Universidad&Ciencia Vol. 2, No. 2, abril-julio (2014)

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

LA MINERÍA DE DATOS APLICADA A LA GESTIÓN DOCENTE DE LA UNIVERSIDAD NACIONAL EXPERIMENTAL DE LAS FUERZAS ARMADAS (UNEFA)

THE DATA MINING APPLIED TO THE TEACHING MANAGEMENT OF THE NATIONAL EXPERIMENTAL UNIVERSITY OF THE ARMED FORCES (UNEFA)

Autores: José Moreno Rodríguez[1]

Cosme Santiesteban-Toca[2]

Yulkeidi Martínez Espinosa[2]

Institución: [1] Universidad Nacional Experimental de las Fuerzas Armadas

[2] Universidad de Ciego de Ávila Máximo Gómez Báez, Cuba

Correo electrónico: jose.moreno.unefa@gmail.com

RESUMEN

La Universidad Nacional Experimental de las Fuerzas Armadas (UNEFA) cuenta con un sistema estandarizado de gestión para la docencia, que registra toda la información relacionada con los estudiantes, sin embargo, en el momento de tomar alguna decisión sobre el aprovechamiento docente de los estudiantes en los primeros años de la carrera, no tiene en cuenta el conocimiento oculto (información no evidente, desconocida a priori y potencialmente útil) de los datos que mantiene almacenados para sustentar determinadas líneas estratégicas en el proceso enseñanza-aprendizaje. El objetivo de la presente investigación es crear una herramienta de evaluación de los estudiantes de nuevo ingreso, a partir del empleo de técnicas de minería de datos que permita a los profesores mejorar los métodos pedagógicos y didácticos en la formación de profesional altamente calificados en la UNEFA. Se realizan cuatro tareas de minería, relacionadas con la asociación, el agrupamiento, la selección de atributos y la clasificación, desarrollándose un total de 48 experimentos, aplicando la metodología CRISP-

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

MD. Como resultados de la la investigación se obtuvieron un conjunto de reglas que presentan parámetros aceptados para ser consideradas útiles durante la toma de decisiones por la Junta Directiva de la UNEFA. La investigación realizada aporta nuevos conocimientos, que permiten redefinir algunos objetivos del negocio, y replantearse un nuevo proceso de KDD.

Palabras clave: Proceso docente, Minería de datos, KDD, Inteligencia Artificial

ABSTRACT

The Experimental National University of the Armed forces (UNEFA) it has a standardized system of administration however for the docencia that registers all the information related with the students, in the moment to make some decision on the use educational the students in the first years of the career, he/she doesn't keep in mind the hidden knowledge (non evident information, ignored a priori and potentially useful) of the data that it maintains stored to sustain certain strategic lines in the process teaching-learning. The objective of the present investigation is to create a tool of the students' evaluation again entrance, starting from the technical empleode of mining of data that allows the professors to improve the pedagogic and didactic methods in the highly qualified professional formation in the UNEFA. They are carried out four mining tasks, related with the association, the cluster, the selection of attributes and the classification, being developed a total of 48 experiments, applying the methodology CRISP-MD. As results of the present investigation they are obtained a group of rules that you/they present parameters accepted to be considered useful during the taking of decisions for the Directive Meeting of the UNEFA. The carried out investigation contributes new knowledge that allow to redefine some objectives of the business, and to reconsider a new process of KDD.

Keywords: I process educational, Mining of data, KDD, Artificial Intelligence.

INTRODUCCIÓN

El proceso de toma de decisiones está regido, de manera general, por la experiencia de los directivos involucrados en las acciones que se planifican, y el concierto de factores subjetivos y en ocasiones hasta imprecisos (propios del

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

comportamiento humano), que pueden entorpecer el diseño de una estrategia correcta. Por otra parte, el aumento del volumen y variedad de la información que se encuentra almacenada en bases de datos y otras fuentes, ha crecido espectacularmente en las últimas décadas, constituyendo gran parte de ella una reseña de situaciones que se han producido.

