

Implementación de patrones de secuencia en el aula extendida de la Universidad Autónoma del Caribe, para la exploración y análisis de las actividades realizadas por los docentes

Implementation of sequence patterns in the extended class of autonomous university of the caribbean for the exploration and analysis of activities carried out by teachers

Juan Carlos Calabria Sarmiento¹, Ilma Azucena Bonilla Botía²

1-, *Magíster Ingeniería de Sistemas y Computación Especialista en desarrollo de Procesos Cognoscitivos. Docente Tiempo Completo Universidad Autónoma del Caribe. Grupo Ingeniería de software y nuevas tecnologías. Programa Ingeniería de Sistemas. jcalabria@uac.edu.co*

2-, *Candidato a Magíster Informática Educativa. Especialista en Informática y Telemática. Docente de Tiempo Completo Universidad Autónoma del Caribe. Grupo Ingeniería de software y nuevas tecnologías.*

Recibido 6/10/2010, Aceptado 1/12/2010

RESUMEN

Se implementa una herramienta informática que apoya el análisis del comportamiento de las actividades de aprendizaje que el docente registra en el Aula Extendida de la Universidad Autónoma del Caribe. Para esto, se adopta una metodología para desarrollo de proyectos en minería de datos, CRISP- DM que provee las fases de comprensión del negocio, comprensión de los datos y preparación de los datos; teniendo en cuenta la herramienta de ETL (Extracción, Transformación y Carga) Talend Open, que permite depurar los datos para la integración.

Para la búsqueda de patrones del comportamiento en el desempeño docente en el aula extendida, se tienen en cuenta las variables socio demográficas y niveles de estudio. Se utiliza la herramienta WEKA que permite; a través de los datos de entrenamiento y validación, realizar los modelos que determinan el comportamiento de un docente. Los modelos de comportamiento se generan a través de diferentes técnicas [4]; árboles de decisión, redes neuronales y reglas de decisión, que dependiendo de una situación específica y la variable de estudio permiten mostrar y seleccionar el mejor análisis. Estos resultados permiten generar estrategias que apoyan el proceso de formación académica y elevar la calidad de la educación en la Universidad Autónoma del Caribe.

Palabras clave: Metodología, ETL, Técnica, Aula Extendida, TALEND, WEKA.

ABSTRACT

Implemented a software tool that supports the analysis of the behavior of learning activities that teachers register with Extended Classroom at the Autonomous University of the Caribbean. For this, we adopt a methodology for project development in data mining, CRISP-DM phases providing business understanding, data understanding and data preparation, taking into account the ETL tool (Extraction, Transformation and Load) Talend Open, which lets you debug data for integration.

To search for patterns of behavior in the classroom teaching performance extended, taking into account the socio-demographic variables and levels of study. Using the WEKA tool that allows, through the training and validation data, make models that determine the behavior of a teacher. Behavioral models are generated using different techniques [4], decision trees, neural networks and decision rules that depend on a specific situation and the study variable can display and select the best analysis. These results allow us to create strategies that support the academic process and improve the quality of education at the Autónoma University of the Caribe.

Key words: Methodology, ETL, Technique, Extended Classroom, Talend, WEKA

1. Introducción

En la Actualidad las empresas suelen generar grandes cantidades de información sobre sus procesos productivos, desempeño operacional, mercados y clientes. Pero el éxito de los negocios depende; por lo general, de la habilidad para visionar las nuevas tendencias en manejo de información como apoyo a la toma de decisiones.

Se define la minería de datos como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos [10]. Es decir, la tarea fundamental de la minería de datos es encontrar modelos inteligibles a partir de los datos [2] y su objetivo es convertir datos en conocimiento.

Las aplicaciones de minería de datos identifican tendencias y comportamientos; que además de extraer información, descubren las relaciones en bases de datos que muestran comportamientos que no son muy evidentes. [3]

Por otro lado, la minería de datos está contribuyendo significativamente en la toma de decisiones [5], ya que permite determinar el comportamiento de los datos, encontrar patrones de secuencia y sus relaciones, facilitando la creación de modelos predictivos y descriptivos, es decir, representaciones abstractas de la realidad.

