

Business Process Analysis in Advertising: an Extension to a Methodology Based on Process Mining Projects

Aníbal Silva Osses
Departamento de Informática
Universidad Técnica Federico Santa María
Santiago de Chile
anibal.silvao@alumnos.usm.cl

Luiz Quelves Da Silva
Departamento de Ciencias de la Computación
Pontificia Universidad Católica de Chile
Santiago de Chile
lqdasilva@puc.cl

Bernardita Fernández Cobo
Departamento de Ciencias de la Computación
Pontificia Universidad Católica de Chile
Santiago de Chile
bvfernandez@puc.cl

Michael Arias
Departamento de Ciencias de la Computación
Pontificia Universidad Católica de Chile
Santiago de Chile
m.arias@uc.cl

Eric Rojas
Departamento de Ciencias de la Computación
Pontificia Universidad Católica de Chile
Santiago de Chile
eric.rojas@uc.cl

Jorge Muñoz-Gama
Departamento de Ciencias de la Computación
Pontificia Universidad Católica de Chile
Santiago de Chile
jmun@ing.puc.cl

Marcos Sepúlveda Fernández
Departamento de Ciencias de la Computación
Pontificia Universidad Católica de Chile
Santiago de Chile
marcos@ing.puc.cl

Abstract— Nowadays organizations generate large amount of data. Only a few make a good use to optimize the performance of the business. Process mining appears as a branch of the data science that tries to understand the actual operational processes in the organizations through different algorithms, allowing the discovery of process models to give insight of the processes and understand how they can be improved.

In this work different process mining techniques are applied to a company dedicated to the advertisement market, specifically the process of dealing with contract issues with customers. The Process Mining Project Methodology was followed to execute a case study. Additional to the basic methodology, elements from the others areas of studies were added to generate better results and have a better understanding of the problem. The case study includes three scenarios with three different hypotheses that were validated through our method.

Key words: *process mining, data mining, methodology, event logs, performance, organizational analysis, process model.*

I. INTRODUCCIÓN

En la actualidad son muchas las organizaciones que están dedicando recursos a controlar y mejorar sus procesos de negocio a través de diferentes sistemas de información. Es por esto que se requiere trabajar con herramientas adecuadas que

permitan realizar análisis pertinentes, obteniendo mayor provecho de los datos generados por los procesos, de forma tal que la información conseguida permita mejorar la calidad de los procesos analizados.

La gestión de procesos de negocio (BPM) [1] es una metodología que permite administrar los procedimientos que las organizaciones ejecutan a través de diferentes técnicas, de las cuales algunas están soportadas por aplicaciones especializadas. Esta práctica de mejora trae consigo beneficios en la realización de los procesos de negocio [2], como estandarizar flujos, automatizar el procedimiento, optimizar el uso de recursos, entre muchos otros.

Aprovechando los datos que las empresas generan e intentando estudiarlos desde una perspectiva de BPM, aparece la minería de procesos [3], disciplina que intenta comprender los procesos de negocio a través de análisis de la información histórica de ejecución de procesos. Con esto es posible entender el funcionamiento real (y no esperado) de los procedimientos de las organizaciones, además de soportarlos y mejorarlos a través de la información generada por diferentes herramientas especializadas en el área. Estas técnicas precisan de registros de eventos (*event logs*) que se almacenan en los sistemas de información gracias a la ejecución de los diferentes eventos que ocurren en relación a un proceso. También, permite descubrir

modelos que representan los procesos y verificar qué tan acertados son estas representaciones comparadas con los *logs* de eventos, validando si un modelo realmente representa los datos almacenados.

Una de las principales ventajas de aplicar minería de procesos es que permite analizar diferentes aristas de los procedimientos, como lo son las perspectivas de flujo (lo que compete al descubrimiento), temporal (análisis de tiempos empleados en las actividades), y organizacional (interacciones entre ejecutores). Comparando brevemente BPM con minería de procesos, se destaca que la primera metodología crea modelos a partir de un entendimiento acabado del proceso, identificando correctamente cuáles son las actividades. En cambio la segunda logra comprender realmente cómo se llevan a cabo estos a través de la información generada al realizar las actividades.

La minería de procesos [3] es una disciplina que permite el estudio y análisis de los procesos y los datos que éstos generan. Se espera que, con el uso de metodologías para aplicar minería de procesos, se puedan obtener resultados que permitan mejorar los procesos analizados. En este trabajo se propone una extensión a la metodología *Process Mining Project Methodology* (PM²) planteada en [4], con el fin de poder realizar un análisis más detallado y que genere más información mezclando técnicas de minería de datos al estudio de procesos.

La estructura del presente artículo es la siguiente. La sección II corresponde a los antecedentes necesarios para realizar el estudio: conocer el estado del arte de minería de procesos, entender la metodología base que se utilizará, y comprender el proceso de negocio que se analizará. La sección III explica la propuesta de investigación. En la sección IV se hablará de la metodología utilizada y cuál ha sido la extensión a ella. En la sección V se explicará el caso de estudio ejecutado. En la sección VI se ahondará en todo el análisis y las discusiones generadas. Por último, en la sección VII, se establecerán las conclusiones del trabajo y se propondrá el trabajo futuro.

II. ANTECEDENTES

El paradigma de desarrollo organizacional orientado en procesos ya se ha instaurado en entidades que impulsan una mejora continua en su negocio, por lo que ciertos elementos como los sistemas de información son requeridos. En base a los datos que estas aplicaciones generan nace la minería de procesos [3], una disciplina que intenta extraer conocimiento a partir de los registros de eventos que se generan en las diferentes actividades que componen los procesos, permitiendo producir análisis a partir de los datos.

Si bien el desarrollo de este artículo se centra en la minería de procesos y cómo ésta herramienta del análisis de datos ha sido aplicada en un caso de estudio real, es fundamental mencionar que esta disciplina nace en el marco de la gestión de procesos de negocios, o BPM como se conoce por sus siglas en inglés (*business process management*). Como se menciona en [5], dichos procedimientos están conformados por una secuencia de

actividades que se realizan de manera ordenada, las que deben comprender el funcionamiento de alguna organización que busca administrar los recursos involucrados en cada una de estas tareas. Entendiendo lo anterior, se puede explicar más en lo que involucra la minería de procesos.

A. Minería de Procesos

Corresponde a una disciplina de la ciencia de los datos que intenta obtener información sobre los procesos de negocio de una organización a partir de datos generados en las distintas actividades que lo componen. Suele mencionarse que es una combinación entre minería de datos y BPM.

El elemento fundamental de minería de procesos son los registros de evento (*event log*), los que guardan características de un evento realizado en un momento determinado. Entre los atributos comúnmente utilizados están: identificador del caso, nombre de actividad, ejecutor que la realiza y fechas de inicio y término (*timestamps*). Un conjunto ordenado de estos registros conforma un caso (traza), que corresponde a una instancia del proceso. La creación de los *event log* debe ser pensada desde la arista del problema del negocio, es decir, se debe comprender éste para determinar qué elementos aportarán información al estudio. En la actualidad, existen sistemas de información que permiten almacenar las operaciones de los procesos en un formato enfocado en el análisis de registros de eventos (*process aware information systems*) [6], lo que puede facilitar mucho el preprocesamiento de los datos. Muchas otras organizaciones cuentan con fuentes de datos convencionales, como bases de datos, planillas de cálculo, o cualquier plataforma comúnmente usada para el almacenamiento de datos, lo que en ocasiones genera un mayor tiempo de trabajo ya que los datos deben ser transformados a una estructura de registros de eventos. La minería de procesos ha sido aplicada en diversas industrias, por ejemplo, en el área de la salud [7] y en el área de educación [8]. Además de distintos tipos de sistemas como en sistemas legados de datos [9], plataformas colaborativas como *Sharepoint* [10] y sistemas recomendadores para asignación de recursos en procesos de negocio [11].

