

# Ciclos Autónomos de Análisis de Datos basados en la Minería de Procesos para el Estudio del Comportamiento Curricular de los Estudiantes

*Autonomous Cycles of Data Analysis based on Process Mining for the Study of the Curricular Behavior of Students*

## ARTICLE HISTORY

Received 01 September 2020  
Accepted 02 November 2020

### **Sonia Duarte**

Universidad Pedagógica Experimental Libertador  
Instituto Pedagógico Rural Gervasio Rubio Rubio, Venezuela.  
sduarte@iprgr.upel.edu.ve

### **Jose Aguilar**

CEMISID, Universidad de los Andes, Mérida, Venezuela.  
GIDITIC, Universidad EAFIT, Medellín, Colombia  
aguilar@ula.ve

## Ciclos Autónomos de Análisis de Datos basados en la Minería de Procesos para el Estudio del Comportamiento Curricular de los Estudiantes

### *Autonomous Cycles of Data Analysis based on Process Mining for the Study of the Curricular Behavior of Students*

**Sonia Duarte**

Universidad Pedagógica Experimental  
Libertador.  
Instituto Pedagógico Rural Gervasio  
Rubio.  
Rubio, Venezuela.  
sduarte@iprgr.upel.edu.ve

**Jose Aguilar**

CEMISID, Universidad de los Andes,  
Merida, Venezuela.  
GIDITIC, Universidad EAFIT,  
Medellin, Colombia  
aguilar@ula.ve

**Resumen—** En este trabajo se evalúa el comportamiento curricular de los estudiantes de una carrera de maestría, a través de la Minería de Procesos. Específicamente, se analiza lo relacionado a la determinación de los factores internos y externos que inciden en la prosecución de sus estudios. Para comprender el comportamiento del estudiante, se usa la metodología MIDANO, la cual ha sido usada para el desarrollo de aplicaciones de analítica de datos. En particular, se especifican los Ciclos Autónomos de tareas de análisis de datos que permiten estudiar el abandono de la maestría durante la escolaridad o durante el desarrollo del trabajo de grado, con el fin de determinar las causas o problemas que se presentan durante la prosecución de los estudios. Se obtuvieron resultados muy alentadores sobre las causas del abandono de la maestría que descubren los ciclos autónomos.

**Palabras claves—** Minería de Procesos, Analítica de Datos, comportamiento curricular, Analítica de Aprendizaje

**Abstract—** In this work, the curricular behavior of the students of a master's degree program is evaluated through Process Mining. Specifically, what is related to the determination of the internal and external factors that affect the pursuit of their studies is analyzed. To understand student behavior, the MIDANO methodology is used, which has been used for the development of data analytics applications. In particular, it is specified the Autonomous Cycles of data analysis tasks that allow studying the dropout of the master's degree program during schooling or during the development of graduate thesis, in order to determine the causes or problems that arise during the

pursuit of the studies. Very encouraging results were obtained on the causes of the dropout of the master's degree program, which discover the autonomous cycles..

**Keywords—** Process Mining, Data Analytics, curricular behavior, Learning Analytics.

## I. INTRODUCCIÓN

El currículo es diseñado por las Instituciones Educativas, para lograr objetivos que se han planteado. En tal sentido, para su logro se establecen restricciones, como, por ejemplo, los estudiantes deben considerar unos cursos específicos antes de ver otros, determinadas unidades de crédito por lapso académico, la disposición de unidades curriculares electivas, entre otros aspectos. Aunado a lo expuesto, se establecen las ofertas de las unidades curriculares por lapso académico, lo que puede ocasionar que estudiantes tomen diferentes caminos una vez se inicie la carrera. De esta manera, para algunos estudiantes, el proceso de prosecución de sus estudios puede resultar más exitoso que para otros. En consecuencia, resulta conveniente recomendar que los estudiantes tomen caminos más apropiados. Para ello, se puede recurrir al análisis de datos de los historiales académicos, con la finalidad de extraer conocimiento.

Por otra parte, a partir de esos datos, su análisis es una necesidad, a través de la aplicación de las diferentes técnicas de minerías, bien sea de datos, semántica, de proceso, entre otras, ya que apertura posibilidades para responder a interrogantes como: ¿Son necesarios los requisitos exigidos para cursar una materia u

otra?, ¿Qué tan probable es que un discente finiquite su carrera o la abandone?, ¿Existen otras posibilidades para terminar la carrera? Desde estas perspectivas, este artículo presenta una herramienta de analítica de datos, para la evaluación del comportamiento curricular de una carrera de maestría, basada entre otras cosas, en la minería de procesos, con el fin de determinar los factores que pudieran incidir y coartar el finiquito de su carrera. Para ello, en este trabajo se lleva a cabo la evaluación comparando el modelo curricular formal del programa mencionado con el modelo real que siguen los estudiantes. Con ello, se determinan los factores a los cuales están expuestos los estudiantes, que en definitiva marcan su prosecución.

En general, es difícil encontrar trabajos cercanos a nuestra propuesta. Algunos trabajos similares al nuestro son los siguientes: el trabajo de Bogarín [1] propone patrones seguidos por los estudiantes durante el proceso de aprendizaje, aplicando técnicas de la minería de procesos sobre los datos generados de la interacción de los estudiantes con el entorno virtual de aprendizaje (EVA) Moodle. Por su parte, Wang et al [2] examinan los caminos exitosos que los estudiantes deben tomar para lograr la consecución de sus estudios, descubriéndolos por medio de simulaciones, y considerando los cursos tomados por los estudiantes.

Por su parte, en Pechenizkiy et al. [3] descubren diferentes estilos individuales de navegación de los estudiantes, aplicando dos pruebas en dos EVAs diferentes, Moodle y Sakai. Para ello, aplican la minería de procesos, y proponen estrategias para reducir la carga cognitiva, y mejorar la facilidad de uso y la eficiencia de aprendizaje de los sistemas de e-Learning. Por su parte, en [4] aplican la minería de procesos a la capacitación profesional, y definen métodos genéricos para ser aplicados en materia de formación profesional en e-learning. Por otro lado, también se ha estudiado el rendimiento del aprendizaje, como el trabajo de Sedrakyan et al. [5], quienes identificaron patrones del proceso de aprendizaje, identificando los peores o mejores resultados del rendimiento de aprendizaje. En [6] presentan el proceso de desarrollo del plan de estudios de una carrera de Ingeniería del Tecnológico de Monterrey, en México, mediante el uso de mapas conceptuales para ayudar a caracterizar los cursos y sus interconexiones. Usan una herramienta llamada STAUNCH, que permite evaluar la cobertura de los cursos y su contribución individual y colectiva al plan de estudios desde una perspectiva sistémica.