Toda esta «memoria histórica» puede procesarse no solo para explicar el pasado, sino también para entender el presente y predecir el futuro. El empleo de técnicas de minería de datos para tratar toda esta información, permite a las empresas, organizaciones e instituciones tener mayores elementos, esta vez de carácter científico, para apoyar la toma de decisiones (Díaz, 2013).

La minería de datos es un término relativamente moderno que integra numerosas técnicas de análisis de datos y construcción de modelos, y permite extraer patrones, describir tendencias y regularidades, predecir comportamientos y en general, sacar partido a la información digitalizada que nos rodea hoy en día, generalmente heterogénea y en grandes cantidades, ayudando a los individuos y a las organizaciones a comprender y modelar, de una manera más eficiente y precisa, el contexto en el que deben actuar y tomar decisiones.

La UNEFA cuenta con un sistema de gestión para la docencia que registra datos personales, relativos al ingreso, matrícula actual y pasada de sus estudiantes, así como información sobre las materias vencidas y en curso, notas, tipo de evaluación, períodos académico, entre otros. En un estudio realizado en la UNEFA se observa como manifestaciones que: a) Poca información brindada por los reportes generados del sistema de gestión docente, desde el punto de vista administrativo, b) La toma de decisiones en el ámbito académico se apoya fundamentalmente en la experiencia de los directivos, docentes y las observaciones del proceso que estos realizan y c) Dificultad para hacer estudios demográficos sobre la población estudiantil.

Al profundizar en el estudio se reflejan como principales causas: a) Incapacidad general de las personas para procesar grandes volúmenes de información e identificar ciertos patrones de comportamiento útiles para apoyar sus

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

decisiones, b) Falta de experiencia de los directivos en el empleo de herramientas de análisis para la identificación de patrones, que puedan apoyar determinadas líneas estratégicas trazadas para la dirección, c) No aprovechamiento del conocimiento subyacente de los datos que mantiene almacenados en el sistema de gestión para la docencia, etc.

El proceso de extracción de conocimiento en bases de datos KDD (del inglés Knowledge Discovery in Databases) (Orallo, 2004), posibilita la extracción de conocimiento oculto en los datos a través de técnicas de minería y contribuye a la mejor compresión de la situación docente en la UNEFA, y de apoyo a la toma de decisiones en este sentido. La investigación en su conjunto persigue como objetivo crear una herramienta de evaluación de los estudiantes de nuevo ingreso, a partir del empleo de técnicas de minería de datos que permita a los profesores mejorar los métodos pedagógicos y didácticos en la formación de profesional altamente calificados en la UNEFA.

DESARROLLO

El término de Minería de Datos (MD, del inglés Data Mining) deriva de la similitud que se encuentra entre buscar valiosa información de negocio en grandes bases de datos, y la búsqueda de vetas de metales preciosos dentro de una montaña, pues ambos procesos requieren examinar inteligentemente una inmensa cantidad de material, hasta encontrar algo que pueda resultar realmente útil e interesante (Pisón, 2003; Han et al. 2006).

Una definición ampliamente aceptada de la minería de datos«…es el proceso que tiene como propósito descubrir, extraer y almacenar información relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones y relaciones globales, tendencias, desviaciones y otros indicadores aparentemente caóticos que tienen una explicación que puede descubrirse mediante diversas técnicas de esta herramienta» (Larrieta, 2009).

La tarea fundamental de la minería de datos es descubrir conocimiento (reglas, patrones) a partir de grandes volúmenes de datos, apoyados en técnicas o herramientas (automáticas o asistidas), de tal manera que su uso ayude a

Universidad&Ciencia Vol. 2, No. 2, abril-julio (2014)

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

tomar decisiones más seguras que reporten algún tipo de beneficio a las organizaciones.

El KDD (Knowledge Discovery in Databases) según (Fayyad, 1996) es la extracción automatizada de conocimiento o patrones interesantes, no triviales, implícitos, previamente desconocidos, potencialmente útiles y predictivos de la información de grandes Bases de Datos.

