

Programa Minería de Datos

Agosto 2025

Página de Curso

Importante: la página oficial del curso se encuentra en

<https://github.com/MaxOjeda/IIC2433>.

Allí se puede acceder a la información de contacto del profesor y de los ayudantes.

Descripción

El desarrollo de la tecnología ha hecho que la mayoría de los datos almacenados de forma física ahora lo estén de forma digital. Esto permite extraer información de estos datos mediante algoritmos computacionales, ya sea para descubrir patrones, construir modelos de predicción o identificar anomalías. En este curso se revisa una batería de técnicas para lidiar con información a través de algoritmos computacionales: manejo de datos, construcción de clasificadores o predictores, clusterización y técnicas de transformación de datos orientadas al trabajo con datos semi-estructurados o multidimensionales.

Objetivo General

El objetivo de este curso es proporcionar al estudiante los elementos necesarios para comprender las principales teorías y prácticas en Minería de Datos. Al finalizar, el alumno será capaz de aplicar técnicas para extraer conocimiento desde distintas fuentes y tipos de datos, y contará con fundamentos teóricos para decidir qué herramienta utilizar, conociendo sus potencialidades y limitaciones. Finalmente, los estudiantes vivirán una experiencia real de aplicación de estas herramientas en un entorno realista.

Metodología

Las clases son los martes bloques 5-6 (K202). Usaremos el bloque de los días jueves (J:5) para las presentaciones de los proyectos u otras actividades excepcionales. Se avisarán por Canvas con anticipación. Las clases son teórico/prácticas, por lo que se recomienda llevar un notebook con Python o bien acceso a Google Collab. Las clases se apoyan en ejemplos de jupyter notebooks los cuales se encontraran en el github del curso-

Evaluación

El curso se evalúa mediante:

1. Tres tareas individuales.
2. Actividades formativas semanales.
3. Un proyecto final en grupos de cuatro personas, estructurado en tres etapas (Planificación, Avances y Entrega Final).

Tareas

Las tareas se publican los martes (horario de clases) y se entregan hasta el viernes de la semana siguiente (23:59 hrs).

Proyecto

El proyecto consta de tres etapas: Propuesta (10 %), Avance de proyecto (30 %) y Entrega Final (60 %). Los detalles sobre el proyecto se darán a conocer en las fechas contempladas en la planificación de la asignatura.

Actividades Formativas

Las actividades formativas se realizarán en clases. Consideran resolución de problemas cortos de la materia del curso. Las actividades se evalúan como logradas (L) o no logradas (N). Los resultados obtenidos en las actividades formativas se bonifican en la nota final de la asignatura en décimas. Una décima por actividad, son máximo 10 actividades al semestre dependiendo del avance del curso.

Cálculo de la nota final

Sea AF la suma de uno por cada actividad *Lograda*. Sean T_i las notas de las tres tareas y P_i las de las tres etapas del proyecto. Entonces:

$$\text{Nota Final} = 0,6 \left(\frac{T_1 + T_2 + T_3}{3} \right) + 0,04P_1 + 0,12P_2 + 0,24P_3 + AF$$

Criterio de Aprobación

Para aprobar el curso, es necesario aprobarlo de forma individual y grupal. Es decir:

- El promedio de las tareas debe ser igual o superior a 4, sin considerar los puntos de AF. Importante, la nota de T_3 debe ser mayor a 4 por si sola.
- El promedio ponderado de las entregas del proyecto debe ser igual o superior a 4.

Calendario del Curso

Semana	Tema	Evaluación
Semana 1	Introducción, Preprocesamiento	-
Semana 2	Distancia y Proximidad / PCA	Actividad
Semana 3	TSNE y UMAP	Actividad, Tarea 1
Semana 4	Regresión Lineal y Logística	Actividad
Semana 5	Vecinos más cercanos - KNN	Actividad, Entrega Tarea 1
Semana 6	Árboles de decisión	Actividad
Semana 7	Inferencia Causal	Actividad, Tarea 2
Semana 8	K-Means, DBSCAN y HDBSCAN	Actividad
Semana 9	Mixture of Gaussians	Actividad, Entrega Tarea 2
Semana 10	Reglas de Asociación	Actividad, Propuesta Proyecto
Semana 11	Redes Bayesianas	Actividad, Tarea 3
Semana 12	Presentaciones de Proyecto I	Entrega Avance Proyecto
Semana 13	Procesamiento de texto	Entrega Tarea 3
Semana 14	Embeddings de texto.	-
Semana 15	Libre	-
Semana 16	Presentaciones de proyectos II	Entrega Final Proyecto

Contenidos

1. Preparación y transformación de la información

- a) Preprocesamiento de datos
- b) Distancias y Análisis de Componentes Principales (PCA)
- c) TSNE y UMAP

2. Modelos de predicción, clasificación y explicación

- a) Regresión lineal y logística
- b) Vecinos más cercanos (K-NN)
- c) Árboles de decisión
- d) Inferencia causal

3. Clustering

- a) K-means, DBSCAN
- b) Modelos de mixturas de distribuciones
- c) Evaluación de clústeres

4. Información semi-estructurada

- a) Procesamiento y Representación de Texto
- b) Embeddings de Texto

Otros

El Departamento de Ciencias de la Computación mantiene una política de tolerancia cero frente a copias o plagios. Se recomienda revisar las políticas y penalidades vigentes. La universidad y la escuela están suscritas a un código de honor que incluye a profesores, ayudantes y estudiantes.

Reflexión sobre copias y plagios. Este curso permite un acercamiento al conocimiento teórico y práctico de técnicas reales que se utilizan tanto en la Minería de Datos como otras áreas y cursos. Realizar las tareas y actividades de forma consciente trae un gran beneficio propio. Los ejercicios están diseñados para que aprendas mientras te evaluamos. Siempre estaremos dispuestos a resolver tus dudas. ¡Aprovecha esta oportunidad para aprender!

Los canales oficiales de comunicación son las clases y la página web del curso. Se asume que toda la información entregada por ambos canales llega a todos los estudiantes.

La asistencia a clases es obligatoria. En caso de ausencia, es responsabilidad del alumno ponerse al día. No se eliminan evaluaciones, pero se aceptará rendir evaluaciones atrasadas dependiendo de cada caso.

Bibliografía mínima

1. Rajaraman, A. & Ullman, J.D. *Mining of Massive Datasets*. Cambridge University Press, 2011 (disponible en línea).
2. Aggarwal, C. *Data Mining: The Textbook*. Springer, 2015.
3. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, 2009.