En cuanto al uso de la analítica de datos para entornos educativos, actualmente existen una variedad de estudios, entre los que se puede mencionar el que analiza el rendimiento de la plataforma Moodle, de Moreno, et al. [7], quienes usaron MIDANO para especificar Ciclos Autónomos (CAs) de tareas de análisis de datos para estudiar los subprocesos de Moodle relacionados con la carga de datos y descarga de archivos, con el fin de mejorar el rendimiento de los medios de almacenamiento. Por otro lado, Aguilar et. al [8] incorporaron la Analítica Social de Aprendizaje (SLA) en aulas inteligentes (SaCI), para mejorar sus procesos de enseñanza-aprendizaje, usando el concepto de CAs de análisis de datos y MIDANO, lo que permite construir un mejor perfil de aprendizaje de un estudiante, y de esta manera, detectar estilos de aprendizaje para grupos de cursos, e incluso carreras. Por otro lado, en [9] se examina el aprendizaje y el análisis académico y su relevancia para la educación a distancia en programas de pregrado y posgrado. El objetivo del trabajo es explorar los datos como predictores del éxito de los estudiantes e impulsores del plan de estudios del programa. [10] explora la aplicabilidad de la analítica de aprendizaje para la predicción del desarrollo de dos competencias transversales: trabajo en equipo y compromiso, basado en el análisis de los registros de datos de interacción de Moodle en un programa de Maestría en la Universidad a Distancia de Madrid (UDIMA).

Particularmente, en este trabajo se especifica la arquitectura computacional del sistema de análisis del comportamiento curricular de los estudiantes, y se realiza el desarrollo de un prototipo del mismo. Seguidamente, se toma como caso de estudio, la evaluación del comportamiento curricular de un Programa de maestría, denominado: Innovaciones Educativas, para descubrir los cuellos de botella existentes en el proceso educativo, y los nodos problemáticos que presenta el estudiante en la carrera. En dicho caso, se realizan experimentos que permitan evaluar el comportamiento curricular del programa educativo con nuestra herramienta. Para el desarrollo de este trabajo se utilizó la metodología de MIDANO [11], la cual ha sido usada para el desarrollo de aplicaciones de analítica de datos, la cual permite el desarrollo de CAs de tarea de análisis de datos (AdD), introducidos en [8, 12], con el fin de integrar y automatizar las actividades de analítica de datos que permiten descubrir y utilizar el conocimiento del proceso analizado, para incidir en el mismo.

## II. CICLO AUTONÓMICO DE TAREAS DE ANÁLISIS DE DATOS

La analítica de datos es usada para examinar los datos, con la finalidad de buscar conocimiento. Al respecto, en [13] se señala que “es la ciencia de la recogida, almacenamiento, extracción, limpieza, transformación, agregación y análisis de datos con el fin de descubrir información y conocimiento”. Por otro lado, toda actividad de analítica de datos debe integrarse y automatizarse. Es por ello que se ha desarrollado el concepto de CAs de tareas de AdD para cada objetivo estratégico del proceso a estudiar. Al respecto, en Aguilar et al. [14, 15] señalan que un “Ciclo Autónomo de Tareas de Analítica de Datos puede definirse como un conjunto de tareas de Análisis de datos cuyo propósito es mejorar el proceso que estudia. Este conjunto de tareas interactúa entre ellas, y tienen diferentes roles: observar el proceso, analizar e interpretar lo que pasa en él, y tomar decisiones para mejorarlo” (ver Fig. 1).

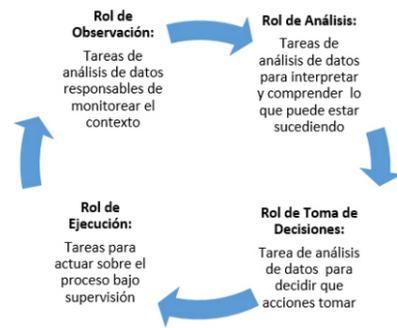


Fig. 1. Ciclo autónomo de tareas de análisis de datos.

La integración de las tareas en un ciclo, visto como un lazo cerrado, permite la resolución autónoma de problemas. En general, los posibles roles de cada tarea de AdD del ciclo son [12, 15]:

- Tareas de monitoreo: son responsables de obtener las variables del proceso a estudiar. Estas tareas monitorean el proceso, ven su comportamiento, y extraen los datos que los describen.
- Tareas de Análisis del sistema: son las responsables de la interpretación de lo que sucede en el proceso. Con ellas, se puede diagnosticar, entender, analizar, entre otras cosas, lo que sucede. Para ello, se construyen modelos de conocimiento (por ejemplo, de predicción, descripción, etc.).
- Tareas de toma de decisiones: definen las acciones a ejecutar a efectos de mejorar el proceso, considerando para ello el objetivo planteado para el CA.

## III. CASO DE ESTUDIO: COMPORTAMIENTO CURRICULAR DE LOS ESTUDIANTES

Para el desarrollo de la arquitectura computacional del sistema, para analizar los problemas de abandono, se usa la metodología MIDANO. Como un primer paso, se analizan los procesos del programa educativo, con la finalidad de determinar los procesos viables para la aplicación de analítica de datos, vinculados al problema a analizar. Los procesos del postgrado de innovaciones educativas se muestran en la Tabla I.

De la Tabla I, claramente se identifica el proceso de prosecución del estudiante (modelo curricular), como el proceso objeto de estudio. Se procede a realizar los Ciclos autónomos para dicho proceso.

Tabla I. procesos de innovaciones educativas.

PROCESOS	
Pre-inscripción	PI
Inscripción curso introductorio	ICI
Inscripción	INSC
Prosecución del estudiante (Modelo curricular)	PE
Asesorías	AS
Emisión constancias	EC
Retiro materias	RM
Retiro semestre o Universidad	RSU
Reingreso	REI
Intención investigativa	IIN
Transcripción de calificaciones	TC
Culminación de estudios Grado	CEG

### A. Ciclo Autónomo “abandono de maestría durante la escolaridad”

Este ciclo tiene como objetivo descubrir el patrón de comportamiento curricular que sigue un estudiante (camino exitoso y no exitoso), destacando las causas o problemas que se le presentan en su prosecución. En tal sentido, sus tareas de análisis de datos son especificadas en la Tabla II y Fig. 2.

Tabla II. Grupo de tareas del ciclo autónomo: abandono de la maestría durante la escolaridad.

