

IN720 TÉCNICAS AVANZADAS DE MINERÍA DE DATOS (10 UD)

REQUISITOS : IN60E / AD
CARÁCTER : Electivo para el Magíster en Gestión de Operaciones
PROFESORES: Richard Weber - Fabián Medel
PROF. AUX. : Jaime Miranda
SEMESTRE : Primavera 2004
HORARIO : C: 1.4 - 3.5 A: 3.4

OBJETIVOS:

Profundizar los conceptos de la Minería de Datos (Data Mining) y comprender las características de las técnicas más relevantes

ESPECIFICOS:

- Profundizar los pasos previos a la aplicación de los métodos de la minería de datos.
- Conocer técnicas avanzadas de la minería de datos.
- Entender los efectos que tienen tanto la preparación de los datos como los parámetros de las técnicas para los resultados obtenidos.
- Conocer varias formas de evaluación de resultados e interpretarlas adecuadamente.
- Desarrollar la capacidad de llevar a cabo proyectos aplicando la minería de datos.

TEMARIO:

- Revisión del proceso KDD (Knowledge Discovery in Databases).
- Métodos de selección y extracción de atributos.
- Transformación de los datos y generación de atributos derivados.
- Métodos avanzados de la minería de datos (redes neuronales de Kohonen, Support Vector Machines, entre otros)
- Evaluación de resultados
- Desarrollo futuro: minería de datos dinámica.

ACTIVIDADES:

Los alumnos, en forma adicional a las clases de cátedra y clases auxiliares, desarrollarán tres tareas y un CTP semanal

Para las 3 tareas los alumnos trabajarán en grupos resolviendo un problema real aplicando herramientas computacionales (Matlab entre otras). Se seguirá durante las tareas con este problema y los datos del mismo.

Tarea 1: Selección de atributos

En esta tarea se aplicará los métodos vistos en clase para la selección de atributos.

Tarea 2: Transformación de los datos y generación de atributos de rivados
Los alumnos aplicarán las técnicas introducidas en clase y desarrollarán enfoques propios para la transformación de datos y generación de atributos.

Tarea 3: Técnicas avanzadas de la Minería de Datos
Se aplicarán métodos avanzados de la minería de datos vistos en clase al problema dado. Cada grupo puede mantener sus resultados de las dos tareas anteriores. También hará que verificar la calidad de los modelos desarrollados con técnicas de evaluación.

Para cada tarea hay que entregar un breve informe describiendo el enfoque central de la tarea respectiva y el sistema desarrollado. Habrá una nota para cada tarea y cada grupo.

Cada semana se asignará un paper a un alumno del curso. La semana siguiente éste debe presentar la idea central de la lectura en una presentación de 15 minutos en el curso. Enseguida se tomará un CTP al curso entero de 15 minutos sobre la materia expuesta. La nota de un CTP puede ser 1 (aprobado) o 0 (reprobado).

EVALUACION

Se evaluará el rendimiento de los alumnos con los siguientes elementos: CTPs semanales, 3 tareas, 2 controles y un examen.

Nota CTPs:

La nota de los CTPs se define como:

$$6 * (\text{número de CTPs aprobados/número total de CTPs}) + 1$$

Nota tareas:

Las tareas deberán ser realizadas en grupos de tres personas. El profesor se guarda el derecho de interrogar a cualquier integrante de un grupo para evaluar el trabajo del grupo completo, esto consiste en la elección de algún integrante al azar, al cual se le harán preguntas sobre la tarea o tendrá que mostrar manejo del software.

La ponderación de la nota de cada tarea será:

$$\text{Nota Tarea} = 0.8 * \text{Nota Informe} + 0.2 * \text{Nota Interrogación.}$$
$$\text{Nota Tarea Final} = \text{Promedio de las notas de cada tarea.}$$

Nota controles:

Cada control tendrá una nota. La nota de los controles será el promedio de las notas de cada uno de los controles.

Al final del semestre habrá un examen.

Para aprobar el curso se necesita lo siguiente:

Nota controles ≥ 4.0

Nota tareas ≥ 4.0

Nota CTPs ≥ 5.0

La nota final del curso se determinará con la siguiente ponderación:

Nota controles: 40%

Nota tareas: 40%

Nota CTPs: 20%

LECTURA BÁSICA DEL CURSO

- Berry, M. J. A., Linoff, G. (2004): Data Mining Techniques – For Marketing, Sales, and Customer Support. Sec. ed., Wiley Computer Publishing, New York
- Cristianini, N., Shawe-Taylor, J. (2003): Support Vector and Kernel Methods. In: Berthold, M., Hand, D. (eds.): Intelligent Data Analysis. Sec. ed., Springer, 169-197
- Duda, R. O., Hart, P. E., Stork, D. G. (2001): Pattern Classification. Sec. ed., John Wiley & Sons, New York, Chichester
- Edelstein H. (1998): Introduction to Data Mining and Knowledge Discovery
- Famili, A., Shen, W.-M., Weber, R., Simoudis, E. (1997): Data Preprocessing and Intelligent Data Analysis. Intelligent Data Analysis Vol. 1, No. 1, 3-23
- Han, J., Kamber, M. (2001): Data Mining – Concepts and Techniques. Morgan Kaufmann Publishers, San Francisco
- Hastie, T., Tibshirani, R., Friedman, J. (2001): The Elements of Statistical Learning. Springer Series in Statistics. Springer Verlag, Basel
- Quinlan, J. R., C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, 1993
- Rojas, R. (1996): Neural Networks – A Systematic Introduction. Springer-Verlag, Berlin Heidelberg
- Weber, R. (2000): Data Mining en la Empresa y en las Finanzas Utilizando Tecnologías Inteligentes. Revista Ingeniería de Sistemas XIV, No 1, 61-78
- Zimmermann, H.-J. (2001): Fuzzy Set Theory - and Its Applications. 4th Edition. Kluwer Academic Publishers, Boston Dordrecht London

Se entregará literatura específica para los CTPs.

Más información sobre los temas del curso es disponible en U-Cursos <http://docencia.ing.uchile.cl>