



Ministerio de Cultura y Educación  
 Universidad Nacional de San Luis  
 Facultad de Ciencias Físico Matemáticas y Naturales  
 Departamento: Informatica  
 Area: Area V: Automatas y Lenguajes

(Programa del año 2012)  
 (Programa en trámite de aprobación)  
 (Presentado el 02/08/2012 22:18:44)

### I - Oferta Académica

| Materia   | Carrera      | Plan   | Año  | Período         |
|---|--------------|--------|------|-----------------|
| (OPTATIVAS) APRENDIZAJE AUTOMATICO Y MINERIA DE DATOS | LIC.CS.COMP. | 006/05 | 2012 | 1° cuatrimestre |

### II - Equipo Docente

| Docente                     | Función                 | Cargo     | Dedicación |
|-----------------------------|-------------------------|-----------|------------|
| ERRECALDE, MARCELO LUIS     | Prof. Responsable       | P.Adj Exc | 40 Hs      |
| LEGUIZAMON, MARIO GUILLERMO | Prof. Colaborador       | P.Adj Exc | 40 Hs      |
| CAGNINA, LETICIA CECILIA    | Responsable de Práctico | JTP Exc   | 40 Hs      |

### III - Características del Curso

| Credito Horario Semanal |          |                   |                                       |       |
|-------------------------|----------|-------------------|---------------------------------------|-------|
| Teórico/Práctico        | Teóricas | Prácticas de Aula | Práct. de lab/ camp/ Resid/ PIP, etc. | Total |
| Hs                      | 2 Hs     | 2 Hs              | 3 Hs                                  | 7 Hs  |

| Tipificación                                   | Periodo         |
|--|-----------------|
| B - Teoria con prácticas de aula y laboratorio | 1° Cuatrimestre |

| Duración   |            |                     |                   |
|------------|------------|---------------------|-------------------|
| Desde      | Hasta      | Cantidad de Semanas | Cantidad de Horas |
| 14/03/2012 | 22/06/2012 | 15                  | 100               |

### IV - Fundamentación

A partir de la popularización de Internet y el crecimiento del número de bases de datos accesibles, se ha producido en los últimos años un gran crecimiento en los volúmenes y disponibilidad de datos alrededor del mundo. Se ha estimado que la cantidad de datos almacenados en las bases de datos del mundo se duplica cada veinte meses. Este crecimiento cuantitativo de los datos no se ha reflejado en un crecimiento cualitativo de la información disponible. Existe un "gap" creciente entre la generación de los datos y nuestro entendimiento de los mismos.

Una alternativa para solucionar este problema que ha recibido un interés creciente es el área de Minería de datos. La Minería de datos (en ingles Data Mininig y de ahora en más MD) es el proceso de descubrir patrones en los datos. Este proceso debe ser automático (o más usualmente) semi-automático. Los patrones descubiertos deben ser significativos en el sentido que deben conducir a alguna ventaja (generalmente económica).

### V - Objetivos / Resultados de Aprendizaje

Introducir al alumno en los principales conceptos vinculados a la Minería de datos. Al finalizar el curso los alumnos deberán ser capaces de identificar las principales etapas y procesos de la Minería de datos, los mecanismos estadísticos y de

aprendizaje automático frecuentemente utilizados en este área. Se espera además que los mismos sean capaces de detectar y aplicar las técnicas de Minería de datos en aplicaciones concretas del mundo real.

## VI - Contenidos

### Unidad 1. Aspectos Generales

Definición de Minería de Datos (MD). Origen. Relación con otras disciplinas. Áreas de aplicación. Aprendizaje de Máquina (AM). Etapas del proceso KDD. AM versus MD. MD en el contexto del proceso KDD. Principales tareas de la minería de datos.

### Unidad 2: Aprendizaje de Conceptos.

Definición del Aprendizaje de Conceptos. Representación de hipótesis. Espacio de Hipótesis. Aprendizaje como Búsqueda. Sesgo Inductivo.

### Unidad 3: Aprendizaje de árboles de decisión.

Algoritmo ID3. Medidas de entropía y ganancia de información. Sesgo inductivo en ID3. Métodos para evitar el sobre-entrenamiento.

