

Más que minería de procesos y predicción de comportamientos. Un nuevo enfoque.



More than mining processes and prediction of behaviors. A new approach.

Lázaro Luis Acosta Quintana.¹, Orlenys López Pintado.², Yasser Vázquez Alfonso.³ & Velasteguí López Efraín⁴.

Recibido: 13-12-2017 / Revisado: 07-02-2018 Aceptado: 05-03-2018/ Publicado: 01-04-2018

Abstract.

DOI: <https://doi.org/10.33262/cienciadigital.v2i2.70>

As a result of the boom and technological advancement of computing, the limitation in the manual analysis of the data has increased. On a daily basis, information systems store large amounts of respective events to the processes they represent. This situation motivated researchers to look for new alternatives. As a result, process mining arises, with a great impact on the organizational world. This discipline aims to extract non-trivial information, in processes of a varied domain of applications. In this branch, several techniques have been developed to model, extend and monitor processes. One of the key achievements of process mining is the Audit 2.0 model. Unifying most of the results obtained in this discipline, Audit 2.0 substantially changes the role of auditors. However, the scarce results in some areas of process mining limit the implementation of this strategy. For example, in predicting behaviors, being one of the least exploited areas within the discipline. The fundamental motivation of the research focuses precisely on this problem. The study focuses on the limitations of predicting behavior. In that sense, the first stage is devoted to the general analysis of process mining. It shows the imbalance between the fundamental stages of this branch and the isolated stages. Finally, a critical analysis of various machine learning techniques is performed that can be used in the problems de tected.

¹ Departamento de Informática, Universidad Agraria de La Habana, Mayabeque, cubaluis92aq@unah.edu.cu

² Institute of Computer Science, University of Tartu, Tartu, Estonia, orlenyslopez@gmail.com

³ Facultad de Turismo, Universidad de La Habana, La Habana, Cuba yalfos1@gmail.com

⁴ Universidad Técnica de Cotopaxi Ext. La Maná, Latacunga, Ecuador, luis.velastegui7838@utc.edu.ec

Keywords: Process Mining, Auditing 2.0, Predicting Behaviors, Machine Learning

Resumen.

Consecuentemente con el auge y avance tecnológico de la computación, ha aumentado la limitación en el análisis manual de los datos. Diariamente, los sistemas de información almacenan grandes cantidades de eventos respectivos a los procesos que representan. Esta situación, motivó a los investigadores para buscar nuevas alternativas. Como resultado surge la minería de procesos, con un gran impacto en el mundo organizacional. Esta disciplina, tiene como objetivo extraer información no trivial, en procesos de un variado dominio de aplicaciones. En esta rama, diversas técnicas han sido desarrolladas para modelar, extender y monitorear procesos. Uno de los logros fundamentales de la minería de procesos lo constituye el modelo Auditoría 2.0. Unificando la mayoría de los resultados obtenidos en esta disciplina, Auditoría 2.0 cambia sustancialmente el rol de los auditores. No obstante, los resultados escasos en algunas áreas de la minería de procesos, limitan la puesta en práctica de esta estrategia. Por ejemplo, en la predicción de comportamientos, siendo una de las áreas menos explotada dentro de la disciplina. La motivación fundamental de la investigación se enfoca precisamente en esta problemática. El estudio realizado, se centra en las limitaciones de la predicción de comportamientos. En ese sentido, la primera etapa está dedicada al análisis general de la minería de procesos. Se demuestra, el desequilibrio existente entre las etapas fundamentales de esta rama y las etapas aisladas. Finalmente, se realiza un análisis crítico de variadas técnicas de aprendizaje de máquinas que pueden ser utilizadas en la problemática detectada.

Palabras Claves: Minería de Procesos, Auditoría 2.0, Predicción de Comportamientos, Aprendizaje de Máquinas.

Introducción.

1. Minería de Procesos: Etapas Fundamentales y Trabajos Relacionados.

La minería de procesos es una disciplina en crecimiento dentro de la ciencia de la computación. Su objetivo, se enfoca en la extracción de conocimiento sobre procesos de un variado dominio de aplicaciones. En la actualidad, el crecimiento de la informatización ha provocado que las organizaciones almacenen grandes cantidades de datos diariamente. Esto imposibilita el análisis de los procesos sin la utilización de herramientas automatizadas. Como resultado, de manera general se detectan desviaciones basadas en suposiciones y no en los datos, siendo estos los que representan el comportamiento ocurrido. En ese sentido, la minería de procesos es una solución efectiva, ofreciendo técnicas de modelación y

mejora de procesos de manera automática. Los resultados obtenidos en esta área han tenido tanto impacto que en el año 2012 se plasmaron varios de los retos fundamentales y paradigmas que rigen la minería de procesos en un manifiesto (Van der Aalst et al, 2011).

Lanzado por un grupo de expertos conocido como IEEE Task Force on Process Mining, dicho manifiesto constituye uno de los documentos fundamentales de esta disciplina. Además, es un estándar soportado por 53 organizaciones y bajo la contribución de 77 especialistas en el tema. Todas las técnicas de minería de procesos asumen que es posible recuperar los eventos secuencialmente de un log⁵. Cada evento representa una actividad⁶ que fue ejecutada en una de las ocurrencias (instancias) del proceso en cuestión. Comúnmente, los registros de eventos incorporan información asociada a los eventos o las instancias, variando en función del proceso. Los atributos utilizados con más frecuencia son: persona o dispositivo que ejecuta la actividad, fecha de ocurrencia, estado en que se encuentra la actividad (iniciada, suspendida, terminada, etc.), entre otros. Teniendo en cuenta esta información, los datos pueden ser analizados desde diferentes perspectivas (Van der Aalst, 2011).

La perspectiva de control de flujo: Analiza el orden y relación entre las actividades. Tiene como objetivo caracterizar el proceso intentando representar todas las posibles secuencias de actividades.

La perspectiva organizacional: Se centra en los actores involucrados y sus relaciones con las actividades. Su objetivo es estructurar el conocimiento obtenido atendiendo a roles o para obtener el grafo social.

La perspectiva de los casos: Se refiere a los atributos de las instancias. Aun cuando esta representa una ejecución del proceso, puede enriquecerse con información que la caracterice, atendiendo a datos generales de dicha ejecución.

La perspectiva del tiempo: Utiliza la información relacionada con los tiempos de ocurrencia y frecuencias de los eventos. Tiene como objetivo detectar posibles cuellos de botella, así como intentar predecir la duración de los futuros eventos.

Tradicionalmente, las diferentes técnicas de minería de procesos toman como punto de entrada un registro de eventos. A su vez, según sus características, dividen en tres etapas fundamentales esta disciplina (Van der Aalst, 2011) descubrimiento de modelos, chequeo de correspondencia y extensión de modelos. Estas etapas interactúan entre sí con, el objetivo de obtener modelos que representen en la medida de lo posible la realidad de los

⁵ Estándar utilizado en la minería de procesos para el almacenamiento de la información.

