

Estudio de Perfil de Rendimiento Académico: Un Abordaje desde Data Warehousing

David La Red Martínez¹, Marcelo Karanik², Mirtha Giovannini³, Noelia Pinto⁴

^{1,2,3,4}Facultad Regional Resistencia, Universidad Tecnológica Nacional

French 414, Resistencia, Chaco

¹laredmartinez@gigared.com, ²mkaranik@gmail.com, ³meg_c51@yahoo.com.ar, ⁴ns.pinto@gmail.com

Resumen: Con la necesidad de definir perfiles de estudiantes y determinar patrones que conduzcan al éxito o fracaso académico, se propone evaluar el rendimiento de los estudiantes mediante técnicas de Minería de Datos. La propuesta no se enfoca sólo en analizar el perfil del estudiante a través de sus calificaciones sino también estudiar el desempeño académico en base a otras variables. En este trabajo se describe un modelo de Data Warehouse para determinar los perfiles de rendimiento académico en la asignatura Algoritmos y Estructuras de Datos de la carrera Ingeniería en Sistemas de Información de la UTN-Facultad Regional Resistencia (UTN-FRRe) con el objetivo de instrumentar estrategias que ayuden a aquellos estudiantes con perfil tendiente al fracaso o deserción. La característica principal del modelo es que combina aspectos relacionados a las calificaciones del estudiante junto a otras variables tales como factores socioeconómicos, demográficos, culturales, entre otros, en base a lo cual permite clasificar diferentes perfiles de alumnos.

Palabras Claves: Rendimiento académico, perfiles de alumnos, almacenes de datos, minería de datos.

Abstract: With the need to define profiles of students and identify patterns that lead to academic success or failure, it is proposed to evaluate the performance of students using data mining techniques. The proposal focuses not only on analyzing the profile of the student through their grades, but also study the academic performance based on other variables. In this paper, a model of Data Warehouse to determine the profiles of academic achievement in the course Algorithms and Data Structures, in Information Systems Engineering from UTN-Facultad Regional Resistencia (UTN-FRRe), is described with the aim of implement strategies to help those students with profile tending to failure or dropout. The main feature of the model is that it combines aspects of a student's grades with other variables such as socioeconomic, demographic, cultural factors, among others, based on allowing to classify different student profiles.

Keywords: Academic performance, student profiles, data warehouses, data mining.

INTRODUCCIÓN

La Universidad enfrenta actualmente el desafío de mejorar su calidad académica sin incremento presupuestario en cantidad de cargos docentes y de tutorías y, al mismo tiempo, adaptarse a las demandas de nuevos escenarios producto del contexto globalizado del cual forma parte. Este reto no implica enfocarse sólo en el sistema de enseñanza-aprendizaje sino que

debe contemplar otras tantas variables, como por ejemplo la sistematización de procesos de evaluación permanentes que permitan monitorear cuestiones ligadas a la calidad académica y retroalimente la propuesta de mejora para la Universidad (Briand et al., 1999). En este sentido uno de los factores más críticos que debe evaluarse continuamente es el rendimiento académico.

Se define al rendimiento académico como la produc-

tividad del sujeto, matizado por sus actividades, rasgos y la percepción más o menos correctade los cometidos asignados (Maletic et al., 2002).

Generalmente al evaluar el rendimiento académico se analizan en mayor o menor medida los elementos que influyen en el desempeño como, entre otros, factores socioeconómicos, la amplitud de programas de estudio, las metodologías de enseñanza, conocimientos previos del alumno (Marcus, 2003). Por esto no resulta adecuado evaluar el desempeño general de los alumnos a través de porcentajes de aprobación, notas obtenidas, etc., ya que este proceso de evaluación no brinda información que pueda ser utilizada para detectar, y corregir, problemas cognitivos, de aprehensión, de discernimiento, etc.

Surge, entonces, la necesidad de implementar un mecanismo que permita determinar las características propias del estudiante analizando la existencia de relaciones y patrones de comportamiento estudiantiles que posibiliten la definición clara de perfiles de alumnos. Para ello una alternativa viable es desarrollar los métodos de evaluación en base a las tecnologías de información disponibles aprovechando los beneficios y avances de éstas al momento de resolver problemas en diversos ámbitos.

Actualmente existen varios métodos para determinar y clasificar patrones que se utilizan en el área de la Inteligencia Artificial (del inglés Artificial Intelligence – AI) y del Aprendizaje de Máquinas (del inglés Machine Learning – ML) (Marcus and Maletic, 2003). La Minería de Datos (del inglés Data Mining - DM) es un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias en grandes volúmenes de datos utilizando técnicas de AI y ML. Estas técnicas permiten extraer patrones y tendencias para describir y comprender mejor los datos y predecir comportamientos futuros.