	Nombres	Fuentes generales de datos requeridas	Indicadores generados	Efectos esperados sobre objetivo estratégico
Tareas de AdD de observación	Caracterizar los estudiantes de maestría.	Base de datos (histórico académico) con registro de ID del estudiante, materias cursadas, fecha de inscripción de materia, fecha de finiquito de materia, reingresos, retiros de materias o semestre.		Organización y definición de los datos para la observación del comportamiento real de los estudiantes
Tareas de AdD de análisis	Identificar los estudiantes exitosos y no exitosos.	Base de datos (datos externos): movilidad o transporte, acceso a internet, electricidad, influencia de otros compañeros, apoyo y motivación por parte de terceros.	Modelo de Minería de procesos- descubrimiento o conformidad	Agrupación de los estudiantes e identificar patrones
	Estudiar el flujo académico.	Base de datos (datos externos): movilidad o transporte, acceso a internet, electricidad, influencia de otros compañeros, apoyo y motivación por parte de terceros.		Descubrimiento de cuellos de botella, nodos críticos.
Tareas de AdD de toma de decisiones	Comparar los flujos para determinar el patrón de no exitosos.	Base de datos personales: Si trabaja o no, si tiene recursos económicos o no, si es cabeza de hogar, si recibe ayuda de los padres, si tiene recursos computacionales.		Confrontación de las situaciones generadas en los caminos no exitosos que considera el estudiante en la prosecución de sus estudios.
	Determinar las causas o problemas que aparecen en el flujo de los no exitosos.			Generar la necesidad de documentar al estudiante para inducirlo por el camino más corto y correcto para el feliz término de la maestría.



Fig. 2. CA abandono de maestría durante la escolaridad.

Siguiendo la metodología de MIDANO, se describen las tareas de análisis de datos (Ver Tablas III, IV, V, VI, VII) para los diferentes CAs.

Tabla III. Tarea de observación 1 del ca abandono temporal o definitivo de la maestría durante la escolaridad.

Nombre de la tarea	Caracterizar los estudiantes de maestría
Descripción	Se seleccionarán los estudiantes, determinando mecanismos de extracción, muestra de estudiantes a usar, organizando los datos en formatos óptimos para ser utilizados en Minería de Procesos.
Tipo de tarea de analítica de datos	Clasificación
Técnicas de analítica de datos	Minería de datos
Tipo de modelo de conocimiento	Modelo de clasificación

Tabla IV. Tarea de análisis 2 del ca: abandono temporal o definitivo de la maestría durante la escolaridad.

Nombre de la tarea	Identificar los estudiantes exitosos y no exitosos
Descripción	Se obtiene información de lo que sucede realmente con los caminos exitosos y no exitosos tomados por los estudiantes durante la maestría (comportamiento real del estudiante).
Tipo de tarea de analítica de datos	Clasificación
Técnicas de analítica de datos	Minería de datos y Minería de Procesos
Tipo de modelo de conocimiento	Modelo de clasificación

Tabla V. Tarea de análisis 3 del ca: abandono temporal o definitivo de la maestría durante la escolaridad.

Nombre de la tarea	Estudiar el flujo académico
Descripción	Se obtiene información de lo que sucede realmente durante la prosecución del estudiante (modelo curricular real), en que momentos se quiebra la continuidad de los estudios de maestría.
Tipo de tarea de analítica de datos	Asociación
Técnicas de analítica de datos	Minería de procesos
Tipo de modelo de conocimiento	Descubrimiento

Tabla VI. Tarea de análisis 4 del ca: abandono temporal o definitivo de la maestría durante la escolaridad.

Nombre de la tarea	Comparar los flujos para determinar el patrón de no exitosos.
Descripción	Verificar el modelo formal y el patrón que sigue el estudiante a fin de diagnosticar el comportamiento del estudiante y las realidades que le aqueja.
Tipo de tarea de analítica de datos	Comparación
Técnicas de analítica de datos	Minería de procesos
Tipo de modelo de conocimiento	Conformidad

Tabla VII. Tarea de toma de decisiones 1 del ca: abandono temporal o definitivo de la maestría durante la escolaridad.

Nombre de la tarea	Determinar las causas o problemas que aparecen en el flujo de los no exitosos.
Descripción	Se obtiene información de los elementos que inciden en los estudiantes que abandonan la maestría durante la escolaridad.
Tipo de tarea de analítica de datos	Asociación
Técnicas de analítica de datos	Análisis de datos social-estadística
Tipo de modelo de conocimiento	Estadística

El modelo de datos multidimensional del ciclo autónomo de abandono temporal o definitivo de la maestría durante la escolaridad se muestra en la Fig. 3. La tabla de hechos es “abandono\_ escolaridad” y a su vez, tiene cinco tablas de dimensiones que definen características específicas.

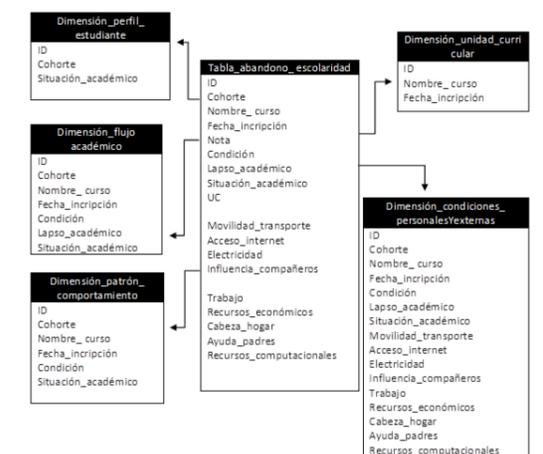


Fig. 3. Modelo de datos CA: abandono temporal o definitivo de la maestría durante la escolaridad.

El modelo de la Fig. 3 incluye datos del postgrado, y datos extraídos de Facebook, WhatsApp (redes sociales). Cada una de esta información está incluida en una dimensión diferente en el modelo de datos. Así, tenemos:

- Dimensión perfil estudiante: Son almacenados datos que definen al estudiante, como la cohorte en que inicia sus estudios y en qué situación académica se encuentra.
- Dimensión flujo académico: Permite conocer el camino que el estudiante recorre durante la prosecución de sus estudios.
- Dimensión patrones de comportamiento: Muestra los grupos de estudiantes en sus caminos exitosos y no exitosos
- Dimensión unidades curriculares: Muestran las unidades curriculares que ha tomado el estudiante en su recorrido durante la maestría.

- Dimensión experimentos: almacena las condiciones de los estudiantes: su situación académica, tiempos de permanencia en la maestría, entre otros.
- Dimensión condiciones externas y personales: almacena los factores del entorno que pueden afectar la prosecución del estudiante.

**B. Ciclo Autónomo “Abandono temporal o definitiva de la maestría durante la intención investigativa y trabajo de grado**

Este ciclo tiene como objetivo descubrir el patrón de comportamiento que sigue el estudiante durante el trabajo de grado, a fin de llevar a cabo el análisis del modelo curricular, destacando las causas o problemas que se presentan en el finiquito de la maestría. La Tabla VIII y Fig. 4 muestran las tareas de análisis de datos.