⁶ Se corresponde con un paso bien definido dentro del proceso en que se ejecuta.

procesos. Para ello, se deben tener en cuenta las 4 dimensiones utilizadas para evaluar modelos (Adriansyah et al, 2015., Buijs et al, 2012., Dongen et al, 2014., Pintado, 2015., Rozinat y Van der Aalst, 2008 & Van der Aalst et al, 2010)

Compleitud: Mide cuánto del comportamiento ocurrido en el proceso es representado por el modelo obtenido. Un modelo completo debe ser capaz de reproducir todo el comportamiento observado en el log.

Simplicidad: Esta dimensión se centra en determinar de todos los modelos que reproduzcan el comportamiento del proceso, el que sea estructuralmente más simple.

Precisión: Evalúa el modelo atendiendo a cuánto del comportamiento representado no ocurre realmente en el proceso. Tiene como objetivo evitar modelos muy generales, en ese sentido, serán penalizados modelos que reflejen comportamiento no observado en el registro de eventos.

Generalización: Determina cuánto del comportamiento no ocurrido, pero que puede ser correcto, debe ser representado, evitando modelos demasiado precisos.

El principal desafío de los algoritmos de descubrimiento, radica en detectar las diferentes relaciones entre las actividades del proceso, teniendo en cuenta los principales patrones estructurales (Medeiros, 2006). En esta área se han realizado numerosas investigaciones, dando lugar a disimiles algoritmos para descubrir modelos de variadas características. Algunas investigaciones asumen condiciones como completitud y ausencia de ruidos en los registros de eventos, muy difíciles de lograr en entornos reales (Weerdt, Backer, Vanthienen & Baesens, 2012). Entre estos, se encuentra el algoritmo Alpha (Van der Aalst, Weijters & Maruster, 2004), siendo uno de los pioneros en esta rama.

Varias investigaciones se centran en la utilización de heurísticas para el descubrimiento de modelos. Por ejemplo, el algoritmo Heuristic Miner (Burattin A. and Sperduti, 2010., Greco et al, 2006., Weijters y Ribeiro, 2011 & Weijters et al, 2006), que es capaz de lidiar con registros de eventos con ruido, no obstante, falla al determinar tareas duplicadas y no recupera el patrón selección no libre. Sin embargo, el Heuristic Miner se encuentra entre los de mayor aceptación en esta área. Otro algoritmo que emplea heurísticas es el Genetic Miner (Bratosin, 2011., MEDEIROS et al, 2007., Medeiros, 2006 & Turner, 2009), es robusto ante ruidos y es capaz de recuperar todos los patrones estructurales más comunes.

Al igual que el Heuristic Miner, este algoritmo ha servido de guía para otras investigaciones posteriores. Un resultado relacionado (Pintado, 2015) intenta aprovechar las ventajas de ambos y eliminar sus limitaciones, mostrando una nueva estrategia para

obtener modelos robustos. El algoritmo Evolutionary Tree Miner (Buijs et al, 2012 & Buijs, 2014), es una técnica reciente relacionada con los algoritmos genéticos, el cual obtiene resultados considerables utilizando alineamientos entre el log y el modelo (Buijs et al, 2012). Los algoritmos mencionados relacionan estrechamente las tres etapas fundamentales de la minería de procesos. En sus enfoques utilizan técnicas que evalúan modelos, luego, basándose en los resultados obtenidos, construyen extensiones hasta lograr el modelo final como resultado del algoritmo. Esto permite la obtención de modelos con índices de calidad elevados.

Desde otra perspectiva, se encuentra el algoritmo Inductive Miner (Leemans S., Fahland D. & Van der Aalst, 2013), que utiliza teoría de grafos para la obtención del modelo. En su versión inicial, presenta problemas para detectar ciclos complejos, tareas invisibles y no construye el patrón selección no libre. No obstante, en su extensión (Leemans, 2014), se tienen en cuenta los efectos de las secuencias de eventos incompletas para intentar solucionar los problemas anteriores. Otro resultado (Leemans S., Fahland D. & Van der Aalst, 2015), utiliza la información relacionada con los tiempos de ejecución de los eventos, para diferenciar tareas concurrentes de tareas intercaladas.

1.1. Trabajos aislados.

Aunque no incluidas, existen diversas técnicas que contribuyen y se relacionan estrechamente con las tres etapas fundamentales de la minería de procesos. Entre ellas, las técnicas de agrupamiento, que a pesar de no ser propias de esta disciplina, han sido aplicadas con diversos objetivos. El principal, ha sido contribuir al descubrimiento de modelos, intentando solucionar el problema de los llamados modelos espagueti (Van der Aalst, 2011). En ese sentido, se han realizado diversas investigaciones, algunas proponen determinar la similitud de las trazas (instancias) atendiendo a la frecuencia de ocurrencia de sus actividades (Song M., Günther C. & Van der Aalst, 2008). No obstante, este enfoque no es eficaz en presencia de ciclos complejos, construcciones de selección no libre y tareas invisibles. Otros resultados Chandra J y Van der Aalst (2009), asumen que las actividades son letras y las instancias son palabras. Luego, utilizan esta información para determinar la similitud de las trazas basándose en patrones. Otra investigación relacionada (Chandra J y Van der Aalst, 2009), determina la similitud de dos trazas basándose en la cantidad mínima de cambios que se necesitan para igualarlas. Aunque generalmente se utiliza el agrupamiento de trazas como soporte para el descubrimiento de modelos, existen otros enfoques. Por ejemplo, el agrupamiento puede ser utilizado para detectar desviaciones en el proceso (Hompes, s/f). Otra variante puede ser, utilizar el agrupamiento en combinación con las reglas propuestas en DecSerFlow (Chesani et al.2007), permitiendo analizar el proceso con otro enfoque. Además, el agrupamiento de trazas puede ser utilizado desde la perspectiva del tiempo. Un ejemplo de ello (Luengo D y Sepúlveda M., 2011), utiliza los tiempos de ejecución de los eventos para agrupar desde esta perspectiva.

Otro grupo de técnicas se definen bajo el nombre de soporte operacional. Una gama relativamente nueva dentro de esta disciplina que es de gran interés para el análisis de procesos. Dividiéndose en tres etapas: detección, predicción y recomendación de comportamientos, el soporte operacional es una de las áreas de investigación menos explotada en la minería de procesos. Las técnicas descritas previamente, son utilizadas en el análisis de trazas finalizadas⁷ o casos históricos. A diferencia, los resultados en esta área se centran en el análisis de trazas no finalizadas⁸ o casos actuales.