Para dar soporte a los grandes volúmenes de información se necesita un almacén de datos (del

inglés Data Warehouse - DW). Un DW es una colección de datos orientada a un dominio, integrada, no volátil y variante en el tiempo para ayudar a tomar decisiones (Salton, 1989). Los DW surgieron por dos razones: primero, la necesidad de proporcionar una fuente única de datos limpia y consistente para propósitos de apoyo para la toma de decisiones; segundo, la necesidad de hacerlo sin afectar a los sistemas operacionales (Molina López and García Herrero, 2006).

Existen algunos trabajos relacionados con la utilización de DW y DM para la determinación de rendimiento académico (La Red Martínez and Podestá Gómez, 2014; La Red Martínez et al., 2012; La Red Martínez et al., 2011). En ellos se modela al estudiante tomando como base ciertas características que le son propias.

En este trabajo se propone la utilización de técnicas de DM sobre información del desempeño de los alumnos de la cátedra Algoritmos y Estructura de Datos de la carrera Ingeniería en Sistemas de Información que se dicta en la Facultad Regional Resistencia de la Universidad Tecnológica Nacional.

El artículo se organiza de la siguiente manera: en la sección 2 se describen el estado del arte y conceptos introductorios relacionados a DW. En la sección 3 se describe el diseño del modelo propuesto de DW junto al proceso de DM utilizado para la clasificación y determinación de perfiles.

Luego en la sección 4 se describe el caso de estudio sobre el cual se validará a futuro el diseño del DW Y, finalmente, en la última sección se exponen conclusiones y trabajos futuros a realizarse.

DATA WAREHOUSE

Un DW es un repositorio de datos que proporciona una visión global, común e integrada de los datos y presenta las siguientes características (Curto Días, 2010):

- Orientado a un tema: organiza una colección de información alrededor de un tema central.
- Integrado: incluye datos de múltiples orígenes y presenta consistencia de datos.
- Variable con el tiempo: se realizan fotos de los datos basadas en fechas o hechos.
- No volátil: sólo de lectura para los usuarios finales.

El término DW puede utilizarse indistintamente para hacer referencia a la arquitectura en sí misma o a uno de los componentes que la integran.

Los bloques funcionales que se corresponden con un sistema de información completo que utiliza un DW se resumen en la Figura 1 (Kimball and Ross, 2013):

1. Nivel Operacional: hace referencia a los sistemas operacionales/transaccionales de la organización y a fuentes que forman parte del proceso de Data Warehousing.
2. Nivel de acceso a la información: es la capa de interacción del usuario cuya finalidad es la conversión de los datos almacenados en información fácil y transparente para las herramientas de los usuarios finales.
3. Nivel de acceso a los datos: comunica el nivel de acceso a la información con el nivel operacional de forma universal.
4. Nivel de directorio de datos (metadatos): repositorio de metadatos de los datos almacenados que proporcionan información sobre el origen y la transformación de los mismos en el proceso de Data Warehousing.
5. Nivel de gestión de procesos: planificación de las tareas y procesos para la construcción y mantenimiento actualizado del Data Warehouse.
6. Nivel de mensaje de la aplicación: determina el transporte de información a lo largo del entorno de computación de la organización a modo de middleware pero más allá de meramente protocolos de red.
7. Nivel Data Warehouse (físico): es el repositorio central altamente flexible de información

donde residen copias de los datos operacionales y/o externos optimizados para su acceso para la consulta.

8. Nivel de organización de datos: incluye todos los procesos necesarios para seleccionar, editar, resumir (normalmente sumarizar), combinar y cargar en el Data Warehouse y en la capa de acceso a la información los datos operacionales y/o externos.

Detrás de la arquitectura de componentes del DW existe un conjunto de procesos básicos asociados: los ETL (del inglés Extract, Transform, Load – Extracción, Transformación y Carga). Los procesos ETL hacen referencia a la recuperación y transformación de los datos desde las fuentes orígenes cargándolos en el DW.

En primer lugar los datos se analizan desde las fuentes y se extraen aquellos que serán de utilidad para el proceso en ejecución. El proceso de extracción debe ser sumamente cuidadoso para evitar perturbar el normal funcionamiento de los sistemas orígenes desde dónde se está recuperando información (Rainardi, 2008).