La minería de procesos permite realizar tres tipos de análisis: el descubrimiento, la verificación de conformidad y el mejoramiento de procesos. El descubrimiento consiste en encontrar modelos que describan los procedimientos, siendo las notaciones más utilizadas: redes de Petri [12], BPMN (*business process model and notation*) [13], YAWL (*yet another workflow language*) [14], *workflow language* [15], y grafos dirigidos [16]. Para conseguir estos modelos existen distintos algoritmos de descubrimiento, como *alpha miner* [17], *heuristic miner* [18], *genetic mining* [19], *fuzzy miner* [20], *inductive miner* [21], entre otros. Cada uno de ellos poseen características diferentes y generan modelos en las notaciones mencionadas, las que son de fácil lectura. El segundo tipo de análisis es el de verificación de conformidad, el cual consiste en contrastar un modelo ya creado (puede ser a través de técnicas de descubrimiento) con un registro de eventos, con el objetivo de entender qué tan similar son los datos que el proceso genera a la ejecución esperada de éste. Esto se mide a través de

dos principales indicadores: el ajuste (*fitness*) y la precisión. El primero permite medir si el comportamiento del *log* de evento puede ser replicado por el modelo. Y la precisión mide si el modelo no permite comportamientos que no están en los datos. Estas dos métricas son obtenidas a través del algoritmo *token replay* presentado en [22]. Por último, el mejoramiento permite extender o mejorar procesos existentes a partir del uso de información actual de ejecución del proceso, almacenada en *logs* o registros de eventos.

Otro punto importante de la minería de procesos es que permite entender diferentes aristas del proceso a través de distintas perspectivas, siendo éstas, de control de flujo, de caso, de temporalidad y organizacional [23]. La primera consiste en analizar el orden y relación de las actividades, lo que puede asociarse con el descubrimiento de modelos. La perspectiva de caso permite identificar las características de cada traza del proceso, para así poder establecer similitudes o diferencias entre éstos. La perspectiva organizacional analiza diferentes relaciones que tienen los ejecutores de las actividades que componen el proceso, como pueden ser el trabajo en conjunto, traspaso de tareas, subcontratación (entregar y recibir la realización de una actividad a un mismo ejecutor) y tareas similares (ejecutores con mismo rol); siendo respectivamente para cada una de estas relaciones los algoritmos: *working together*, *handover of work*, *subcontracting* y *similar task*. Por último, la perspectiva temporal permite analizar los tiempos de ejecución de las distintas etapas del proceso, por ejemplo, saber cuánto demora una actividad en realizarse o conocer el tiempo de espera entre dos ramas del proceso que se deben efectuar en paralelo.

Existen herramientas enfocadas en la aplicación de técnicas de los distintos tipos de análisis de minería de procesos. Las dos principales herramientas son ProM [24] y Disco¹. ProM es un *framework* de código abierto que permite realizar el descubrimiento de modelos, la verificación de la conformidad, y otros tipos de análisis bajo las perspectivas de minería de procesos. En el caso de Disco, ésta es una aplicación comercial, la cual permite generar un entendimiento inicial del proceso gracias a su simple interfaz y fácil uso. También, brinda otras funcionalidades para analizar más en detalle el proceso, permitiendo realizar el descubrimiento de éste, aplicar filtros, o analizar el desempeño temporal, entre otras características.

B. Metodología

Desde la aparición de la minería de procesos, han surgido metodologías para realizar la aplicación de esta disciplina en distintos contextos. Una de estas es *Process Mining Project Methodology* (PM^2) [4], la que está conformada por 6 etapas: planificar (1), extraer (2), procesar (3), analizar (4), evaluar (5), y mejorar y soportar (6), como lo muestra la figura 1. Esta metodología se basa en 4 elementos fundamentales, crear *preguntas de investigación* que se derivan de los *objetivos del proyecto*, las que serán respondidas en el grado que se obtenga

de la *calidad y conformidad del análisis*, lo que proporcionará *ideas de mejora* para lograr los objetivos propuestos.

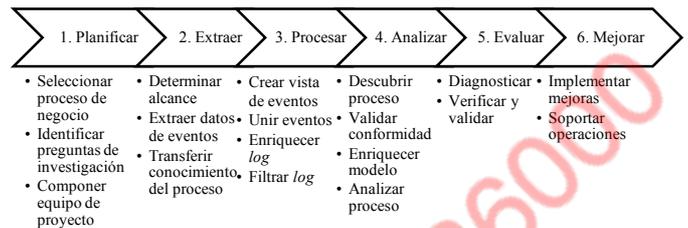


Figura 1. Desglose de la metodología *Process Mining Project Methodology* considerada en este trabajo, especificando tareas principales para cada una de las etapas.

En este artículo se propone una variación a la metodología PM^2 recién mencionada, que consistirá en complementar la aplicación de minería de procesos utilizando técnicas de minería de datos. Así, se generará un nuevo ciclo entre las etapas (3), (4) y (5) de la metodología, en donde el análisis de datos le aporta información al de procesos y viceversa. En la sección IV se detallará en profundidad este método.

C. Sobre la Organización en Estudio

Para aplicar la metodología planteada en este trabajo, se ha considerado el proceso de *ventas y anulaciones de contratos* de una empresa dedicada al rubro de publicidad, la cual tiene como negocio la creación e impresión de directorios de personas y compañías. Hasta mediados del año 2000, la publicidad para las pequeñas y medianas empresas era su fuerte, sin embargo, el creciente mercado de publicidad digital ha generado pérdidas de clientes. Por lo mismo, a partir del año 2010 la organización comenzó a desarrollar un nuevo modelo de negocio basado en tecnologías digitales. Los productos y servicios principales que ofrece actualmente son:

- Servicios Digitales:
 - Creación de sitios *web*.
 - Tiendas *online* y funcionalidades de autoservicio vía *eCommerce*.
 - Servicios para móviles, comunicación y ventas de PyMEs hacia sus consumidores.
- Servicios Impresos:
 - Directorio de páginas amarillas.
 - Directorio de páginas blancas.
 - Revista propia de la empresa.
 - Marketing directo.

A raíz de esta transformación y dicotomía que se ha generado entre el modelo antiguo y el nuevo, se están presentando muchos problemas para lograr la transformación de la empresa y poder dar un servicio de calidad, además de lograr que se genere el retorno de la inversión hecha. Por lo que es imperante detectar los problemas y crear planes de acción para solucionar estos conflictos.

¹ www.fluxicon.com

Si bien, hay muchos procesos con problemas en la compañía, en este artículo se analizará específicamente el proceso de *ventas y anulaciones de contratos*, el cual comienza cuando la propuesta de la PyME es recibida, y termina cuando el contrato es cerrado (venta exitosa) o anulado. Este último escenario es en el que se ha profundizado, intentando entender el porqué de tantas anulaciones por parte de los clientes, conflicto que viene siendo analizado y monitoreado por la compañía hace 4 años, incluso aplicando cambios en sus sistemas de ventas y gestión. Sin embargo, éstos no ayudaron a solucionar el problema. Se sabe que cerca del 20% de los clientes terminan cancelando la orden del servicio, de los cuales se desconoce el motivo concreto por el que ocurre. Así, la principal pregunta a resolver es ¿por qué los clientes anulan los contratos? En segundo lugar, se pretende comprender cómo se está ejecutando el proceso y entender si el resultado obtenido está relacionado con la alta tasa de anulaciones o no. Luego de entender el proceso de negocio y la dificultad que presenta, se han propuesto tres diferentes causales que podrían explicar este fenómeno, las que se listan a continuación:

1. *Elevados tiempos de espera en confección y revisión de contrato.*

Surge revisando el proceso desde la perspectiva temporal, en donde en algunas ocasiones hay tiempos muy largos para atender la solicitud del cliente. Dada esta situación se analizará la hipótesis de que muchos de los contratos anulados son consecuencia de los largos tiempos de espera por una modificación o mejora a éste.

2. *Ejecutores del proceso no tienen roles claros; no existen especialistas para cada tarea.*

Se basa en la perspectiva organizacional de la empresa. Se observa en el funcionamiento de ésta que hay diversos recursos (ejecutores) que realizan varias actividades, no dedicándose a un número pequeño de ellas. Así puede entenderse que no existen roles, o que, si los hay, estos no se respetan.

3. *Cientes con diferentes características reaccionan de diversas formas al funcionamiento del proceso.*

Se estima que clientes nuevos anulan más contratos que antiguos. Esta hipótesis permitirá un mayor desarrollo del trabajo mezclando disciplinas de análisis de datos (minería de datos) con las técnicas de minería de procesos.