El descubrir conocimiento en bases de datos o KDD, puede definirse como «el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia, comprensibles a partir de los datos» (Fayyad, 1996). Un proceso clásico de KDD se organiza en torno a cinco fases fundamentales (Orallo, 2004): 1) Integración y Recopilación, 2) Selección, Limpieza y Transformación, 3) Minería de Datos, 4) Evaluación e Interpretación y 5) Difusión y Uso (Ver figura 1).

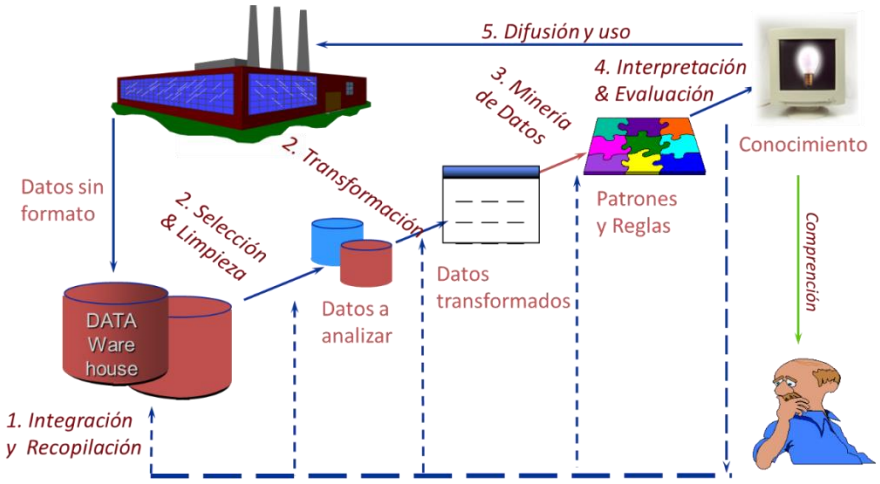

Figura 1. Fases de un proceso clásico de KDD (Orallo, 2004).

En la fase de integración y recopilación (1) se determinan las fuentes de información que pueden ser útiles y dónde conseguirlas, se transforman todos los datos a un formato común, y se detectan y resuelven las inconsistencias. La segunda fase (2), referente a la selección, limpieza y transformación, es donde se eliminan o corrigen los datos incorrectos, y se decide la estrategia a seguir con los datos incompletos, además, se consideran únicamente aquellos

Universidad&Ciencia Vol. 2, No. 2, abril-julio (2014)

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

atributos que van a ser relevantes, con el objetivo de hacer más fácil la tarea propia de minería.

La fase de minería de datos (3), se aplica el modelo, la tarea, la técnica y el algoritmo seleccionado para la obtención de reglas y patrones. Posteriormente se procede a la fase de evaluación e interpretación (4), donde se evalúan los patrones y se analizan por expertos, y si es necesario, se vuelve a las fases anteriores para una nueva iteración. Finalmente, en la fase de difusión (5) se hace uso del nuevo conocimiento y se hace partícipe de él a todos los posibles interesados.

Minería de Datos en el entorno educacional.

En el ámbito educacional, eje central de esta investigación, las técnicas de minería de datos pueden descubrir información útil para ser usada en el momento de establecer las bases docentes para las decisiones estratégicas, una vez que se diseñe o modifique el entorno de aprendizaje de un proceso educativo.

La aplicación de la minería de datos en los sistemas educacionales es un ciclo interactivo de formación de hipótesis, pruebas y refinamiento. Como se observa en la (Figura 2), los profesores y responsables académicos son los encargados de diseñar, planificar, y mantener el Sistema Educacional. Los estudiantes por su parte, usan, interactúan y de manera general participan en el mismo.

