

UNIVERSIDAD NACIONAL DE ASUNCIÓN
FACULTAD POLITÉCNICA
LICENCIATURA EN CIENCIAS INFORMÁTICAS
ÉNFASIS EN PROGRAMACIÓN DE COMPUTADORAS
PLAN 2009
PROGRAMA DE ESTUDIOS

Resolución N° 17/04/50-00 Acta N° 992/20/02/2017 - ANEXO 10

IDENTIFICACIÓN

- | | |
|-------------------------------------|----------------------------|
| 1. Asignatura | : Electiva I - Data Mining |
| 2. Código | : 8.2.A |
| 3. Horas semanales | : 5 horas |
| 4. Total real de horas disponibles: | 80 horas |

I. - JUSTIFICACIÓN

No toda la información está a la vista. Si bien muchas veces se puede llegar a conclusiones analizando datos a través del armado de reportes, pueden existir patrones ocultos, correlaciones o tendencias que pasan desapercibidos. La minería de datos o data mining es una herramienta que ayuda a generar conocimiento para la toma de decisiones, descubriendo patrones ocultos en los datos y realizando predicciones.

En esta asignatura el estudiante examinará los principios, metodologías y técnicas de KDD y Data Mining, así como su uso y aplicación en Inteligencia de Negocios. Se analizará una metodología para el diseño e implementación de un proyecto de Data Mining junto con las principales técnicas estadísticas y de Inteligencia Artificial empleada en tareas de pronóstico, clasificación, diagnóstico y optimización.

Se evaluarán las principales herramientas computacionales disponibles en el mercado para el desarrollo de aplicaciones correspondiente al ámbito de la asignatura. El estudiante aplicará los principios, metodología de desarrollo y técnicas de Minería de Datos a través de proyectos que serán desarrollados durante el semestre.

II. - OBJETIVOS GENERALES

Al término de la asignatura el estudiante podrá:

- Reconocer el potencial y la problemática del análisis de sistemas de información para la toma de decisiones.
- Describir el modelo multidimensional de datos de los almacenes de datos y los operadores de refinamiento asociados.
- Distinguir las fases del descubrimiento de conocimiento de Bases de Datos y la importancia de las mismas en el éxito del proceso.
- Reconocer los beneficios del uso sistemático de técnicas de extracción de conocimiento para la obtención de modelos y patrones predictivos o descriptivos.

III. - OBJETIVOS ESPECÍFICOS

Conocimientos

1. Describir los conceptos básicos y las etapas del proceso de extracción de conocimiento.
2. Identificar técnicas para exploración y preparación de datos
3. Identificar diferentes técnicas de minería de datos.
4. Comparar técnicas y definir las más apropiadas de minería de datos, para resolver problemas concretos.
5. Registrar ejemplos de problemas resueltos con Data Mining que sirvan como futuros disparadores.

Habilidades

1. Reconocer el potencial y la problemática del análisis de sistemas de información para la toma de decisiones.
2. Reconocer los beneficios del uso sistemático de técnicas de extracción de conocimiento para la obtención de modelos y patrones predictivos o descriptivos.
3. Generar modelos con conjuntos de datos sencillos utilizando herramientas de software libre.
4. Evaluar la calidad de los modelos.
5. Resolver casos prácticos que permitan aplicar gran parte de las técnicas analizadas.

Competencias

1. Desarrollar las competencias necesarias para brindar soluciones de Minería de Datos (Data Mining) que permitan apoyar la toma de decisiones en sus organizaciones.
2. Capacidad para la formulación y gestión de proyectos de Data Mining
3. Capacidad de aplicar los conocimientos en la práctica

IV. - PRE-REQUISITO

1. Programación de Aplicaciones en Redes

V. - CONTENIDO

6.1. Unidades programáticas

1. Introducción a Almacenes de Datos y Minería de Datos
2. Planeamiento de proyectos de Almacenes de Datos
3. Modelado Dimensional
4. Arquitectura de los Almacenes de Datos
5. Opciones de Configuración Física
6. Selección de Productos
7. Diseño Físico de la Base de Datos
8. Procesos y Técnicas de Preparación de Datos
9. Despliegue y perspectiva de Crecimiento
10. Entrada de Datos para Minería de Datos
11. Salida de Datos: Representación del Conocimiento
12. Algoritmos de Minería de Datos: Métodos básicos



