**Propuesta de Tesis de Maestría en Ingeniería de Software**

1. Nombre y apellido del Alumno/Tesista. Victor Martinez
2. Nombre y apellido del Director/es y si correspondiera Codirector/es.

Directores: Dra. Laura Lanzarini – Dr. Franco Ronchetti

1. Título del Tema de Tesis propuesto

**Minería de Procesos y Minería de Datos en la Ingeniería de Software - Análisis del proceso de distribución de piezas postales.**

1. **Objetivo**

*En este punto se indicará claramente y con una extensión no mayor a 400 palabras el objetivo general de la Tesis, los temas particulares que abordará y el aporte que resultará de su concreción.*

La minería de procesos es un conjunto de técnicas que permiten extraer conocimiento y facilitar el análisis de los procesos basados en registros de eventos.   
Se considera un evento una actividad realizada por un caso en particular. Por ejemplo, en una operación postal tradicional se deja registro del ingreso de la pieza a correo y todos los pasos relacionados a su distribución, el registro de los movimientos de un producto puede ser analizado a través de la Minería de Procesos para descubrir modelos que describan el recorrido realizado por dicho producto desde el ingreso a correo hasta ser entregado al cliente.

La modelización del proceso no sólo permite comprender el funcionamiento real y contrastarlo con el diseñado originalmente, 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.

El objetivo de esta tesis es utilizar la minería de procesos para analizar los registros de eventos correspondientes a la distribución de piezas realizadas por la empresa Correo Argentino durante los años 2017 hasta 2020. Por las características propias del proceso, en ocasiones, debido a errores humanos u otras fallas, se producen retrasos en la distribución o desvíos operativos que impactan en la calidad del servicio.

En esta tesis se propondrán estrategias para identificar desvíos o retrasos en el proceso y posibles mejoras. Para ello, se modelizará el proceso actual con técnicas de Minería de Procesos y luego se comparará contra el ideal o el propuesto originalmente para encontrar diferencias, cuellos de botella y sugerir posibles mejoras.

Además se agruparán los casos problemáticos utilizando técnicas de Minería de Datos para detectar características similares que describan el motivo del retraso.

Se verificará en base a la media de los casos finalizados aquellos que estén en curso y puedan ser desvíos, enviando alertas sobre posibles casos problemáticos antes de su finalización facilitando la solución y mejorando la calidad del servicio.

El conocimiento adquirido será de ayuda tanto en el diagnóstico correcto del problema como en la formulación de estrategias de solución.

Los resultados alcanzados y los modelos obtenidos estarán dirigidos a personas que no estén familiarizadas la minería de procesos y que no tengan un perfil técnico expresados de una forma clara que facilite su interpretación, además podrán servir de base para estudios posteriores que contribuyan al seguimiento del problema planteado y al análisis de resultados luego de la aplicación de las medidas que pueda adoptar la empresa.

1. Motivación /Estado del Arte del Tema

*En este punto se resumirá el contexto científico/tecnológico/académico que justifica el desarrollo de una Tesis en la temática. La extensión de este punto no debe exceder las 2 páginas. En ella se pueden hacer referencias/citas que refuercen la motivación que origina la propuesta.*

Según la definición de la IEEE 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 es una disciplina madura que 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.

La minería de procesos es una técnica relativamente nueva que se enfoca en mejorar procesos de las organizaciones, se busca obtener conocimiento a través registros o logs de eventos.

En sus comienzos la minería de procesos se encontró con el problema de obtener cantidad y calidad en los datos, en los últimos años debido al avance tecnológico creció exponencialmente el volumen de datos disponibles.

Actualmente una gran cantidad de herramientas de software dejan logs de las tareas realizadas ya sea en archivos de texto o en bases de datos, estas herramientas pueden ser desarrollos propios o empresariales como software de gestión ERPs (SAP), de ventas CMRs, de procesos productivos MRPs, de almacenamiento WMSs, de mantenimiento (bug tracking), de control de calidad, o correos electrónicos, incluso con las nuevas tendencias aparece un gran volumen de datos que se puede explotar por ejemplo las redes sociales como Facebook, Instagram, Twitter, etc. proporcionan un montón de información de la actividad de los usuarios (existen técnicas de process mining dedicadas a analizar redes sociales), o el rastro dejado por internet de las cosas (IOT) en dispositivos del hogar (como Alexa de Amazon) o aparatos domésticos inteligentes.