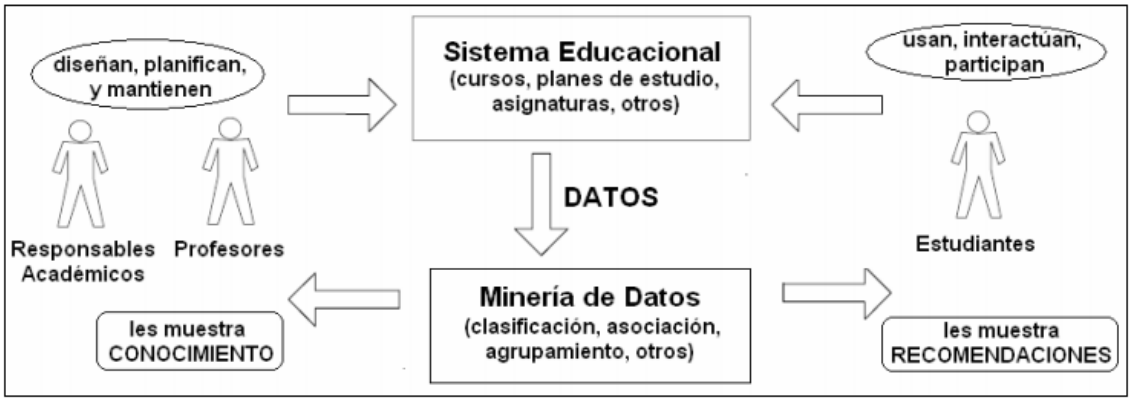

Figura 2. Ciclo de aplicación de la Minería de Datos a un Sistema Educacional (Romero and Ventura 2007)

A partir de toda la información disponible sobre los estudiantes, los profesores, los cursos, y las interacciones entre todos, las diferentes técnicas de minería de

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

datos pueden ser aplicadas a fin de descubrir conocimiento útil que ayude a mejorar el proceso docente educativo. El conocimiento descubierto puede ser usado no solo por los educadores (profesores y responsables académicos) sino también por los estudiantes, así la aplicación de la minería de datos a un sistema educacional puede estar orientada a diferentes actores, cada uno con un punto de vista particular (Romero and Ventura, 2007).

- Orientado hacia los Estudiantes: El objetivo es recomendar actividades, deberes y tareas que puedan favorecer y mejorar su aprendizaje, sugerir experiencias, o guiar por un camino más corto el avance, basado en actividades antes realizadas por estudiantes similares.
- Orientado hacia los Profesores: El objetivo es la retroalimentación de la enseñanza, la evaluación de la estructura de los cursos y su efectividad en el proceso de aprendizaje, la clasificación de estudiantes en grupos, basados en sus necesidades específicas, sus errores más frecuentes, y las actividades que resultan más efectivas para formarlos.
- Orientado hacia los Responsables Académicos: El objetivo es contar con parámetros para mejorar la eficiencia en el aprovechamiento de los estudiantes, optimizar los recursos institucionales (humanos y materiales) y organizar de manera eficiente, los programas de aprendizaje y la construcción institucional de planes de estudio educacionales.

Existen varios trabajos relacionados con la aplicación de la minería de datos en el entorno educacional (Sanjeev and Zytkow, 2007), (Becker et al. 2000), (Ma et al. 2000), (Luan, 2000), que empleaba técnicas de minería para comprender la matrícula de los estudiantes a fin de apoyar la toma de decisiones respecto a las políticas institucionales, proactivamente manejar sus resultados docentes, identificaban a los estudiantes que presentaban dificultades en el aprendizaje, para lo cual empleaba técnicas como la asociación y clasificación, agrupamiento y predicción.

Más recientemente se encuentran los trabajos de (Espinosa and Pérez, 2007) que utiliza técnicas de agrupamiento y asociación para encontrar patrones

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

entre los resultados académicos del primer año y el origen social de los estudiantes. Ese mismo año (Acosta, 2007) empleaba fundamentalmente técnicas de clasificación para predecir el resultado de los estudiantes en su primer año, en dependencia de sus características de ingreso y la especialidad que cursaban. (Brito, 2008) desarrolla un proceso de Descubrir Conocimiento en Bases de Datos KDD en el Instituto Superior Politécnico José Antonio Echeverría, como ayuda a la toma de decisiones de su Vice Rectoría Docente (VRD), que permiten apoyar determinadas actividades orientadas a la docencia.