Tabla VIII. Grupo de tareas del ciclo autónomo de intención investigativa y trabajo de grado.

	Nombres	Fuentes generales de datos requeridas	Indicadores generados	Efectos esperados sobre objetivo estratégico
Tareas de AdD de observación	Caracterizar los estudiantes de maestría que están en la última etapa curricular	Base de datos con registro de fechas de inscripción de intención investigativa o trabajo de grado, reintrosos.	Modelo de minería de procesos-descubrimiento	Organización y definición de los datos para la observación del comportamiento real de los estudiantes.
Tareas de AdD de análisis	Identificar los estudiantes exitosos y no exitosos de la última parte			Agrupación de los estudiantes e identificar patrones
	Estudiar la última parte curricular		Estadísticas de minería de procesos	Descubrimiento de cuellos de botella, nodos críticos.
Tareas de AdD de toma de decisiones	Determinar las causas que afectan el finiquito de la intención investigativa o trabajo de grado.		Implicaciones	Confrontación de las situaciones generadas en los caminos no exitosos que considera el estudiante en la prosecución de sus estudios.  Generar la necesidad de documentar al estudiante para inducirlo por el camino más corto y correcto para el feliz término de la maestría.

**Tarea 3. Descubrir patrones en los flujos académicos de los estudiantes:** A partir del flujo general de la Fig. 6, donde se consideran todos los flujos de los estudiantes exitosos, se trata de construir el patrón de flujos exitosos en la Fig. 7. Se considera como patrón de los exitosos al flujo de los casos que tomaron más el mismo camino (es lo que muestra la Fig. 9). El promedio de duración es equivalente a 1,99 años.

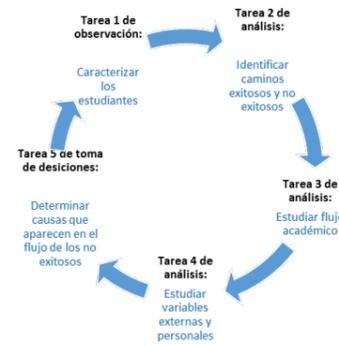


Fig. 4. Representación gráfica CA abandono de maestría durante trabajo de grado o intención investigativa.

La especificación detallada de las tareas, y del modelo de datos, se hace de una manera similar a la indicada en la sección CA “abandono de maestría durante la escolaridad” para el primer ciclo autónomo. Para las siguientes secciones, se especifica cómo implementar el primer ciclo autónomo.

**IV. EXPERIMENTACIÓN**

Para los experimentos, se trabaja con la base de datos SIPOST del Instituto Pedagógico Rural Gervasio Rubio, alimentada por los eventos que se registran durante la prosecución de los estudiantes desde el año 2015 al 2019, considerándose en su totalidad 166 casos o estudiantes y 1955 eventos. Por otra parte, se consideran datos personales y del entorno del estudiante, obtenidos del sistema académico o redes sociales.

Se consideran, simulaciones para tres escenarios. Un primer experimento analiza si los factores externos son causas considerables en los tiempos de permanencia del estudiante durante la escolaridad de la maestría. En un segundo experimento, se busca determinar la incidencia de las características personales en el éxito o no del estudiante; para finalizar con un tercer experimento, donde se determina la incidencia de las unidades curriculares y unidades crédito cursadas por lapso académico.

**A. Escenario experimental 1**

En este experimento, se determina si el flujo curricular de cada estudiante está condicionado por su entorno: transporte, situación social: luz, internet, gas, agua, gasolina, y situación política del país: marchas, concentraciones, paros. La hipótesis es: ¿Son los factores externos causas considerables en los tiempos de permanencia del estudiante durante la escolaridad de la maestría? A continuación, se presenta el comportamiento de cada tarea del ciclo en este escenario experimental.

**Tarea 1: Caracterizar los datos de los estudiantes:**

Para la selección de estudiantes, se considera su situación académica: egresado, graduando, regular, participante especial, sin inscripción, retiro permanente. De esta manera, determinar su estatus.

**Tarea 2. Identificar caminos exitosos y no exitosos:**

se identifican los estudiantes exitosos (que culminaron sus estudios). Primero que nada, se determinan los caminos tomados por los estudiantes; sin embargo, se recurre a varios filtros. Se considera el tiempo promedio de 2 años efectivos para el finiquito de unidades curriculares, para ser considerado un camino como exitoso (finiquito de la escolaridad de manera exitosa). De esta manera, se logra determinar los casos que cumplen con el requerimiento de haber cursado todas las unidades curriculares hasta trabajo de grado inscrito, generando el modelo de procesos (descubrimiento) que se muestra en la Fig. 5 para los flujos exitosos.

En la Fig. 5, se puede visualizar en los círculos

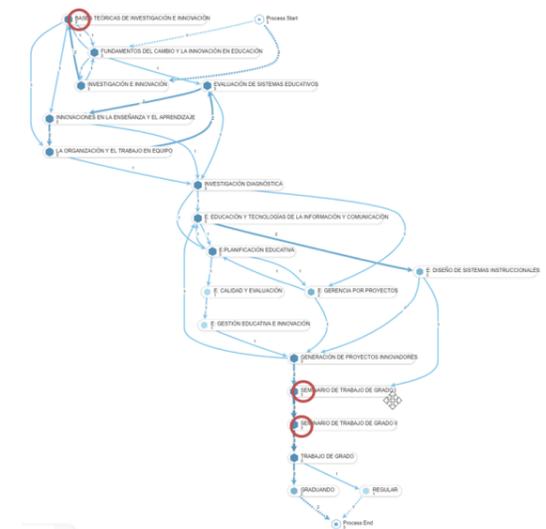


Fig. 5. Modelo de Procesos general de los caminos exitosos de los estudiantes.

rojos los nodos (cursos) más comunes seguidos por los estudiantes. El resto de flujos son los flujos no exitosos. Se consideran caminos no exitosos, los casos con más de 2 años efectivos en la escolaridad. En la Fig. 6, se muestra el modelo de procesos (descubrimiento) de los casos no exitosos definidos por esta tarea. Igualmente, los círculos rojos representan los nodos (cursos) más seguidos en los caminos no exitosos.