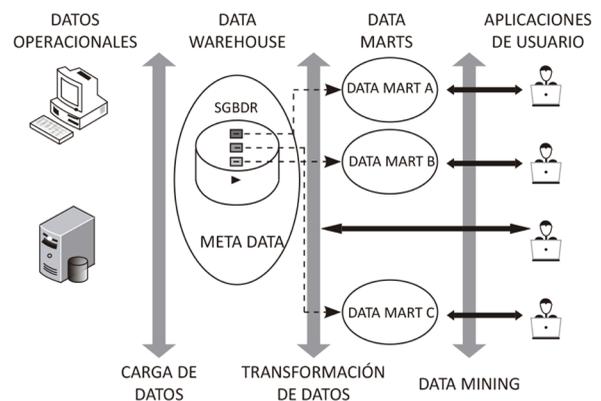


Figura 1 - Arquitectura Básica de un DW.

Luego de extraer los datos se los carga al DW pero, en muchas ocasiones, éstos requieren pasar por un proceso de transformación. La transformación de los datos significa un formateo y/o estandarización de los mismos convirtiendo ciertos números en fechas, eliminando campos nulos, etc. Es necesario que antes de completar el DW con los

datos se realicen varios controles para enviar información cualitativamente correcta.

Finalmente se procede a aplicar alguna técnica para realizar el análisis de los datos almacenados en el DW. El método más utilizado es el proceso de DM que aplica la inteligencia artificial para encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos, es decir, representaciones abstractas de la realidad, pero es el descubrimiento del conocimiento (KDD, por sus siglas en inglés) quien se encarga de la preparación de los datos y la interpretación de los resultados obtenidos, los cuales dan un significado a estos patrones encontrados (Valcárcel Ascencios, 2004).

Existen varias alternativas del DM, por ejemplo la Minería de Datos en Educación (Educational Data Mining, EDM).

El objetivo de la EDM es el desarrollo de métodos para la exploración de tipos de datos únicos provenientes de plataformas educativas y usándolos para entender mejor a los estudiantes en el aprendizaje (Baker and Yaceff, 2009).

Existen diversos estudios y publicaciones que abordan la evaluación de rendimiento académico utilizando técnicas de Minería de Datos (Formia et al., 2013; Timarán Pereira, 2010).

MODELO PROPUESTO

La estructura del DW, como se muestra en la Figura 2, consta de una tabla de hechos y varias tablas de dimensión.

Una tabla de hechos o una entidad de hecho es una tabla o entidad que almacena medidas para medir el negocio como las ventas, el coste de las mercancías o las ganancias (IBM Knowledge Center, 2015). Cada medida se corresponde con una intersección de valores de las dimensiones y generalmente se trata de cantidades numéricas, continuamente evaluadas y aditivas. Se pueden distinguir dos tipos

de columnas en una tabla de hechos, columnas de hechos y columnas llaves. Las columnas de hechos almacenan las medidas del negocio que se quieren controlar y las columnas llaves forman parte de la clave de la tabla.

Una tabla de dimensiones o entidad de dimensiones es una tabla o entidad que almacena detalles acerca de hechos. Por ejemplo una tabla de dimensión de hora almacena los distintos aspectos del tiempo como el año, trimestre, mes y día. Además incluye información descriptiva sobre los valores numéricos de una tabla de hechos. Por ejemplo, las tablas de dimensiones para una aplicación de análisis de mercado pueden incluir el tipo de período de tiempo, región comercial y producto.

Asimismo las tablas de dimensiones describen los distintos aspectos de un proceso de negocio. Por ejemplo, si se desea determinar los objetivos de ventas, se pueden almacenar los atributos de dichos objetivos en una tabla de dimensiones. Cada tabla de dimensiones contiene una clave simple y un conjunto de atributos que describen la dimensión (La Red Martínez et al., 2010).

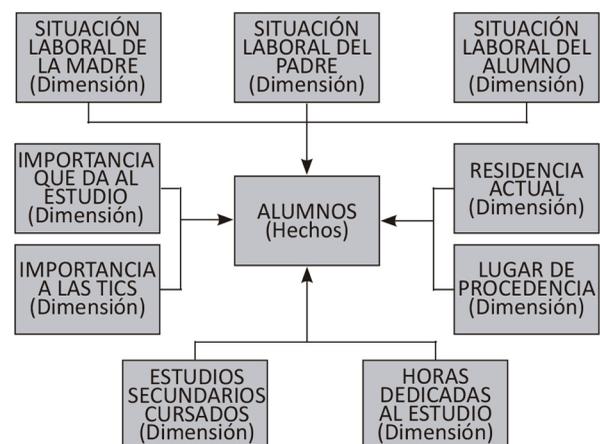


Figura 2 - Modelo Propuesto del DW.

