

Planificación de
Tópicos de Minería de Datos



Código/s: Electiva

Identificación y características del Espacio Curricular

Carrera/s:	Licenciatura en Ciencias de la Computación		
Plan de Estudios:	2010, TO2024	Carácter:	Electiva
Bloque/Campo:		Área:	Algoritmos y Lenguajes
Régimen de cursado:	Cuatrimestral		
Cuatrimestre:	- [LCC], 1º [LCC]		
Carga horaria:	75 hs. / 5 hs. semanales	Formato curricular:	Asignatura
Escuela:	Ciencias Exactas y Naturales	Departamento:	
Docente responsable:	GRANITTO, Pablo		

Programa Sintético

Introducción. Preprocesamiento y Extracción de Características. Métodos no supervisados. Clasificadores avanzados.

Espacios Curriculares Relacionados

Previos Aprobados:

Simultaneos Recomendados:

Posteriores:

Vigencia desde 2024

Firma Profesor

Fecha

Firma Aprob. Escuela

Fecha

Con el aval del Consejo Asesor:

Fundamentación

La minería de datos o ciencia de datos es una de las tecnologías claves de la época actual. En este curso proponemos dar un vistazo inicial a las distintas técnicas que se usan en esta área, discutiendo las bases conceptuales que guían el desarrollo y los problemas generales que se tratan de resolver, dando una mirada a las tecnologías actuales.

Esta es una materia optativa que se realiza en el último año de cursado, y en el que se continua la formación en el área de la inteligencia artificial agregando conceptos fundamentales introducidos en Métodos numéricos, Probabilidad y estadística y Modelos Físicos, y complementando los fundamentos de Aprendizaje Automatizado.

Resultados del aprendizaje

Al finalizar el cursado los/las estudiantes serán capaces de:

RA1 Reconocer y conceptualizar fundamentos específicos de la ciencia de datos.

RA2 Comprender el proceso integral de la minería o ciencia de datos.

RA3 Realizar y reportar análisis de datos reales.

Competencias / Ejes transversales y Resultados del Aprendizaje

Competencia/Eje transversal al que tributa	Nivel	Resultados del Aprendizaje
CGT1-Identificación, formulación y resolución de problemas de informática	Medio	RA1
CGT4-Utilización de técnicas y herramientas de aplicación en la informática	Alto	RA2
CGS2-Fundamentos para la comunicación efectiva	Medio	RA3

Programa Analítico

1. Introducción

- 1.1. Formulación general del problema de minería de datos.
- 1.2. Historia y desarrollo.
- 1.3. Ejemplos introductorios.
- 1.4. Introducción al entorno R. Librerías específicas

2. Preprocesamiento y Extracción de Características

- 2.1. Introducción.
- 2.2. Pre- y post- procesado de datos. Normalización y codificación de la información. Datos faltantes.
- 2.3. Visualización de datos
- 2.4. Proyecciones en subespacios: lineales (PCA) y no-lineales (ICA). Ejemplos de aplicación.
- 2.5. Selección de variables. Filtros y Wrappers. Métodos embebidos.

3. Métodos No Supervisados

- 3.1. Introducción a Clustering.
- 3.2. Métodos aglomerativos. K-means.
- 3.3. Métodos jerárquicos.
- 3.4. Otros métodos

3.5. Gap y estabilidad de soluciones.

4. Clasificadores avanzados

4.1. Métodos de ensamble: Introducción. Filosofía y taxonomía de los ensambles.

4.2. Modelos jerárquicos: Cascadas.

4.3. Modelos promediados: Bagging y Random Forest.

4.4. Modelos aditivos: Boosting.

4.5. Métodos de kernel: Introducción.

4.6. Support Vector Machines: deducción.

4.7. Kernels

4.8. Extensiones: problemas de muchas clases, regresión y detección de novedades.

Modalidades de enseñanza

Clases teóricas y prácticas en laboratorio de computación para la primera parte. Trabajos prácticos individuales extensivos de lo tratado en clase.