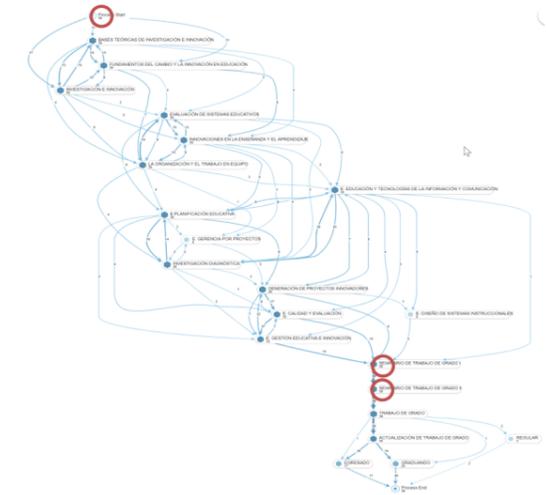


Fig. 6. Modelo de Procesos general de los caminos no exitosos durante la escolaridad.

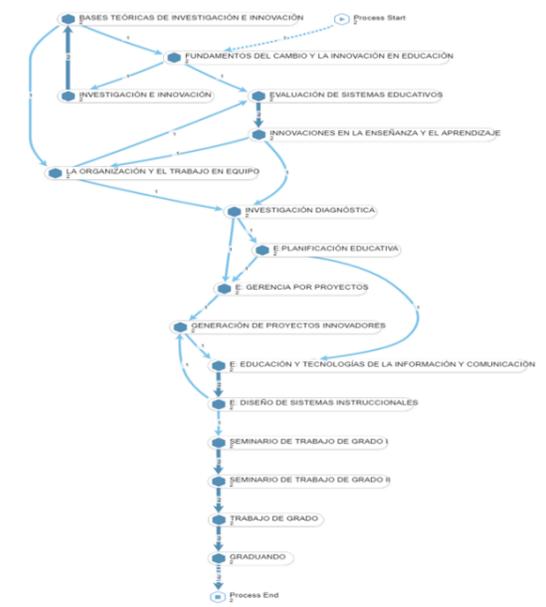


Fig. 7. Patrón casos estudiantes exitosos.

El mismo procedimiento se sigue para determinar el patrón de los caminos no exitosos.

**Tarea 4. Comparar con el modelo curricular:**

En esta tarea se realiza la comparación del patrón del flujo de los estudiantes no exitosos con el del pensum de estudio. Para ello, se contrasta el grafo del pensum con el grafo que siguen los estudiantes no exitosos, como se muestra en la Fig. 8.

En la Fig. 8, se aprecia la condición de los estudiantes, reflejándose en círculos azules cuantos egresados y cuantos en condición de preparación del trabajo de grado (lado izquierdo). Además, se compara ambos grafos o modelos, determinándose discrepancia en la cantidad de unidades curriculares que se deberían cursar por lapso académico (círculos azules del grafo del lado derecho)

y las que realmente son tomadas por los estudiantes (círculos rojos del grafo del lado izquierdo, resaltándose el lapso académico adicional). En este sentido, se determina que existe un desplazamiento de un lapso académico adicional, que genera incidencia en la prosecución del estudiante durante los estudios de la escolaridad.

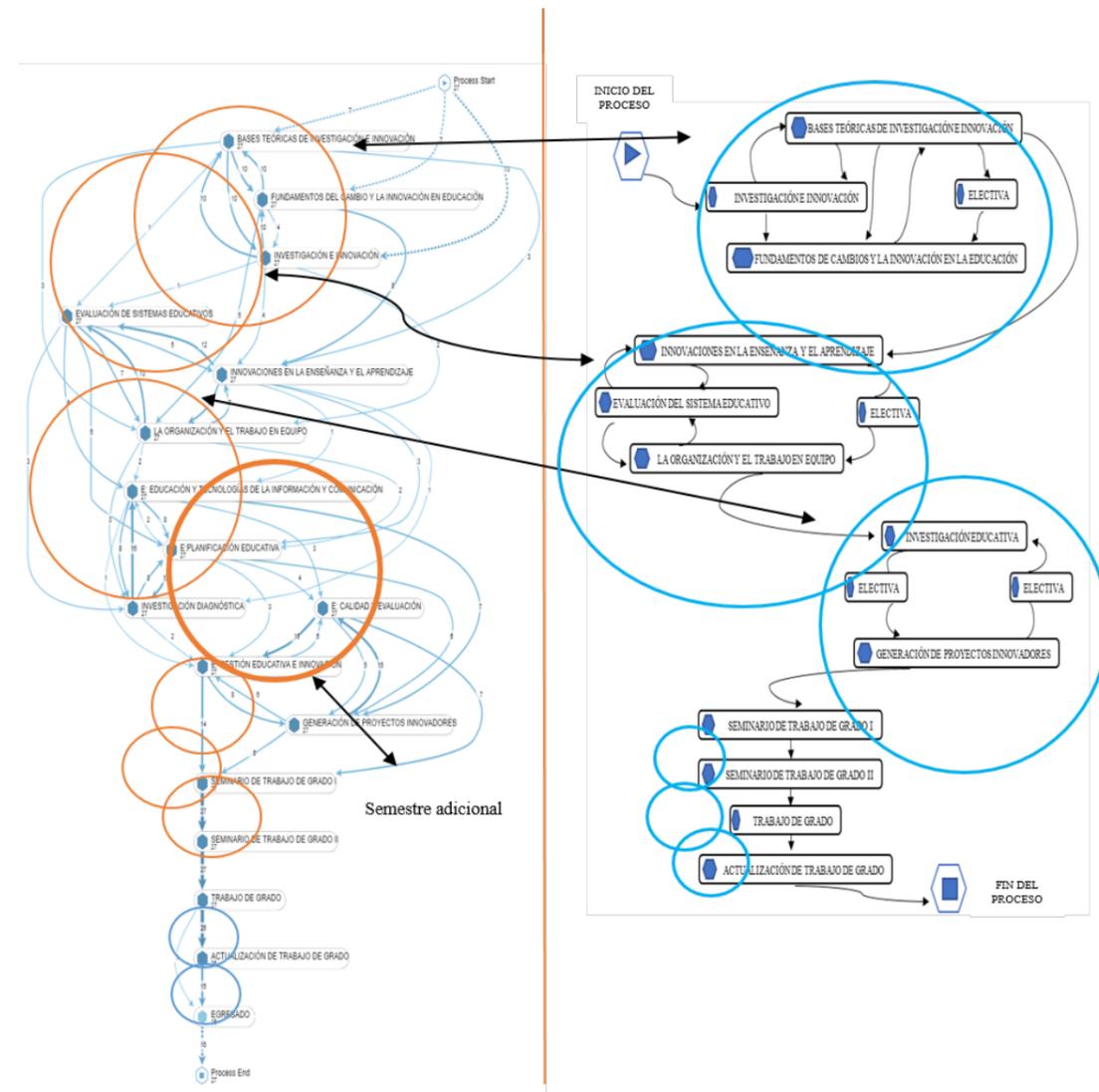


Fig. 8. Patrón de los no exitosos Vs Modelo del pensum de estudio.