La Universidad Autónoma del Caribe brinda el Aula Extendida como plataforma para que los docentes agreguen actividades de enseñanza aprendizaje. Actualmente no posee un mecanismo que permita el análisis de las actividades secuenciales que se producen en una determinada sesión, por lo tanto no se tiene conocimiento del desempeño y buen uso este recurso como orientación a la formación académica.

Teniendo en cuenta lo anterior, se implementaran Patrones de Secuencia en el Aula Extendida de la Universidad Autónoma del Caribe, para la exploración y análisis de las actividades realizadas por los docentes, como apoyo en la toma de decisiones o plantear estrategias para potenciar el recurso en las actividades de enseñanza aprendizaje por parte de la Dirección Académica.

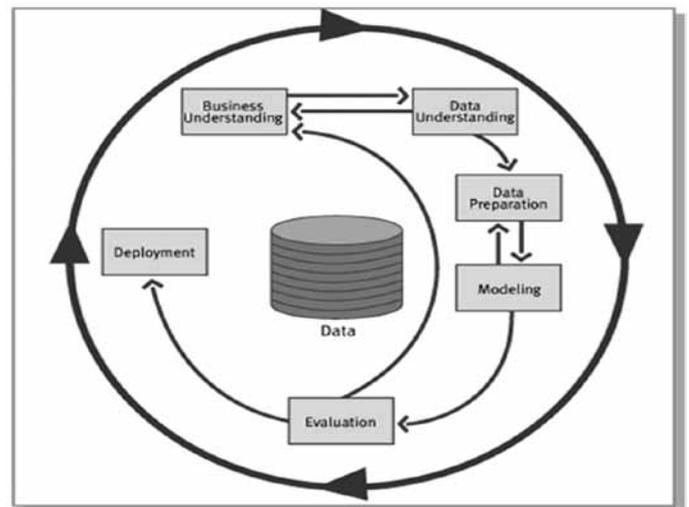
Para tal fin, se hace uso de la metodología CRISP-DM [10] [12] y haciendo uso la herramienta WEKA [7] [11], que permite a través de los datos de entrenamiento, validación y predicción realizar los modelos que permiten hacer las predicciones para determinar el comportamiento de un docente.

2. Metodología

La metodología de CRISP-DM [10] [12], está descrita en términos de un modelo de proceso jerárquico, consistente en un conjunto de tareas descritas en cuatro niveles de abstracción (de lo general a lo específico): fase, tarea genérica, tarea especializada, e instancia de procesos.

Figura 1. Etapas de un Proyecto de Minería de Datos. Metodología CRISP-DM

Figure 1. Stages of a Data Mining Project. CRISP-DM Methodology



2.1 Comprensión del Negocio

Determinar los Objetivos del Negocio:

La Universidad Autónoma del Caribe como Institución Educativa cuenta con un Aula Virtual para el proceso de enseñanza – aprendizaje. Su estructura para la realización de actividades docentes esta dada por:

Zona de descarga: Diseñada para que los docentes suban material de estudio relacionado al tema del curso asignado.

Zona de actividades de aprendizaje: Permite asignar actividades a los estudiantes, las cuales deben realizar y subir al aula.

Zona de Evaluación: en la cual se colocan en línea quiz tipo E-CAES, foros de discusión, entre otros para que los estudiantes desarrollen en determinado tiempo.

La plataforma actual le permite a la Dirección Académica de la Universidad Autónoma del Caribe obtener información de los docentes que utilizan el aula virtual, es decir, los que ingresan, pero no sobre los contenidos y actividades que se generan.