Las columnas de una tabla de dimensiones se utilizan para crear informes o para mostrar resultados de consultas. Por ejemplo las descripciones

textuales de un informe se crean desde las etiquetas de las columnas de una tabla de dimensiones.

El modelo que se presenta en este trabajo se compone de la tabla de hechos “ALUMNOS” y varias tablas de dimensiones asociadas a la misma que incluyen características que se desean estudiar. En la Figura 2 se representa gráficamente esta estructura.

Para llevar a cabo este trabajo se ha utilizado la suite de herramientas incluidas en el IBM Data Warehouse Edition (DWE) V.9.5, un paquete de productos que combina la potencia de DB2 Universal Database (DB2 UDB) con la robusta infraestructura de inteligencia empresarial de IBM (IBM, 2005). Se compone de varios productos pero a los fines de este trabajo se han usado los siguientes componentes:

- El Sistema de Gestión de Base de datos DB2 Enterprise Server Edition (DB2 ESE) que permite conexiones multiusuarios y presenta alta escalabilidad.

- Design Studio (DS): Una plataforma a modo de herramienta que usan los analistas de negocios y los administradores de almacenes de datos para diseñar las reglas de las cargas de trabajo, los flujos de transformación de datos y los flujos analíticos para la minería de datos y el análisis de textos.

- Intelligent Miner (IM): Conjunto de funcionalidades que permiten realizar análisis de información de acuerdo a conceptos de Business Intelligence (BI).

A continuación se describen los campos de cada una de las tablas que componen el DW con el que se trabajó.

DESCRIPCIÓN DE LAS TABLAS

En la Tabla 1 se observa la estructura de la tabla de hechos “Alumnos” junto al detalle del significado de cada campo.

En la Tabla 2 se enumeran las variables de la tabla “Importancia que da al estudio” junto a la descripción correspondiente de cada una.

Nombre de Variable	Significado
LU_ALUM	Número de Legajo
DNI_ALUM	Número de DNI
CARRERA_ALUM	Carrera
SEXO_ALUM	Sexo
EDAD	Edad
ESTCIVIL_ALUM	Estado Civil
FECHA_NACIM	Fecha de Nacimiento
PAIS_NACIM	País de Nacimiento
PROV_NACIM	Provincia de Nacimiento
CIUDAD_NACIM	Ciudad de Nacimiento
FECHA_ENCUESTA	Fecha de la Encuesta
GRUPO_SANGUINEO	Grupo Sanguíneo
NOTA_PP_ALUM	Nota de Primer Parcial
NOTA_PR_ALUM	Nota de Primer Recuperatorio
NOTA_SP_ALUM	Nota de Segundo Parcial
NOTA_SR_ALUM	Nota de Segundo Recuperatorio
NOTA_EXT_ALUM	Nota de Extraordinario
SF	Situación Final del Alumno
ANIO	Año Lectivo

Tabla 1 - Variables de la Tabla “Alumnos”.

Nombre de Variable	Significado
LU_IES	Número de Legajo
DNI_IES	Número de DNI
IMPORTANCIA_IES	Importancia dada al estudio

Tabla 2 - Variables de la Tabla de “Importancia que da al Estudio”.

A continuación se detalla y explica en la Tabla 3 el significado de la tabla “Lugar de Procedencia”.

Nombre de Variable	Significado
LU_P	Número de Legajo
DNI_P	Número de DNI
PROVINCIA_P	Provincia de procedencia
CIUDAD_P	Ciudad de procedencia

Tabla 3 - Variables de la Tabla de “Lugar de Procedencia”.

En la Tabla 4 se pueden observar las variables que conforman la tabla “Importancia otorgada a las TICS” junto a su correspondiente descripción.

Nombre de Variable	Significado
LU_TIC	Número de Legajo
DNI_TIC	Número de DNI
UTIL_TIC	Utilización de TICS

Tabla 4 - Variables de la Tabla de "Importancia Otorgada a las TICS".

En las Tablas 5 y 6 se enumera y se explica el significado correspondiente a las tablas "Estudios Secundarios cursados" y "Residencia Actual".