Recursos

Se utiliza un laboratorio informático para el desarrollo de la actividad. Para las clases teóricas se utiliza un proyector.

El material de trabajo teórico/práctico se pone a disposición de los alumnos en la plataforma MOODLE de la facultad.

Actividades de Formación Práctica

Nº	Título	Descripción
1	TP1 Programación en R	En R, crear generadores de datos apropiados para ejemplos. Aplicar y evaluar clasificadores. Familiarizarse con notebooks.
2	TP2 Selección de variables	Implementar métodos de selección de variables, evaluarlos, proyectar datos con PCA.
3	TP3 Clustering	Analizar datos con clustering, implementar métodos para descubrir la cantidad de clusters, evaluarlos
4	TP4 Clasificadores avanzados	Aplicar clasificadores avanzados, detectar sus límites, evaluarlos
5	TF Análisis	Realizar un análisis completo de un conjunto de datos reales.

Evaluación

El alumno entrega todos los trabajos prácticos al nivel mínimo requerido y completa una evaluación conceptual de los conceptos discutidos. Realiza una presentación oral individual de un tema avanzado.

Resultado de Aprendizaje Actividades/Modalidad de Enseñanza Modalidad de Evaluación

RA1	Actividades 1 a 6. Modalidad Teórico - práctica	Se evalúa la resolución del problema planteado, incluyendo la identificación de la situación y la elección y análisis del tipo de solución. Se evalúa la comprensión de los conceptos teóricos mediante un examen escrito conceptual.
RA2	Actividades 1 a 6. Modalidad Teórico - práctica	Se evalúa la utilización práctica de las herramientas introducidas y el análisis del resultado obtenido. Se evalúa la comprensión de los conceptos teóricos asociados mediante un examen escrito conceptual.
RA3	Actividades 1 a 6. Modalidad práctica	Se evalúa en los informes de prácticos la capacidad de comunicar resultados mediante escritura, figuras y gráficas. En la presentación individual se evalúa la capacidad de explicar conceptos avanzados.

Bibliografía básica

Autores (Apellido, Inicial nombre)	Año de edición	Título de la obra	Editorial o Revista	Ejemplares disponibles o sitio web
I. Witten y E. Frank	2005	Data Mining: Practical Machine Learning Tools and Techniques. Second Edition	Elsevier	1
Hastie, T y otros	2001	The Elements of Statistical learning	Springer	1

Bibliografía complementaria

Autores (Apellido, Inicial nombre)	Año de edición	Título de la obra	Editorial o Revista	Ejemplares disponibles o sitio web
------------------------------------	----------------	-------------------	---------------------	------------------------------------

Distribución de la carga horaria

Presenciales

Teóricas		32 Hs.
Prácticas	Formación Experimental	
	Resolución de Problemas vinculados a la Profesión	20 Hs.
	Resolución de Problemas y Ejercicios	20 Hs.
	Actividades de Proyecto y Diseño	
	Formación en la Práctica Profesional	
Evaluaciones		3 Hs.

Total 75 Hs.

Dedicadas por el alumno fuera de clase

Preparación Teórica 10 Hs.

Preparación Práctica 10 Hs.

Elaboración y redacción de informes, trabajos, presentaciones, etc. 10 Hs.

Total 30 Hs.

Cronograma de actividades

Semana	Unidad	Tema	Actividad
1	1	1.1 a 1.3	
2	1•	1.4	
3	1	Trabajo Práctico	1
4	2	2.1 a 2.3	
5	2	2.4	
6	2	2.5	
7	2	Trabajo Práctico	2
8	3	3.1 a 3.4	
9	3	3.5	
10	3	Trabajo Práctico	3
11	4	4.1 a 4.3	
12	4	4.4 a 4.8	
13	4	Trabajo Práctico	4
14	Todas	Presentaciones individuales	
15	Todas	Trabajo final	5