Los estudios anteriores dejan abierta la investigación encaminada al refinamiento de sus resultados, y en la profundización y ampliación de los mismos. Quedando pendientes, conocer por especialidades, qué relaciones se establecen entre las características de ingreso de los estudiantes, qué grupos de ellos se comportan con rasgos similares, cuáles características influyen en el desempeño de cada uno de los años académicos y cómo lo hacen, o cómo inciden, por ejemplo, los resultados de determinadas asignaturas en el desempeño de otras.

Métodos y herramientas

El desarrollo de un proyecto de KDD provoca usualmente desviaciones y retrasos en su planificación, lo que se debe de forma general a que no es posible extraer conclusiones por adelantado, unido al hecho de que una gran parte del esfuerzo se produce en la preparación de los datos (Gondar, 2004).

Para el desarrollo del proceso se utiliza la metodología CRISP-DM (Chapman et al. 2000), una de las metodologías más difundidas y utilizadas, estructura el proyecto de KDD en fases que se encuentran interrelacionadas entre sí, y lo describen de forma iterativa e interactiva. La herramienta seleccionada para el análisis de los datos es WEKA (Waikato Enviromentfor Knowledge Analysis) (WEKA, 2005) es una aplicaciones de código abierto y de libre distribución y no compromete su uso con una metodología en particular. Además exhibe altas prestaciones en lo referente al pre-procesado de los datos y a la modelación de los mismos (Acosta, 2007).

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

Por otro lado, el proceso de minería de datos requiere en un principio, establecer los objetivos para el análisis de los datos disponibles (Orallo, 2007), de ahí que para su cumplimiento sean necesarias varios tipos de tareas, entre las que sobresalen por su uso: Clasificación, Regresión (Predicción o Estimación), Agrupamiento (Clustering o Segmentación), Asociación, y la Correlación, entre otros. (Pisón, 2003; López and Herrero, 2004).

Los algoritmos constituyen la forma de implementar, paso a paso, cada una de las tareas y técnicas de minería de datos. De este modo, es preciso conocer sus parámetros de entrada y sus características, para preparar los datos sujetos al análisis. Entre los algoritmos más significativos pertenecientes a tareas propias de minería se pueden mencionar a: K-medias (Agrupamiento), Apriori (Asociación), ID3 (Árboles de Decisión. Clasificación) y C4.5 (Árboles de Decisión. Clasificación) (López and Herrero, 2004).

Minería de datos aplicados a la carrera de Ingeniería de sistemas en la UNEFA.

El proceso de KDD se utiliza para extraer los datos almacenados en el sistema docente en la Universidad Nacional Experimental de las Fuerzas Armadas (UNEFA), utilizando la metodología CRISP-DM y la herramienta WEKA. Para ello se emplean técnicas de asociación, agrupamiento y clasificación con árboles de decisión. Se pretende encontrar, tanto a nivel de Universidad como de Carrera, reglas que describan las relaciones entre las características de los estudiantes al entrar a la universidad y grupos de ellos con atributos similares, de igual modo se intenta predecir el promedio de cada año académico en función de las características y resultados anteriores que presenten los educandos.

El modelo de proceso de CRISP-DM (Chapman *et al*., 2000) proporciona una descripción del ciclo de vida de un proyecto de minería de datos, que contiene las fases de un proyecto, sus tareas respectivas y las relaciones entre ellas. Dicho ciclo, cuya secuencia no es rígida, consta de seis fases: Comprensión del Negocio, Comprensión de los Datos, Preparación de los Datos, Modelado, Evaluación y Despliegue

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

La información recopilada fue obtenida de una única fuente, que de manera centralizada es la empleada en la UNEFA para registrar todos los datos relacionados con la docencia. Los resultados más significativos del reporte exploratorio de los datos por campos son: Municipio, Sexo, Raza, Notas de 1er año, Notas de 2do Año, Notas de 3ro Año, Notas de 4to Año y Notas de 5to Año, Promedio, Tipo de Curso, Vía de Ingreso, Situación del Padre, Situación de la Madre, Nivel Educacional del Padre, Nivel Educacional de la Madre, Centro de Procedencia, Carrera.

Resultados Experimentales Obtenidos.