La figura 1 muestra una comparación entre ambos enfoques, teniendo en cuenta las técnicas de soporte operacional y las etapas clásicas de minería de procesos. Nótese que, la detección, predicción y recomendación de comportamientos, están orientadas al trabajo con casos actuales. Sin embargo, la detección también es aplicable a las trazas finalizadas. Por otro lado, las técnicas de descubrimiento, chequeo y extensión de modelos están destinadas al trabajo con casos históricos. Es importante señalar que, el soporte operacional, a pesar de utilizar modelos de procesos en la mayoría de los casos, tiene como objetivo el análisis de trazas. Esto ocurre, debido a que las técnicas de esta área no contribuyen directamente a la modelación. No obstante, pudiese considerarse que el soporte operacional está estrechamente relacionado con la extensión de modelos. El soporte operacional, en conjunto con las etapas fundamentales de la minería de procesos dan lugar al modelo Auditoría 2.0 (Van Aalst et al, 2010), como nueva estrategia para la auditoria de procesos.

enfoque	trazas			modelo				
	accion	chequear	predecir	recomendar	descubrir	chequear	modificar	extender
casos actuales	✓	✓	✓					
casos históricos	✓			✓	✓	✓	✓	✓

costo
tiempo

Figura 1. Comparación de la utilización de las diferentes técnicas de minería de procesos en casos actuales y casos históricos. (Tomada de: Van der Aalst et al. 2010).

⁷El análisis de trazas finalizadas, suele identificarse con el nombre de fuera análisis de línea (del inglés offline).

⁸ El análisis de trazas no finalizadas, también aparece bajo el nombre de análisis en línea (del inglés online).

2. Modelo Auditoría 2.0.

El modelo Auditoría 2.0, presentado en el año 2010 por van der Aaslt y colaboradores, cambia drásticamente el rol de los auditores. Típicamente, el trabajo de los analistas ha estado limitado, debido a los grandes volúmenes de datos almacenados. Con la utilización de esta nueva estrategia de auditoría, las elevadas cantidades de eventos registrados no son una limitante. En consecuencia, posibilita la extracción de conocimiento a mayor escala. En Auditoría 2.0, las diferentes funcionalidades aparecen divididas, debido a las fuentes de datos para las cuales están dirigidas. En un primer grupo se encuentran las técnicas de descubrimiento, chequeo y extensión de modelos, utilizadas a partir de los datos históricos. Luego, a estas técnicas contribuye la comparación entre modelos reales y modelos ideales.

De esta forma, a partir de los casos históricos, se pueden detectar violaciones o inconsistencias en el proceso. Además de detectar violaciones, los modelos pueden ser evaluados con otros objetivos. Por ejemplo, si el modelo que se obtiene debe ser descriptivo, en este caso una traza no representada por el mismo no necesariamente sería una violación. En ese sentido, el modelo podría no estar correctamente construido.

Basándose en el conocimiento adquirido a partir de las trazas finalizadas, y enfocadas en los casos actuales, se encuentran las técnicas de soporte operacional. En este punto, orientadas al funcionamiento en tiempo real, dichas técnicas habilitan a los auditores para el análisis de procesos desde otro enfoque. Aunque el modelo Auditoría 2.0 propone utilizar la mayoría de las técnicas de minería de procesos, fuese conveniente adicionar otros enfoques. Por ejemplo, el agrupamiento de datos, con el objetivo de detectar desviaciones, así como obtener varios modelos que se relacionen en lugar de un único modelo. Otra posible extensión, tendría lugar vinculando otras disciplinas como minería de datos con el objetivo de mejorar las técnicas de soporte operacional (De Leoni M. y Van der Aalst, 2014)

3. Predicción de Comportamientos: Análisis Crítico y Trabajos Relacionados.

Los resultados obtenidos hasta el momento en la predicción de comportamientos, se basan fundamentalmente en la utilización de análisis estadísticos. Entre ellos, destaca una investigación realizada por (Van der Aalst et al. 2011), en la cual se propone una estrategia general para el soporte operacional. En ese caso, se analiza la duración de los eventos y su tiempo de espera antes de comenzar. Con ello, se extiende un modelo que será utilizado para predecir comportamientos. Otras investigaciones (Van der Aalst, Pesic y Minseok, 2010., Van der Aalst, Schonenberg y Song, 2011), trabajan igualmente desde la perspectiva del tiempo utilizando datos estadísticos y modelos de procesos básicos.

En ninguna de las investigaciones mencionadas se analiza la influencia de los atributos de los eventos, así como la influencia de las posibles secuencias de los mismos. Por ejemplo, un modelo de procesos debe tener cierto nivel de generalización, permitiendo comportamientos no reflejados en los datos (pero que pudieran ser correctos). En ese sentido, una predicción puede indicar la ocurrencia de un evento determinado. Sin embargo, para la secuencia de eventos ocurrida puede que esta predicción sea improbable.

Un caso similar ocurre cuando el evento que se predice, representa un comportamiento reflejado por el proceso. No obstante, debido al análisis de las frecuencias de ejecución de los eventos, puede ocurrir que la probabilidad con la que se estima dicha predicción sea incorrecta. Igualmente ocurre con la predicción de atributos, en este caso, lo que sucede comúnmente es que se estiman los valores utilizando métodos puramente estadísticos. Esto imposibilita el análisis de los atributos según las secuencias de eventos ocurridas. Otra variante a utilizar sería representar el proceso utilizando árboles de prefijos, de esta forma solo serían consideradas las secuencias de eventos ocurridas. En comparación con las técnicas mencionadas, este enfoque brinda mayor precisión en las predicciones, no obstante, se pierde la capacidad de generalización. Además, persiste la problemática mencionada con respecto a la influencia de las secuencias de eventos y los atributos.

Recientemente, una investigación realizada por van der Aalst y colaboradores (Nakatumba, Westergaard & Van der Aalst, 2012), propone un meta-modelo para el soporte operacional. Dicho enfoque puede resultar útil para generalizar estrategias que permitan diagnosticar, comparar, predecir y recomendar comportamientos. No obstante, dada la naturaleza de dicha investigación, continúan las limitaciones detectadas.

Según el análisis realizado, el principal reto a la hora de realizar predicciones, radica en la utilización de técnicas capaces extraer conocimiento de los datos de cada suceso. Además, que permitan analizar la influencia de los mismos en las diferentes etapas del proceso. Esta problemática constituye uno de los principales desafíos planteados en el manifiesto de minería de procesos.

