FORMATO 4

Programas de Estudios Modalidad Escolarizada

NOMBRE DE LA ASIGNATURA: Minería de Datos

CICLO, ÁREA O MÓDULO: Optativa

CLAVE: COM 23106

OBJETIVO(S) GENERAL(S) DE LA ASIGNATURA:

Que el alumno adquiera conocimientos y dominio de los métodos de la minería de datos que le permitan desarrollar aplicaciones en el análisis y la organización de grandes conjuntos de datos.

TEMAS Y SUBTEMAS:

I. Introducción a la minería de datos y al descubrimiento de conocimiento en grandes bases de datos.
- I.1. Discusión de los conceptos de dato, información y conocimiento.
- I.2. Origen y propósito de la minería de datos.
- I.3. Áreas relacionadas con la minería de datos.
- I.4. Métodos de consulta y métodos de exploración de datos. Métodos de análisis gráfico.
- I.5. Plataforma tecnológica de la minería de datos. “Data Warehouse” y “On-Line Analytical Processing”.
- I.6. Preparación, selección, depuración, enriquecimiento y codificación de los datos.
- I.7. Aplicación de la minería de datos a problemas que se presentan en las organizaciones

II. Marco conceptual de los métodos de la minería de datos
- II.1.Los métodos inductivo y deductivo. Principios de la formulación de modelos. Ejemplos de métodos de la minería de datos
- II.2.Experiencia, algoritmo y medida del aprendizaje.
- II.3. Aprendizaje de un concepto
- II.4. Representación de hipótesis y búsqueda en el espacio de hipótesis. Sesgo inductivo.

III. Modelos de árboles de decisión y clasificación
- III.1. Árboles, clasificación y aproximación de funciones discretas.
- III.2.- El algoritmo ID3. Función de evaluación: ganancia de información.
- III.3.- Construcción y validación del modelo. Curva de aprendizaje. Podas y “sobreajuste”
- III.4.- Extensiones del algoritmo ID3, C4.5: atributos continuos, otras funciones de evaluación, información faltante, incorporación de costos.

IV. Aplicación de los modelos de redes neuronales
- IV.1.- El “Perceptron”. Redes con conexiones hacia adelante. Aproximación de funciones vectoriales.
- IV.2.- Entrenamiento con “backpropagation” y validación del modelo. “Sobreajuste”
- IV.3.- Otros modelos. Redes de Kohonen. Redes que aproximan mediante funciones gaussianas (“Radial Basis Functions”)

V Aplicación de métodos basados en la estimación de funciones de probabilidad
- V.1 Revisión de la teoría de la probabilidad. Espacio de probabilidad. Fórmula de Bayes.
- V.2 Principios de búsqueda de la hipótesis más probable. Hipótesis de máxima verosimilitud. Estimación cuadrática media y curva de regresión. Principio de descripción mínima.
- V.3 Clasificador óptimo de Bayes y clasificador “Naive”. Clasificación de textos.
- V.4 Modelos de redes causales (redes de Bayes)

V.5 Algoritmo de “Expectation Maximization” (EM) y variables ocultas.

VI Aplicación de métodos de aproximación con base a una función de cercanía y métodos de agregación (“clustering”).
V.1 Aproximación de funciones continuas y discretas con K vecinos cercanos.
V.2 Regresión local mediante funciones lineales.
V.3 Métodos “clustering”

ACTIVIDADES DE APRENDIZAJE:

El método de enseñanza consiste en exponer la teoría, discutir los algoritmos y proponer al estudiante ejercicios simples que completen el aprendizaje. Los ejercicios que requieran análisis de grandes conjuntos de datos deben ser solucionados sobre algún ambiente de cómputo de propósito general que permita la programación, el cálculo numérico, el cálculo simbólico y la realización de gráficas. En ocasiones también se puede disponer de productos de “software” o “paquetes” con los algoritmos de minería de datos programados que sólo requieran familiaridad con la interfase gráfica y la elección de parámetros para ser aplicados en problemas.
La realización de trabajos y proyectos por parte de los estudiantes constituye un elemento importante del aprendizaje. En estos trabajos se puede pedir a los estudiantes realizar actividades que exijan un mayor dominio de la teoría como la aplicación de los algoritmos a conjuntos de datos con información real socioeconómica, transacciones en negocios, mediciones en ingeniería o textos, la modificación de los métodos estudiados para extraer otra información útil de los datos y el desarrollo de nuevos métodos

EVALUACIÓN DEL CURSO:

La calificación se asignará conforme a la siguiente expresión:

C = 30%ex1 + 35%ex2 + 35%f

ex1 y ex2	Exámenes Parciales 1 y 2
f	Examen Final

BIBLIOGRAFÍA:

- Hand, David, Mannila, Heikki,Smyth, Padhraic, “Principles of data mining”, MIT Press, Cambridge, MA, 2001
- Witten, Ian H. and Frank, Eibe, “Data mining: Practical Machine Learning Tools and Techniques with Java Implementations”, Morgan Kaufmann Publishers, San Francisco, CA, 2000
- Berry, Michael J. A. y Linoff, Gordon, “Mastering data mining : the art and science of customer relationship management”, John Wiley & Sons, New York, 2000
- Pyle, Dorian, Data preparation for data mining, San Francisco, Calif. : Morgan Kaufmann Publishers, 1999
- Berry, Michael J.A. y Linoff, Gordon, “Data mining techniques for marketing, sales, and customer support“,Wiley, New York , 1997
- Mitchell, Tom M., “Machine Learning”, McGraw-Hill, Boston, MA, 1997
- Adriaans, Pieter y Zantinge, Dolf, “Data Mining”, Addison-Wesley, 1996
- Fayyad, Usama M., et. al. “Advances in Knowledge Discovery and Data Mining”, AAAI Press y MIT Press, Cambridge, MA, 1996
- Goonatilake, Suran and Treleaven, Philip (editors), “Intelligent Systems for Finance and Business”, John Wiley & Sons Ltd, 1995