El validar las hipótesis requerirá utilizar las perspectivas de flujo, temporal y organizacional que la minería de procesos tiene a disposición. De esta forma, se podrá generar un análisis detallado que permita entender las dificultades en el proceso que generan el conflicto.

III. PROPUESTA DE INVESTIGACIÓN

En este trabajo se busca determinar las posibles causas que influyen en que una organización no pueda ejecutar adecuadamente uno de sus principales procesos de negocio, el

de *ventas y anulaciones de contratos*. La empresa vende anuncios publicitarios en diferentes medios, y está siendo afectada por una alta tasa de anulación de los servicios que sus clientes habían contratado (comprado) con anterioridad. Considerando este problema, se utilizó minería de procesos para entender cómo se está ejecutando el proceso descrito. A través de esta disciplina es posible realizar análisis de diferentes tipos, como estudiar el flujo del proceso, y analizar el proceso desde la perspectiva temporal y organizacional. Para concretar este análisis, fue necesaria la aplicación de varios algoritmos, principalmente para generar descubrimiento de procesos, los que permitieron determinar la conformidad de los modelos entregados. Además, se utilizó minería de datos para complementar el análisis de minería de procesos, con el fin de obtener información relevante sobre el mismo proceso estudiado.

Un objetivo secundario es poner en práctica técnicas fundamentales de minería de procesos apoyadas de un análisis tradicional de datos, a través de técnicas de minería de datos. Para ello, se plantea una extensión a la metodología PM² la que facilita la realización de proyectos de minería de procesos. Con esto, es posible generar información útil para la organización estudiada, teniendo la oportunidad de mejorar el desempeño de su principal proceso.

IV. METODOLOGÍA

Si bien el objetivo principal de este trabajo es entender la problemática de un caso de estudio, también se ha propuesto extender la metodología *Process Mining Project Methodology* [4]. Esto implicó en primera instancia establecer una perspectiva de negocio, teniéndose que elegir un proceso que fuera importante para la organización y que esté soportado por su plan estratégico (etapa de planificación). En segundo lugar, a partir de la visión ya establecida se continuó el trabajo con un enfoque técnico, basado en minería de procesos y minería de datos, lo que contempla las etapas (3), (4) y (5) presentadas en la figura 1. La figura 2 muestra el nuevo método propuesto.

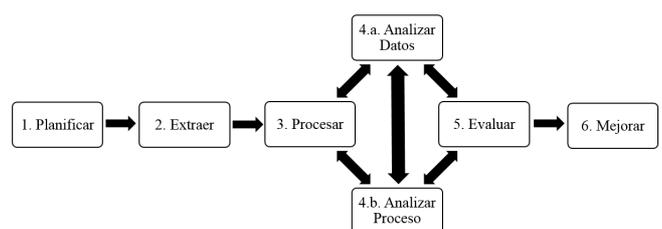


Figura 2. Variación a la metodología *Process Mining Project Methodology* propuesta, especificando sus etapas y el flujo de ellas.

A continuación, se explicitan las etapas de la metodología:

1. Planificar

Se debe tener un entendimiento acabado del problema en cuestión, lo que implica comprender la organización y el proceso seleccionado que ésta ejecuta. Para ello, es requerido estudiar el negocio y cuáles son los datos que generan, con el

objetivo de poder identificar los datos que serán adecuados para crear el registro de eventos y que permitirá realizar todo el análisis. Lo que se hace es *seleccionar el proceso de negocio* que interesa analizar y mejorar. Luego, se deben *identificar las preguntas de investigación* entendiendo los objetivos del análisis; y por último se debe *crear el equipo adecuado de individuos*, tanto analistas de procesos como expertos del negocio que sean capaces de generar información que permita responder las interrogantes antes planteadas.

2. Extraer

La segunda etapa comprende la extracción de los datos. Lo primero es *determinar el alcance* que éstos tendrán considerando los objetivos establecidos en la etapa anterior. Luego se deben *extraer los datos* desde los diferentes sistemas de información. En paralelo, los expertos del proceso deben *traspasar los conocimientos* a los analistas, para así comenzar el estudio.

Los registros descargados para este trabajo fueron filtrados y se seleccionaron los más relevantes para el análisis. Es necesario en este paso entender adecuadamente el objetivo del análisis a realizar, además de comprender la causa de un problema específico que la organización posee.

3. Procesar

En esta etapa se deben *crear vistas de los registros de eventos* necesarias para responder las interrogantes, con lo que es posible *generar agregaciones* entre éstos con el fin de reducir la complejidad del análisis y mejorar los resultados de la minería. Posteriormente se requiere *enriquecer el registro de eventos* con el objetivo de generar mayor información al momento de realizar el análisis. Por último, es necesario *filtrar el log* para reducir nuevamente la complejidad del análisis y que los resultados sean más claros.

4. Analizar

Como se ve en el diagrama de la figura 2, esta etapa está constituida por dos elementos: análisis de datos o minería de datos (4.a) y análisis de proceso o minería de procesos (4.b). Lo que se hace en esta fase de la metodología es generar información a partir de las dos disciplinas de forma no simultánea. Así, cada resultado obtenido servirá para mejorar el análisis del otro enfoque propuesto; desde datos a procesos y viceversa. Por ejemplo, a través de árboles de decisión es posible entender cuál es la variable con mayor representatividad dentro de todas las que son complementarias al proceso, como el monto del contrato o el tipo de cliente. Lo anterior permite crear un *feedback* mutuo que genere un ciclo entre estos dos análisis, lo que se complementa con la etapa anterior de procesamiento para modificar los datos, y con la etapa siguiente de evaluación para verificar si el análisis hecho aporta información relevante.

En (4.a) se obtiene información a partir de técnicas de minería de datos, siendo cualquiera de éstas válida siempre que logre aportar valor al análisis, ya que existen muchos algoritmos dentro de esta rama de estudio. En (4.b) en cambio, se hace análisis directamente a partir del proceso con técnicas de

descubrimiento, análisis de conformidad, y mejora de procesos, con lo que se espera *analizar el proceso* de forma acabada, más aun considerando lo que pueda aportar la etapa (4.a).

5. Evaluar

Luego de generar todo lo que implica el análisis, se deben entender estos resultados. Para ello en primera instancia es requerido *diagnosticar el desarrollo*, lo que implica interpretar correctamente los resultados, distinguir cuáles de éstos son interesantes o inusuales, e identificar si las preguntas de estudio iniciales requieren ser iteradas para un mejor entendimiento. Luego se debe *verificar y validar*, por lo que respectivamente es necesario: comparar los resultados con la información original y con los elementos claves del proceso o ejecutores de él. Esta etapa además permitirá intuir qué ideas son posibles para mejorar el proceso.

6. Mejorar

Consiste en *implementar las mejoras* detectadas para el proceso y *soportar las operaciones* de éste. Uno de los principales motivos de que exista la minería de procesos es el lograr mejoras, por lo que esta última etapa es muy relevante. Además, es posible realizar mediciones de estas mejoras hechas al proceso a través de otros proyectos de análisis. Por último, apoyar las operaciones a través de minería de procesos es fundamental ya que permite detectar posibles problemas en la ejecución de los casos. Lo anterior se logra a través de sistemas de información que soportan el proceso, (midiéndolo y gestionándolo) mejorando la calidad de los resultados del procedimiento.

En las dos secciones siguientes se presenta la implementación de este método sobre el caso de estudio seleccionado. En la sección V desde la etapa de planificación (1) hasta la de procesamiento de los datos (3); en la sección VI está el análisis, (4.a) y (4.b), y la evaluación (5). Se destaca que la etapa de mejora (6) aún no se ha concretado por lo que no forma parte del alcance de este trabajo, pero sí se mencionen posibles mejoras al proceso en base a la información inferida.

V. CASO DE ESTUDIO

Para probar la extensión a la metodología propuesta se realizó un caso de estudio específico, el cual consiste en analizar el proceso de contratación de servicio de una empresa del rubro publicitario. En concreto, se intentó entender la razón de por qué actualmente existen tantas anulaciones del servicio por parte de los clientes cuando ya lo han contratado.