Nombre de Variable	Significado
LU_ES	Número de Legajo
DNI_ES	Número de DNI
COLEGIO_ES	Nombre de Escuela
DEPENDENCIA_ES	Dependencia de Escuela
PROVINCIA_ES	Provincia a la que pertenece la Escuela
CIUDAD_ES	Ciudad a la que pertenece la Escuela
TITULO_ES	Título otorgado
FEGRESO_ES	Fecha de Egreso

Tabla 5 - Variables de la Tabla de "Estudios Secundarios Cursados".

Nombre de Variable	Significado
LU_RA	Número de Legajo
DNI_RA	Número de DNI
TIPO_RA	Tipo de Residencia
CALLE_RA	Dirección actual
PROVINCIA_RA	Provincia donde reside
CIUDAD_RA	Ciudad donde reside

Tabla 6 - Variables de la Tabla de "Residencia Actual".

En la Tabla 7 se observa la descripción de la tabla de dimensión "Horas dedicadas al estudio".

A continuación se describen, en las Tablas 8 y 9, las dimensiones "Situación Laboral de la Madre" y "Situación Laboral del Padre".

Nombre de Variable	Significado
LU_ES	Número de Legajo
DNI_ES	Número de DNI
HS_ES	Horas dedicadas al estudio

Tabla 7 - Variables de la Tabla de "Horas Dedicadas al Estudio".

Nombre de Variable	Significado
LU_SLM	Número de Legajo
DNI_SLM	Número de DNI
ULT_EST_SLM	Grado de escolaridad de la madre
SIT_LAB_SLM	Situación laboral de la madre
HS_TRAB_SLM	Horas semanales trabajadas
REL_TRAB_SLM	Rama de la actividad laboral de la madre
CATOCUP_SLM	Categoría ocupacional de la madre

Tabla 8 - Variables de la Tabla de "Situación Laboral de la Madre".

Nombre de Variable	Significado
LU_SLP	Número de Legajo
DNI_SLP	Número de DNI
ULT_EST_SLP	Grado de escolaridad del padre
SIT_LAB_SLP	Situación laboral del padre
HS_TRAB_SLP	Horas semanales trabajadas
REL_TRAB_SLP	Rama de la actividad laboral del padre
CATOCUP_SLP	Categoría ocupacional del padre

Tabla 9 - Variables de la Tabla de "Situación Laboral del Padre".

En la Tabla 10 se enumeran y describen las variables que conforman la tabla "Situación Laboral del Alumno".

Nombre de Variable	Significado
LU_SL	Número de Legajo
DNI_SL	Número de DNI
TIPO_SL	Categoría ocupacional
ACTIV_ECONOM_SL	Rama actividad económica
HS_TRAB_SL	Horas semanales trabajadas
REL_TRAB_SL	Relación del trabajo con la carrera elegida
OBRA_SOCIAL_SL	Tiene Obra Social
CATOCUP_SL	Categoría ocupacional
SIT_LAB_SL	Situación Laboral

Tabla 10 - Variables de la Tabla de "Situación Laboral del Alumno".

CASO DE ESTUDIO

El estudio se realizó sobre alumnos regulares que estaban cursando el 2°, 3°, 4° o 5° año de la carrera de Ingeniería en Sistemas de Información de la

Facultad Regional Resistencia, Universidad Tecnológica Nacional (FRRe-UTN) y particularmente se trabajó sobre el rendimiento que estos estudiantes tuvieron en la materia Algoritmos y Estructuras de Datos.

ETAPA DE RECOLECCIÓN DE DATOS

Tal como se planteó en la sección 1, el estudio del desempeño académico de los estudiantes no sólo debe evaluarse teniendo en cuenta los resultados de las instancias de evaluaciones previstas por asignatura sino que también deben analizarse otros factores culturales, sociales y/o económicos que afecten el rendimiento del alumno.

Por ello para este trabajo resultó determinante la participación directa del estudiante, pues era necesario conocer datos sobre aspectos personales que no se podían obtener de otra manera que no fuera a través de respuestas directas por parte de cada alumno.

A tal fin se dispuso la elaboración de una aplicación web que permitió contar con una Encuesta On-Line compuesta por preguntas relacionadas a situación familiar, historial de estudios secundarios, entre otras cuestiones. En la Figura 3 se muestra un ejemplo de pantalla correspondiente al Sistema de Encuesta Web que contiene un formulario que fue utilizado por los estudiantes para responder sobre diversas cuestiones.

