



CÓDIGO ASIGNATURA
1131-3

DEPARTAMENTO: *Ingeniería e Investigaciones Tecnológicas*

ASIGNATURA: DATA MINING y DATA WAREHOUSE
Plan 2009

Ingeniería en Informática
Año: 5 (Electiva -
Ingeniería de Software)

1. OBJETIVOS

OBJETIVOS GENERALES

- ❑ Que el estudiante conozca y comprenda la metodología de descubrimiento de conocimiento (KDD) en grandes bases de datos a partir de extracción de información desde ellas.
- ❑ Que el estudiante aprenda a utilizar las técnicas de explotación de datos (Data Mining) que permiten reconocer patrones, agrupar elementos y predecir comportamientos sobre grandes bases de datos.
- ❑ Que el estudiante conozca y comprenda el concepto de almacén de datos (Data Warehouse) como colección de datos orientada a sujetos, integrada, no volátil y variante en el tiempo.
- ❑ Que el estudiante forme idea de la diversidad de dominios sobre los cuales pueden usarse estas técnicas.

OBJETIVOS PARTICULARES

- ❑ Que el estudiante se familiarice con los principios del aprendizaje automático aplicados a la explotación de datos.
- ❑ Que el estudiante sepa identificar el modelo estadístico y el algoritmo adecuado para cada tipo de problema en el que sea necesario reconocer patrones, clasificar o predecir.
- ❑ Que el estudiante se habitúe al trabajo con grandes volúmenes de datos.
- ❑ Que el estudiante conozca y comprenda las etapas de análisis de requerimientos, diseño conceptual, diseño lógico y diseño físico que involucra la construcción de un almacén de datos (Data Warehouse).
- ❑ Que el estudiante se familiarice con la relación entre las técnicas de la minería de datos y las de almacenamiento de los mismos.

2. CLASIFICACIÓN DE LA ACTIVIDAD CURRICULAR, FORMACIÓN PRÁCTICA Y CARGA HORARIA



2.1

| | Carga horaria en horas reloj |
|--|------------------------------|
| Bloque de Ciencias Básicas | |
| Bloque de Tecnologías Básicas | |
| Bloque de Tecnologías Aplicadas | 4 |
| Bloque de Complementarias | |
| Otros Contenidos | |
| Carga horaria total de la actividad curricular | |

2.2

| Disciplina | Carga Horaria |
|---|---------------|
| Matemática | |
| Física | |
| Química | |
| Sistemas de representación y fundamentos de informática | 1 |
| Biología | |
| Otros (ciencia de la tierra, geología, etc.) | |
| Total | |

2.3

| Formación Práctica | | | | |
|------------------------|---------------------------------------|----------------------------------|----------------------------------|-------|
| Formación Experimental | Resolución de problemas de ingeniería | Actividades de proyecto y diseño | Práctica profesional supervisada | Total |
| 1 | 1 | 1 | | 3 |

2.4

| | |
|--|---|
| Carga horaria semanal | 4 |
| Carga horaria semanal dedicada a la formación práctica | 3 |

3. CONTENIDOS

Unidad 1: El proceso de descubrimiento del conocimiento. Datos: fuentes, limpieza, integración y almacenamiento. Exploración de datos: consultas y análisis estadísticos. Descubrimiento de información: algoritmos inteligentes, entrenamiento y testeo. Visualización de información: técnicas de presentación multidimensional. Dominios de aplicación.

Unidad 2: Introducción al Data Warehouse. Definición y ejemplos. Arquitectura: fuentes de datos, proceso ETL, data marts y herramientas de análisis.. Visualización de datos. Cubos. Operaciones OLAP: roll-up, drill-down , pivoting y slice & dice.

Unidad 3: Diseño de Data Warehouse. Fases. Análisis de requerimientos: estrategias. Diseño conceptual: DFM, Árbol de atributos Diseño lógico. Construcción del esquema lógico y diseño del proceso de carga. Molap, Rolap y Holap. Diseño físico. Carga y actualización. Aplicaciones.