4. Aprendizaje de Máquinas: Basamentos Fundamentales y Técnicas de Interés.

El aprendizaje, al igual que la inteligencia, tiene diferentes concepciones según el área de la ciencia en la que se analiza. Del aprendizaje de máquinas o aprendizaje automático, específicamente, se puede decir que tiene grandes similitudes con el aprendizaje de los animales. Existe un gran número de técnicas de esta disciplina que tienen sus basamentos en descubrimientos de carácter biológico, creando a partir de ellos modelos computacionales. De igual manera, se puede considerar que los resultados obtenidos por los investigadores del aprendizaje automático, pueden aportar a las teorías del aprendizaje

animal. Desde el punto de vista de la computación, se puede decir que una máquina aprende cuando cambia su estructura, programa o información, debido a cambios en el entorno que estudia. Similarmente, se dice que una máquina ha aprendido cuando ha sido capaz de estructurar la información, detectando las relaciones significativas. Con ello, debe poder analizar nuevos datos basándose en el conocimiento obtenido. En muchas ocasiones, el campo del aprendizaje automático se solapa con el de la minería de datos y la estadística.

En un inicio, se pudiese considerar la creación de programas de computadoras, ya diseñados y optimizados para resolver las tareas concretas que se necesiten. De esta forma, se evitaría la utilización del aprendizaje automático, sustituyéndolo por algoritmos que resuelvan la misma problemática sin la necesidad de inducir conocimiento. No obstante, existen varias razones para rechazar esta hipótesis, demostrando la necesidad del aprendizaje de máquinas, a continuación, se describen algunas:

1. Las grandes cantidades de datos almacenados respectivos a las problemáticas que se estudian, pueden imposibilitar su análisis de forma manual. Debido a ello, sería complejo diseñar programas desconociendo total o parcialmente el problema a resolver. En ese sentido, las computadoras poseen capacidades que les permiten analizar grandes volúmenes de información en menor tiempo. Además, existen varias técnicas de aprendizaje automático diseñadas para aprender a medida que los datos son obtenidos.
2. Algunas tareas no pueden ser definidas de forma precisa, excepto, atendiendo a ejemplos de entrada y salida. Sin embargo, se desconocen las relaciones entre los datos. En esos casos, convienen técnicas capaces de analizar la información y modificar su estructura interna para representar el conocimiento.
3. Las condiciones del ambiente donde se desarrolla la problemática pueden cambiar constantemente. En ese sentido, las técnicas de aprendizaje automático pueden adaptarse a nuevas condiciones reduciendo la necesidad de rediseñar los programas.
4. Nuevos conocimientos son obtenidos constantemente acerca de las tareas a resolver. El rediseño de los sistemas en esos casos sería una solución poco práctica. No obstante, el aprendizaje de máquinas puede estar habilitado para manejar dichos cambios sin necesidad de rediseño.

Teniendo en cuenta la tipología de los datos de entrada, los algoritmos de aprendizaje automático se dividen en varios grupos. A continuación, se describen de manera general los fundamentales (Harrington, 2012 & Nilsson, 1996)

Aprendizaje supervisado: Se refiere a los algoritmos que parten de ejemplos en forma de tuplas, refiriéndose cada una a los datos de la entrada y su correspondiente salida. El objetivo de estas técnicas es obtener funciones o estructuras que establezcan una correspondencia entre las entradas y las salidas deseadas. Esta gama, a su vez, se divide en

dos grupos: clasificación y predicción de valores numéricos (también conocido como regresión).

Aprendizaje no supervisado: El aprendizaje se lleva a cabo a partir ejemplos formados solamente por datos de entrada, para los cuales se desconocen sus categorías. Los algoritmos de esta área intentan reconocer patrones dentro de los datos, con el objetivo de etiquetarlos y poder etiquetar nuevas entradas. Entre estas técnicas se encuentran algoritmos de agrupamiento, algoritmos para encontrar reglas de asociación y conjuntos de datos frecuentes.

Atendiendo a los dos grupos descritos, se realizó un estudio de las técnicas pertenecientes a cada uno, que se prevé pueden ser utilizadas en la predicción de comportamientos. Lo que resta de esta sección, se dedicó al análisis crítico de las estructuras y algoritmos identificados.

4.1. Aprendizaje Supervisado. Clasificación.

Las técnicas de clasificación tienen lugar cuando la variable a predecir (variable objetivo) es de tipo nominal, aunque una variable nominal puede ser mapeada a una variable discreta. En ese sentido, se espera que un clasificador una vez terminado su aprendizaje, sea capaz de asociar a una entrada su variable objetivo correspondiente. Además, debe ser capaz de generalizar los comportamientos presentes en los datos. La mayoría de las técnicas de esta gama, cuentan con dos etapas: una etapa de entrenamiento y una etapa de predicción de nuevos valores. En la primera etapa, un conjunto de datos es suministrado al algoritmo, para ser utilizados en el proceso de aprendizaje. Es importante señalar que, dichos datos deben reflejar de manera general todo el espacio de entrada, de lo contrario, la estructura puede no aprender correctamente. Algunos de los clasificadores más comunes son los árboles de decisión, el método k-nn conocido como k vecinos más cercanos (del inglés K nearest neighbours), las redes neuronales y las máquinas de soporte vectorial.

Los árboles de decisión (Harrington, 2012., Nilsson, 1996 & Shwartz y Shai, 2014), son estructuras que dividen los datos de entrada atendiendo a sus atributos, hasta llegar a un estado final o decisión. Formalmente, se puede decir que cada nodo representa un estado, mientras que los arcos representan posibles decisiones dado un estado. Los nodos hoja, particularmente, se refieren al resultado final del conjunto de decisiones tomadas. Bajo este principio, los algoritmos que construyen árboles de decisión realizan cortes secuencialmente en los datos, hasta llegar al valor de la variable objetivo. En cada estado, las divisiones en la información se realizan teniendo en cuenta el concepto entropía, basándose en los atributos.

De esta forma, el árbol se construye eligiendo primero aquellos atributos que contengan

mayor cantidad de información. Una vez finalizado el proceso de entrenamiento, la estructura resultante puede ser utilizada tanto para predecir como para tomar decisiones. Los árboles de decisión pueden ser construidos utilizando atributos tanto discretos como continuos. No obstante, el resultado de una predicción siempre es nominal, de ahí que se encuentre en el grupo de los clasificadores. Una de las mayores ventajas de los árboles de decisión, es su facilidad de comprensión por los humanos, convirtiéndolos en un paradigma en la toma de decisiones. Además, pueden tratar de forma consistente con datos que presenten ruidos o atributos irrelevantes. También, se destaca que su construcción es computacionalmente factible. Sin embargo, es necesario señalar que esta estructura es propensa a sobreajuste⁹ (comúnmente se utiliza el término en inglés *overfitting*).