### Unidad 4: Análisis de Clusters.

Conceptos básicos y Algoritmos. Grupos significativos vs. Grupos útiles. Tipos de agrupamientos. Particional vs Jerárquico. Difuso vs. No Difuso. Tipos de clusters. Medidas de similitud. Validación de grupos. Algoritmos de clustering.

### Unidad 5: Análisis de Asociaciones.

Conceptos básicos y algoritmos. Regla de asociación. Soporte y confianza. Itemsets frecuentes. El algoritmo Apriori.

### Unidad 6: Otras componentes del proceso de KDD. Preparación de los datos. Evaluación, interpretación y visualización de patrones.

### Unidad 7: Aspectos prácticos involucrados en el desarrollo de aplicaciones de MD para un problema particular. Herramientas de desarrollo de aplicaciones para MD. El sistema Weka. RapidMiner. Bases de datos públicas.

## VII - Plan de Trabajos Prácticos

Práctico 1: Aspectos generales.

Práctico 2: Aprendizaje Automático.

Práctico 3: Aprendizaje de conceptos. Aprendizaje de Árboles de decisión.

Práctico 4: Minería de textos.

Práctico 5: Desarrollo de un proyecto de programación a ser entregado a final de cuatrimestre.

## VIII - Regimen de Aprobación

La materia sólo tiene régimen Promocional: se aprueba con la entrega de los trabajos prácticos, un proyecto de programación propuesto más un informe adicional vinculado a los principales conceptos de MD. La nota de promoción es el promedio de las notas obtenidas en cada una de estas tareas.

## IX - Bibliografía Básica

[1] "Introduction to Data Mining". Tan, Steinbach y Kumar. Pearson. 2006.

[2] "Data Mining: Practical Machine Learning Tools and Techniques". Third Edition. Witten y Frank. Morgan Kaufmann.

2011.

[3] "Introducción a la Minería de datos". Orallo, J. H., Ramírez Quintana, Ma. J. y Ramírez, C.F.. Pearson Prentice Hall. 2004.

[4] "Machine Learning". Tom Mitchell. McGraw-Hill Series in Computer Science, 1997.

[5] "Readings in Machine Learning". Shavlik, J. y Dietterich, Thomas. Morgan Kaufmann, 1990.

[6] "The Elements of Statistical Learning". T. Hastie, R. Tibshirani, J. H. Friedman. Springer, 2002.

[7] "Principles of Data Mining (Adaptive Computation and Machine Learning)". David J. Hand, Heikki Mannila, Padhraic Smyth. MIT Press, 2001.

## **X - Bibliografía Complementaria**

[1] "The Text Mining HandBook: Advanced Approaches in Analyzing Unstructured Data". Feldman and Sanger. Cambridge University Press. 2007.

[2] Publicaciones y apuntes complementarios sobre los distintos tópicos del curso que estarán disponibles en <http://www.dirinfo.unsl.edu.ar/~aamd>

## **XI - Resumen de Objetivos**

Introducir al alumno en los principales conceptos vinculados a la Minería de datos. Al finalizar el curso los alumnos deberán ser capaces de identificar las principales etapas y procesos de la Minería de datos y los mecanismos de aprendizaje automático frecuentemente utilizados en este área. Se espera además que los mismos sean capaces de detectar y aplicar las técnicas de Minería de datos en aplicaciones concretas del mundo real.

## **XII - Resumen del Programa**

Unidad 1: Introducción al Aprendizaje Automático y la Minería de Datos.

Unidad 2: Aprendizaje de Conceptos.

Unidad 3: Aprendizaje de Árboles de Decisión.

Unidad 4: Análisis de Clusters.

Unidad 5: Análisis de Asociaciones.

Unidad 6: Otras componentes del proceso de KDD.

Unidad 7: Desarrollo de aplicaciones de Minería de Datos.

## **XIII - Imprevistos**

## **XIV - Otros**

| <b>ELEVACIÓN y APROBACIÓN DE ESTE PROGRAMA</b> |                             |
|--|-----------------------------|
|  | <b>Profesor Responsable</b> |
| Firma:   |                             |
| Aclaración:                                    |                             |
| Fecha:   |                             |