

UNIVERSIDAD NACIONAL DE ASUNCIÓN
FACULTAD POLITÉCNICA
INGENIERIA EN INFORMÁTICA
PLAN 2008
PROGRAMA DE ESTUDIO
Resolución N° 18/14/19-00 Acta N° 1029/16/07/2018 - ANEXO 02

I. - IDENTIFICACIÓN

- | | |
|-------------------------------------|--------------------------|
| 1. Asignatura | : Electiva 3 Data Mining |
| 2. Semestre | : Octavo |
| 3. Horas semanales | : 7 horas |
| 3.1. Clases teóricas | : 2 horas |
| 3.2. Clases prácticas | : 2 horas |
| 3.3. Clases de laboratorio | : 3 horas |
| 4. Total real de horas disponibles: | 112 horas |
| 4.1. Clases teóricas | : 32 horas |
| 4.2. Clases prácticas | : 32 horas |
| 4.3. Clases de Laboratorio | : 48 horas |

II. FUNDAMENTACION

En este curso el estudiante conocerá los principios, metodologías y técnicas de KDD y Data Mining, así como su uso y aplicación en Inteligencia de Negocios.

Se presentará una metodología para el desarrollo de aplicaciones 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.

III. OBJETIVOS:

1. Reconocer el potencial y la problemática del análisis de sistemas de información para la toma de decisiones.
2. Identificar el modelo multidimensional de datos de los almacenes de datos y los operadores de refinamiento asociados.
3. Describir los conceptos específicos de KDD y DM, como base necesaria para desarrollar cualquier proyecto de KDD en el ámbito empresarial.
4. Describir las fases del Descubrimiento de Conocimiento de Bases de Datos y la importancia de las mismas en el éxito del proceso.
5. Identificar las diferentes tareas y técnicas disponibles hoy en día.
6. Implementar de algunos algoritmos utilizados en técnicas de Clasificación, Segmentación y de Asociación.
7. Reconocer los beneficios del uso sistemático de técnicas de extracción de conocimiento para la obtención de modelos y patrones predictivos o descriptivos

IV. PRE-REQUISITOS:

El alumno debe haber aprobado el 70 % de los créditos.

V. CONTENIDO

5.1. Unidades programáticas

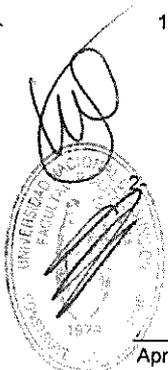
- 1- Introducción al proceso KDD(Knowledge Discovery in Database)
- 2- Introducción a la Minería de Datos. (DM – Data Mining)
- 3- Técnicas de preparación y limpieza de datos.
- 4- Técnicas de Minería de Datos
- 5- Algoritmos de Data Mining
- 6- Generación y validación de Modelos de Data Mining.

5.2. Desarrollo de las unidades programáticas

- 1- Introducción al proceso KDD(Knowledge Discovery in Database)
 - 1.1. Concepto y definición de KDD
 - 1.2. Proceso de KDD.
 - 1.3. Agentes del KDD.
- 1.4. Metodologías de KDD (CRISP-DM y SEMMA).
- 1.5. Herramientas de KDD (KNime, SPSS Modeler, Weka)

Introducción a la Minería de Datos. (DM – Data Mining)

- 2.1. Objetivos del DM (Predicción y Descripción).
- 2.2. Tareas de DM (Clasificación, Modelos de Dependencias, Agrupamiento o clusterización, Regresión, Visualización)
- 2.3. Tareas vs. Técnicas vs. Algoritmos



- 3- Técnicas de preparación y limpieza de datos.
 - 3.1. Tratamiento de datos ausentes
 - 3.2. Tratamiento de anomalías
 - 3.3. Discretizaciones
 - 3.4. Técnicas de Normalización de Datos
 - 3.5. Técnicas de Reducción de Dimensiones
- 4- Técnicas de Minería de Datos
 - 4.1. Introducción al Árbol de Decisión.
 - 4.2. Introducción a las Reglas de Inducción
 - 4.3. Introducción a las Redes Neuronales
 - 4.4. Introducción a las Redes Bayesianas
 - 4.5. Introducción a los Metodos del vecino mas cercano
 - 4.6. Introducción a los Modelos de Recomendación
- 5- Algoritmos de Data Mining
 - 5.1. Algoritmos de Clasificación basados en Tecnicas de Arbol de Decisión (J48 y CART – Gini)
 - 5.2. Algoritmos de Reglas de Asociacion (A priori, AprioriAll, PredictiveApriori)
 - 5.3. Algoritmo de Clustering (K-nn, K- means)
 - 5.4. Algoritmos de Recomendación (Collaboriting Filtering, High Utility Itemset,...)
- 6- Generación y validación de Modelos de Data Mining.
 - 6.1. Técnicas de entrenamientos y pruebas de modelos.
 - 6.2. Técnicas de validación de modelos y patrones.
 - 6.2.1. Matriz de Clasificación
 - 6.2.2. Matriz de Contingencia
 - 6.2.3. Curva ROC
 - 6.2.4. Curva Lift
 - 6.3. Métricas de validación de Modelos (Lift, Recall, Now, Confidence, Sensibility,...)

VI. ESTRATEGIAS METODOLÓGICAS

1. Exposición de la teoría con diferentes técnicas.
2. Técnicas grupales e individuales para la resolución de ejercicios.
3. Elaboración y presentación de trabajos prácticos.
4. Trabajo en Laboratorio.