El algoritmo k-nn (Harrington et al, 2014), a diferencia de los árboles de decisión, no construye una estructura en concreto, aunque pueden utilizarse estructuras de datos para su optimización. El objetivo de esta técnica es clasificar un nuevo valor, para el cual se desconoce a qué grupo pertenece. Para ello, el algoritmo se basa en los k vecinos más cercanos al él, donde la distancia se mide en base a los valores de sus atributos. Este enfoque se basa en la premisa de que, si los k elementos más cercanos al ejemplo a predecir se encuentran en su mayoría en un grupo determinado, es probable que el caso a predecir también pertenezca a esta categoría. El valor seleccionado de k, supone en esta técnica, cambios en cuanto a la generalización-especificación, el crecimiento de k es directamente proporcional a la capacidad de generalización. Una extensión de este algoritmo puede ser la utilización de distancias ponderadas, dando mayor peso a los elementos más cercanos. El algoritmo k-nn, pese a que realiza la clasificación con altos valores de precisión y es robusto ante ruidos, es computacionalmente costoso y no recomendado para grandes volúmenes de datos.

Una amplia gama de algoritmos y estructuras se encuentran bajo el nombre de redes neuronales artificiales, constituyendo un paradigma en el aprendizaje computacional. Basadas en las características del sistema nervioso animal, las redes neuronales están compuestas por un conjunto de nodos, interconectados estructuralmente por aristas de pesos ajustables. Diversos tipos de redes neuronales han sido diseñadas, diferenciándose entre sí tanto en estructura como en mecanismos de entrenamiento. Entre las redes más básicas se encuentran Adaline y Perceptrón (Harrington, 2012 & Ben Krose et al, 1993), las cuales no son capaces de resolver problemas no lineales. Por otro lado, el Perceptrón Multicapa (Ben Krose et al, 1993 & Nilsson, 1996), en conjunto con el método de entrenamiento propagación hacia atrás (del inglés *backpropagation*), es capaz de aproximar relaciones no lineales entre datos de entrada y salida. Esta red se ha convertido en una de las arquitecturas más utilizadas actualmente. Las redes neuronales son robustas ante datos que presenten

⁹Es el resultado de sobreentrenar un algoritmo de aprendizaje con ciertos datos para los cuales se conoce su resultado, perdiendo la capacidad de generalización.

ruidos. Además, es importante señalar que son capaces de ajustar su estructura interna para representar la información. De esta forma, no es necesario conocer detalladamente las relaciones entre los datos. No obstante, algunos parámetros importantes deben ser seleccionados por el usuario, por ejemplo: cantidad de capas ocultas, cantidad de nodos por capa, umbrales, etc. En ese sentido, el diseño de redes neuronales ha sido apoyado por la utilización de algoritmos evolutivos, para la obtención de valores que arrojen resultados satisfactorios. La limitación fundamental en la utilización de redes neuronales radica en el alto coste computacional requerido para su entrenamiento. Sin embargo, esta gama de técnicas continúa siendo utilizada e investigada debido a su capacidad de aprendizaje.

Las máquinas de soporte vectorial (Andrew, 2013. & Zhang, 2001), son un conjunto de algoritmos utilizados para problemas de clasificación. Un elevado número de investigadores consideran esta técnica como uno de los mejores clasificadores. Una máquina de soporte vectorial, intenta dado un conjunto de datos de entrada obtener un modelo de predicción, basándose en la obtención de un conjunto de hiperplanos¹⁰ que los dividan. La manera más simple de realizar una separación es utilizando una recta o un hiperplano N-dimensional, no obstante, en la mayoría de los casos esto no es posible. Esta problemática trae como consecuencia un elevado coste computacional para este tipo de algoritmos. Como solución, se plantea la utilización de funciones Kernel para proyectar la información a un espacio de características de mayor dimensión. Las máquinas de soporte vectorial constituyen un mecanismo aceptado en sustitución de clasificadores como las redes neuronales. En ese sentido, cabe señalar que el entrenamiento de estos algoritmos es muy eficiente. Sin embargo, se necesitan definir parámetros como la función Kernel.

Pese a que todas las técnicas analizadas tienen ventajas y desventajas, existe una limitación en común: la problemática conocida como clasificación desequilibrada. Considérese, una situación donde algunas clasificaciones estén representadas por cantidades de elementos potencialmente menores que el resto. Cuando esto ocurre, de manera general los clasificadores presentan problemas para etiquetar las entradas pertenecientes a estos grupos. Diversas investigaciones han sido realizadas para intentar solucionar esta problemática. De los resultados obtenidos, uno de los que ha tenido mayor impacto es el meta-algoritmo AdaBoost (Harrington, 2012 & Shwartz & Shai, 2014).

Utilizando múltiples clasificadores de un mismo tipo, dicho algoritmo los entrena secuencialmente, teniendo en cuenta en cada momento los errores detectados en el entrenamiento anterior. Finalmente, en presencia de un caso de entrada desconocido, éste es sometido a cada uno de los clasificadores, donde cada clasificador influye en el resultado final según haya sido su error en el entrenamiento.

¹⁰ Un hiperplano, en geometría, es una extensión del concepto plano.

4.2. Aprendizaje Supervisado. Regresión.

Al igual que el enfoque de clasificación, la regresión se encuentra en el área de los algoritmos que utilizan aprendizaje supervisado. Esta variante, es utilizada cuando la variable a predecir es numérica y continua, a diferencia de las técnicas de clasificación. El método de regresión lineal (Harrington, 2012 & Montgomery, 2015) es comúnmente utilizado en esta área. La regresión lineal tiene como objetivo modelar las relaciones existentes entre los atributos del caso de análisis y la variable a predecir. Para lograr este resultado, se plantea una ecuación que contiene los atributos multiplicados por coeficientes.

Los algoritmos de regresión intentan determinar los valores de dichos coeficientes. La limitación fundamental de estas técnicas, es que, pese a que reducen el error global, los errores locales influyen notablemente en las predicciones. Una forma de resolver esto es sobrecargando en la función los datos cercanos a los de interés, de esta forma, se logra una predicción más precisa (Harrington, 2012). No obstante, tiene como desventaja, que para cada predicción debemos aplicar el algoritmo de regresión, lo que aumenta considerablemente el costo computacional. En ocasiones, la variable a predecir muestra cambios considerables en algunos segmentos dentro del espacio de valores. En ese sentido, convendría utilizar el método de regresión segmentada (Oosterbaan et al,1990), que plantea analizar los datos por intervalos. Este enfoque mejora los resultados en comparación con la regresión lineal clásica. Otra problemática ocurre cuando las características de los datos, no permiten encontrar una ecuación lineal que los represente. Debido a ello, varias investigaciones han sido llevadas a cabo utilizando regresiones no lineales. En muchos casos, los métodos de regresión son factibles, dado su bajo costo computacional. Además, atendiendo a las características de los datos, existen varios métodos de regresión que pueden ser seleccionados, aumentando el alcance de dicha técnica.