A continuación en las Tablas 11, 12, 13, 14, 15, 16, 17 y 18 se enumeran y describen los datos solicitados en la Encuesta web

5. Nacionalidad:

6. Sexo:

7. Tipo de Documento:

8. Número:

9. Estado Civil:

10. Datos Sanguíneos:

11. Email:

12. Orientación Vocacional:

DATOS DE NACIMIENTO

13. Fecha de Nacimiento:

14. País:

Figura 3 - Formulario inicial del Sistema de Encuesta Web.

Campo	Significado	Opciones
Facultad	UTN-FRRE (por defecto)	
Carrera	Ing. en Sist. de Inf. (por defecto)	
Año Académico	Año en que se evalúa el desempeño del alumno	Ingresar
Apellidos	Del Alumno	Ingresar
Nombres	Del Alumno	Ingresar
Nacionalidad	Del Alumno	Argentino Extranjero Naturalizado Por opción
Sexo	Género del Alumno	Femenino Masculino
Tipo de Documento	Del Alumno	DNI LC LE PASAPORTE
Número	Número de DNI del Alumno	Ingresar

Tabla 11 - Datos de la Encuesta Web – 1ra. Parte.

Campo	Significado	Opciones
Estado Civil	Del Alumno	Soltero Casado Divorciado Viudo Unión de hecho
Grupo Sanguíneo	Del Alumno	Ingresar
Email	Del Alumno	Ingresar
Orientación Vocacional	¿El alumno obtuvo orientación vocacional antes del ingreso a la carrera?	Ninguna Entidad oficial Entidad privada Profesional
DATOS DE NACIMIENTO		
Fecha	Del Alumno	Ingresar
País	Del Alumno	Seleccionar en lista de países
Provincia	Del Alumno	Seleccionar en lista de provincias
Departamento	Del Alumno	Ingresar
Localidad	Del Alumno	Ingresar
Código Postal	De la Localidad	Ingresar

Tabla 12 - Datos de la Encuesta Web – 2da. Parte.

DATOS DEL SECUNDARIO		
Nombre	Nombre del Colegio de donde egresó el alumno	Ingresar
Título Obtenido	Del Alumno	Ingresar
Dependencia	Del Establecimiento	Seleccionar entre la lista
Calle	Del Establecimiento	Ingresar
País	Del Establecimiento	Seleccionar entre la lista
Provincia	Del Establecimiento	Seleccionar entre la lista
Departamento	Del Establecimiento	Ingresar
Localidad	Del Establecimiento	Ingresar
Código Postal	Del Establecimiento	Ingresar
Año de Egreso	Del Alumno	Ingresar

Tabla 13 - Datos de la Encuesta Web – 2da. Parte.

DATOS FAMILIARES		
Apellido/s del Padre	Del Alumno	Ingresar
Nombre/s del Padre	Del Alumno	Ingresar
Apellido/s de la Madre	Del Alumno	Ingresar
Nombre/s de la Madre	Del Alumno	Ingresar
DOMICILIO PERIODO LECTIVO ACTUAL		
Tipo de Residencia	Del Alumno	Con Familiares Independiente Residencia Universitaria Otra Situación
Calle	Donde vive actualmente el Alumno	Ingresar
País	Donde vive actualmente el Alumno	Seleccionar entre la lista
Provincia	Donde vive actualmente el Alumno	Seleccionar entre la lista
Departamento	Donde vive actualmente el Alumno	Ingresar
Ciudad	Donde vive actualmente el Alumno	Ingresar
Código Postal	Donde vive actualmente el Alumno	Ingresar

Tabla 14 - Datos de la Encuesta Web – 3ra. Parte.

DOMICILIO DE PROCEDENCIA FAMILIAR		
Calle	De residencia familiar del Alumno	Ingresar

País	De residencia familiar del Alumno	Seleccionar entre la lista
Provincia	De residencia familiar del Alumno	Seleccionar entre la lista
Departamento	De residencia familiar del Alumno	Ingresar
Localidad	De residencia familiar del Alumno	Ingresar
Código Postal	De residencia familiar del Alumno	Ingresar
Teléfono	De residencia familiar del Alumno	Ingresar
SITUACIÓN DEL ALUMNO		
Vive alguno de los padres		Si
		No
Cantidad de Familiares a cargo		Ingresar
Situación Laboral		Ocupado
		No trabaja
		Sub-Ocupado
		Desocupado
		Jubilado
Categoría Ocupacional		Seleccionar entre la lista
Rama de Actividad Económica		Seleccionar entre la lista
Horas semanales trabajadas		No trabaja
		Hasta 20
		De 21 a 35
		De 36 o más
Relación del trabajo con la carrera elegida		Total
		Parcial
		No se relaciona
Tiene Obra Social		Ninguna
		De los padres
		Del cónyuge
		Propia
		Otros

Tabla 15 - Datos de la Encuesta Web – 4ta. Parte.