Dada la gran variedad de logs disponibles las técnicas de minería de procesos se pueden aplicar en un amplio espectro de actividades por ejemplo en la producción de productos, distribución, almacenamiento, logística, medicina, desarrollo de software, gestión de empresas u organizaciones públicas o privadas.

Las técnicas se adaptan prácticamente a cualquier proceso de negocio debido a que son genéricas, se pueden aplicar siempre que los datos posean las siguientes características:

1. Deben estar almacenados en uno o más registros de eventos
2. En el registro debe haber datos de varias instancias
3. Cada instancia debe tener sus eventos
4. Cada evento debe tener el identificador de la instancia a la que pertenece, una descripción, y un timestamp de cuando se realizó
5. Opcionalmente se puede agregar información acerca de quién realizó el evento, la fecha de finalización y cualquier dato relacionado al negocio

Hoy uno de los desafíos es utilizar de forma eficiente los datos para realizar análisis de calidad, dada la gran variedad y complejidad de los procesos operativos es importante vincular la ingeniería de software a la minería de procesos generando herramientas obtengan resultados rápidos y faciliten el análisis para las personas que tienen que tomar decisiones.  
La amplia variedad de datos que hay actualmente almacenada en diversas fuentes requiere también un procesamiento previo de los mismos para ajustarlos a la forma requerida.

La minería de procesos utiliza técnicas de la minería de datos y se combina con esta para obtener información relevante de un proceso de negocio utilizando las trazas dejadas en los logs de eventos.

Se puede realizar la minería de procesos de tres formas diferentes:

1. Analizando un histórico de eventos se descubre el proceso real utilizando diferentes algoritmos (Alpha, Inductive Miner, Heursitics Miner) sin información previa.
2. Realizando una verificación de conformidad, para ello se contrasta un proceso (puede ser el descubierto o uno existente) con el log de eventos para comprobar cuántos casos se ajustan al proceso y cuales son desvíos
3. Buscando mejoras al proceso, luego del análisis del log, se identifican cuellos de botella y se sugieren cambios en función de los resultados.

Basándose en estos datos se pueden aplicar las técnicas de esta metodología para encontrar y verificar el proceso real, no el que se diseñó originalmente sino el que se usa diariamente.

La minería de procesos se ubica como un nexo entre la gestión de procesos y la minería de datos, buscando generar información de fácil interpretación para personas que no tengan un gran conocimiento técnico, es decir no hace falta ser un científico de datos para interpretar los resultados obtenidos de un buen análisis basado en minería de procesos.

Se busca aplicar los conceptos de la ingeniería de software a la minería de procesos en un caso práctico, demostrando la problemática encontrada al descubrir un proceso real, interactuar con un volumen importante de datos y generar una herramienta que permita mostrar resultados claros de forma eficiente.Para ello se analizarán los datos de distribución de piezas postales.

El proceso de distribución de piezas postales tiene una gran complejidad debido a las características de cada producto, por ejemplo algunos tienen diferentes estándares de entrega (en los tiempos requeridos), o distinta cantidad de intentos de entrega o diferente forma de ingreso (algunos se imprimen en un centro de impresión nacional, otros se imprimen localmente en cada lugar de distribución). Sumado a que el correo debe llegar a todo el país con diferente equipamiento tecnológico (en algunos lugares no hay red ni internet), todo esto presenta un gran desafío para dar un servicio de calidad.

Las herramientas de calidad actuales se enfocan en los datos almacenados de las distribuciones finalizadas y si bien detectan varios problemas en la distribución lo hacen con un análisis posterior, siempre sobre casos finalizados.

Se generará una herramienta que analice los datos, detecte los desvíos y de apoyo en los mismos, enviando alarmas antes de su finalización para mejorar el servicio en este caso particular.