En entornos reales, es común encontrar situaciones en las que no se puedan determinar claramente las ecuaciones que las representen. En ese sentido, existe una variante de regresión que puede ser de utilidad, conocida como regresión basada en árboles (Buja A. & Lee, 2001 & Harrington, 2012).

Lo que propone dicha estrategia, es dividir los datos consecutivamente, hasta lograr que cada división pueda ser representada con un modelo de regresión lineal. Luego, llevar a cabo el entrenamiento para cada división obtenida. Finalmente, la estructura construida, funcionaría de forma similar a los árboles de decisión analizados en el apartado anterior, con la diferencia de que las hojas contendrán modelos de regresión lineal. Esta nueva variante, eleva el costo computacional en comparación con el uso de regresión lineal solamente. Sin embargo, habilita dichas técnicas para analizar datos con relaciones complejas entre sus variables.

4.3. Aprendizaje no Supervisado.

En muchas situaciones reales, a diferencia de los casos analizados previamente, no se tiene una variable respuesta para los datos de entrada. En estos casos, en lugar de clasificar o predecir valores numéricos, lo que se busca comúnmente es extraer información útil, por ejemplo: posibles agrupamientos. El objetivo de estas técnicas es encontrar subconjuntos dentro del espacio de entrada, minimizando las diferencias entre elementos pertenecientes a un mismo grupo. La similitud entre dos elementos de la entrada es computada mediante una función de distancia atendiendo a los valores de sus atributos. Diversos algoritmos han sido desarrollados para automatizar este procedimiento. Uno de los más conocidos es el algoritmo K-Means (Alsabti et al, 1997., Harrington, 2012., Morissette y Chartier, 2013. & Shwartz y Shai, 2014) el cual divide los datos de entrada en K subgrupos, tomando K como parámetro. La principal dificultad para utilizar este algoritmo, es que para entornos reales difícilmente se conoce el número óptimo de K. Sin embargo, algunas investigaciones han sido llevadas a cabo para determinar métricas que aseguren una buena selección de K (Pham, Dimov & Nguyen, 2005)

El algoritmo basado en encadenamiento (Shwartz S. & Shai, 2014), es otra variante clásica para realizar agrupamientos. A diferencia del algoritmo K-means, tiene en cuenta las distancias internas y externas de los grupos. Dicho algoritmo propone analizar iterativamente el espacio de entrada, uniendo cada vez los dos grupos más cercanos en un nuevo grupo.

Diferentes métricas de distancia pueden ser utilizadas: basándose en la distancia mínima, distancia máxima y distancia promedio. Al igual que K-means, se puede contar con un parámetro K, el cual determinaría la finalización del algoritmo una vez obtenidos K grupos. No obstante, existe otra variante utilizando un umbral, que se obtiene teniendo en cuenta la distancia máxima entre dos elementos del espacio de entrada. En este caso, el algoritmo une dos grupos siempre que no sobrepasen el umbral.

Existe otra estrategia que propone agrupar en dos etapas. La primera etapa utiliza una red neuronal conocida como SOM (del inglés Self-Organizing Map) (Richard et al, 1999., Vesanto & Alhoniemi, 2000 & Yin, 2008), para reducir el espacio de entrada en un número menor de prototipos que lo represente de manera aproximada. Luego, los prototipos obtenidos son agrupados utilizando alguno de los algoritmos de agrupamiento analizados previamente. Esta estrategia tiene como ventaja, que puede ser utilizada en grandes volúmenes de datos, minimizando el coste computacional. Una posible extensión de esta propuesta, consiste en utilizar funciones de densidad para agrupar utilizando solamente la primera etapa (Cabanés G, & Bennani, 2010)

En el ámbito de las redes neuronales existe una problemática conocida como Dilema de la Estabilidad y Plasticidad del Aprendizaje. La plasticidad se refiere a la capacidad de una red de aprender nuevos patrones, mientras que la estabilidad permite a una red entrenada poder retener los conocimientos aprendidos. Diseñar redes que cumplan con una de estas características es sencillo, el principal reto consiste en la creación de redes que cumplan ambos objetivos. En las redes más comunes, como Perceptrón Multicapa, Adaline y SOM, el aprender nuevas características puede suponer el olvido del conocimiento ya obtenido, de ahí una limitación considerable. En ese sentido, una nueva gama de redes neuronales ha sido desarrollada bajo el nombre de Teoría de la Resonancia Adaptativa (ART, del inglés Adaptive Resonance Theory). Este modelo de redes se encuentra en el grupo conocido como redes de aprendizaje competitivo, al igual que la red SOM. Estas redes poseen variantes de aprendizaje supervisado (Carpenter A., Grossberg S. & Reynolds J.,1991), y no supervisado (Carpenter & Stephen, 1987., Carpenter & Stephen, 1990., Carpenter & Stephen, 1991 & Grossberg S.1987).

Basándose en las propuestas básicas de estas redes, se han obtenido variantes que utilizan lógica difusa (Carpenter A, & Grossberg ,1992. & Carpenter A, & Grossberg, 1991), con el objetivo de brindar mayor realismo. Además, como característica fundamental se tiene que crean su propia clasificación de lo que aprenden. El principal impacto de las redes ART radica en su capacidad de utilización en tiempo real. Esto viene dado porque la red es capaz de detectar nuevas características en los datos, lo que supone una nueva clasificación, sin llevar al olvido del conocimiento previamente obtenido. A diferencia de los algoritmos de agrupamiento analizados, utilizando las redes de la familia ART, se pueden desarrollar técnicas orientadas a agrupar en tiempo real, sin necesidad de nuevos entrenamientos. Sin embargo, persiste la problemática general en el diseño de redes neuronales, la elección de los parámetros correctos, dígase umbrales, cantidad de capas, etc.

Conclusiones.

- El estudio realizado, en su propósito de analizar el estado actual de las técnicas de predicción de comportamientos, fue dividido en 4 etapas lógicas. La primera parte de la investigación, se dedicó a la minería de procesos, demostrando su impacto en el mundo organizacional. Se analizaron las principales etapas en que se divide esta disciplina, así como algunos trabajos aislados. Por ejemplo, el agrupamiento de trazas y el soporte operacional, que pudiesen considerarse como etapas independientes. La interacción de las diferentes técnicas de esta rama da lugar al modelo Auditoría 2.0. Esto cambia drásticamente la forma de trabajo de los auditores, consolidando la aplicabilidad de la minería de procesos. Específicamente, la predicción de comportamientos, juega un papel fundamental dentro del modelo

Auditoría 2.0, analizando los procesos desde una perspectiva en línea. De esta parte de la investigación, es importante señalar, que la predicción es una de las áreas de investigación menos explotada dentro de la disciplina. A su vez, los resultados obtenidos no logran establecer la predicción de comportamientos como paradigma en la auditoría de procesos. Además, la carencia de investigaciones relacionadas en los últimos años ha contribuido al estancamiento de este enfoque. Finalizando, en concordancia con los investigadores reconocidos en el área, se concluye con uno de los principales retos de la minería de procesos: vincular disciplinas externas a la predicción de comportamientos.

