UNIVERSIDAD NACIONAL DE ASUNCIÓN
FACULTAD POLITÉCNICA
INGENIERÍA EN INORMÁTICA
PLAN 2008
PROGRAMA DE ESTUDIOS

Resolución 25/07/09-00 Acta 1215/07/04/2025
ANEXO 05

I. - IDENTIFICACIÓN

1.	Asignatura	: Electiva - Data Mining
2.	Horas semanales	: 7 horas
2.1.	Clases teóricas	: 2 horas
2.2.	Clases prácticas	: 2 horas
2.3.	Clases de laboratorio	: 3 horas
3.	Total real de horas disponibles	: 112 horas
3.1.	Clases teóricas	: 32 horas
3.2.	Clases prácticas	: 32 horas
3.3.	Clases de laboratorio	: 48 horas

II. - JUSTIFICATIVO

La creciente informatización del mundo permite recoger grandes volúmenes de datos y surge la necesidad de explotarlos eficaz y eficientemente. El descubrimiento de conocimiento en bases de datos (del inglés Knowledge Discovery in Databases, KDD) es un proceso no trivial de identificación de patrones válidos, novedosos, parcialmente útiles y, en última instancia, comprensibles a partir de los datos. El KDD consta de varias etapas donde la Minería de Datos (Datamining) es la más importante y cuyo objetivo es encontrar modelos inteligibles a partir de estos datos.
La incorporación de Datamining como electiva en la carrera obedece al propósito de ayudar al futuro ingeniero en informática en la concreción de sus conocimientos técnicos, adquiriendo conocimientos sobre las diversas técnicas existentes en cada una de las fases del KDD y desarrollando habilidades que le permitan extraer patrones, describir tendencias y regularidades, así como también predecir comportamientos.

III. - OBJETIVOS

3.1. Objetivo general

El principal objetivo de esta asignatura es capacitar al estudiante para reconocer el potencial y la problemática del análisis de sistemas de información para la toma de decisiones, conocer el modelo multidimensional de datos de los almacenes de datos y los operadores de refinamiento asociados. Del mismo modo, se busca que el estudiante comprenda las fases del descubrimiento de conocimiento de Bases de Datos y la importancia de las mismas en el éxito del proceso. Por último, el estudiante deberá reconocer los beneficios del uso sistemático de técnicas de extracción de conocimiento para la obtención de modelos y patrones predictivos o descriptivos.

3.2. Objetivos específicos

3.2.1 Generar nuevos conocimientos mediante la obtención y recopilación de información valiosa utilizando las matemáticas y técnicas de minería de datos.

3.2.2 Aplicar soluciones de minería de datos a problemas de datos en una amplia gama de entornos, como logística, medicina, agricultura, pecuaria, robótica entre otros.

3.2.3 Formar profesionales con alto nivel competitivo en la búsqueda y aplicación de métodos y técnicas para la óptima toma de decisiones.

3.2.4 Contribuir con investigación e innovación en minería de datos.

3.2.5 Formar profesionales con habilidades para diseñar, preparar, analizar y manejar grandes volúmenes de información, estructurada y no-estructurada usando técnicas de minería de datos.

3.2.6 Formar líderes y emprendedores en tecnología.

IV. - PRE-REQUISITO

Para cursar esta asignatura, que se ofrece en la carrera como Electiva, el estudiante deberá cumplir con los prerrequisitos establecidos según la electiva en la que se inscribe, de acuerdo con la siguiente tabla:

Alternativa	Porcentaje de créditos aprobados	Cantidad de créditos requeridos
Electiva 1	55%	184
Electiva 2, Electiva 3, Electiva 4, Electiva 5	70%	235
Electiva 6, Electiva 7	80%	268

V. - CONTENIDO

5.1. Unidades programáticas

5.1.1 Introducción a Almacenes de Datos y Minería de Datos
5.1.2 Planeamiento de proyectos de Almacenes de Datos
5.1.3 Modelado Dimensional
5.1.4 Arquitectura de los Almacenes de Datos
5.1.5 Opciones de Configuración Física
5.1.6 Selección de Productos
5.1.7 Diseño Físico de la Base de Datos
5.1.8 Procesos y Técnicas de Preparación de Datos
5.1.9 Despliegue y perspectiva de Crecimiento
5.1.10 Entrada de Datos para Minería de Datos
5.1.11 Salida de Datos: Representación del Conocimiento
5.1.12 Algoritmos de Minería de Datos: Métodos básicos

5.2. Desarrollo de las unidades programáticas

5.2.1 Introducción a Almacenes de Datos y Minería de Datos
5.2.1.1 Definiciones de almacenes de datos. Motivación y características. Arquitectura.
5.2.1.2 Sistemas ROLAP y MOLAP. Carga y Mantenimiento de un Almacén de Datos. Diseño de un almacén de datos.
5.2.1.3 Introducción a la Minería de Datos. Motivación. Problemas tipo y aplicaciones.
5.2.1.4 Técnicas de Minería de Datos.
5.2.1.5 Líneas de investigación abiertas.