**Tarea 5: Determinar posibles causas de no éxito:** Con el patrón de flujo de los estudiantes no exitosos determinados en la tarea 4, se pasa a obtener información de tipo social a través de datos personales y externos de estos estudiantes, los cuales se obtienen de las planillas de inscripción que reposan en control de estudios y en las diferentes redes sociales

(Facebook, WhatsApp, Twitter). Para ello, se consideran aspectos del entorno del estudiante, como: situación social: transporte público, luz, internet, gas, agua, gasolina y situación política del país: marchas, concentraciones, paros. Las estadísticas de los estudiantes de los casos no exitosos se muestran en la Tabla 9, ordenada por su grado de afectación.

De la Tabla IX, se puede destacar que los factores electricidad, transporte público, acceso a internet, agua, y paros tienen una incidencia alta (entre 85 y 100%). Por otra parte, el factor gasolina incide en un 74%. Además, la falta del gas tiene una incidencia del 44%. Por último, la incidencia de marchas y concentraciones es baja, con un 37%. En conclusión, en este escenario se determina que la mayoría de los factores del entorno del estudiante considerados como posible causa han incidido en la prosecución del estudiante de manera alta durante su carrera. Ahora bien, el uso del gas, marchas y concentraciones tienen menor incidencia. Por otra parte, los factores de agua, y electricidad tienen la incidencia más alta. En cuanto a la electricidad, significa que los estudiantes se ven más afectados y no logran desarrollar sus trabajos a tiempo, y en cuanto a los paros, atrasa sus tiempos de prosecución, extendiéndose sus periodos académicos.

Tabla IX. Incidencia del entorno del estudiante.

Factor	% Incidencia
Electricidad	100
Paros	96
Agua	93
Acceso a internet	89
Transporte público	85
Gasolina	74
Gas	44
Marchas	37

**B. Escenario experimental 2**

El objetivo de este escenario es determinar si el flujo de cada estudiante está condicionado por los hijos, trabajo, alimentación, ubicación, cabeza de hogar, dependencia económica, computador. La hipótesis es ¿Son las situaciones personales causas de los tiempos de permanencia del estudiante durante la escolaridad de la maestría?

En general, se recaba la información de la misma manera que para el experimento 1 (las primeras 4 tareas son iguales al caso anterior), evaluándose en esta oportunidad los aspectos personales y su incidencia, obteniéndose la Tabla X, que corresponde al estudio estadístico que representan el patrón de los estudiantes no exitosos.

En la Tabla X, se puede apreciar el porcentaje que representa cada uno de los factores personales del estudiante, destacándose que los factores de poseer computador y si trabaja tienen una incidencia del 96%, las cuales son más altas, infiriéndose que casi en su totalidad los estudiantes deben trabajar para mantenerse, presentando dificultad para desarrollar sus actividades académicas. También, el hecho de poseer un computador, inciden mucho en la prosecución del estudiante. En cuanto al hecho de tener hijos, cabeza de hogar, dependencia de su alimentación, dependencia de la

alimentación de la familia, representan del 70 al 78%. Por último, el aspecto de dificultad en alguna unidad curricular sólo representa el 11 % de incidencia, que significa que hubo un pequeño porcentaje de estudiantes que presentaron problemas al momento de cursar materias.

Tabla X. Incidencia de los aspectos personales del estudiante.

Factor	% Incidencia
Trabajo	96
Computador	96
Cabeza de hogar	78
Dependencia de la alimentación de la familia	78
Dependencia de su alimentación	76
Hijos	70
Dificultad en unidad curricular	11

**C. Escenario experimental 3**

El objetivo es determinar si el flujo que sigue cada estudiante no exitoso está condicionado por unidades curriculares y/o unidades de créditos cursadas por lapso académico. La hipótesis es ¿Es el flujo que sigue el estudiante el que repercute en su éxito?

Las cuatro primeras tareas son las mismas de los experimentos 1 y 2. Así en esta sección solo se presenta el desarrollo de la tarea 5. Para ello, se analiza el flujo académico de los estudiantes que resultaron no exitosos. Particularmente, se comparan los patrones de los estudiantes no exitosos de la tarea 4 con el modelo formal del pensum, detectándose que existe discrepancia entre lo que se establece en el pensum y la prosecución real del estudiante, como se muestra en la Fig. 9.

En la Fig. 9, se pueden apreciar los lapsos académicos cursados por los estudiantes no exitosos durante la escolaridad (grafo a la derecha, círculos rojos intensos). Se determina que cursaron 6 lapsos académicos, lo que incide notoriamente en los tiempos de permanencia durante la escolaridad, ya el estudiante debe cursar un semestre adicional, un tiempo no previsto según el pensum de estudios, el cual establece 5 lapsos académicos (grafo a la izquierda, círculos rojo pálido). Ahora bien, viendo más de cerca el pensum, se logra detectar que en los primeros lapsos académicos debería verse una unidad curricular electiva (círculos en verde), y ésta no se ve en los dos primeros lapsos académicos. Igualmente, en el tercer lapso deberían verse 2 electivas y se ve 1, lo que desenlaza en un definitivo mayor tiempo durante la escolaridad.

Así, efectivamente el flujo que sigue el estudiante repercute en su éxito, ya que se evidencia que los estudiantes cursan menos unidades curriculares de las que se debería en los primeros lapsos, de acuerdo al Reglamento de Estudios de Postgrado [16].