Por lo mencionado anteriormente los objetivos de negocio [6] del proyecto son los siguientes:

- Establecer perfiles de los docentes de acuerdo a las actividades realizadas en el Aula Virtual de la Universidad Autónoma del Caribe.
- Realizar estadísticas con base al desempeño de los docentes en la utilización de las herramientas que brinda el aula virtual de la Universidad Autónoma del Caribe.
- Establecer cuáles son las herramientas más y menos utilizadas por los docentes en el Aula Virtual de la Universidad Autónoma del Caribe.
- Predecir el desempeño de los docentes de acuerdo a variables socio demográficas y a su formación académica.

Comprensión de los Datos

2.2.1 Recolección de los datos Iniciales

Los datos se recolectaron a partir de las diversas tablas de la base de datos que hacen referencia a información relevante de los docentes y datos del aula extendida de la Universidad Autónoma del Caribe. Todos los datos identifican y relacionan las actividades que realiza un docente en el aula.

Los datos recolectados son los que se muestran a continuación:

- Nombres y Apellidos del docente
- Cédula
- Sexo
- Fecha Nacimiento
- Fecha de Ingreso a la Universidad
- Dirección
- Periodo Actual en la Universidad
- Horas laboradas
- Estudios realizados por el docente
- Las materias que dicta un docente en un periodo
- La actividad que realiza en el aula extendida
- La fecha en que ingreso a realizar la actividad
- El curso donde realizó la actividad

2.2.2 Descripción de los Datos

Los datos mencionados anteriormente tienen uno o varios campos asignados a la base de datos y en diversas tablas. A continuación se hace una descripción a detalle de toda esta información discriminado por tablas y por campos:

TABLA EMPLEADO: Se presentan los campos escogidos para la tabla profesores que hace referencia a la tabla EMPLEADO y a la tabla CATEDRÁTICO las cuales se unieron en una sola tabla diferenciando el tipo de docente de acuerdo a un nuevo campo que se agregó en esta nueva tabla. Los nombres de los campos se modificaron para un mayor entendimiento.

TABLA CATEDRÁTICOS: Esta tabla hace referencia a los datos de los docentes de la universidad, esta tabla se integró con la tabla empleados, ya que manejaban datos similares.

TABLA ESTUDIOS DOCENTES: Esta tabla muestra la información del grado de estudio de un docente.

TABLA MATERIAS DOCENTES: Esta tabla contiene información de las materias que tiene a cargo un determinado profesor. Se modificaron los nombres de los campos para un mayor entendimiento.

TABLA MDL_LOG: Esta tabla guarda los registros de las actividades realizadas por los docentes en una sesión determinada en el aula extendida. Dicha tabla se encuentra en la base de datos del moodle.

TABLA MDL_USERS: La cual tiene la información de los usuarios que entran al aula extendida.

TABLA MDL_COURSE: Esta tabla hace parte de la base de datos del moodle. En esta tabla se encuentra la información del curso la cual se relaciona con la tabla materia a través del IdNumber.

TABLA UAC_TODOS: En esta tabla se encuentra la relación de la información del docente en la base de datos de la universidad con la información del docente en el log.

TABLA MDL_ROLE_ASSIGNMENTS: A continuación se muestra la tabla mdl_role_assignments la cual hace parte de la base de datos del moodle es decir del aula virtual. Dicha tabla nos brinda la información del rol del usuario es decir si el usuario es profesor o estudiante.

TABLA MDL_ROLE: A continuación se muestra la tabla mdl_role la cual hace parte de la base de datos del moodle es decir del aula virtual. Dicha tabla nos brinda la información de los roles con que cuenta el moodle.

TABLA MINERIA_DATOS: A continuación se presenta la tabla minería de datos la cual contiene las variables necesarias para la evaluación de los datos, los cuales son obtenidos de las diversas tablas descritas anteriormente. La

tabla Minería de datos contiene un campo éxito que es la variable que nos determina si el docente cumplió con las actividades necesarias para subir en el aula extendida en un corte específico.

2.2.3 Verificación de los Datos

Al realizar una exploración minuciosa de los datos recolectados se analizó que en algunas tablas hay datos inconsistentes o están vacíos. Es el caso de la tabla empleados y catedráticos en la que se encontraron datos vacíos en la fecha de nacimiento de los docentes, sin embargo el número de registros con esta situación no era considerable.