Comenzando con la etapa de *planificación* (1), se debe comprender el estado actual del proceso en lo que respecta a la parte técnica de éste y el almacenamiento de los datos. La plataforma en donde están los datos utilizada por la empresa corresponde a una base de datos que guarda todos los registros generados. El modelo de ésta posee un gran número de entidades, las que se reflejan en miles de tablas, planillas y

ficheros planos que lo componen (1.740 elementos en total). Sin embargo, se ha seleccionado sólo un subdominio de todos estos, por lo que quedaron 107 tablas para realizar el desarrollo con el objetivo de enfocar el trabajo en las hipótesis del análisis. Las principales entidades son *solicitud*, *cuenta*, *cliente*, *ficha* y *mov_solicitud*. En la figura 3 se presenta el modelo de datos que describe las transacciones de ventas de la compañía en una versión resumida.

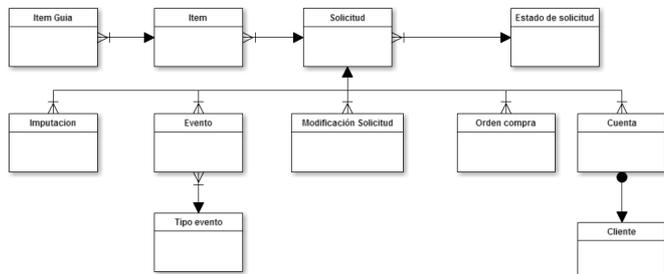


Figura 3. Modelo de datos resumido de las transacciones de venta que genera la organización.

Como gran parte de los procesos de la compañía se encuentran automatizados, y muchos son registrados en las diversas bases de datos, se utilizaron consultas SQL para la extracción del *log* de eventos, lo que corresponde a la segunda etapa de la metodología (*extraer*). Una de las principales dificultades al aplicar minería de procesos es la calidad de los datos a utilizar [25]. En otras palabras, que los eventos que conforman el proceso se registren de forma tal que puedan ser usados en diversas herramientas de minería de procesos. Este caso no es la excepción, en algunas ocasiones los datos no estaban almacenados en un formato que permitiera aplicar minería de procesos directamente. Para trabajar con este conflicto, se utilizaron técnicas para la creación de columnas dinámicas a los resultados de las consultas. En particular, lo anterior se hizo para la generación de los *timestamps* de inicio y fin de cada actividad, ya que no era una información registrada de forma directa por el sistema de la organización, pero sí era requerida para los posteriores análisis de temporalidad.

Para dar un mayor nivel de detalle sobre la consulta SQL y los elementos que ésta extrae, se pueden mencionar que corresponde a una unión entre las tablas: *solicitud*, *mov_solicitud*, *mae_est_solicitud*, *cuenta* y *ficha*. Así se obtienen columnas relevantes como *nu_solicitud* (número del contrato y *case_id* del proceso), *to_neto_solicitud* (monto del contrato) desde la tabla *solicitud*; los atributos *no_est_solicitud* (cambios de estado del contrato, actividades del proceso), *fe_movimiento* (fecha del cambio) y *co_usuario* (usuario que lo realizó, ejecutor) se extrajeron desde la tabla *mov_solicitud*. La tabla I muestra un fragmento del *log* de eventos generado con los atributos anteriormente mencionados, los que no corresponden a la totalidad de los utilizados en el análisis.

Tabla I. Extracto del registro de eventos utilizado compuesto por los principales atributos necesarios para realizar el análisis.

NU_SOLICITUD (case_id)	NO_EST_SOLICITUD (actividad)	FE_MOVIMIENTO (timestamp)	CO_USUARIO (ejecutor)	TO_NETO_SOLICITUD (valor contrato)
2451460	CONTRATOS ANULADOS	24-04-2015	R_PEREGRIN	155093
2451461	CONTRATOS EN PRIMER INGRESO	26-03-2015	C_ROJAS	600000
2451461	CONTRATOS VENDIDOS	30-03-2015	J_PAREDES	600000
2451462	CONTRATOS EN PRIMER INGRESO	30-03-2015	C_ROJAS	789600
2451462	PENDIENTE EN VALIDACIÓN	31-03-2015	R_FLORES	789600
2451462	CONTRATOS VENDIDOS	31-03-2015	MAR_REYES	789600
2451463	CONTRATOS EN PRIMER INGRESO	31-03-2015	J_PAREDES	503602
2451463	CONTRATOS VENDIDOS	01-04-2015	P_CONCHA	503602
2451465	CONTRATOS EN PRIMER INGRESO	27-03-2015	C_ROJAS	766503
2451465	CONTRATOS RECHAZADOS	30-03-2015	J_PAREDES	766503
2451465	PENDIENTE EN VALIDACIÓN	31-03-2015	J_PAREDES	766503
2451465	CONTRATOS VENDIDOS	31-03-2015	J_PAREDES	766503
2451466	CONTRATOS EN PRIMER INGRESO	30-03-2015	C_ROJAS	925503
2451466	CONTRATOS VENDIDOS	31-03-2015	J_ROJAS	925503
2451467	CONTRATOS EN PRIMER INGRESO	31-03-2015	C_ROJAS	2442000

Con la extracción de los datos lista es posible avanzar a la etapa de *procesamiento* (3). Fue necesario cargar el *log* de eventos en Disco para poder tener una noción global del proceso, el que está constituido por: 112.573 eventos, 32.105 casos en 235 variantes, 11 actividades, y 274 ejecutores. Fue necesario aplicar un filtro antes de continuar, el que estuvo compuesto en dos partes. Primero se filtraron todos los casos que no estaban completos a través de atributo *nu_secuencial*, el que para cada actividad indica la posición de ésta en el caso, entonces, todo caso que no tuviera alguna de sus tareas con *nu_secuencial* igual a 1 debía descartarse. El modelo resultante aún era demasiado complejo con 214 variantes todavía, por lo que se decidió reducir a un 10% el total de éstas a través del filtro de Disco que permite seleccionar un tramo de todas las variantes, las que se encuentran ordenadas en base a su frecuencia dentro del registro de eventos. La tabla II muestra un resumen de los registros utilizados para todo el análisis posterior luego de aplicar los dos filtros.

Tabla II. Resumen de los datos del proceso luego de aplicar los filtros.

Característica del Proceso	Cantidad
Casos	25.823
Eventos	92.960
Variantes	23
Actividades	7
Ejecutores	263
Tiempo Promedio de Duración de Caso (en días)	22,1

Se detectó en el *log* de eventos inicial (sin filtro) que todas las actividades podían iniciar algún caso, por lo que se decidió reducir éstas a las dos más relevantes (esto ocurrió al aplicar el segundo filtro). Así, se produjo una disminución de cuatro actividades, las que no eran relevantes en las variantes más frecuentes del proceso. Las que sí se mantuvieron en el modelo son: *contratos en firma*, *contratos en primer ingreso*, *vendido*

televentas, pendiente en validación, contratos rechazados, contratos vendidos y contratos anulados, siendo las dos primeras actividades de inicio y las dos últimas de término. La figura 5 muestra el mapa del proceso correspondiente a la tabla II.

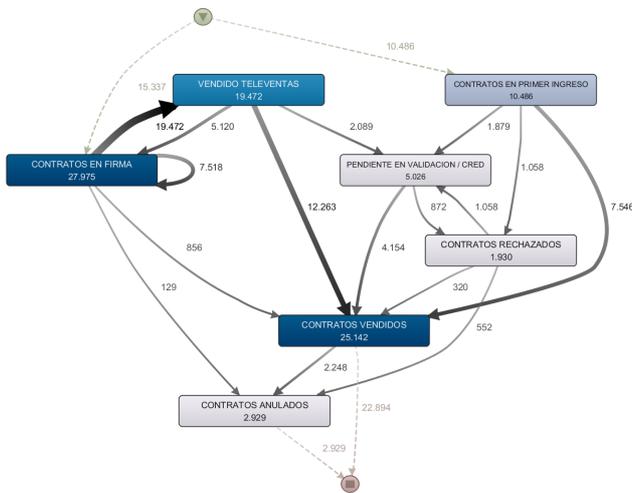


Figura 5. Modelo del proceso entregado por Disco considerando la totalidad de actividades y rutas de flujo.