5.2.2 Planeamiento de proyectos de Almacenes de Datos
5.2.2.1 Marco dimensional del ciclo de vida del negocio.
5.2.2.2 Preparación y evaluación de riesgos.
5.2.2.3 Priorización de alcance y justificaciones.
5.2.2.4 Roles y responsabilidades del equipo de desarrollo.
5.2.2.5 Desarrollo y mantenimiento del plan de proyecto.
5.2.2.6 Definición de los requisitos del negocio.
5.2.2.7 Técnicas para la recolección de requisitos

5.2.3 Modelado Dimensional
5.2.3.1 El rol del modelado dimensional.
5.2.3.2 Características de la tabla del hecho y de la dimensional.
5.2.3.3 Cuatro pasos para diseñar modelos dimensionales.
5.2.3.4 Desnormalización de jerarquías dimensionales. Llaves sustitutas.
5.2.3.5 Tablas del hecho sin hechos.
5.2.3.5.1 Variaciones del copo de nieve.
5.2.3.5.2 Dimensiones degeneradas.
5.2.3.6 Implicancias de cadena de valores.
5.2.3.7 Arquitectura de bus y la matriz de integración de modelos dimensionales.
5.2.3.8 Dimensiones confortantes.
5.2.3.9 El rol de las dimensiones.
5.2.3.10 Hechos semi aditivos y no aditivos.
5.2.3.11 Productos heterogéneos.
5.2.3.12 Consideraciones para el diseño de numerosas dimensiones.

5.2.4 Arquitectura de los Almacenes de Datos
5.2.4.1 Definiciones y descripciones.
5.2.4.2 Arquitectura Del AD.
5.2.4.3 Modelos, niveles de detalle.
5.2.4.4 Componentes y servicios comunes.
5.2.4.5 Sitio Trasero.
5.2.4.6 Consideraciones sobre el área de presentación.
5.2.4.7 Sitio Delantero.
5.2.4.8 Tipos de usuarios del almacén de los datos.
5.2.4.9 Reportes estándar.
5.2.4.10 Minera de Datos.
5.2.4.11 Depósito de los meta datos.
5.2.4.12 Contenido. Estándares – XML.
5.2.4.13 El proceso de desarrollo de la arquitectura en ocho pasos

5.2.5 Opciones de Configuración Física
5.2.5.1 Infraestructura.
5.2.5.2 Almacén de los datos vs. Data Marts.
5.2.5.3 Data Marts aislados.
5.2.5.4 Almacén de Datos empresariales.
5.2.5.5 Almacén de Datos confortantes
5.2.5.6 Tipos de almacenes de datos operacionales(ODS).
5.2.5.7 Arquitecturas de presentación - OLAP, ROLAP, MOLAP, HOLAP

5.2.6 Selección de Productos
5.2.6.1 Proceso de selección basado en la arquitectura.
5.2.6.2 Matriz de evaluación de productos

5.2.7 Diseño Físico de la Base de Datos
5.2.7.1 Estándares y convenciones de nombramiento.
5.2.7.2 Desarrollo del modelo físico.
5.2.7.3 Navegación y estrategia de la agregación.
5.2.7.4 Tipos y pautas de índices.
5.2.7.5 Gerencia de datos

5.2.8 Procesos y Técnicas de Preparación de Datos
5.2.8.1 Diseño de la aplicación para preparación de los datos.

5.2.8.2 Preparación de las tablas de dimensionales.
5.2.8.3 Preparación de las tablas hechos.
5.2.8.4 Validación de procesos.
5.2.8.5 Implementación de las operaciones del almacén de datos.
5.2.8.6 Técnicas básicas de extracción y transformación de datos.
5.2.8.7 Creación y mantenimiento de claves sustitutas.
5.2.8.8 Aceleración del ciclo de carga de dato.
5.2.8.9 Integración de datos desde múltiples fuentes.
5.2.8.10 Calidad y limpieza de los datos