6.2. Desarrollo de las unidades programáticas

1. Introducción a Almacenes de Datos y Minería de Datos
 - 1.1. Definiciones de almacenes de datos. Motivación y características. Arquitectura.
 - 1.2. Sistemas ROLAP y MOLAP. Carga y Mantenimiento de un Almacén de Datos. Diseño de un almacén de datos.
 - 1.3. Introducción a la Minería de Datos. Motivación. Problemas tipo y aplicaciones.
 - 1.4. Técnicas de Minería de Datos.
 - 1.5. Líneas de investigación abiertas.
2. Planeamiento de proyectos de Almacenes de Datos
 - 2.1. Marco dimensional del ciclo de vida del negocio.
 - 2.2. Preparación y evaluación de riesgos.
 - 2.3. Priorización de alcance y justificaciones.
 - 2.4. Roles y responsabilidades del equipo de desarrollo.
 - 2.5. Desarrollo y mantenimiento del plan de proyecto.
 - 2.6. Definición de los requisitos del negocio.
 - 2.7. Técnicas para la recolección de requisitos
3. Modelado Dimensional
 - 3.1. El rol del modelado dimensional.
 - 3.2. Características de la tabla del hecho y de la dimensional.
 - 3.3. Cuatro pasos para diseñar modelos dimensionales.
 - 3.4. Desnormalización de jerarquías dimensionales. Llaves sustitutas.
 - 3.5. Tablas del hecho sin hechos.
 - 3.5.1. Variaciones del copo de nieve.
 - 3.5.2.1 Dimensiones degeneradas.
 - 3.6. Implicancias de cadena de valores.
 - 3.7. Arquitectura de bus y la matriz de integración de modelos dimensionales.
 - 3.8. Dimensiones confortantes.
 - 3.9. El rol de las dimensiones.
 - 3.10. Hechos semi aditivos y no aditivos.
 - 3.11. Productos heterogéneos.
 - 3.12. Consideraciones para el diseño de numerosas dimensiones.
4. Arquitectura de los Almacenes de Datos
 - 4.1. Definiciones y descripciones.
 - 4.1.1. Arquitectura Del AD.
 - 4.1.2. Modelos, niveles de detalle.
 - 4.1.3. Componentes y servicios comunes.
 - 4.1.4. Sitio Trasero.
 - 4.1.5. Consideraciones sobre el área de presentación.
 - 4.1.6. Sitio Delantero.
 - 4.1.7. Tipos de usuarios del almacén de los datos.
 - 4.1.8. Reportes estándar.
 - 4.1.9. Minera de Datos.
 - 4.2. Depósito de los meta datos.
 - 4.3. Contenido. Estándares – XML.
 - 4.4. El proceso de desarrollo de la arquitectura en ocho pasos
5. Opciones de Configuración Física
 - 5.1. Infraestructura.
 - 5.2. Almacén de los datos vs. Data Marts.
 - 5.3. Data Marts aislados.
 - 5.4. Almacén de Datos empresariales.
 - 5.5. Almacén de Datos confortantes
 - 5.6. Tipos de almacenes de datos operacionales (ODS).
 - 5.7. Arquitecturas de presentación - OLAP, ROLAP, MOLAP, HOLAP
6. Selección de Productos
 - 6.1. Proceso de selección basado en la arquitectura
 - 6.2. Matriz de evaluación de productos
7. Diseño Físico de la Base de Datos
 - 7.1. Estándares y convenciones de nombramiento.
 - 7.2. Desarrollo del modelo físico.
 - 7.3. Navegación y estrategia de la agregación.
 - 7.4. Tipos y pautas de índices.
 - 7.5. Gerencia de datos
8. Procesos y Técnicas de Preparación de Datos
 - 8.1. Diseño de la aplicación para preparación de los datos.
 - 8.2. Preparación de las tablas de dimensionales.
 - 8.3. Preparación de las tablas hechos.
 - 8.4. Validación de procesos.
 - 8.5. Implementación de las operaciones del almacén de datos.
 - 8.6. Técnicas básicas de extracción y transformación de datos.
 - 8.7. Creación y mantenimiento de claves sustitutas.
 - 8.8. Aceleración del ciclo de carga de dato.
 - 8.9. Integración de datos desde múltiples fuentes.
 - 8.10. Calidad y limpieza de los datos
9. Despliegue y perspectiva de Crecimiento
 - 9.1. Rol de las aplicaciones para usuarios finales.
 - 9.2. Diseño de plantilla de aplicaciones y navegación.