2.3 Preparación de los Datos

2.3.1 Selección de los Datos

En esta etapa se seleccionan los datos que realmente serán utilizados para el modelo de minería de datos excluyendo los datos que no son relevantes para este propósito.

Estos datos fueron seleccionados de acuerdo a las características socio demográfico, y grado de estudio de los docentes tales como edad, sexo, nivel de estudio, tipo de profesor.

2.3.2 Limpieza de Datos

Como se dijo en la fase de verificación de datos se encontraron registros con campos vacíos, como fue el caso de

la fecha de nacimiento de la tabla catedráticos y empleados. Para solucionar esta inconsistencia se tomó la moda estadística de los registros y se le asignó a los fecha de nacimiento que se encontraban vacías. Al completar estos datos el impacto fue significativo en el modelo de minería de datos que se obtuvo, porque los resultados fueron lo esperado y se dejaron preparados para su transformación.

2.3.3 Transformación de Datos

Al analizar las tablas empleado y catedrático, se notó que tenían los mismos campos, a partir de estas se generó la tabla profesores con los campos de las tablas antes mencionadas. Además, se le agregó un nuevo campo llamado "Tipo" el cual toma dos dominios N para Ptc y S para catedráticos. Como se muestra a continuación:

Los campos que tienen asignadas fechas en las tablas se les modificará el tipo de datos convirtiéndose a tipo date con un formato dd/mm/yyyy. Esto se debe a que en la realización del modelo se hacen operaciones con campos tipo date ya que con cadena no se pueden realizar. Se añadirá un nuevo campo que es la edad del docente y se incluirá en la tabla profesor. Este dato se obtendrá a partir de una operación que se realice con la fecha de nacimiento del docente.

Para esto se utilizó la herramienta Talend Open Source como se muestra en la figura 5.

Figura 2. Tabla Empleado (Table employee)

Campo	Tipo de Dato	Tamaño	Descripción	Domino de Dato
EMPAPL2	Cadena	30	Segundo apellido del docente	Cadena
EMPAPL1	Cadena	30	Primer apellido del docente	Cadena
EMPNMB2	Cadena	30	Segundo nombre del docente	Cadena
EMPNMB1	Cadena	30	Primer nombre del docente	Cadena
CTDDCM	Númerico	12	Cedula del Docente.	Este campo hace referencia a otras tablas para otros datos suministrados.
EMPSXO	Char	1	Sexo del Docente	Sólo puede tomar el valor de F que es femenino y M masculino.
EMPNCMFCH	Númerico	8	Fecha de Nacimiento del Docente.	Tiene la siguiente estructura: los cuatro primeros dígitos son el año, los otros dos que le siguen son el mes y los últimos dos dígitos son el día.
EMPDRC	Cadena	60	Dirección del Docente	
RSLPRDANO	Númerico	6	Periodo Semestre	Periodo en el que se encuentra laborando el docente. Los cuatro primeros dígitos se encuentran formado por el año y los dos últimos por el mes.
CNTFCHINC	Númerico	8	Fecha de Inicio de Ingreso del Docente.	Los cuatro primeros dígitos son el año, los otros dos que le siguen son el mes y los últimos dos dígitos son el día.
EMPJRNLAB	Númerico	4	Horas que labora el docente.	Horas que labora el docente en la universidad en un semestre