Luego de obtener el mapa del proceso con la herramienta Disco, fue posible exportar la información que contenía el archivo CSV de los registros de eventos a los formatos MXML y XES, extensiones necesarias para poder utilizar las dos versiones de la herramienta ProM (5.2 y 6.5).

Lo último que se hizo antes de comenzar el análisis fue crear la red de Petri que representa el flujo esperado del proceso, la que corresponde a la figura 6. Ésta se construyó en la plataforma WoPeD en formato PNML, a través del entendimiento del proceso de negocio con la información entregada por los ejecutores de éste.

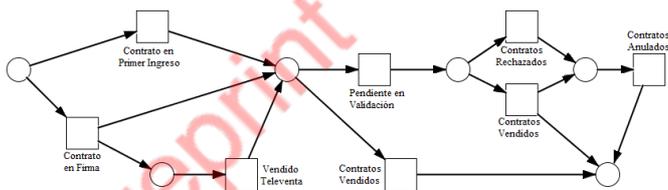


Figura 6. Red de Petri que representa el proceso ideal de ventas y anulaciones de contratos que ejecuta la empresa.

Es posible destacar las distintas herramientas que se han utilizado en este trabajo en la figura 4, mostrando cuáles fueron y cómo se relacionaron entre sí. Los elementos cuadrados son aplicaciones y los elípticos son archivos, los que se entregaban como salida o utilizaban como entrada. Para la generación del registro de eventos, se utilizaron las herramientas Oracle Database como fuente de datos y Oracle Developer para generar las consultas, así se creó un archivo CSV (*comma separated values*). Esto permitió realizar consultas y crear un formato sobre los datos de forma tal de obtener las actividades y los

eventos, y se pudo utilizar Excel para hacer filtros y *data cleansing*. Luego, se utilizó la herramienta Disco para filtrar y exportar los mismos datos anteriores en archivos XES (*eXtensible Event Stream*) y MXML (*process mining XML*), y en paralelo se utilizó la herramienta Rapid Miner para generar árboles de decisión (*decision trees*), lo que aportó información relevante para continuar con el análisis de minería de procesos. Finalmente, con la herramienta WoPeD (*Workflow Petri Net Designer*) se generó la red de Petri [3] de referencia del proceso en extensión PNML (*petri net markup language*), y más los archivos XES y MXML fue posible aplicar todos los algoritmos requeridos para un completo análisis a través de ProM en sus dos versiones. Esta parte del desarrollo hecho puede comprenderse en las etapas 2 y 3 de la extensión a la metodología propuesta.

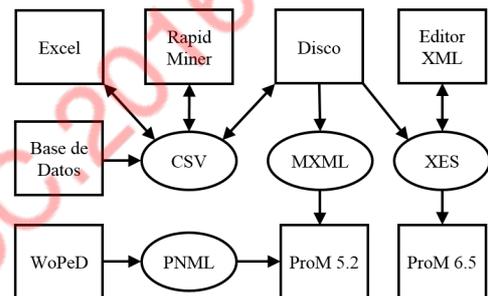


Figura 4. Herramientas utilizadas para concretar todo el desarrollo realizado, desde la extracción hasta la creación y comparación de los modelos.

VI. ANÁLISIS Y DISCUSIÓN

Esta sección está compuesta por dos partes. La primera es un análisis general al proceso utilizando diferentes técnicas de minería de procesos, específicamente de descubrimiento y de conformidad. La segunda, contempla el desarrollo requerido para estudiar las hipótesis planteadas en base al problema inicial. Así, las etapas (4.a) y (4.b) más la etapa (5) de la metodología están presentes de manera implícita en esta sección.

Con la red de Petri (figura 6) más el registro de eventos en formato MXML, fue posible aplicar el algoritmo *token replay* [22] en ProM 5.2 para realizar análisis de conformidad, determinando qué tan relacionado están el modelo propuesto con los datos del *log* de eventos. Así, los resultados obtenidos para el ajuste (*fitness*) y la precisión respectivamente son 0,78 y 0,65. De éstos se destaca el primero, el que indica que prácticamente un 80% del comportamiento total de los casos del registro de eventos concuerda con el modelo predefinido. A través del resultado que este algoritmo entrega, se puede ver que las actividades: *contratos en firma*, *contratos rechazados*, *contratos anulados* y *contratos vendidos* fallan en *token replay*, por lo que deben ser consideradas en lo que resta del análisis. Para comprender el flujo real del proceso lo que se hizo fue aplicar algoritmos de descubrimiento sobre el *log* de eventos, los que fueron: *alfa miner* [17], *heuristic miner* [18], *genetic mining* (con tamaño población de 20 y 30 generaciones) [19], utilizando ProM 5.2; y por último *inductive miner* [21] en ProM

6.5. El resultado de los dos primeros algoritmos generó fuertes inconsistencias con el modelo esperado, mientras que *genetic* e *inductive* sí eran similares a lo deseado. Se volvió a aplicar el algoritmo de conformidad *token replay* a estos cuatro nuevos modelos, resultados que se muestran en la tabla III.

Tabla III. Valores de ajuste y precisión tras aplicar *token replay* a los cuatro modelos entregados por los diferentes algoritmos de descubrimiento.

Algoritmo	Ajuste	Precisión
Alpha	0,49	0,00
Heuristic	0,62	0,91
Genético	0,69	0,81
Inductive	0,98	0,90

De la tabla anterior se puede identificar que el mejor resultado se obtuvo al aplicar el algoritmo *inductive miner*, por lo que el modelo entregado por éste fue considerado para continuar con el análisis de temporalidad. En la figura 7 se puede observar dicho modelo.

Nótese que el modelo de la figura 7 es diferente al presentado en la figura 5 (modelo esperado). Uno de los motivos es la cantidad de actividades “invisible” que tiene (coloreadas en negro). Se debe destacar que éste será utilizado para comprender la arista temporal del proceso, determinando en dónde están los principales “cuellos de botella” y no en definir si es válido o no.

Luego de realizar el descubrimiento del proceso y el análisis de conformidad pertinente, es posible comenzar a establecer si las hipótesis serán validadas o refutadas. Antes de continuar, se recuerda que el problema que presenta la organización es la alta tasa de anulaciones de contratos ya vendidos, por lo que se han propuesto las siguientes tres hipótesis como causas de éste: *tiempos muy elevados en la generación del contrato del servicio, ejecutores no tienen roles claros, y diferentes tipos de clientes reaccionan de diferente forma al flujo del proceso.*

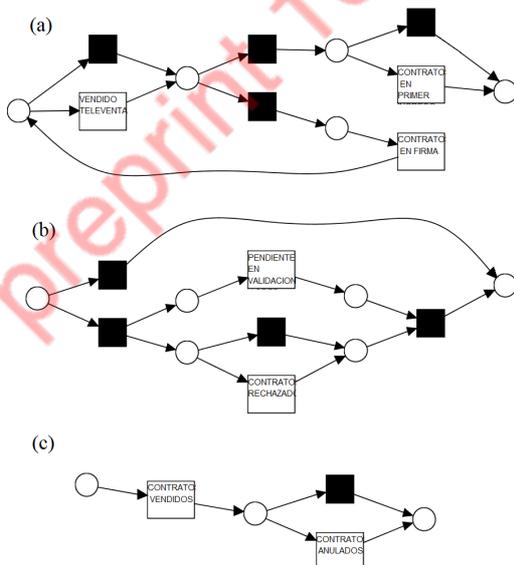


Figura 7. Modelo obtenido a través del algoritmo *inductive miner* aplicado al log de eventos segmentado en 3 partes ordenadas respecto al flujo del proceso.

Hipótesis 1: Elevados tiempos de espera en confección y revisión de contrato.