SITUACIÓN DEL PADRE		
Ultimo Nivel de estudios cursados	Del Padre	Seleccionar entre la lista
Situación Laboral	Del Padre	Ocupado
		No trabaja
		Sub-Ocupado
		Desocupado
		Jubilado

Situación Laboral	Del Padre	Ocupado
		No trabaja
		Sub-Ocupado
		Desocupado
		Jubilado
Categoría Ocupacional	Del Padre	Seleccionar entre la lista
Rama de Actividad Económica		Seleccionar entre la lista
Horas semanales trabajadas		No trabaja
		Hasta 20
		De 21 a 35
		De 36 o más

Tabla 16 - Datos de la Encuesta Web – 5ta. Parte.

SITUACIÓN DE LA MADRE		
Ultimo Nivel de estudios cursados	De la Madre	Seleccionar entre la lista
Situación Laboral	De la Madre	Ocupado
		No trabaja
		Sub-Ocupado
		Desocupado
		Jubilado
Categoría Ocupacional	De la Madre	Seleccionar entre la lista
Rama de Actividad Económica		Seleccionar entre la lista
Horas semanales trabajadas		No trabaja
		Hasta 20
		De 21 a 35
		De 36 o más

Tabla 17 - Datos de la Encuesta Web – 6ta. Parte,

ACTITUD HACIA EL ESTUDIO		
Prioridad otorgada al estudio		Más que al trabajo
		Más que a la Familia
		Más que a la Diversión
Número de horas semanales dedicadas al estudio	Sin considerar horas de clases	Hasta 10 inclusive
		Más de 10 y hasta 20 inclusive
		Más de 20
Estudia para		Aprobar las asignaturas
		Aprender integralmente

Utiliza TICS en el estudio porque		y aprobar
		Aprender acerca de los contenidos
		Otras motivaciones
		Están de moda
		Facilitan la enseñanza
		Son una realidad de la actualidad
		Para la vida profesional

Tabla 18 - Datos de la Encuesta Web – 7ma. Parte.

ETAPA DE DEPURACIÓN Y PREPARACIÓN DE DATOS

Para la realización de una correcta explotación del DW se debe asegurar que los datos obtenidos en la etapa anterior sean consistentes y mantengan la coherencia entre ellos. De esta forma, en esta etapa se procedió a un proceso de limpieza en los datos que significó la eliminación de aquellos registros con todos sus campos en blanco, corrección de errores tipográficos, llenado de algunos campos nulos, entre otros. Además la Encuesta no permite la carga, por parte de los estudiantes, de calificaciones de la asignatura en estudio. Esto se dispuso así para evitar errores en los datos ya sea por olvido, o confusión al momento de ingresar los valores. Por ello la carga de notas correspondientes al 1° Parcial, 2° Parcial, 3° Parcial, Recuperatorios, Examen Final y Situación del alumno (Regular, Promocionado, Libre), fue realizada por el equipo responsable de este trabajo de investigación. La información se obtuvo a partir de la base de datos histórica de la cátedra respecto a calificaciones de los alumnos.

Con esta información depurada se deberá proceder a trabajar en las próximas etapas:

- Carga de Datos al DW: Mediante la ejecución del flujo de datos, la información almacenada en la tabla ENCUESTA se distribuirá a las tablas pertenecientes al modelo del DW.

- Obtención de Resultados finales: Mediante el diseño y ejecución de flujos de minería a partir de la información disponible respecto a diversos aspectos (Lugar de Procedencia, Horas dedicadas al estudio, Situación Laboral de los padres, etc.).

DISCUSIONES Y TRABAJOS FUTUROS

En este artículo se propone un modelo para la determinación de perfiles de rendimiento académico de alumnos utilizando técnicas de DM. Hasta el momento se ha completado la primera etapa que implicó el diseño del modelo del DW sobre el cual se implementarán técnicas de DM a fin de determinar perfiles de estudiantes vinculados a su desempeño académico en la asignatura Algoritmos y Estructuras de Datos de la carrera Ingeniería en Sistemas de Información de la UTN-FRRe.

En el avance que aquí se presenta respecto del Proyecto se pudo comprobar que la etapa de depuración y preparación de los datos ha demandado mucho tiempo y esfuerzo debido principalmente a la poca integridad y coherencia que existía en la información que se utilizará para realizar la evaluación final.