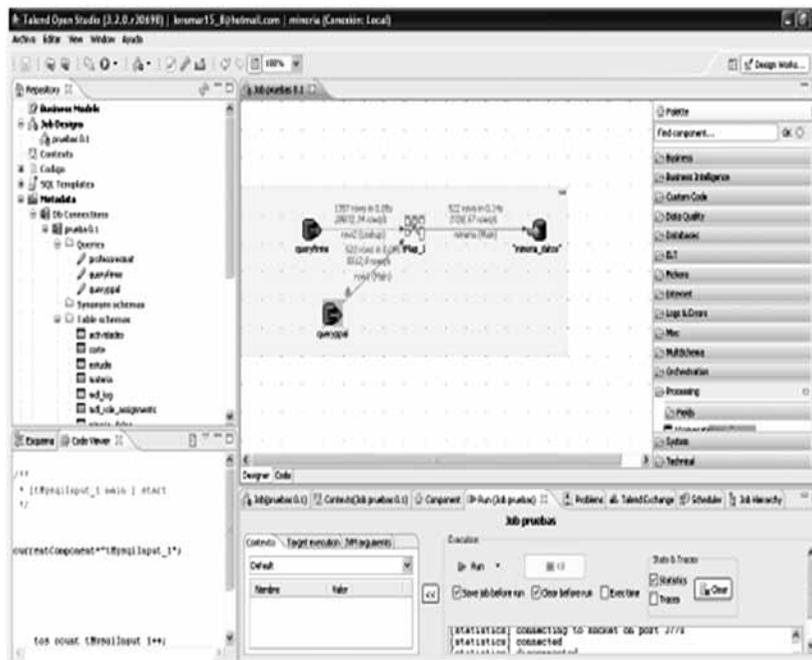
Figura 3. Tabla Catedrático (Proffesor)

Campo	Tipo de Dato	Tamaño	Descripción	Domino de Dato
CTDAPL2	Cadena	30	Segundo apellido del docente	Cadena
CTDAPL1	Cadena	30	Primer apellido del docente	Cadena
CTDNMB2	Cadena	30	Segundo nombre del docente	Cadena
CTDNMB1	Cadena	30	Primer nombre del docente	Cadena
CTDDCM	Númérico	12	Cedula del Docente.	Este campo hace referencia a otras tablas para otros datos suministrados.
CTDSXO	Char	1	Sexo del Docente	Sólo puede tomar el valor de F que es femenino y M masculino.
CTDNMFCH	Númérico	8	Fecha de Nacimiento del Docente.	Tiene la siguiente estructura: los cuatro primeros dígitos son el año, los otros dos que le siguen son el mes y los últimos dos dígitos son el día.
CTDDRC	Cadena	60	Dirección del Docente	
RSLPRDANO	Númérico	6	Periodo Semestre	Periodo en el que se encuentra laborando el docente. Los cuatro primeros dígitos se encuentran formado por el año y los dos últimos por el mes.
EMPJRNLAB	Númérico	4	Horas que labora el docente.	Horas que labora el docente en la universidad en un semestre
RSLINCPRD	Númérico	8	Perido Inicio de contrato.	Los cuatro primeros dígitos son el año, los otros dos que le siguen son el mes y los últimos dos dígitos son el día.

Figura 4. Tabla Integrada Profesores (Table Integrated Teaching)

Campo	Tipo Dato	Tamaño	Descripción	Dominio de Datos
Cedula	Cadena	60	Cedula del Docente	Valor Numéricos
Primer_nombre	Cadena	60	Primer Nombre del Docente	Cadena
Segundo_nombre	Cadena	60	Segundo Nombre del Docente	Cadena
Primer_apellido	Cadena	60	Primer Apellido del Docente	Cadena
Segundo_apellido	Cadena	60	Segundo Apellido del Docente	Cadena
Sexo	Cadena	6	Sexo del Docente	M (Masculino) - F (Femenino)
Fecha_nacimiento	Cadena	36	Fecha de Nacimiento	Fecha
Direccion	Cadena	135	Dirección Residencia del Docente	Cadena
Tipo	Cadena	6	Tipo de Docente	S para Catedráticos y N para PTC

Figura 5. Aplicación de Talend Open Source (Talend Open Source application)



2.4. Modelamiento

Al aplicar la herramienta Talend Open Source a las tabla catedráticos y empleados da como resultado la tabla Profesores, como se muestra en la figura **Figura 4** que junto con la tabla Minería _ datos conforman el modelo a evaluar:

4. Resultados

Después de generar el modelo mostrado en la figura 6. teniendo en cuenta datos de entrenamiento, validación y predicción, se evaluó la efectividad en cada una de las técnicas; Redes Neuronales, Árboles de Decisiones y Reglas de decisiones, para determinar cual es mejor una con respecto a la otra para este tipo de modelo.