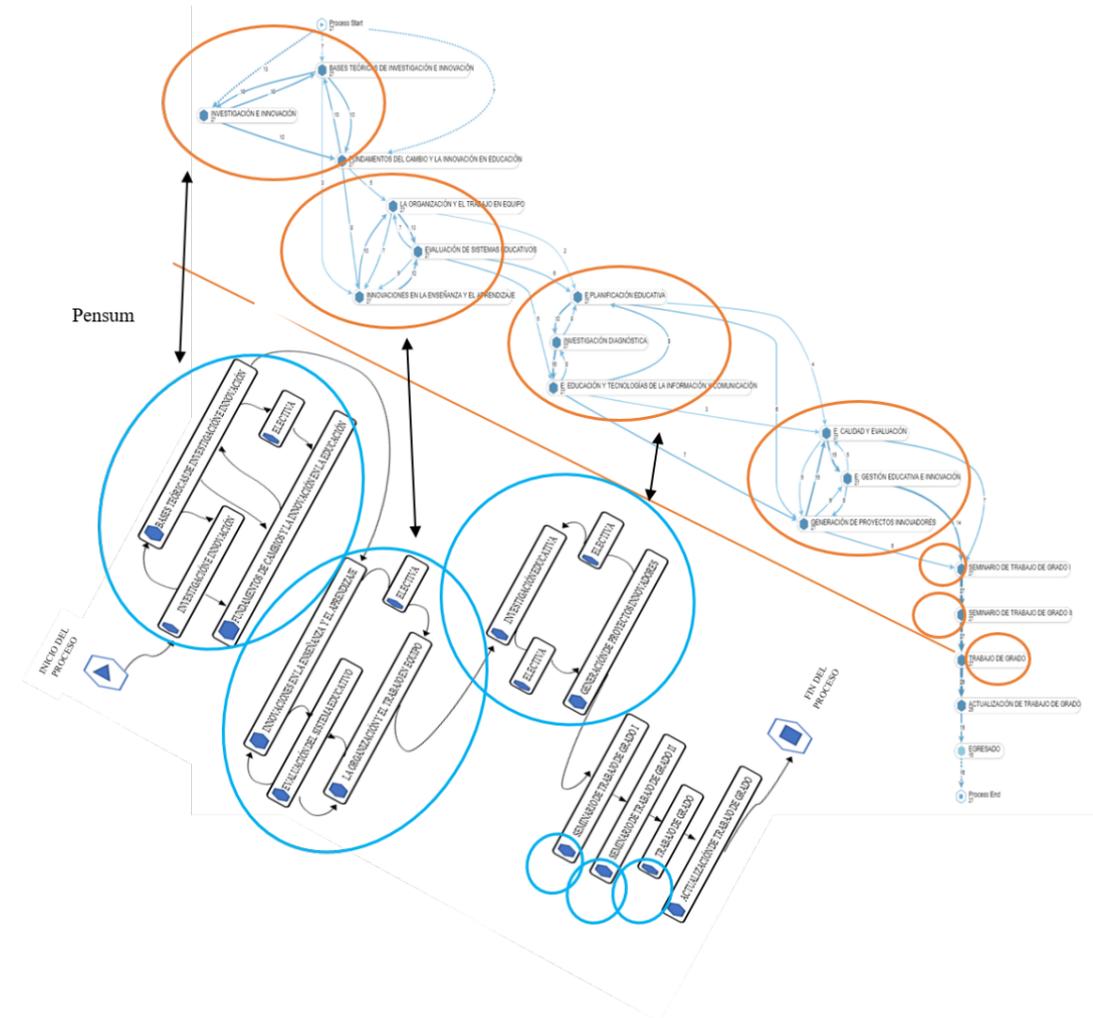


Fig. 9. Comparación de patrón no exitoso Vs Penum formal.

**D.Comparación con otros Trabajos similares**

En general, los modelos curriculares no han sido estudiados ampliamente, a través de la minería de procesos. En cuanto al estudio de modelos curriculares por medio de minería de procesos, sólo un estudio previo consideró examinar caminos exitosos en la consecución de los estudios a través de la minería de procesos [4], sin embargo, no considera ninguna otra minería. El uso de minería de procesos ha sido empleada como única minería en varias investigaciones previas. En [1] se limitan a sólo mostrar diferentes modelos de descubrimiento aplicando varios algoritmos, y posteriormente los comparan. En [1] aplican minería de procesos para descubrir modelos y flujos de trabajo, para investigar los comportamientos de los usuarios. Por otro lado, en [2] se aplica minería de procesos a fin de determinar caminos exitosos. Asimismo, [5] sólo la aplican para analizar los registros de eventos de los trabajos grupales de los estudiantes.

Por otro lado, los ciclos autónomos de tareas de análisis de datos han sido empleados en [7], para desarrollar las implementaciones específicas de optimización de un EVA, para acelerar los procesos de transmisión de archivos y evitar la pérdida de datos. A su vez, [8] lo emplean para mejorar el proceso de enseñanza-aprendizaje que se da en un salón inteligente, al obtener el estilo de aprendizaje para un grupo de cursos. Finalmente, [12, 13] diseñan ciclos autónomos de tareas de analítica de aprendizaje.

Existen propuestas de un enfoque de minería de cualquier cosa (minería de datos, MP, minería semántica), en el proceso de aprendizaje. En [6] ha sido propuesta la minería de datos, tales como árboles de decisión, reglas de asociación, redes neuronales, entre otros, porque se requiere la optimización del proceso de evaluación. También, [8] emplea

minería de cualquier cosa para obtener el estilo de aprendizaje de los estudiantes en un salón de clases. Se aplica la analítica social de aprendizaje para determinar el estilo de aprendizaje adecuado para los estudiantes del curso seleccionado, usando datos de redes sociales [8].

Por otro lado, investigaciones previas usan fuentes externas de datos, como por ejemplo [7], quienes al analizar un EVA pueden considerar para la optimización, variables personales y externas que inciden en el proceso. Por su parte, [1], al presentar la rutas de aprendizaje, pueden escalar la investigación. Al comparar el presente estudio con investigaciones previas, se pueden destacar las diferencias de nuestra propuesta:

- Hace un análisis de modelos curriculares, lo cual ha sido poco estudiado, encontrándose sólo una propuesta.
- No sólo se aplica técnicas de minería de procesos, sino que va más allá, incorporando minería de datos, analítica de datos social-estadística, entre otras tareas.
- Al incorporar datos personales y del entorno para determinar las causas, se permite no sólo analizar los grafos de la prosecución personalizados, como lo hacen la mayoría de las investigaciones, pero además, se consideran otros elementos que afectan a los estudiantes y que influyen en el éxito de sus estudios.

Por último, la investigación es escalable porque se puede extender con otras tareas de análisis de datos para estudiar otros aspectos más específicos, como por ejemplo, factores locales de donde se vive.

**V. CONCLUSIONES**

En el trabajo se propuso un enfoque que usa la minería de procesos, entre otros enfoques de minería, para evaluar el comportamiento curricular que siguen los estudiantes de la maestría de innovaciones educativas. El estudio se desarrolla incorporando el concepto de CA de tareas de análisis de datos, el cual permite alcanzar objetivos específicos de análisis, para ayudar a tomar decisiones estratégicas.

La incorporación de la Minería de cualquier cosa, incluyendo la MP, en los CAs de tareas para analizar el abandono de la maestría durante la escolaridad, permite descubrir los modelos de los flujos que sigue el estudiante en la prosecución de sus estudios. En particular, para la experimentación se seleccionó el Ciclo "Abandono de la maestría durante la

escolaridad". Se desarrollaron 3 escenarios, un primer escenario tenía como objetivo determinar si el flujo de cada estudiante está condicionado por su contexto social, un segundo objetivo tenía como objetivo determinar si el flujo de cada estudiante está condicionado por entorno personal, y un tercer escenario tenía como objetivo determinar si el flujo que sigue cada estudiante está condicionado por el modelo académico de la maestría.