Como trabajos futuros se prevé finalizar con el proceso de minería de datos a fin de evaluar y comparar patrones que se obtengan para definir los perfiles de estudiantes. La evaluación, análisis y utilidad de estos patrones permitirá soportar la toma de decisiones eficaces por parte del cuerpo docente de la asignatura analizada quienes, en el marco de los programas institucionales que adhieren al Plan Estratégico para la Formación de Ingenieros, deben enfocarse en formular políticas y estrategias relacionadas con los programas de retención estudiantil, que actualmente se encuentran establecidos, mejorando las estadísticas de regularización y aprobación aportando al ciclo de mejora continua respecto a desempeño académico en la UTN –FRRe.

AGRADECIMIENTOS

Este trabajo ha sido desarrollado en el marco del proyecto “Determinación de Perfiles de Estudiantes y de Rendimiento Académico Mediante la Utilización de Minería de Datos”. UTN, FRRe. Código EIINRE571.

REFERENCIAS

Briand, Daly and Wüst, “A unified framework for coupling measurement in objectoriented systems”, IEEE Transactions on Software Engineering, 25, 1, pp. 91-121, (1999).

Maletic, Collard and Marcus, “Source Code Files as Structured Documents”, in Proceedings 10th IEEE International Workshop on Program Comprehension (IWPC’02), Paris, France, June 27-29, pp. 289-292, (2002).

Marcus, Semantic Driven Program Analysis, Kent State University, Kent, OH, USA, Doctoral Thesis, (2003).

Marcus and Maletic, “Recovering Documentation-to-Source-Code Traceability Links using Latent Semantic Indexing”, in Proceedings 25th IEEE/ACM International Conference on Software Engineering (ICSE’03), Portland, OR, May 3-10, pp. 125-137, (2003).

Salton, Automatic Text Processing: The Transformation, Analysis and Retrieval of Information by Computer, Addison-Wesley, (1989).

Molina López, García Herrero, “Técnicas de Análisis de Datos”. Universidad Carlos III. Madrid, España. (2006).

La Red Martínez, Podestá Gómez, “Contributions from Data Mining to Study Academic Performance of Students of a Tertiary Institute American”. Journal of Educational Research, Vol. 2, No. 9, 713-726, (2014).

La Red Martínez, Acosta, Uribe, Rambo, “Academic Performance: An Approach from Data Mining”. Journal of Systemics, Cybernetics and Informatics, Vol. 10, No. 1, 66-72, (2012).

La Red Martínez, Acosta, Agostini, Uribe, Rambo, "La importancia otorgada al estudio y su relación con el rendimiento académico". *Revista Documentación*, No 24, 54-63, (2011).

Curto Días, "Introducción al Business Intelligence". Editorial UOC. Barcelona, España. ISBN: 978-84-9788-886-8, (2010).

Kimball and Ross, "The Data Warehouse Toolkit, 3rd Edition". John Wiley & Sons. USA, (2013).

Rainardi, "Building a Data Warehouse: With Examples in SQL Server". Apress. USA, (2008).

Valcárcel Asencios, "Data Mining y el Descubrimiento del Conocimiento". *Revista de la Facultad de Ingeniería Industrial*. Vol. (7) 2: pp. 83-86. Universidad Nacional Mayor de San Marcos, UNMSM, Perú, (2004).

Baker, Yaceff, "The State of Educational Data Mining in 2009: A Review and Future Visions". *Journal of Educational Data Mining*, Artículo 1, Vol 1, No 1, (2009).

Formia, Lanzarini, Hasperué, "Characterization of University Drop-Out at UNRN Using Data Mining. A study case", CACIC, Mar del Plata, Buenos Aires, Argentina, (2013).

Timarán Pereira, "Una lectura sobre deserción universitaria en estudiantes de pregrado desde la perspectiva de la minería de datos". *Revista Científica Guillermo de Ockham*. Vol. 8, N° 1, pp. 121-130, (2010).

IBM Knowledge Center: "InfoSphere Data Architect 7.5.3". USA, (2015).

La Red Martínez, Acosta, Cutro, Uribe, Rambo, "Data Warehouse y Data Mining Aplicados al Estudio del Rendimiento Académico". CISCI 2010 (9na. Conferencia Iberoamericana en Sistemas, Cibernética e Informática); Memorias, Volumen I, págs. 289-294; Orlando, Florida, USA, (2010).

IBM Press: "IBM DB2 Universal Database Data Warehouse Edition V. 8.2.1". USA, (2005).