A continuación tenemos una tabla comparativa con las cantidades de registros con éxito y no éxito de los 3 periodos analizados, 2008-01(Entrenamiento), 2008-01 (Validación) y 2009-01 (Predicción).

Periodos		
2008-01	2008-02	2009-01
3178	3301	3586
3062	2939	2654

Sí
No

4.1 Algoritmos “Redes Neuronales”

En la siguiente tabla relacionamos cada uno de los Algoritmos pertenecientes a las Redes Neuronales por cada período, teniendo en cuenta la predicción de registros no exitoso.

	MultilayerPerceptron	RBFNetwork	VotedPerceptron
2008-01	79,80%	73,50%	72,60%
2008-02	79,30%	79,70%	74,40%
2009-01	92,00%	60,30%	85,10%

4.2 Algoritmos “Árboles de Decisión”

En la siguiente tabla relacionamos cada uno de los Algoritmos de Arboles de Decisión por cada período, teniendo en cuenta la predicción de registros no exitoso.

	ADTree	BFTree	ecisionStum	FT	Id3
2008-01	84,10%	79,60%	16,20%	79,30%	79,60%
2008-02	86,70%	82,10%	16,80%	82,10%	82,10%
2009-01	92,90%	91,90%	94,30%	92,90%	87,30%

	J48	LADTree	NBTree	REPTree
2008-01	84,10%	84,10%	84,10%	84,40%
2008-02	86,70%	86,70%	86,70%	86,20%
2009-01	92,90%	92,90%	88,20%	91,90%

Se concluye del análisis realizado por cada Algoritmo, el resultado con mayor efectividad para la predicción de las variables de éxito. Podemos observar que hay 3 Algoritmos que cumplen con el criterio de efectividad. Por interpretación del Árbol que genera cada uno, para un mejor entendimiento de la información se escoge el J48 (C4.5).

4.3 Algoritmos “Reglas de Decisión”

En la siguiente tabla relacionamos cada uno de los Algoritmos de Reglas de Decisión por cada período, teniendo en cuenta la predicción de registros no exitoso.

Figura 6. Tabla Minería _ datos(Table Data Mining)

Campo	Tipo Dato	Tamaño	Descripción	Dominio de Datos
Cedula	Cadena	60	Cedula del Docente	Valor Numérico
Primer_nombre	Cadena	60	Primer Nombre del Docente	Cadena
Segundo_nombre	Cadena	60	Segundo Nombre del Docente	Cadena
Primer_apellido	Cadena	60	Primer Apellido del Docente	Cadena
Segundo_apellido	Cadena	60	Segundo Apellido del Docente	Cadena
Edad	Double		Edad del Docente	Valor Numérico
Sexo	Cadena	6	Sexo del Docente	M (Masculino) - F (Femenino)
Tipo	Cadena	6	Tipo de Docente	S (Catedratico) - N (PTC)
Estudio	Cadena	15	Estudio del Docente	Cadena
Curso	Cadena	30	Curso del Docente	Cadena
Exito	Cadena	15	Cumplimiento con las actividades	0 (No) - 1 (Si)
Corte	Cadena	30	Corte del Periodo	Cortes C1 - C2 - C3
Rango_edad	Cadena	45	Rango de Edades discretizados	Cadena

	ConjunctiveR	DecisionTabl	DTNB	Jrip
2008-01	16.20%	79.60%	79.60%	83.50%
2008-02	16.80%	82.10%	82.10%	86.10%
2009-01	94.30%	87.30%	92.00%	92.90%

	Nnge	OneR	PART	Ridor	ZeroR
2008-01	83.90%	51.00%	83.60%	83.30%	0.00%
2008-02	83.50%	52.10%	82.10%	85.10%	0.00%
2009-01	81.40%	56.10%	91.90%	91.90%	0.00%