De acuerdo con la revisión de la literatura, no hay precedentes del uso de la metodología MIDANO para estudios en modelos curriculares. En particular, el ciclo de abandono de la maestría durante la escolaridad permite estudiar plenamente las causas que aquejan al estudiante. Con esta investigación, se crea el precedente de que con el uso de ciclos basado en tareas de minería de cualquier cosa (incluyendo la MP), es posible generar modelos de comportamiento curricular que emprenden los estudiantes en instituciones educativas. Todo ello, basado en los registros de datos académicos, eventualmente enriquecidos con datos del entorno (redes sociales, etc.). Así, es posible determinar los recorridos curriculares más exitosos o no, aspectos que puedan influir en esos recorridos exitosos o no, con el fin de determinar eventuales acciones correctivas. Un ciclo autónomo es una gran ayuda en la toma de decisiones correctivas, debido a que es capaz de generar conocimiento, y con ello, ayuda a determinar las decisiones que favorezcan el desempeño estudiantil.

Trabajos futuros estarán dedicado a estudiar diferentes estrategias de Minería de Procesos en Mallas Curriculares de carreras de pregrado basadas en el concepto de CAs, para automatizar tareas de detección de recorridos académicos de estudiantes anormales, proponer nuevos modelos curriculares según comportamientos de los estudiantes o necesidades en el mercado, entre otras cosas. Para ello, esos CAs serán enriquecidos con analítica de aprendizaje, análisis de redes sociales, etc., para enriquecer la información que usarán.

**REFERENCIAS**

[1] A. Bogarín, C. Romero, R. Cerezo and M. Sánchez-Santillán. "Clustering for Improving Educational Process Mining". Proc. 4th Intl Conf. Learning Analytics and Knowledge, pp. 11-15. 2014

[2] R. Wang and O. Zaiane. "Discovering Process in Curriculum Data to Provide Recommendation." EDM. Pp. 580-581, 2015

- [3] M. Pechenizkiy, N. Trcka, E. Vasilyeva, W. Van der Aalst and P. De Bra. "Process Mining Online Assessment Data". International Working Group on Educational Data Mining. 2009
- [4] A. Cairns, B. Gueni, M. Fhima, S. David and N. Khelifa. "Process mining in the education domain". International Journal on Advances in Intelligent Systems, vol. 8, pp. 219-232. 2015.
- [5] G. Sedrakyan, M. Snoeck and J. De Weerd. "Process mining analysis of conceptual modeling behavior of novices-empirical study using JMermaid modeling and experimental logging environment". Computers in Human Behavior, vol. 41, pp. 486-503. 2014
- [6] F. Lozano and R. Lozano, "Developing the curriculum for a new Bachelor's degree in Engineering for Sustainable Development", Journal of Cleaner Production, vol. 64, pp. 136-146, 2014.
- [7] Y. Moreno, C. Aguilar and F. Hidrobo. "Análisis de los Problemas de Rendimiento en un EVA a través de la Extracción de Conocimiento". Revista Ingeniería al Día, vol. 4, pp. 3-24. 2018
- [8] J. Aguilar, O. Buendía, A. Pinto and J. Gutiérrez. Social learning analytics for determining learning styles in a smart classroom, Interactive Learning Environments. 2019.
- [9] K. Mattingly, M. Rice and Z. Berge, "Learning analytics as a tool for closing the assessment loop in higher education", Knowledge Management & E-Learning: An International Journal, vol.4, pp. 236-247, 2012.
- [10] S. Iglesias-Pradas, C. Ruiz-de-Azcárate, and Á. Agudo-Peregrina, "Assessing the suitability of student interactions from Moodle data logs as predictors of cross-curricular competencies", Computers in Human Behavior, vol. 47, pp. 81-89, 2015.
- [11] F. Pacheco, C. Rangel, J. Aguilar, M. Cerrada and J. Altamiranda, "Methodological framework for data processing based on the Data Science paradigm", XL Latin American Computing Conference (CLEI). 2014
- [12] J. Aguilar, O. Buendía, K. Moreno and D. Mosquera, "Autonomous cycle of data analysis tasks for learning processes, In International Conference on Technologies and Innovation, Communications in Computer and Information Science Series", vol. 658, pp. 187-202. 2016
- [13] J. Aguilar, M. Sánchez, M, J. Cordero, P. Valdiviezo-Díaz, L. Barba-Guamán and L. Chamba-Eras. "Learning analytics tasks as services in smart classrooms". Universal Access in the Information Society, vol. 17, pp. 693-709, 2018
- [14] J. Aguilar, J. Cordero and O. Buendía. "Specification of the Autonomic Cycles of Learning Analytic Tasks for a Smart Classroom", Journal of Educational Computing Research, vol. 56, pp. 866-891. 2018
- [15] M. Sánchez., J. Aguilar, J. Cordero, P. Valdiviezo-Díaz, L. Barba-Guamán, and L. Chamba-Eras "Cloud Computing in Smart Educational Environments: Application in Learning Analytics as Service". In: Rocha Á., Correia A., Adeli H., Reis L., Mendonça Teixeira M. (eds) New Advances in Information Systems and Technologies. Advances in Intelligent Systems and Computing, vol. 444., pp. 993-1002 2016.
- [16] Reglamento de Estudios de Postgrado. Universidad Pedagógica Experimental Libertador. Gaceta 1-2018. Resolución N° 2018.488.056. 2018

# AUTHORS



## Jose Aguilar

Ingeniero de Sistemas, Universidad de los Andes-Mérida-Venezuela, Maestría en Informática, Universidad Paul Sabatier-Toulouse-France, y Doctorado en Ciencias Computacionales, Universidad Rene Descartes-Paris-France. Postdoctorado en el Departamento de Ciencias de la Computación de la Universidad de Houston, en el Laboratoire d'Automatique et Analyses de Systemes-CNRS-Toulouse-France y en el Departamento de Automática, Universidad de Alcalá-España (actualmente). Profesor Titular del Departamento de Computación de la Universidad de los Andes-Venezuela, Profesor Contratado de la Universidad EAFIT-Colombia. Miembro de la Academia de Mérida.



## Sonia Duarte

Ingeniero de Sistemas, Universidad de los Andes-Mérida-Venezuela, Especialista en Planificación Educacional, Universidad Valle del Momboy, Magister en Innovaciones Educativas, Universidad Pedagógica Experimental Libertador. Profesora en la Especialidad de informática, Universidad Pedagógica Experimental Libertador