Luego de tener el modelo seleccionado (entregado por *inductive miner*) conectado con el registro de eventos a través del algoritmo de análisis de desempeño de ProM 5.2 *performance analysis with Petri net*, fue posible identificar cuáles son las principales actividades que enlentecen el proceso. Como resultado, se detectaron las actividades *vendido televentas*, *contratos rechazados*, *pendientes en validación* y *contratos vendidos*, en donde los únicos lugares (elementos circulares en el modelo) de la red de Petri con tiempos elevados fueron los que se encontraban entre el inicio y fin de las actividades ya mencionadas. En la figura 8 se puede ver el estado de los lugares unidos a las actividades recién mencionadas, así, si ésta coloreado con verde el tiempo fue menor a 12 horas, con amarillo si demoró entre 12 y 24 horas, y con rojo si tardó más de 24 horas. Sumado a lo anterior, se destaca que no hubo tiempos de sincronización (en donde una rama del proceso debe esperar a que otra concluya para poder continuar el flujo), sino sólo de espera. Como primera afirmación se establece que, de las cuatro actividades, las tres primeras son parte de ciclos que el algoritmo difuso (*fuzzy miner*) de Disco detectó (ver figura 9), los que corresponden a la generación del contrato (entre actividades *contratos en firma* y *vendido televentas*) y la revisión de éste (entre actividades *pendiente en validación* y *contratos rechazados*). Así, lo primero que podría concluirse es que dichos ciclos dentro del proceso son los que extienden el tiempo del proceso, afirmación errada como se verá en el siguiente párrafo.

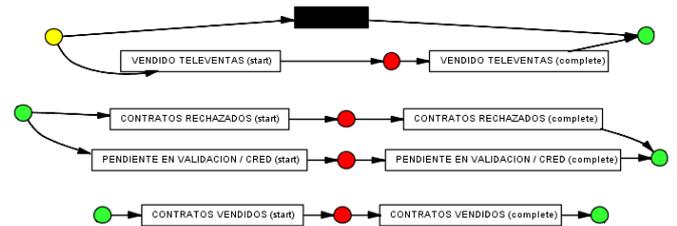


Figura 8. Segmentos del modelo entregado por *inductive miner* al aplicarle el algoritmo de *performance*, en donde se muestran los lugares que generan demoras en el proceso.



Figura 9. Segmento del modelo generado en Disco desde la perspectiva temporal, en donde se muestran los ciclos entre las actividades que el proceso posee.

Para comprender de mejor forma el flujo del proceso, lo que se hizo fue dividir el modelo que entrega Disco en dos, cuando concluye en contratos vendidos (CV) y en anulados (CA) (ver figura 10). Al comparar éstos, se percibe que para que los

contratos no sean anulados los ciclos sí o sí deben existir. En CA, cuando se llega a la actividad *contratos rechazados* el proceso se queda estancado hasta que el contrato finalmente es anulado, lo que claramente es una falencia en cómo se está ejecutando el proceso. Se debe revisar el porqué de estar tanto tiempo detenido en esta actividad, siendo en los casos exitosos aproximadamente 7 días, y en los anulados más de 24 días. Entonces, si no se genera el paso entre rechazar y revisar el contrato, éste será anulado. Otro hecho que se destaca es que en los dos modelos siempre la actividad *contratos vendidos* demora mucho tiempo, 15 días aproximadamente cuando la venta es exitosa y 70 cuando no. Este último número es extremadamente alto, por lo que también debe ser revisada esta etapa del proceso y entender por qué cuando ya se ha concretado la venta ésta demora en finalizar formalmente.

Se puede establecer entonces que la hipótesis 1 es válida. Existen tiempos elevados dentro del proceso en generar y revisar el contrato, lo que desencadena en que el cliente anule el servicio solicitado.

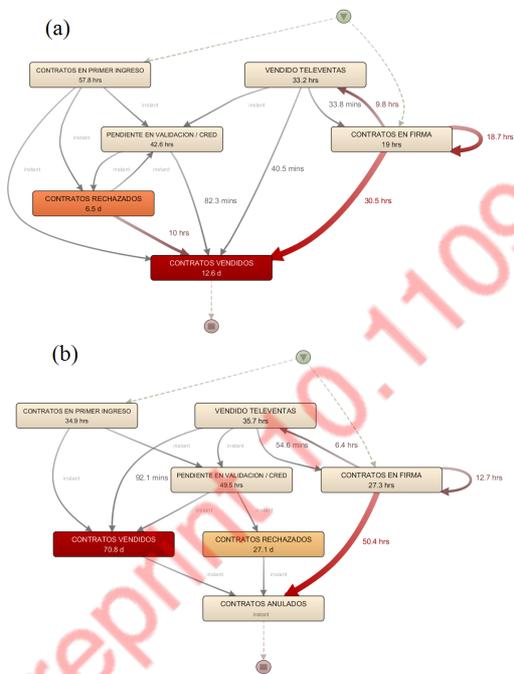


Figura 10. (a) Modelo generado por Disco para los casos que terminan en la actividad *contratos vendidos*. (b) Modelo generado por Disco para los casos que terminan en la actividad *contratos anulados*.

Hipótesis 2: Ejecutores del proceso no tienen roles claros; realizan diversas actividades sin ser especialistas.

Para poder validar esta hipótesis lo que se hizo fue aplicar algoritmos de análisis organizacional de ProM 6.5, específicamente fueron tres:

1. *Similar task*: agrupa a los ejecutores en base a quiénes realizan actividades similares. Permite identificar roles dentro del proceso.
2. *Working together*: relaciona a los ejecutores sobre la base de quiénes trabajan en los mismos casos.
3. *Handover of work*: permite observar qué recursos entregan trabajo a otros recursos (quiénes son consecutivos).

Los resultados de los tres fueron similares, ninguno consiguió agrupar a los ejecutores, siendo siempre grafos conexos (todos los nodos unidos entre sí por algún camino) y densos (grados de los nodos elevados). La figura 11 muestra lo que entregaron los tres algoritmos, ordenando los nodos en una distribución de grado radial, mientras mayor sea el grado del nodo más cerca del centro del plano estará.

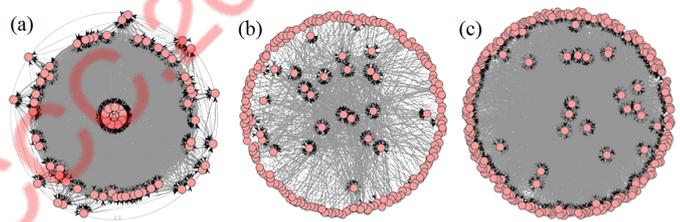


Figura 11. Grafos dirigidos obtenidos a partir de los algoritmos: (a) *similar task*, (b) *handover of work* y (c) *working together*.

En base a lo anterior, es posible determinar que la hipótesis 2 es verdadera, es decir, no existen roles definidos para los ejecutores dentro de la organización. Esto influye directamente en el desempeño del proceso: no existen ejecutores especialistas en las diversas tareas que se realizan, lo que implica una no estandarización de las actividades y que las tareas que componen el proceso no se realicen de la mejor forma. Por el contrario, si existieran roles definidos, el desempeño de los ejecutores sería mucho mejor ya que realizarían menos actividades, pero conocerían muy bien cada una de éstas.

En el desarrollo por validar esta hipótesis se generó información extra. Utilizando las configuraciones de vista de resultado que traen consigo los algoritmos de análisis organizacional de ProM 6.5, fue posible detectar qué nodos son los que tienen más conexiones con los demás (qué ejecutores se relacionan más con el resto) a través de aplicar una distribución de grados sobre el grafo, la que genera niveles radiales sobre el plano en donde se sitúan los nodos, así los con mayor grado se sitúan más cerca del centro (los que tienden a ser un número reducido), por lo que es fácil identificarlos considerando además que han eliminado los arcos del grafo. La figura 12 muestra el resultado de lo generado por el algoritmo *working together*.

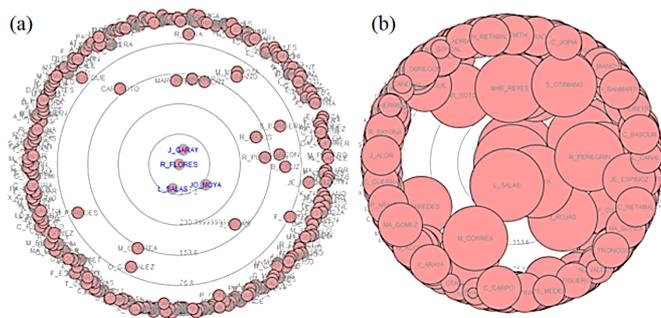


Figura 12. Grafo dirigido obtenido a partir del algoritmo *working together* utilizando una distribución de grado como ubicación de nodos y eliminando los arcos de la imagen. El grafo (a) corresponde a su vista normal, y en (b) el tamaño del nodo representa su grado (mientras más grande éste, mayor es su grado).