5.2.9 Despliegue y perspectiva de Crecimiento

5.2.9.1 Rol de las aplicaciones para usuarios finales.
5.2.9.2 Diseño de plantilla de aplicaciones y navegación.
5.2.9.3 Proceso de entrega de la aplicación.
5.2.9.4 Soporte y entrenamiento.
5.2.9.5 Documentación.
5.2.9.6 Planes para el crecimiento

5.2.10 Entrada de Datos para Minería de Datos

5.2.10.1 Conceptos
5.2.10.2 Ejemplos
5.2.10.3 Instancias
5.2.10.4 Atributos
5.2.10.5 Preparación de los datos

5.2.11 Salida de Datos: Representación del Conocimiento

5.2.11.1 Tablas de Decisión
5.2.11.2 Arboles de decisión
5.2.11.3 Reglas de Clasificación
5.2.11.4 Reglas de Asociación
5.2.11.5 Reglas con Excepciones
5.2.11.6 Reglas que involucran relaciones
5.2.11.7 Arboles con predicciones numéricas
5.2.11.8 Representación basada en instancias
5.2.11.9 Clusters

5.2.12 Algoritmos de Minería de Datos: Métodos básicos

5.2.12.1 Inferencia de reglas rudimentarias
5.2.12.2 Modelado estadístico
5.2.12.3 Divide y vencerás. Construcción de Arboles de decisión
5.2.12.4 Algoritmos de cobertura: Construyendo Reglas
5.2.12.5 Minería de reglas de asociación
5.2.12.6 Modelos lineales
5.2.12.7 Aprendizaje basado en instancias
5.2.12.8 Clustering

VI. - ESTRATEGIAS METODOLÓGICAS.

6.1 Exposición oral
6.2 Laboratorio
6.3 Formación de grupos y trabajo en equipo para resolver problemas durante las clases.
6.4 Trabajos prácticos.

VII. - MEDIOS AUXILIARES

7.1 Pizarra
7.2 Proyector Multimedia
7.3 Sala de máquinas para clases prácticas
7.4 Bibliografía de apoyo.

VIII. - EVALUACIÓN

La evaluación sobre el aprendizaje y conocimiento adquiridos por el alumno se realizará de acuerdo a lo establecido en el reglamento de la Facultad Politécnica de la UNA.

IX. - BIBLIOGRAFÍA

- Hernández, J. Ramírez, M.J. & Ferri, C. (2004). *Introducción a la Minería de Datos*. Pearson Prentice Hall.
- Kimball, R., Ross, M., Thornthwaite, W, Mundy, J. & Becker, B. (2008). *The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing, and Deploying Data Warehouses*. (2° Ed.). Washington: WILEY.
- Kimball, Ralph. (2004). *The Data Warehouse ETL Toolkit: Practical Techniques for Extracting*. WILEY.
- Torgo, L. (2011). *Data Mining with R: Learning with Case Studies*. Boca Raton: CRC Press. Recuperado de: http://www.bagualu.net
- Witten, I. H. & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. (2° Ed.). Amsterdam: Elsevier.

MATERIALES BIBLIOGRÁFICOS DISPONIBLES EN LA BIBLIOTECA DE LA FACULTAD POLITÉCNICA

- Beynon Davies, P. (2014). *Sistemas de bases de datos*. Barcelona: Reverté.
- Colle, R. (2002). *Explotar la información noticiosa: data mining aplicado a la documentación periodística*. Madrid: Universidad Complutense de Madrid.

- Silberschatz, A., Korth, H. F. & Sudarshan, S. (2006). Fundamentos de bases de datos. (5° Ed.). Madrid: McGraw-Hill.

RECURSOS DISPONIBLES A TRAVÉS DE CICCO

- Cleve, J., & Lämmel, U. (2016). *Data Mining.* Berlin: De Gruyter Oldenbourg. Recuperado de: http://eds.b.ebscohost.com.
- Lancheros-Cuesta, D., Mojica Ospina, A. D., & Barbosa Vargas, D. A. (2015). Sistema de información para administración de inventarios basado en técnicas de minería de datos. Revista Épsilon. Recuperado de: http://eds.b.ebscohost.com.
- Makhabel, B. (2015). *Learning Data Mining with R. Birmingham.* England: Packt Publishing. Recuperado de: http://eds.b.ebscohost.com.