* van der Aalst W. et al. (2012) Process Mining Manifesto. In: Daniel F., Barkaoui K., Dustdar S. (eds) Business Process Management Workshops. BPM 2011. Lecture Notes in Business Information Processing, vol 99. Springer, Berlin, Heidelberg. <https://doi.org/10.1007/978-3-642-28108-2_19>

Process Mining Manifiesto: breve descripción de los objetivos de la minería de procesos, la metodología, técnicas utilizadas y objetivos.

* van der Aalst, W., Weijters, A., Maruster, L.: Workflow mining: discovering process models from event logs. IEEE Trans. Knowl. Data Eng. 16(9), 1128–1142 (2004)

En este trabajo se describe la forma de extraer datos de los archivos de eventos y las principales técnicas de minería para encontrar la información e interpretarla dentro de los datos.

* P. Markowski and M. R. Przybyłek, "Process mining methods for post-delivery validation," 2017 Federated Conference on Computer Science and Information Systems (FedCSIS), Prague, 2017, pp. 1199-1202, doi: 10.15439/2017F372.

En este paper se brinda un ejemplo de implementación de minería de procesos para realizar una validación en parte del proceso de construcción de piezas mecánicas y su almacenamiento realizando controles de calidad.

1. Temas de Investigación

*Deben indicarse sintéticamente los temas centrales que el Tesista investigará en el desarrollo de la Tesis.*

* Estudio y análisis de diferentes técnicas de representación de secuencias de eventos.
* Estudio de técnicas de Minería de Datos para determinar similitudes entre trazas (secuencias de eventos) utilizando técnicas de agrupamiento. Se trabajará con técnicas de clustering aplicadas sobre distintas representaciones.
* Estudio de técnicas de Minería de Procesos para descubrir el proceso real utilizado en la distribución de piezas postales, realizar una verificación de conformidad para verificar cuales son las que se ajustan a este proceso y cuales son desvíos, evaluar posibles mejoras al proceso y alertar en forma temprana posibles desvíos para su tratamiento.
* Estudio e implementación de distintas técnicas de visualización buscando destacar las partes del proceso que pueden ser mejoradas.

1. Desarrollos/Trabajo Experimental a Realizar

*En el caso que la Tesis contemple la realización de trabajo experimental debe indicarse sintéticamente cual sería y el producto final (prototipo, mediciones, evaluaciones comparativas, etc) que resultará del trabajo propuesto, así como el impacto en el ámbito concreto de aplicación.*

Las tareas experimentales y de desarrollo de esta tesis son las siguientes:

* Construcción de la vista minable a partir de la recopilación e integración de los datos de secuencia de eventos a un formato adecuado para las técnicas de minería de datos y de procesos.
* Implementación de una técnica de agrupamiento que identifique similitudes entre trazas.
* Construcción del modelo del proceso a partir de los registros de eventos y teniendo en cuenta el resultado de las técnicas de agrupamiento.
* Visualización del modelo destacando las zonas más congestionadas.
* Construcción y aplicación de dispositivos de evaluación del modelo obtenido para establecer alarmas.

1. Esquema de Plan de Trabajo C/Actividades y Tiempos

*Se indicarán las actividades principales del desarrollo de la Tesis y una distribución tentativa de tiempos. Debiera servir como un documento de control de la ejecución de la propuesta.*

ACTIVIDAD / Tiempo Estimado

|  |  |
| --- | --- |
| Lectura bibliografía Minería de Procesos y Minería de Datos | 5 meses |
| Búsqueda y análisis de investigaciones similares | 1 mes |
| Recopilación e integración de datos | 1 mes |
| Agrupamiento de secuencias de eventos | 3 meses |
| Elección de técnicas y construcción de modelos | 3 meses |
| Visualización del modelo del proceso | 3 meses |
| Evaluación e interpretación de los patrones descubiertos | 3 meses |
| Redacción de Tesis | 6 meses |