VII. MEDIOS AUXILIARES

1. Pizarra
2. Marcadores y borrador de pizarra.
3. Bibliografía de apoyo.
4. Equipo multimedia
5. Equipo de Laboratorio

VIII. EVALUACIÓN

Según Normativas y Reglamentaciones vigentes en la Facultad Politécnica.

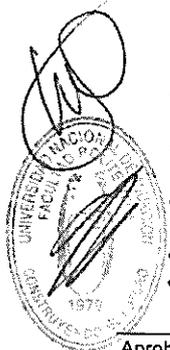
IX. BIBLIOGRAFÍA

PRINCIPAL:

- Larose D. (2005). *Discovery Knowledge in Data. An Introduction in Data Mining*. Wiley.
- Hernández, J., Ramírez, M.J.; Ferrí, C. (2004). *Introducción a la Minería de Datos*. Pearson Prentice Hall.
- Molinas, J.M. & García Herrero, J. (2004). *Técnicas de análisis de dato*.

BIBLIOGRAFÍA ALTERNATIVA:

- Berry M.J.A. & Linoff, G.S. (2000). *Mastering Data Mining*. Wiley.
- Berthold, M., Hand, D.J. (ed). (1999). *Intelligent Data Analysis*. Springer.
- Dunham, M.H. (2003). *Data Mining: Introductory and Advanced Topics*. Prentice Hall.
- Dzeroski, S. & Lavrac, N. (2001). *Relational Data Mining*. Springer.
- Elmasri, R. & Navathe, S. B. (2002). *Almacenes de datos (data warehousing) y minería de datos: En su Fundamentos de sistemas de bases de datos*. (3° Ed.) Madrid: Addison Wesley.
- Fayyad, U.M., Grinstein, G. & Wierse, A. (2001). *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann, Harcourt Intl.
- Fayyad, U.M., Piatetskiy-Shapiro, G., Smith, P. & Ramasamy, U. (1996). *Advances in Knowledge Discovery and Data Mining*. AAAI Press / MIT Press.
- Han, J. & Kamber, M. (2001). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Hand, D.J., Mannila, H. & Smyth, P. (2000). *Principles of Data Mining*. The MIT Press.



- Mitchell, T.M. (1997). Machine Learning. McGraw-Hill.
- Pyle, D. "Data Preparation for Data Mining" Morgan Kaufmann, Harcourt Intl., 1999.
- Ricardo, M. R. (2009). *Almacenes de datos (data warehousing) y minado de datos (data mining)*. México: McGraw-Hill.
- Silberschatz, A. & Korth, H. F. (2006). *Análisis y minería de datos*. (5° Ed.). Madrid: McGraw-Hill.
- Thuraisingham, B. (1999). Data Mining. Technologies, Techniques, Tools, and Trends. CRC Press.
- Witten, I.H. & Frank, E. (1999). Tools for Data Mining. Morgan Kaufmann.
- Wong, P. C. (1999). Visual Data Mining. Special Issue of IEEE Computer Graphics and Applications.

BIBLIOGRAFÍA COMPLEMENTARIA EXISTENTE EN EL CENTRO DE INFORMACIÓN Y CULTURA

- Abadal Falgueras, E. & Codina Bonilla, L. (2009). *Bases de datos documentales: características, funciones y método*. Madrid: Síntesis.
- Análisis y diseño de sistemas. (2001). México: McGraw-Hill
- Balabanian, N. & Carlson, B. (2002). *Principios de diseño lógico digital*. México: Compañía Editorial Continental.
- Barrios Poquioma, J. (2008). *Diseño de bases de datos con Winisis: nivel básico*. Lima: Biblioteca Nacional
- Beynon-Davies, P. (2014). *Sistemas de bases de datos*. Barcelona: Reverté
- Burgos, A. (2009). *Seguridad PC: proteja sus datos y privacidad de los ataques de hackers y virus*. Banfield: Gradi.
- Centro de Computación Profesional de México (2001). *Análisis y diseño de sistemas*. México: McGraw-Hill
- Colle, R. (2002). *Explotar la información noticiosa: data mining aplicado a la documentación periodística*. Madrid: Universidad Complutense de Madrid.
- Elmasri, R. & Navathe Shamkant (2002). *Fundamentos de sistemas de bases de datos*. (3° ed.). Madrid: Pearson Educación
- Mendelzon, A. (2000). *Introducción a las bases de datos relacionales*. Buenos Aires: Prentice Hall.
- Miguel, Castaño, A., Plattini Velthuis, M. & Marcos Martínez, E. (2000). *Diseño de bases de datos relacionales*. México: Alfaomega.
- Plattini Velthuis, M. (2007). *Tecnología y diseño de bases de datos*. México: Alfaomega
- Ricardo, C. M. (2009). *Bases de datos*. México: McGraw-Hill
- Silberschatz, A., Korth, H. F. & Sudarshan, S. (2006). *Fundamentos de bases de datos*. (5° Ed.). Madrid: McGraw-Hill.
- Whitten, J. L., Bentley, L. D. (2008). *Análisis de sistemas: diseño y métodos*. (7° ed.)México: McGraw-Hill

LIBROS ELECTRONICOS DISPONIBLES EN LA COLECCIÓN MGH

- Whitten, J. L., & Bentley, L. D. (2008). *Análisis de sistemas: diseño y métodos (7a. ed.)*. Retrieved from <http://ebookcentral.proquest.com>