- Con una probabilidad aproximada del 81% se cumple que si el centro de procedencia del estudiante es un Liceo (Liceo Bolivariano Dionisio López Origuela, Néstor Luis Pérez o Anibal Rojas), el concepto de ingreso a la universidad es directo, y el promedio es menor o igual a 17 puntos, entonces el promedio de 1er año será entre [13;15] (esta regla resulta poco significativa, respecto al otro intervalo del promedio); sin embargo, si el estudiantes es técnico medio (proveniente del IUT Delfín Mendoza o de la propi UNEFA), entonces el promedio de 1er año será entre (15;17] (esta regla es la que mayor cantidad de instancias clasifica correctamente).
- Con una probabilidad aproximada del 75% se cumple que si el promedio de 1er año del estudiante es entre [13;15], entonces el promedio de 2do será también entre [13;15] (esta regla resulta poco significativa, respecto al otro intervalo del promedio); sin embargo, si el promedio de 1er año es entre (13;15] y es blanco, entonces el promedio de 2do será entre (13;15] (esta regla es la que mayor cantidad de instancias clasifica correctamente).
- Con una probabilidad aproximada del 80% se cumple que si el promedio de 2do año del estudiante es entre [13;15], el nivel educacional de uno de los padres es universitario, entonces el promedio de 3er año será entre [13;15] (esta regla resulta poco significativa, respecto al otro intervalo del promedio); sin embargo, si el promedio de 2do es entre

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

(15;17], entonces el promedio de 3ro también será entre (15;17] (esta regla es la que mayor cantidad de instancias clasifica correctamente).

- Con una probabilidad aproximada del 89% se cumple que si el promedio de 3er año del estudiante es entre [13;15], la provincia de residencia es en Tucupita, Estado Delta Amacuro, el promedio de 2do es entre [13;15], y no pertenece a ninguna organización política, entonces el promedio de 4to año será entre [13;15] (esta regla resulta poco significativa, respecto al otro intervalo del promedio); sin embargo, si el promedio de 3er año es entre (15;17], entonces el promedio de 4to será igualmente entre (15;17] (esta regla es la que mayor cantidad de instancias clasifica correctamente).
- Con una probabilidad aproximada del 93% se cumple que si el promedio de 4to año del estudiante es entre [13;15], la provincia de residencia es en Tucupita, Estado Delta Amacuro, si el estudiante es graduado de técnico, entonces el promedio de 5to año será entre [15;17] (esta regla resulta poco significativa, respecto al otro intervalo del promedio); sin embargo, si el promedio de 4to año es entre (15;17], entonces el promedio de 5to será igualmente entre (15;17] (esta regla es la que mayor cantidad de instancias clasifica correctamente).

CONCLUSIONES

Se considera que los objetivos planteados se han alcanzado, dando respuesta al problema de la investigación formulado. Teniendo en cuenta los resultados obtenidos, se llegó a la conclusión de que el empleo de una metodología y una herramienta de análisis de datos en un proceso de minería resultan convenientes, si se desea obtener resultados satisfactorios, en este sentido, la elección de CRISP-DM y WEKA respectivamente, brindan las características necesarias para ser empleadas en el proyecto. El proceso de KDD se desarrolló sobre un conjunto inicial de datos formado por 25 atributos y 935 instancias. Luego del pre-procesado de los mismos, y producto de su calidad, se decide trabajar con dos vistas minables independientes. Además, enfocados

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

en los objetivos del negocio, fueron planteadas cuatro tareas de minería, relacionadas con la asociación, el agrupamiento, la selección de atributos y la clasificación, desarrollándose un total de 48 experimentos. Las reglas obtenidas presentan parámetros aceptados para ser consideradas útiles durante la toma de decisiones por la Junta Directiva de la UNEFA, que aporta nuevos conocimientos, que permiten redefinir algunos objetivos del negocio y replantearse un nuevo proceso de KDD.