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *ACTIVIDAD* | *M1* | *M2* | *M3* | *M4* | *M5* | *M6* | *M7* | *M8* | *M9* | *M10* | *M11* | *M12* |
| Lectura bibliografía de Minería de procesos y de datos |  |  |  |  |  |  |  |  |  |  |  |  |
| Búsqueda y análisis de investigaciones similares |  |  |  |  |  |  |  |  |  |  |  |  |
| Recopilación e integración de datos |  |  |  |  |  |  |  |  |  |  |  |  |
| Agrupamiento de secuencias de eventos |  |  |  |  |  |  |  |  |  |  |  |  |
| Elección de técnicas y constr. de modelos |  |  |  |  |  |  |  |  |  |  |  |  |
| Visualización del modelo del proceso |  |  |  |  |  |  |  |  |  |  |  |  |
| Evaluación e interpretación de los patrones descubiertos |  |  |  |  |  |  |  |  |  |  |  |  |
| Redacción de Tesis |  |  |  |  |  |  |  |  |  |  |  |  |

1. Posibilidades de Realización en el Ámbito del Tesista

*Se puede indicar sintéticamente las posibilidades que tiene el Tesista en su contexto laboral (académico, profesional) para el desarrollo de la Tesis y si la misma está inserta en un proyecto de I/D específico apoyado por un organismo académico/científico/privado. En este punto se puede señalar algún aspecto metodológico que se considere importante para el desarrollo de la Tesis. También deben explicitarse los recursos con los que cuente para poder llevar adelante el desarrollo de la Tesis (por ejemplo, equipamiento, acceso a bibliografía específica, datos para realizar un muestreo particular para la investigación, etc.)*

Trabajo en desarrollo de sistemas en el correo desde hace más de 14 años, realizando nuevos proyectos de software, corrección, mantenimiento y evolutivo de herramientas relacionadas directa o indirectamente con la distribución para el cliente interno o externo. En varias oportunidades detectamos desvíos en los tiempos de entrega, generalmente se descubren por reclamos de algún cliente.

Por ejemplo, encontramos casos de piezas que van dando vueltas por diferentes centros de distribución de forma innecesaria antes de llegar al que les corresponde o algunas que no cumplen con los pasos requeridos para la entrega.

En algunas oportunidades evaluamos los datos relacionados a la distribución pero no logramos encontrar una mejora sustancial, solo se pudo realizar el análisis de casos puntuales, con la minería de procesos y de datos buscamos identificar rápidamente los desvíos para alertar de forma temprana y actuar en consecuencia con el objetivo de reducir los reclamos de los clientes y mejorar la calidad del servicio de distribución.

Contamos con acceso a datos para generar las muestras necesarias, con la infraestructura necesaria para realizar la investigación y con el conocimiento teórico de la metodología utilizada en la minería de procesos (recientemente realicé el curso Process Mining - Data Science in Action en una plataforma online para profundizar el conocimiento en la materia)

Finalmente se propondrá integrar esta herramienta o alguna que incluya su utilización al proceso actual dentro de la empresa

1. Bibliografía Básica Relacionada

*Se citará la bibliografía relacionada más significativa. No se trata de una enumeración extendida de bibliografía sobre el tema general de la Tesis, sino una selección de textos/artículos/sitios WEB de referencia en el tema.*

* Cleiton dos Santos Garcia, Alex Meincheim, Elio Ribeiro Faria Junior, Marcelo Rosano Dallagassa, Denise Maria Vecino Sato, Deborah Ribeiro Carvalho, Eduardo Alves Portela Santos, Edson Emilio Scalabrin (2019). “Process mining techniques and applications – A systematic mapping study”, *Expert Systems with Applications*, Volume 133, Pages 260-295, ISSN 0957-4174.
* M. Pegoraro and W. M. P. van der Aalst (2019), "Mining Uncertain Event Data in Process Mining," *2019 International Conference on Process Mining (ICPM)*, Aachen, Germany, pp. 89-96, doi: 10.1109/ICPM.2019.00023.
* Aalst, W. v. (2016). *Process Mining - Data Science in Action Second edition.* London: Springer.
* Aalst, W. v. (2011). *Process Mining - Discovery, Conformance and Enhancement of Business Processes . London: Springer.*