Unidad 4: Modelado: descripción y predicción. Reglas de asociación y algoritmo apriori. Clusterización: agrupamientos jerárquicos y no jerárquicos, criterios y distancias. Dendrogramas. Algoritmo K-means. Noción de análisis de componentes principales y de correspondencias. Aplicaciones.

Unidad 5: Principios de aprendizaje automático. Clasificación y predicción. Árboles de decisión: algoritmo J48. Técnicas bayesianas. Entrenamiento y testeo. Validación, sobreentrenamiento y "pruning". Curvas ROC. Bootstrap y jackknife. Aplicaciones.

Unidad 6: Modelo de regresión lineal. Supuestos, estimación de parámetros. Adecuación. Validación. Predicción. Modelo de red neuronal. Perceptrón y Adaline. Back Propagation. Clasificación. Aplicaciones.

4. BIBLIOGRAFÍA

| Título | Autor(es) | Editorial | Año Edición | Ejemplares disponibles en UNLaM |
|---|--|-----------------------------|-------------|---------------------------------|
| Data Minig Techniques | Berry,M Linoff,G | MileyNY | 2004 | 1 |
| Data Minig Solutions | Blaxton,T Westphal, Ch | John Wiley NY | 1998 | 1 |
| Data Minig: a hands approach for busines | Groth,R | Prentice Hall | 2000 | 1 |
| Data Mining | Han, J Kam- ber,M Pei,J | Morgan Kaufmann | 2006 | no |
| Data Warehousing: building the corporate knowlwdge base. | Hammergen, T | International Thomsom | 2000 | 1 |
| Data Warehousing: strategies, technologies and techniques | Mattisson,R | McGraw-Hill | 1996 | 1 |
| Cluster analysis | Everitt, B Landau, S Leese, M | Arnold | 2001 | no |
| Introducción al Análisis de Regresión Lii-neal | Montgomery ,D Peck, E Vining, G | Grupo Edi- torial Patria | 2007 | no |
| Métodos Multi-variados Apli-cados al Análi-sis de Datos | Johnson, D | International Thomsom | 2000 | no |
| Redes Neuro-nales | Freeman, M Skapura, D | Addison- Wesley | 1993 | 2 |
| Redes Neuro-nales y Siste-mas Borrosos | Del Brío, M Sanz Molina, A | Alfaomega | 2007 | no |



5. DESCRIPCIÓN DE ACTIVIDAD CURRICULAR

5.1) MODALIDAD DE ENSEÑANZA EMPLEADA

Las clases se llevarán a cabo en un laboratorio informático. Tendrán una primera parte expositiva en la cual el docente explicará los fundamentos de cada tema con el auxilio de pizarra y cañón. Su duración será aproximadamente de 1 hora. Luego tendrá lugar la parte correspondiente a actividades de formación práctica que se prolongará por las siguientes 3 horas, en la cual los estudiantes utilizarán conjuntos de datos y software a efecto de realizar tareas de aplicación de las técnicas expuestas.

5.2) MATERIALES DIDÁCTICOS NECESARIOS

Se utilizarán distintos conjuntos de datos cargados en el servidor al que tengan acceso los estudiantes y el siguiente software disponible bajo licencia en la universidad o libre: Weka, Spss, R, Spad, SQLServer

6. EVALUACIÓN

La evaluación será de tipo práctico debiendo el estudiante desarrollar y/o resolver un conjunto de problemas sobre la computadora, confeccionando un informe en formato .doc que enviará una vez terminado a la dirección de correo electrónico que el docente indique para su corrección. El examen tendrá lugar en el mismo laboratorio en que se desarrollaron las clases y durará un tiempo máximo de 4 horas. El régimen de aprobación, promoción y cursada se ajustará a las resoluciones dictadas por el Consejo Académico del DIIT.

7. COMPOSICIÓN DEL EQUIPO DOCENTE ACTUAL

7.1 Responsable a cargo de la actividad curricular:

Cristóbal R. Santa María, Especialista en Data Mining. (UBA) Profesor Asociado. Interino. Dedicación Exclusiva.