- Atendiendo a la problemática detectada, se realizó un estudio de las técnicas de la rama aprendizaje automático, que puedan ser utilizadas en la predicción de comportamientos. Centrándose específicamente en las técnicas de aprendizaje supervisado y no supervisado, se analizan parcialmente estos grupos, teniendo en cuenta los algoritmos y estructuras de interés. Atendiendo a las características de estas técnicas se concluye, que es de vital importancia su utilización en la predicción de comportamientos. Además, se estima que la combinación del aprendizaje automático y la predicción de comportamientos, puede aportar nuevas funcionalidades y enfoques en la auditoría de procesos.
- Diversas líneas de investigación pueden ser identificadas a partir del estudio realizado. Pudiese ser de interés, tanto la aplicación de otras técnicas de esta disciplina, así como otras disciplinas. Además, los resultados que se obtengan a partir de esta investigación, pueden arrojar nuevas teorías en el soporte operacional en general, dígase detección, predicción y recomendación de comportamientos.

Referencias bibliográficas.

- Arya Adriansyah, Jorge Munoz-Gama, Josep Carmona, Boudewijn F van Dongen, and Wil MP van der Aalst, Measuring precision of modeled behavior, *Information systems and e-Business Management* 13 (2015), no. 1, 37–67.
- Khaled Alsabti, Sanjay Ranka, and Vineet Singh, An efficient k-means clustering algorithm, (1997).
- Alex M Andrew, An introduction to support vector machines and other kernel-based learning methods, *Kybernetes* (2013).
- RP Jagadeesh Chandra Bose and Wil MP van der Aalst, Abstractions in process mining: A taxonomy of patterns, *International Conference on Business Process Management*, Springer, 2009, pp. 159–175.
- _____, Context aware trace clustering: Towards improving process mining results, *Proceedings of the 2009 SIAM International Conference on Data Mining*, SIAM, 2009, pp. 401–412.

- _____, Trace clustering based on conserved patterns: Towards achieving better process models, *International Conference on Business Process Management*, Springer, 2009, pp. 170–181.
- C. Bratosin, *Grid Architecture for Distributed Process Mining*, Ph.D. thesis, Technische Universiteit Eindhoven, 2011.
- J.C.A.M. Buijs, *Flexible Evolutionary Algorithms for Mining Structured Process Models*, Ph.D. thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2014.
- J.C.A.M. Buijs, B.F. van Dongen, and W.M.P. van der Aalst, A genetic algorithm for discovering process trees, In *IEEE Congress on Evolutionary Computation*, IEEE, 2012, pp. 1–8.
- Joos CAM Buijs, Boudewijn F van Dongen, and Wil MP van der Aalst, A genetic algorithm for discovering process trees, *Evolutionary Computation (CEC)*, 2012 IEEE Congress on, IEEE, 2012, pp. 1–8.
- Joos CAM Buijs, Boudewijn F Van Dongen, and Wil MP van Der Aalst, On the role of fitness, precision, generalization and simplicity in process discovery, *OTM Confederated International Conferences. On the Move to Meaningful Internet Systems*”, Springer, 2012, pp. 305–322.
- Joos CAM Buijs, Boudewijn F van Dongen, and Wil MP van der Aalst, Quality dimensions in process discovery: The importance of fitness, precision, generalization and simplicity, *International Journal of Cooperative Information Systems* 23 (2014), no. 01, 1440001.
- Andreas Buja and Yung-Seop Lee, Data mining criteria for tree-based regression and classification, *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 2001, pp. 27–36.
- A. Burattin and A. Sperduti, *Heuristics miner for time intervals*, ESANN, 2010.
- Gu`ena`el Cabanes and Youn`es Bennani, *Learning the number of clusters in self organizing map*, INTECH Open Access Publisher, 2010.
- Gail A Carpenter and Stephen Grossberg, Art 2: Self-organization of stable category recognition codes for analog input patterns, *Applied optics* 26 (1987), no. 23, 4919–4930.
- _____, Art 3: Hierarchical search using chemical transmitters in self-organizing pattern recognition architectures, *Neural networks* 3 (1990), no. 2, 129–152.
- Gail A Carpenter, Stephen Grossberg, Natalya Markuzon, John H Reynolds, and David B Rosen, Fuzzy artmap: A neural network architecture for incremental supervised learning of analog multidimensional maps, *IEEE Transactions on neural networks* 3 (1992), no. 5, 698–713.
- Gail A Carpenter, Stephen Grossberg, and John H Reynolds, Artmap: Supervised real-time learning and classification of nonstationary data by a self-organizing neural network, *Neural networks* 4 (1991), no. 5, 565–588.

- Gail A Carpenter, Stephen Grossberg, and David Rosen, Art 2-a: An adaptive resonance algorithm for rapid category learning and recognition, *Neural Networks*, 1991., IJCNN-91-Seattle International Joint Conference on, vol. 2, IEEE, 1991,
- Gail A Carpenter, Stephen Grossberg, and David B Rosen, *Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system*, *Neural networks* **4** (1991), no. 6, 759–771.
- Federico Chesani, Paola Mello, Marco Montali, and Sergio Storari, *Towards a dec- serflow declarative semantics based on computational logic*, (2007).
- Massimiliano De Leoni and Wil MP van der Aalst, *The featureprediction package in prom: Correlating business process characteristics.*, BPM (Demos), 2014, p. 26.
- A KA de MEDEIROS, A JMM Weijters, and W MP van der Aalst, *Genetic process mining: an experimental evaluation*, *Data Mining and Knowledge Discovery* **14** (2007), no. 2, 245–304.
- A.K.A. de Medeiros, *Genetic Process Mining*, Ph.D. thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2006.
- Jochen De Weerd, Manu De Backer, Jan Vanthienen, and Bart Baesens, *A multi- dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs*, *Information Systems* **37** (2012), no. 7, 654–676.
- Gianluigi Greco, Antonella Guzzo, Luigi Pontieri, and Domenico Sacca, *Discovering expressive process models by clustering log traces*, *IEEE Transactions on Knowledge and Data Engineering* **18** (2006), no. 8, 1010–1027.
- Stephen Grossberg, *Competitive learning: From interactive activation to adaptive resonance*, *Cognitive science* **11** (1987), no. 1, 23–63.
- Peter Harrington, *Machine learning in action*, vol. 5, Manning Greenwich, CT, 2012.
- BFA Hompes, JCAM Buijs, WMP van der Aalst, PM Dixit, and J Buurman, *Discovering deviating cases and process variants using trace clustering*, *Proceedings of the 27th Benelux Conference on Artificial Intelligence (BNAIC)*, November, pp. 5–6.
- Ben Kröse, Ben Krose, Patrick van der Smagt, and Patrick Smagt, *An introduction to neural networks*, (1993).
- Richard D. Lawrence, George S. Almasi, and Holly E. Rushmeier, *A scalable pa- rallel algorithm for self-organizing maps with applications to sparse data mining problems*, *Data Mining and Knowledge Discovery* **3** (1999), no. 2, 171–195.
- Sander JJ Leemans, Dirk Fahland, and Wil MP van der Aalst, *Discovering block- structured process models from event logs-a constructive approach*, *International Conference on Applications and Theory of Petri Nets and Concurrency*, Springer, 2013, pp. 311–329.
- _____, *Discovering block-structured process models from incomplete event logs*, *International Conference on Applications and Theory of Petri Nets and Concurrency*,

