

Formulario de Aprobación Curso de Actualización 2010.

Asignatura: Introducción al Data Mining

Profesor de la asignatura ¹: Ing, Magister, Gustavo Denicolay, Universidad de Buenos Aires

Profesor Responsable Local ¹: Prof. Alejandro Vaisman, Gr. 4, INCO, UdelAR

Instituto ó Unidad: Computación

Departamento ó Area:

¹ Agregar CV si el curso se dicta por primera vez.

Fecha de inicio y finalización: 25 de octubre al 1 de noviembre de 2010

Horario y Salón: a confirmar

Horas Presenciales: 30

Arancel: \$ 7.154

Público objetivo y Cupos: Profesionales y académicos universitarios, principalmente de las áreas de Informática, ciencias económicas y ciencias biológicas.

Objetivos: Introducir el problema de la canasta de mercado con sus ricas derivaciones junto con las herramientas que lo solucionan. Capacitar en forma práctica y teórica a los alumnos para que puedan detectar y resolver problemas relacionados con reglas de asociación y patrones secuenciales.

Conocimientos previos exigidos: Ninguno.

Conocimientos previos recomendados: Nociones de probabilidad y estadística, bases de datos relacionales y algoritmos.

Metodología de enseñanza:

Se planteará un problema inicial de canasta de mercado a los alumnos junto con su historia y motivaciones para luego ir presentando los elementos teóricos y conceptuales en el justo tiempo que van siendo requeridos para solucionar las distintas partes del problema. La parte teórica demandará el 50% del tiempo. El curso se completará con un trabajo a entregar luego de la finalización del dictado de clases.

Forma de evaluación:

Trabajos realizados en el laboratorio. 10hs.

Trabajo domiciliario a entregar luego de la finalización del dictado de clases. 15hs

Temario:

- Analisis de Asociaciones, historia y motivación.
- El problema de la canasta del mercado
- Generacion de Itemsets frecuentes, soporte y confianza
- Algoritmo *Apriori*
- Generacion de Reglas
- Alternativas al algoritmo *A priori*, algoritmo FP
- Medidas objetivas del grado de interés de un patrón de asociación
- Extensiones
 - Atributos categoricos

- Algoritmo *Apriori* (2hs)
- Generacion de Reglas (1h)
- Alternativas al algoritmo *A priori*, algoritmo FP (1h)
- Medidas objetivas del grado de interés de un patrón de asociación (1h)
- Extensiones (1h)
 - Atributos categoricos
 - Atributos numéricos
 - Jerarquias
- Búsqueda de patrones infrecuentes (2hs)
- Aproximaciones Temporales (2hs)
- Patrones Secuenciales (2hs)
- Patrones en grafos (2hs)
- Problemas resolubles con las metodologías propuestas (1h)

Bibliografía:

Papers

Mining Association Rules between Sets of Items in Large Databases, Agrawal Rakesh

Fast Algorithms for Mining Association Rules, Agrawal Rakesh

Mining Generalized Association Rules, Srikant Ramakrishnan

Mining Sequential Patterns, Agrawal Rakesh

Mining Sequential Patterns : generalization and performance improvements, Agrawal Rakesh

Mining Sequential Patterns with regular expressions constrains, Garofalakis Minos

Beyond Market Baskets: Generalizing Association Rules to Correlations Sergey Brin , Rajeev Motwani , Craig Silverstein

Libros

Introduction to Data Mining, Pang-Ning Tan, Vipin Kumar, Michael Steinbach, , Addison-Wesley, ISBN 0321321367, Mayo 2005

Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms . Jean-Marc Adamo Springer, ISBN 0387950486, Dic 2000