7.2) PROFESORES

| Apellido y Nombre | Grado académico máximo | Cargo Docente | Situación | Dedicación en horas semanales al cargo |
|---|------------------------|---------------|-----------|--|
| Se nombrará un docente más con estudios de posgrado y experiencia profesional sobre la temática de la materia | | | | |

Cantidad total de profesores: 2

7.3) AUXILIARES GRADUADOS



| Apellido y Nombre | Grado académico máximo | Cargo Docente | Dedicación en horas semanales al cargo |
|--|------------------------|---------------|--|
| Se nombrará un auxiliar docente con estudios de posgrado en la temática de la materia. | | | |

Cantidad total de auxiliares:1

10. SUFICIENCIA Y ADECUACION DE LOS ÁMBITOS

Se considera que varios laboratorios de la Unlam son adecuados y suficientes para la tarea. Además el software requerido está disponible.

11. INSCRIPCIÓN Y PROMOCIÓN DE ALUMNOS

12. EVALUACIÓN CAPACIDAD DE CATEDRA

Un equipo docente con las características señaladas estará altamente capacitado para las tareas docentes y de investigación en el campo que abarca la materia.

13. ACCIONES, REUNIONES, COMISIONES

14. CALENDARIO DE ACTIVIDADES

| Nº de Clase | Semana del calendario | Unidad Temática o Actividad |
|-------------|-----------------------|---|
| 1 | 34 | Unidad 1. Completa |
| 2 | 35 | Unidad 2 Completa |
| 3 | 36 | Unidad 3 Diseño de Data Warehouse. Fases. Análisis de requerimientos: estrategias. Diseño conceptual: DFM, Árbol de atributos |
| 4 | 37 | Unidad 3: Diseño lógico. Construcción del esquema lógico y diseño del proceso de carga. Molap, Rolap y Holap. |
| 5 | 38 | Unidad 3: Diseño físico. Carga y actualización. Aplicaciones |
| 6 | 39 | Unidad 4: Modelado: descripción y predicción. Reglas de asociación y algoritmo apriori. |
| 7 | 40 | Unidad 4: Clusterización: agrupamientos jerárquicos y no jerárquicos, criterios y distancias. Dendrogramas. Algoritmo K-means |
| 8 | 41 | Unidad 4: Noción de análisis de componentes principales y de corres- |



| | | |
|----|----|--|
| | | pondencias. Aplicaciones |
| 9 | 42 | Unidad 5: Principios de aprendizaje automático. Clasificación y predicción. Árboles de decisión: algoritmo J48 |
| 10 | 43 | Unidad 5: Técnicas bayesianas |
| 11 | 44 | Unidad 6: Entrenamiento y testeo. Validación, sobreentrenamiento y "pruning". Curvas ROC. Bootstrap y jackknife. Aplicaciones. |
| 12 | 45 | Unidad 6: Modelo de regresión lineal. Supuestos, estimación de parámetros. Adecuación. Validación. Predicción |
| 13 | 46 | Unidad 6: Modelo de red neuronal. Perceptrón y Adaline. Back Propagation. Clasificación. Aplicaciones |
| 14 | 47 | Examen parcial |
| 15 | 48 | Resolución del parcial y notas |
| 16 | 49 | Examan recuperatorio |

INFORMACIÓN PROPIA CÁTEDRA

17. APUNTES ELABORADOS POR LA CÁTEDRA

Se adjunta material de clases preparado para introducir el tema Data Mining

25. ACLARACIÓN, CARGO Y FECHA

“Certifico que el presente programa de estudios de la asignatura....Data Mining y Data Warehouse..... es el vigente para el ciclo lectivo ..2013....., guarda consistencia con los contenidos mínimos del plan de estudios y se encuentra convenientemente actualizado”

Firma
socio

Aclaración Cristóbal R. Santa maría
Fecha 6/10/2010

Cargo Profesor