A continuación se explican los resultados según la técnica aplicada:

Árbol de decisión J48 teniendo en cuenta las variables del negocio.

```

rango_edad = ENTRE 51 Y 60
| tipo = S: No (1425.0/629.0)
| tipo = N: Si (423.0/18.0)
rango_edad = ENTRE 41 Y 50
| tipo = S: No (1338.0/460.0)
| tipo = N: Si (447.0/42.0)
rango_edad = ENTRE 71 Y 80: No (75.0)
rango_edad = ENTRE 61 Y 70: No (495.0)
rango_edad = ENTRE 20 Y 30
| estudio = G
| | sexo = M
| | | tipo = S: Si (15.0)
| | | tipo = N: No (18.0)
| | sexo = F: Si (105.0/12.0)
| estudio = D
| | tipo = S: Si (141.0/12.0)
| | tipo = N
| | | sexo = M: No (63.0)
| | | sexo = F: Si (9.0)
| estudio = H: Si (30.0/2.0)
| estudio = C: Si (0.0)
| estudio = A: Si (0.0)
| estudio = B: No (96.0/4.0)
| estudio = F: Si (0.0)
| estudio = I: Si (0.0)
rango_edad = ENTRE 31 Y 40
| estudio = G: Si (756.0/40.0)
| estudio = D: Si (453.0/41.0)
| estudio = H: Si (291.0/20.0)
| estudio = C: Si (0.0)
| estudio = A: Si (0.0)
| estudio = B: No (48.0)
| estudio = F: Si (0.0)
| estudio = I: Si (12.0/2.0)
    
```

Según el árbol de decisión generado:

Los profesores que tienen una edad entre 61 y 70 en un 100% no suben actividades de aprendizaje en el aula extendida.

Los profesores que tienen una edad entre 41 y 50 que son Catedráticos en su 65,6% no participan en las actividades de aprendizaje para los estudiantes en el aula extendida.

Los profesores que tienen una edad entre 20 y 30 que tienen una especialización realizada que son de sexo masculino y son PTC en un 100% no participan en las actividades de aprendizaje para los estudiantes en el aula extendida.

Los profesores que tienen una edad entre 20 y 30 que tienen una carrera profesional que son PTC y que son de sexo Masculino en un 100% no participan en las actividades de aprendizaje para los estudiantes en el aula extendida.

Los profesores que tienen una edad entre 20 y 30 que tienen estudios de tecnología en un 96% no participan en las actividades de aprendizaje para los estudiantes en el aula extendida.

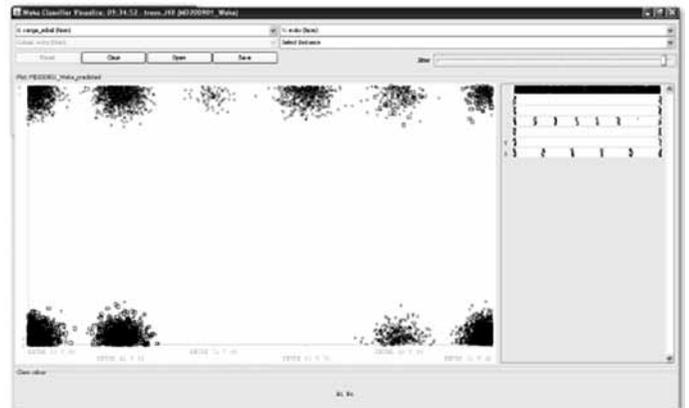
Los profesores que tienen una edad entre 31 y 40 que tienen estudios de tecnología en un 100% no participan en las actividades de aprendizaje para los estudiantes en el aula extendida.

Los profesores que tienen una edad entre 31 y 40 que tienen estudios de tecnología en un 100% no participan en las actividades de aprendizaje para los estudiantes en el aula extendida.