BIBLIGRAFÍA CONSULTADA

ACOSTA, R.; VÁZQUEZ, L.; BRITO, R.; ROSETE, A: *Empleo de Minería de Datos para la obtención de patrones en el Sistema Docente del Instituto Superior Politécnico José Antonio Echeverría.* III Taller de Inteligencia Artificial, Memorias III Conferencia Científica de la Universidad de las Ciencias Informáticas. UCIENCIA pp. 25-27, La Habana, Cuba, 2007.

BECKER, K., GHEDINI, C., & TERRA, E: *Using KDD to analyze the impact of curriculum revisions in a Brazilian university. In Eleventh international conference on data engineering*, Proceedings of the SPIE 14th annual international conference on aerospace/defense, sensing, simulation and controls, 2000.

BRITO R.: *Minería de Datos aplicada a la Gestión Docente del Instituto Superior Politécnico José Antonio Echeverría*, Tesis para optar por la Maestría en Informática Aplicada, Instituto Superior Politécnico José Antonio Echeverría, Ciudad de la Habana, Cuba, 2008.

CHAPMAN, P. [et al.] CRISP-DM 1.0: *Step-by-step data mining guide*, SPSS Inc., CRISP-DM Consortium, United States, 2000.

DÍAZ, F. J.; OSORIO, M. A.; AMADEO, A. P. AND ROMERO, D.: *Aplicando estrategias y tecnologías de Inteligencia de Negocio en sistemas de gestión académica*, XV Workshop de Investigadores en Ciencias de la Computación, 2013.

ESPINOSA, I., PÉREZ, S.: *Obtención de Reglas y Patrones en el Proceso Académico de la Universidad de Ciencias informáticas*, Tesis de Diploma, CEIS, Cuba, 2007.

Pág. 39-51 ISSN: 2227-2690 RNPS: 2450

Recibido: 16/01/2014 Aceptado: 14/03/2014 http://revistas.unica.cu/uciencia

Publicado 30/4/2014

FAYYAD, U. PIATETSKY-SHAPIRO, G. AND SMYTH, P.: *«The KDD process for extracting useful knowledge from volumes of data, Commun»*, en ACM, vol. 39, no. 11, pp. 27–34, 1996.

GONDAR, J. E: *«Metodologías para la Realización de Proyectos de Data Mining»,* Madrid, España, 2004. Data Mining Institute. Disponible en http://www.estadistico.com. Visitado el 12 de julio de 2007.

HAN, J. KAMBER, M. AND PEI, J: *Data mining: concepts and techniques*, Morgan kaufmann, 2006.

LARRIETA, Á.; ISABEL, M. A. AND SANTILLÁN, A: *Minería de datos: concepto, características, estructura y aplicaciones,* Contaduría y Adm., No. 190, 2009.

LÓPEZ, J. M. M. AND HERRERO, J. G: *Técnicas de análisis de datos*, Univ. Carlos III, Madrid, España, 2004.

LUAN, J: *Data mining, knowledge management in higher education, potential applications. In Workshop associate of institucional research international conference*, pp. 1–18. Toronto, Canadá, 2002.

MA, Y., LIU, B., WONG, C., YU, P., & LEE, S: *Targeting the right students using data mining. In KDD* '00, Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 457–464, 2000.

ORALLO, J. H. QUINTANA, M. J. R., AND RAMÍREZ, C. F: *Introducción a la Minería de Datos*, Pearson Prentice Hall, 2004.

PISÓN, F. J. M.: *Optimización mediante técnicas de minería de datos del ciclo de recocido de una línea de galvanizado*, Universidad de la Rioja, 2003.

ROMERO C. AND VENTURA S.: *Educational data mining: A survey from 1995 to 2005, Expert Syst. Appl.*, vol. 33, no. 1, pp. 135–146, 2007.

SANJEEV, P., & ZYTKOW, J. M.: *Discovering enrollment knowledge in university databases,* In KDD. pp. 246–251. 1995.

WEKA, Waikato Environment for Knowledge Analysis. Weka 3: Data Mining Software in Java. [Software]. Weka 3.4.7, 2005. Disponible en http://www.cs.waikato.ac.nz/ml/weka/. Visitado el 20 de septiembre de 2013.