- Springer, 2014, pp. 91–110.
- _____, *Using life cycle information in process discovery*, International Conference on Business Process Management, Springer, 2015, pp. 204–217.
- Daniela Luengo and Marcos Sepúlveda, *Applying clustering in process mining to find different versions of a business process that changes over time*, International Conference on Business Process Management, Springer, 2011, pp. 153–158.
- Douglas C Montgomery, Elizabeth A Peck, and G Geoffrey Vining, *Introduction to linear regression analysis*, John Wiley & Sons, 2015.
- Laurence Morissette and Sylvain Chartier, *The k-means clustering technique: General considerations and implementation in mathematica*, *Tutorials in Quantitative Methods for Psychology* **9** (2013), no. 1, 15–24.
- Joyce Nakatumba, Michael Westergaard, and Wil MP van der Aalst, *A meta-model for operational support*, BPM Center Report BPM-12-05, BPMcenter. org (2012), 1–6
- John Neter, Michael H Kutner, Christopher J Nachtsheim, and William Wasserman, *Applied linear statistical models*, vol. 4, Irwin Chicago, 1996.
- Nils J Nilsson, *Introduction to machine learning. an early draft of a proposed text-book*, Citeseer, 1996.
- RJ Oosterbaan, DP Sharma, KN Singh, and KVGK Rao, *Crop production and soil salinity: evaluation of field data from india by segmented linear regression with breakpoint*, *Proceedings of the symposium on land drainage for salinity control in arid and semi-arid regions*, vol. 3, 1990, pp. 373–383.
- Duc Truong Pham, Stefan S Dimov, and CD Nguyen, *Selection of k in k-means clustering*, *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* **219** (2005), no. 1, 103–119.
- O.L Pintado, *Refinamiento iterativo de matrices causales*, Master's thesis, Universidad de La Habana, Cuba, 2015.
- Anne Rozinat, AK Alves De Medeiros, Christian W Günther, AJMM Weijters, and Wil MP Van der Aalst, *Towards an evaluation framework for process mining algorithms*, BPM Center Report BPM-07-06, BPMcenter. org **123** (2007), 142.
- Anne Rozinat and Wil MP van der Aalst, *Conformance checking of processes based on monitoring real behavior*, *Information Systems* **33** (2008), no. 1, 64–95.
- Shai Shalev-Shwartz and Shai Ben-David, *Understanding machine learning: From theory to algorithms*, Cambridge University Press, 2014.
- Minseok Song, Christian W Günther, and Wil MP Van der Aalst, *Trace clustering in process mining*, International Conference on Business Process Management, Springer, 2008, pp. 109–120.
- C.J.A. Turner, *Genetic Programming Based Business Process Mining Approach*, Ph.D. thesis, Cranfield University, 2009.
- Wil MP van Aalst, Kees M van Hee, Jan Martijn van Werf, and Marc Verdonk, *Auditing 2.0: Using process mining to support tomorrow's auditor*, *Computer* **43**

- (2010), no. 3.
- Wil Van Der Aalst, Arya Adriansyah, Ana Karla Alves De Medeiros, Franco Arcieri, Thomas Baier, Tobias Blickle, Jagadeesh Chandra Bose, Peter van den Brand, Ronald Brandtjen, Joos Buijs, et al., *Process mining manifesto*, International Conference on Business Process Management, Springer, 2011, pp. 169–194.
- Wil Van der Aalst, Ton Weijters, and Laura Maruster, *Workflow mining: Discovering process models from event logs*, IEEE Transactions on Knowledge and DataEngineering **16** (2004), no. 9, 1128–1142.
- Wil MP van der Aalst, Maja Pesic, and Minseok Song, *Beyond process mining: from the past to present and future*, International Conference on Advanced Information Systems Engineering, Springer, 2010, pp. 38–52.
- Wil MP Van der Aalst, Vladimir Rubin, HMW Verbeek, Boudewijn F van Dongen, Ekkart Kindler, and Christian W Günther, *Process mining: a two-step approach to balance between underfitting and overfitting*, Software & Systems Modeling **9** (2010), no. 1, 87.
- Wil MP Van der Aalst, M Helen Schonenberg, and Minseok Song, *Time prediction based on process mining*, Information Systems **36** (2011), no. 2, 450–475.
- W.M.P. van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*, Springer-Verlag, Berlin, 2011.
- Juha Vesanto and Esa Alhoniemi, *Clustering of the self-organizing map*, IEEE Transactions on neural networks **11** (2000), no. 3, 586–600.
- AJMM Weijters and JTS Ribeiro, *Flexible heuristics miner (fhm)*, Computational Intelligence and Data Mining (CIDM), 2011 IEEE Symposium on, IEEE, 2011.
- AJMM Weijters, Wil MP van Der Aalst, and AK Alves De Medeiros, *Process mining with the heuristics miner-algorithm*, Technische Universiteit Eindhoven, Tech. Rep. WP **166** (2006), 1–34.
- Hujun Yin, *The self-organizing maps: background, theories, extensions and applications*, Computational intelligence: A compendium, Springer, 2008, pp. 715–762.
- Tong Zhang, *An introduction to support vector machines and other kernel-based learning methods*, AI Magazine **22** (2001), no.2,103.

Para citar el artículo indexado.

Acosta L., López O., Vázquez Y. & Velasteguí E. (2018). Más que minería de procesos y predicción de comportamientos. Un nuevo enfoque. *Revista electrónica Ciencia Digital* 2(2), 8-28. Recuperado desde:
<http://cienciadigital.org/revistacienciadigital2/index.php/CienciaDigital/article/view/70/65>



artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Ciencia Digital**.

El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Ciencia Digital**.

