Ministerio de Cultura y Educación **(Programa del año 2015)**
Universidad Nacional de San Luis
Facultad de Ciencias Físico Matemáticas y Naturales
Departamento: Informatica
Area: Area V: Automatas y Lenguajes

I - Oferta Académica

Materia	Carrera	Plan	Año	Período
(OPTATIVA) APRENDIZAJE AUTOMATICO Y MINERIA DE DATOS	LIC.CS.COMP.	32/12	2015	2° cuatrimestre

II - Equipo Docente

Docente	Función	Cargo	Dedicación
ERRECALDE, MARCELO LUIS	Prof. Responsable	P.Adj Exc	40 Hs
CAGNINA, LETICIA CECILIA	Prof. Co-Responsable	P.Adj 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	1 Hs	5 Hs

Tipificación	Periodo
B - Teoria con prácticas de aula y laboratorio	2º Cuatrimestre

Duración			
Desde	**Hasta**	**Cantidad de Semanas**	**Cantidad de Horas**
10/08/2015	20/11/2015	15	75

IV - Fundamentación

A partir de la popularización de Internet y el incremento 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 inglés Data Mininig y de ahora en más MD) es el proceso de descubrir patrones en grandes volúmenes de 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 al aprendizaje automático y 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. Aprendizaje Automático (AA).
Definición. Importancia y Aplicaciones del AA. Tarea de Aprendizaje bien definida. Tipos de feedback para el aprendiz. Diseño de un sistema de aprendizaje. Aspectos de estudio relevantes en AA.

Unidad 2. Aprendizaje de Conceptos
Definición del Aprendizaje de Conceptos. Aprendizaje de un concepto a partir de ejemplos. Representación de hipótesis. Espacio de Hipótesis. Ordenamiento "General-a-específico". Los espacios de Versión y el algoritmo de eliminación de candidatos. Aprendizaje como Búsqueda. La necesidad del sesgo inductivo

Unidad 3. Aprendizaje por refuerzos.
Fundamentos de la Teoría de Utilidad. Problemas de decisión secuencial. Procesos de Decisión de Markov (MDPs). Funciones de valor y políticas óptimas. Algoritmos para resolución de MDP´s. Value Iteration. Q-Learning. Aprendizaje por Refuerzo en Robótica.

Unidad 4. Minería de Datos. 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. Clasificación, Aprendizaje de conceptos, Categorización, Regresión, Análisis de grupos (cluster analysis), Análisis de asociaciones (association analysis), Análisis de correlaciones.

Unidad 5: 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 6: Máquinas de vectores soporte (SVM).
Introducción. SVM para clasificación binaria. Justificación teórica. Aplicaciones de SVM. Extensiones y temas avanzados. Software para SVM.

Unidad 7: 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 8: Análisis de Asociaciones.
Conceptos básicos y algoritmos. Regla de asociación. Soporte y confidencia. Itemsets frecuentes. El algoritmo Apriori.

Unidad 9: Otras componentes del proceso de KDD.
Preparación de los datos. Evaluación, interpretación y visualización de patrones.

Unidad 10: Aplicaciones
Aspectos prácticos involucrados en el desarrollo de aplicaciones de MD para un problema particular. Caso de estudio: Minería de Textos y la Web. Herramientas de desarollo de aplicaciones para MD. El sistema Weka. RapidMiner. Bases de datos públicas.

VII - Plan de Trabajos Prácticos

Práctico 1: Aspectos generales. Aprendizaje de conceptos.

Práctico 2: Aprendizaje por Refuerzo.

Práctico 3: Aprendizaje de Arboles de decisión.

Práctico 4: Minería de textos.

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

Práctico de Máquina N ro 1: Q-Learning

Práctico de Máquina Nro 2: Minería de Textos

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 - Bibliografia 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 al aprendizaje automático y 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.
Unidad 2: Aprendizaje de Conceptos.
Unidad 3: Aprendizaje por refuerzos.
Unidad 4: Introducción a la Minería de Datos.
Unidad 5: Aprendizaje de Arboles de Decisión.
Unidad 6: Máquinas de vectores soporte (SVM).
Unidad 7: Análisis de Clusters.
Unidad 8: Análisis de Asociaciones.
Unidad 9: Otras componentes del proceso de KDD.
Unidad 10: Desarrollo de aplicaciones de Minería de Datos.

XIII - Imprevistos

XIV - Otros