En este caso, fue posible identificar a los 4 ejecutores que más se relacionan con el resto (nodos con mayor grado), lo que podría entenderse como los que tienen un “peso” mayor dentro del proceso. El paso siguiente fue volver a Disco para generar un modelo para cada una de estas personas a través de filtros adecuados, y determinar los porcentajes de anulación de contratos de cada uno. Los resultados fueron: Salas 14,5%, Flores 11,5%, Moya 10% y Garay 9,2%, con un 5,3% de diferencia de anulaciones entre Salas (con mayor porcentaje de anulación) y Garay (con el menor). Cabe destacar que este último análisis no sólo aplica para detectar qué ejecutores son los que más anulan contratos cuando realizan actividades en trazas determinadas, sino también puede ser utilizado para detectar quién vende más contratos de forma exitosa, quién demora más/menos tiempo en concluir el proceso, entre muchos otros análisis posibles.

Hipótesis 3: Clientes antiguos anulan menos que los clientes nuevos.

Para poder rechazar o validar la tercera hipótesis, se utilizaron dos herramientas computacionales distintas: Rapid Miner y Disco, poniendo en práctica el ciclo presente en la metodología que se genera entre las etapas de *análisis de datos* y *análisis de proceso*. La primera herramienta es un *software* utilizado en minería de datos, y fue en esta ocasión ocupado para encontrar la variable que más influye en las anulaciones. De manera complementaria, se utilizó Disco para apoyar los resultados obtenidos.

En primer lugar, se tuvo que transformar el *log* de eventos de forma tal de poder discriminar qué casos terminan anulando y cuáles no, lo que se consiguió gracias a la herramienta Excel. Se agregó una nueva columna a la planilla llamada *anulación*, la que puede tener dos valores: *anula* o *no anula*. Se clasificó cada registro en base al atributo *to_neto_anulado*, si éste era igual a 0 no anulaba el contrato, en caso contrario sí lo hacía. Luego de agregar esta columna se quitaron los duplicados de la tabla considerando como único cada *case_id*. Después, se hizo uso del *software* Rapid Miner para generar un árbol de decisión en donde la variable *anulación* fue identificada como variable a predecir. Además, cabe destacar que el criterio ocupado para

la clasificación fue *information gain*, la cual corresponde a elegir variables para ser *spliteadas* de manera que se gane entropía en el árbol. Con respecto a los parámetros utilizados, su máxima profundidad se estableció en 4, con una mínima ganancia de información de 0,15.

Como se mencionó en la sección anterior, el registro de eventos está compuesto por varios atributos, de los que destacan: *origen de venta*, *código de la sucursal*, *región de la sucursal*, *antigüedad del cliente*, *forma de pago*, *ejecutor* y *monto a cobrar*.

Del árbol de decisión obtenido se desprende que la variable más predictiva en cuanto a la anulación es el *origen de venta* (*no_ti_cuenta*). Cabe destacar que existen 9 diferentes orígenes de venta, en donde cada uno define si es de tipo *potencial* o *cartera*, clientes nuevos y antiguos respectivamente. Se infiere del árbol que si la venta proviene de un origen de tipo *cartera* en general no se anula. Ahora bien, si proviene de una de tipo *potencial* va a depender de cuál es el origen. Así, si las ventas proceden de *dirección comercial potencial* se puede predecir que la mayoría serán anuladas; si proviene de *especialista potencial* o *grandes cuentas potencial*, probablemente la venta no se anule; si proviene de *televentas potenciales* o *ventas directas potenciales* y si el cliente es nuevo se puede presagiar que será anulada, y en caso de que sea nuevo no lo será. Esto apoya la hipótesis que indica que los clientes nuevos son más propensos a anular que los clientes antiguos, ya que la mayoría de las anulaciones no provienen de las ventas de tipo *cartera* (clientes antiguos).

Con la información obtenida con Rapid Miner se utilizó Disco para continuar el análisis. Al observar el atributo *fl_cartera* (indica si cliente es antiguo o nuevo) del *log* de eventos, se puede observar que aproximadamente un 56% de los casos considerados son clientes antiguos, mientras que los restantes corresponden a clientes nuevos. Así, se refleja una simetría en cuanto a los clientes nuevos y antiguos atendidos. Ahora, si se filtran los casos en que existe anulación se puede ver que un 88% de los clientes que anulan corresponden a clientes nuevos, mientras que sólo el 12% restante corresponde a los clientes antiguos. También es factible realizar una comparación entre la duración de los casos de los clientes nuevos contra la de los clientes antiguos, como se aprecia en la tabla IV. De ella se deduce que en general el proceso con clientes nuevos dura 3 veces más que con clientes antiguos. Sin embargo, si se comparan los casos en que hay anulación, los clientes antiguos esperan más que los clientes nuevos. Es así como se puede concluir que los clientes antiguos están mayormente dispuestos a esperar antes de anular que los clientes nuevos.

Tabla IV. Comparación de duración promedio del proceso entre clientes nuevos contra clientes antiguos, considerando la totalidad de los casos y cuando terminan anulados.

Casos	Tipo de Cliente	
	Nuevo	Antiguo
Total	41,6 días	10,6 días
Terminan Anulados	72,4 días	90,8 días

La figura 13 muestra el modelo de anulaciones para los clientes nuevos. En ésta se destaca la gran duración de la actividad *contratos vendidos*, que contrastada con el modelo de la figura 10 (a) existe una diferencia de 55,3 días promedio. Es posible identificar que el “cuello de botella” en los casos en que el cliente es nuevo está en *contratos vendidos*.

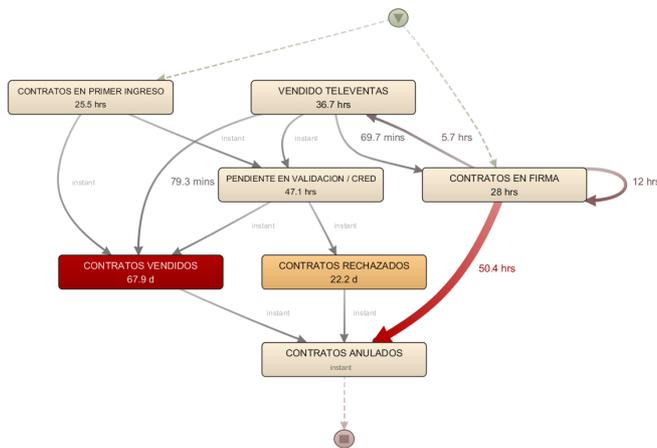


Figura 13. Modelo del proceso desde la perspectiva temporal para los clientes nuevos que concluye anulando el servicio.

Finalmente, se puede ver que existe una alta correlación entre los clientes nuevos y la anulación, lo que se traduce en un gran tiempo de espera. Esto indica que los clientes nuevos anulan porque el proceso se estanca poco antes de terminar. Así, la hipótesis 3 queda validada.

VII. CONCLUSIONES Y TRABAJO FUTURO

En este trabajo se propuso realizar la extensión a una metodología para analizar procesos ya existente llamada *Process Mining Project Methodology* (PM²), la que fue aplicada en un caso de estudio real. A través de ella, fue posible entender las causas de un problema existente en uno de los principales procesos de negocio de una empresa del rubro publicitario, el de *ventas y anulaciones de contratos*. Para ello se definieron diferentes hipótesis que fueron estudiadas utilizando técnicas de minería de procesos, apoyadas algunas con minería de datos.

