.2020 15:56:15.000	INGRESADO #1 15.01.2020 19:53:23.000
ENTREGADO #2 23.01.2020 13:20:00.000	ENTREGADO #2 21.01.2020 10:31:00.000	ENTREGADO #2 23.01.2020 12:01:00.000	SE MUDO #2 27.01.2020 12:30:00.000

Figura 7. Trazas que cumplen el proceso pero con tiempos muy altos

ron en errores operativos. Se identificaron tanto los casos que no cumplieron con el modelo como también aquellos que lo hicieron con redundancia de tareas.

Con el análisis de performance se descubrieron trazas que tienen desvíos operativos debido al excesivo tiempo de finalización pero que se ajustan correctamente al proceso. Estos casos tienen un gran impacto en la calidad del servicio.

Como futura línea de trabajo se propone continuar con las técnicas de minería de procesos y encontrar la forma de insertar en el sistema alertas tempranas que eviten los desvíos y cuellos de botella operativos.

Referencias

1. Process Mining Applied to Postal Distribution, Victor Martinez; Laura Lanzarini; Franco Ronchetti, CACIC 2021, 2021, <http://sedici.unlp.edu.ar/handle/10915/130342>
2. Process Mining: Data Science in Action, Wil van der Aalst, 978-3-662-49850-7, 2016, Springer
3. Research of Postal Data mining system based on big data, Xia Hu1; Yanfeng Jin1; Fan Wang, 3rd International Conference on Mechatronics, Robotics and Automation, 2015, 10.2991/icmra-15.2015.124, https://www.researchgate.net/publication/300483008_Research_of_Postal_Data_mining_system_based_on_big_data
4. Context Aware Process Mining in Logistics, Mitchell M. Tseng; Hung-Yin Tsai; Yue Wang, 2017, The 50th CIRP Conference on Manufacturing Systems, <https://www.sciencedirect.com/science/article/pii/S2212827117303311>
5. PM2: a Process Mining Project Methodology, Maikel L. van Eck; Xixi Lu; Sander J.J. Leemans; Wil M.P. van der Aalst, Eindhoven University of Technology, The Netherlands, http://www.processmining.org/_media/blogs/pub2015/pm2_processminingprojectmethodology.pdf
6. Wil van der Aalst, The Process Mining Manifesto by the IEEE Task Force, 2012, <https://www.tf-pm.org/resources/manifesto>

7. Workflow mining; discovering process models from event logs, W. van der Aalst; T. Weijters; L. Maruster, 1041-4347, 2004, IEEE, <https://ieeexplore.ieee.org/document/1316839>
8. Process mining in flexible environments, Christian Walter Gunther,978-90-386-1964-4,2009,Technische Universiteit Eindhoven, <https://research.tue.nl/en/publications/process-mining-in-flexible-environments>
9. IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams, IEEE Std 1849-2016, 2016, DOI 10.1109/IEEEESTD.2016.7740858
10. inductive visual miner, Sander J.J. Leemans, 2017, <http://leemans.ch/leemansCH/publications/ivm.pdf>
11. Reinventing the Postal Sector in an Electronic Age, Michael A. Crew; Paul R. Kleindorfer,978-1849803601,2011,Edward Elgar Publishing

ANEXO C

Anexo Herramientas utilizadas

Se listan a continuación las principales técnicas o herramientas de Ingeniería de Software utilizadas para el presente trabajo:

Eclipse: Entorno de desarrollo integrado (IDE) que proporciona un conjunto de herramientas de programación de código abierto multiplataforma para desarrollar java y otros lenguajes de programación.

<https://www.eclipse.org/ide/>

GitHub: Repositorio Git en la nube que permite almacenar gratuitamente proyectos de desarrollo de software en una gran cantidad de lenguajes de programación, facilitando el versionado y la distribución de los fuentes.

<https://github.com/>

Java: Lenguaje de programación orientado a objetos que permite desarrollar aplicaciones independientes de la plataforma.

<https://www.java.com/es/>

Notepad++: Editor gratuito avanzado de archivos de texto y código fuente con soporte a una gran variedad de lenguajes de programación.

<https://notepad-plus-plus.org/>

ProM Tools: Conjunto de herramientas open source que permite aplicar técnicas de Minería de Procesos en forma de plugins que se pueden concatenar uno detrás de otro. ProM es un set extensible, independiente de la plataforma, desarrollado en Java con una gran comunidad que lo mantiene en constante crecimiento.

<http://promtools.org/doku.php>

RapidMiner: Conjunto de herramientas para análisis de datos que contempla operadores de entrada - salida, pre procesamiento y visualización de resultados

<https://rapidminer.com/>

Signavio: Herramienta para el análisis y el diseño de procesos de negocio. Permite diseñar, modelar, documentar y simular procesos de negocio.

<https://www.signavio.com/>

python, jupiter, disco, mysql,dbeaver, sublime text. git, postman, docker, qgis, poo de tfi, pandas, R,

Anaconda

Spider

Python

PMPY <https://pm4py.fit.fraunhofer.de/>

jira atlassian

Spring Boot: framework de Java que permite crear aplicaciones web y de microservicios de manera rápida y sencilla, con una configuración mínima.

JPA (Java Persistence API): API de Java que permite gestionar el acceso a bases de datos de manera sencilla y estandarizada.

Swagger: herramienta de documentación de APIs que permite definir, visualizar y probar APIs de manera sencilla y eficiente.

Postgre SQL: base de datos relacional open-source, escalable y altamente disponible, utilizada en muchos proyectos de microservicios.

Docker: plataforma de contenedores que permite empaquetar, distribuir y ejecutar aplicaciones de manera sencilla y reproducible.

Grafana: herramienta de monitoreo y visualización de datos que permite visualizar métricas de una aplicación de manera sencilla y personalizada.

GitFlow

Fin del documento.

Victor Martinez