

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 evento			Descripción del evento	
Identificador de traza			Fecha evento	
	123 trazaID	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 Märušter[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

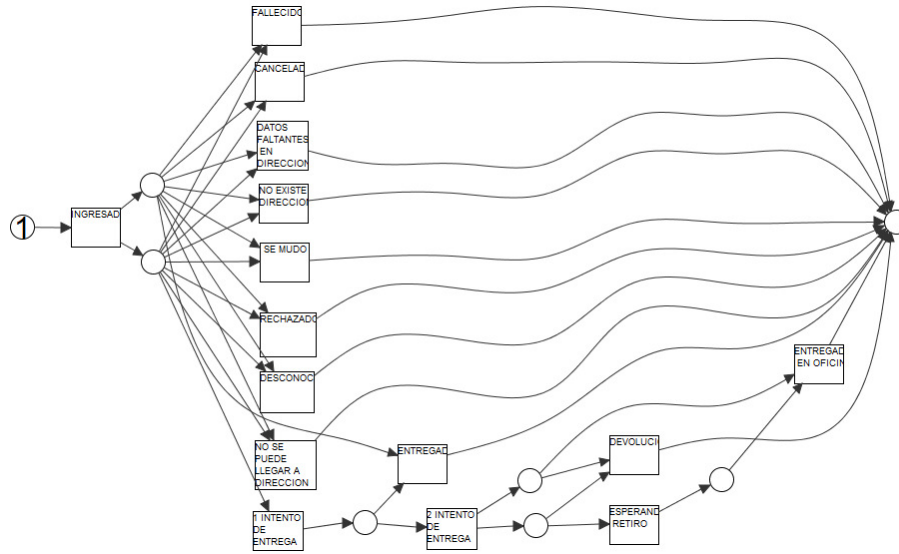


Figura 2. Proceso descubierto que deberán cumplir las trazas

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 cuanto 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.

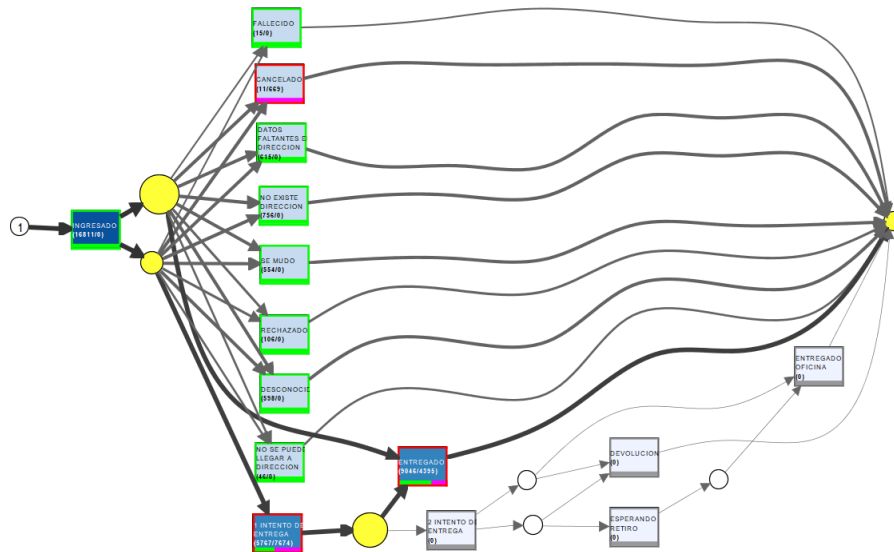


Figura 3. Recorrido de las trazas: eventos más comunes en oscuro y recorridos más frecuentes en línea más gruesa

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 saltan alguno. Estas trazas tienen un ajuste más bajo cuanto más se diferencian 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.

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 16:40:00.000	

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.

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/IEEESTD.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