Árbol de Decisión – Algoritmo J48

Visualización

Rango Edad (eje X) Vs Éxito (eje Y)



	Si	No	Total
ENTRE 20 y 30	278	199	477
ENTRE 31 y 40	1409	151	1560
ENTRE 41 y 50	865	920	1785
ENTRE 51 y 60	1034	814	1848
ENTRE 61 y 70	0	495	495
ENTRE 71 y 80	0	75	75

Reglas de Decisión – Algoritmo JRip

A continuación se muestran los resultados que arrojó el algoritmo de Reglas de decisión JRip teniendo en cuenta las variables del negocio.

```
=== Classifier model (full training set) ===

JRIP rules:
*****

(tipo = S) and (rango_edad = ENTRE 61 Y 70) => exito=No (453.0/0.0)
(tipo = S) and (rango_edad = ENTRE 41 Y 50) => exito=No (1338.0/460.0)
(tipo = S) and (rango_edad = ENTRE 51 Y 60) => exito=No (1425.0/629.0)
(estudio = B) => exito=No (144.0/4.0)
(rango_edad = ENTRE 71 Y 80) => exito=No (75.0/0.0)
(rango_edad = ENTRE 20 Y 30) and (tipo = N) and (sexo = M) => exito=No (81.0/0.0)
(rango_edad = ENTRE 61 Y 70) => exito=No (42.0/0.0)
=> exito=Si (2682.0/189.0)

Number of Rules : 8
```

6. Conclusiones

- El uso de la metodología CRIP-DM facilitó el desarrollo de la investigación, teniendo en cuenta la naturaleza de los datos que se manejan en el aula extendida de la Universidad Autónoma del Caribe.
- Las herramientas Open Sources: Talend y Weka , fueron esenciales en el desarrollo de la investigación. La primera fue fundamental en la preparación de los datos tomados de las tablas originales y transformados para la construcción del modelo y la aplicación de la minería de datos. La segunda permitió evaluar el modelo a través de las técnicas: Redes Neuronales, Árboles de Decisiones y Reglas de Decisiones, proporcionando resultados para la toma de decisiones por parte del departamento de secretaria general.
- Aplicando las técnicas redes neuronales, reglas de decisión con el algoritmo JRip y árbol de decisión con el algoritmo RJ48, se logra determinar que esta última es la que más se ajusta al fenómeno en estudio.
- Con relación a los resultados del estudio, se encontró que los profesores no participan en actividades de aprendizaje a través del Aula Extendida. Se analizaron las variables edad, tipo de contrato y nivel de educación.

Referencias

- [1] KRABARTI, SOUMEN. Mining The Web Discovering Knowledge From Hypertext Data. Morgan Kaufmann, 2002.
 - [2] HERNÁNDEZ, José. Introducción a la Minería de Datos. Pearson Educación, 2004.
 - [3] DUNHAM, Margaret. Data Mining Introductory and Advanced Topic. Prentice Hall, 2003.
 - [4] EIBE, Frank. Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). Morgan Kaufmann, 2005.
 - [5] MARGARET DUNHAM, Data Mining Introductory and Advanced Topic, Prentice Hall, 2003, P 204.
 - [6] MEHMED KANTARDZIC, Data Mining: Concepts, Models, Methods, and Algorithms, Wiley- IEEE Press, 2002, P 180.
 - [7] MORGAN KAUFMANN, Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). 2005
 - [8] PEREZ, Cesar. Data Mining, soluciones con Enterprise Miner. Alfaomega, 2006
 - [9] GIUDICI, Paolo. Applied Data Mining, Wiley, 2006
 - [10] WITTEN, Ian. FRANK, Eibe. Data Mining: Practical Machine Learning Tools and Techniques, Second Edition. Morgan Series in Data Management Systems. 2005
- ENLACES:**
- [11] **Weka 3: Data Mining Software in Java.** [Internet]. WEKA the University of Waikato.
Disponible desde: <<http://www.cs.waikato.ac.nz/ml/weka/>> [Acceso Julio 10 de 2010]
 - [12] **CRISP-DM.** [Internet]. Cross Industry Standard Process for Data Mining. Disponible desde : <http://www.crisp-dm.org/> [Acceso Marzo 5 de 2010]