- 9.3. Proceso de entrega de la aplicación.
- 9.4. Soporte y entrenamiento.
- 9.5. Documentación.
- 9.6. Planes para el crecimiento
10. Entrada de Datos para Minería de Datos
 - 10.1. Conceptos
 - 10.2. Ejemplos
 - 10.3. Instancias
 - 10.4. Atributos
 - 10.5. Preparación de los datos
11. Salida de Datos: Representación del Conocimiento
 - 11.1. Tablas de Decisión
 - 11.2. Árboles de decisión
 - 11.3. Reglas de Clasificación
 - 11.4. Reglas de Asociación
 - 11.5. Reglas con Excepciones
 - 11.6. Reglas que involucran relaciones
 - 11.7. Árboles con predicciones numéricas
 - 11.8. Representación basada en instancias
 - 11.9. Clusters
12. Algoritmos de Minería de Datos: Métodos básicos
 - 12.1. Inferencia de reglas rudimentarias
 - 12.2. Modelado estadístico
 - 12.3. Divide y vencerás. Construcción de Árboles de decisión
 - 12.4. Algoritmos de cobertura: Construyendo Reglas
 - 12.5. Minería de reglas de asociación
 - 12.6. Modelos lineales
 - 12.7. Aprendizaje basado en instancias
 - 12.8. Clustering

VI. - ESTRATEGIAS METODOLÓGICAS.

1. Exposición oral
2. Laboratorio
3. Formación de grupos y trabajo en equipo para resolver problemas durante las clases.
4. Trabajos prácticos.

VII. - MEDIOS AUXILIARES

1. Pizarra
2. Proyector Multimedia
3. Sala de máquinas para clases prácticas
4. Bibliografía de apoyo.

VIII. - EVALUACIÓN

La evaluación se realizará de acuerdo a las reglamentaciones de la Facultad y la Carrera.

IX. - BIBLIOGRAFÍA

Bibliografía Básica

- The data warehouse lifecycle toolkit. / Ralph Kimball... [et al.] -- 2a. ed. -- Washington : Wiley Publishing, 2008. -- 636 p.
- Witten, Frank. Data Mining: Practical Machine Learning Tools and Techniques, Third Edition (The Morgan Kaufmann Series in Data Management Systems), Hall, 2011

Bibliografía Complementaria

- Torgo, Luís. Data Mining with R: Learning with Case Studies (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series), 2011
- Hernández, J. Introducción a la Minería de Datos, J. Hernández; MJ Ramírez,; C. Ferri, Pearson Prentice Hall, 2004
- Ralph Kimball. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, 2004

BIBLIOGRAFÍA COMPLEMENTARIA EXISTENTE EN EL CENTRO DE INFORMACIÓN Y CULTURA

- Elmasri, Ramez. Almacenes de datos (data warehousing) y minería de datos / Ramez Elmasri, Shamant B. Navathe. p. 800-830. -- En su Fundamentos de sistemas de bases de datos -- 3a ed. -- Madrid: Addison Wesley, 2002.
- Ricardo, Catherine M. Almacenes de datos (data warehousing) y minado de datos (data mining) : / Catherine M. Ricardo. -- p.543-562. -- México: McGraw-Hill, 2009.
- Silberschatz, Abrahan. Análisis y minería de datos / Abrahan Silberschatz, Henry F. Korth, . Sudarshan. p. 601-628. En su Fundamentos de bases de datos. -- 5a. ed. -- Madrid: McGraw-Hill, 2006.