Sobre la metodología base PM² que se ha empleado en este trabajo, se puede mencionar que es bastante sencilla e intuitiva. Al extenderla y robustecerla incluyendo un análisis de datos cuando sea requerido, se logra obtener más información del proceso gracias al apoyo de otro tipo de técnicas (las de minería de datos) que pueden ser utilizadas sobre los mismos datos. Así, es posible generar un flujo de *feedback* entre los análisis ejecutados mediante el uso de estas dos disciplinas de minería (de procesos y de datos), como se vio en el análisis hecho para determinar si la hipótesis 3 era válida o no. En este caso sólo fue utilizado el algoritmo de árboles de decisión para aportar información, en este caso detectar cuál de todos los atributos discriminaria de mejor forma qué tipo de clientes anulan más

contratos, pero es claro que muchos de los algoritmos de minería de datos podrían ser aptos para aplicar, como técnicas de asociación, regresión (modelos probabilísticos), agrupamiento, entre otras; siempre con el fin de comprender de mejor forma cuál es el análisis correcto que se debe hacer utilizando las técnicas de minería de procesos.

En base al objetivo principal del trabajo desarrollado (caso de estudio), se puede establecer que las tres hipótesis planteadas en un comienzo fueron validadas, lo que claramente genera información de cómo se realiza el proceso en cuestión. Esto se logró gracias a la obtención de los registros de eventos, lo que fue bastante difícil en un comienzo debido a la cantidad de fuentes de datos, lo que implicó mucho tiempo para obtener un buen *event log*. Lo anterior permitió un óptimo funcionamiento de los algoritmos de minería de procesos en general, tanto en el descubrimiento del proceso como en posibles análisis sobre los modelos, concluyó en un desarrollo exitoso. Para cada hipótesis de puede concluir:

- Sobre la primera hipótesis era de esperar que fuese validada ya que es sabido que el proceso demora mucho. Aun así, fue posible identificar en qué actividades el flujo se estanca y cuáles podrían ser los motivos, intentando relacionar los resultados con el negocio de la empresa. Debe existir una mejora constante en la creación del contrato y en la mejora de éste (fase en que está *pendiente/rechazado*), tanto en acuerdo con el cliente como revisión posterior, esto para que sea vendido exitosamente y no sea anulado después.
- Considerando la segunda hipótesis, fue posible establecer que la organización no tiene roles definidos, lo que claramente afecta su desempeño. Es muy importante que esto pueda corregirse, ya que se cree que eliminando esta falencia el proceso podrá llevarse a cabo de una manera mucho más eficiente ya que existirían ejecutores especialistas en las tareas, por lo que no deberían haber tantas demoras.
- Respecto a los resultados de la tercera hipótesis, se propone que la empresa se enfoque en disminuir los tiempos de espera, en especial cuando se trata de clientes nuevos. Más aún, se debería tener especial atención en la actividad *contratos vendidos*, identificada como “cuello de botella”. También, se sugiere que los orígenes de venta se especialicen en clientes nuevos o antiguos. Esto idealmente es así, ya que cada origen es de tipo cartera o potencial, pero con el análisis hecho en Disco se puede observar que esto no se cumple. Se destaca también que cualquier cambio en beneficio de los clientes nuevos, en específico los que aceleren el proceso de contratación, serán bien acogidos considerando que este segmento de personas es la más propensa a anular si de tiempos elevados se trata.

Como trabajo futuro es imperante realizar la última fase de la metodología planteada, la mejora del proceso. Además, la cantidad de algoritmos de minería de datos es muy grande, por lo que aún quedan muchos que aplicar e inferir nueva

información que oriente desde otra perspectiva el desarrollo del análisis de minería de procesos, para esto se deberá realizar un estudio previo de cómo aplicar estas técnicas y cuáles son los beneficios de cada una. En base a poner más énfase en el uso de minería de datos, se estima que sería muy provecho el analizar e incluir un marco de trabajo ya establecido para esta disciplina, el que pueda complementar el método PM² a través de etapas bien definidas. Por último, en este trabajo se han utilizado datos para atender solamente un proceso de una organización, por lo que para validar la extensión del método se requiere replicarlo en muchas otras instancias.

REFERENCIAS

- [1] M. Weske, Business process management: concepts, languages, architectures. Springer, 2012.
- [2] Learned from Successful BPM Organizations. Business Rules Journal 12.10 (2011).
- [3] Van Der Aalst, W., Adriansyah, A., De Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... & Burattin, A. (2011, August). Process mining manifesto. In International Conference on Business Process Management (pp. 169-194). Springer Berlin Heidelberg.
- [4] van Eck, M. L., Lu, X., Leemans, S. J., & van der Aalst, W. M. (2015, June). PM²: A Process Mining Project Methodology. In International Conference on Advanced Information Systems Engineering (pp. 297-313). Springer International Publishing.
- [5] M. Weske, Business process management: concepts, languages, architectures. Springer, 2012.
- [6] W. M. Van der Aalst, "Process-aware information systems: Lessons to be learned from process mining. In Transactions on Petri Nets and Other Models of Concurrency II", Springer Berlin Heidelberg, 2009, pp 1-26.
- [7] Rojas, E., Munoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal of biomedical informatics*, 61, 224-236.
- [8] Trcka, N., Pechenizkiy, M., & Van der Aalst, W. M. P. (2011). Process mining from educational data (Chapter 9).
- [9] Perez-Castillo, R., Weber, B., Pinggera, J., Zugel, S., de Guzman, I. G. R., & Piattini, M. (2011). Generating event logs from non-process-aware systems enabling business process mining. *Enterprise Information Systems*, 5(3), 301-335.
- [10] Arias, M., & Rojas, E. (2014, September). Deciphering event logs in SharePoint Server: A methodology based on process mining. In *Computing Conference (CLEI), 2014 XL Latin American* (pp. 1-12). IEEE.
- [11] Arias, M., Rojas, E., Munoz-Gama, J., & Sepúlveda, M. (2015, August). A Framework for Recommending Resource Allocation based on Process Mining. In *3th International Workshop on Decision Mining and Modeling for Business Processes (DeMiMoP)*, 2015.
- [12] T. Murata, "Petri nets: Properties, analysis and applications." *Proceedings of the IEEE* 77.4. 1989, pp 541-580.
- [13] S. A. White, "Introduction to BPMN." IBM Cooperation. 2004, pp 2008-029.
- [14] W. M. Van Der Aalst and A. H. T. Hofstede, "YAWL: yet another workflow language." *Information systems* 30.4 . 2005, pp 245-275.
- [15] L. Fischer, ed. *The Workflow Handbook*, 2002. Future Strategies Inc., 2002.
- [16] Koster, A. M., & Munoz, X. (2009). Graphs and algorithms in communication networks on seven league boots. In *Graphs and Algorithms in Communication Networks* (pp. 1-59). Springer Berlin Heidelberg.
- [17] W. M. P. Van der Aalst, B. F. van Dongen, J. Herbst, L. Maruster, G. Schimm, and A. J. M. M. Weijters. *Workflow mining: A survey of issues and approaches*. Data and Knowledge Engineering. 2003.
- [18] A. J. M. M. Weijters and W. M. P. Van der Aalst, "Rediscovering workflow models from event-based data using little thumb." 2003.
- [19] A. K. A. de Medeiros, *Genetic Process Mining* (PhD Thesis). Eindhoven University of Technology, Eindhoven. 2006.
- [20] C. W. Günther and W. M. Van der Aalst, "Fuzzy Mining: Adaptive Process Simplification Based on Multi-Perspective Metrics." In G. Alonso, P. Dadam, and M. Rosemann, editors, *International Conference on Business Process Management (BPM 2007)*, volume 4714 of *Lecture Notes in Computer Science*. Springer, Berlin, 2007, pp 328-343.
- [21] Leemans, S. J., Fahland, D., & van der Aalst, W. M. (2013, June). Discovering block-structured process models from event logs-a constructive approach. In *International Conference on Applications and Theory of Petri Nets and Concurrency* (pp. 311-329). Springer Berlin Heidelberg.
- [22] Rozinat, A., & van der Aalst, W. M. (2008). Conformance checking of processes based on monitoring real behavior. *Information Systems*, 33(1), 64-95.
- [23] W. M. Van der Aalst, *Process mining: discovery, conformance and enhancement of business processes*. Springer, 2011.
- [24] W. M. P. Van der Aalst, et al. *ProM: The Process Mining Toolkit*, 2009.
- [25] Bose, R. J. C., Mans, R. S., & van der Aalst, W. M. (2013, April). Wanna improve process mining results?. In *Computational Intelligence and Data Mining (CIDM), 2013 IEEE Symposium on* (pp. 127-134). IEEE.