



UNIVERSIDAD
NACIONAL
DE LA PLATA

Minería de Datos y Aprendizaje Automático

Año 2026

Carrera/ Plan:

Cientista de Datos en Organizaciones Plan 2024

Año: 3°

Régimen de Cursada: Semestral

Carácter: Obligatoria

Profesor responsable: Dr. Franco Ronchetti y Dr. Waldo Hasperué.

Correlativas: Taller de lenguajes, Matemática D e Inglés (Prueba lecto-comprensión y traducción).

Hs. semanales de teoría: 3

Hs. semanales de práctica: 3

FUNDAMENTACIÓN

La Minería de Datos y el Aprendizaje Automático constituyen dos de las áreas más relevantes dentro de la Inteligencia Artificial moderna. Ambas disciplinas proporcionan métodos y herramientas para descubrir patrones, generar modelos predictivos y extraer conocimiento útil a partir de grandes volúmenes de datos.

En la actualidad, organizaciones públicas y privadas generan enormes cantidades de información que pueden ser aprovechadas para apoyar la toma de decisiones, automatizar procesos y generar ventajas competitivas. La capacidad de transformar datos en conocimiento constituye una competencia fundamental para los profesionales en ciencia de datos.

Esta asignatura presenta los fundamentos teóricos y prácticos de las principales técnicas de Minería de Datos y Aprendizaje Automático, incluyendo métodos supervisados, no supervisados y técnicas modernas de aprendizaje profundo. El curso pone especial énfasis en la preparación de los datos, la evaluación rigurosa de los modelos obtenidos y la interpretación de los resultados alcanzados.

OBJETIVOS GENERALES

Introducir al alumno en los conceptos, métodos y herramientas fundamentales de la Minería de Datos y el Aprendizaje Automático. Se analizarán técnicas supervisadas y no supervisadas para la construcción de modelos predictivos y descriptivos, incorporando además los fundamentos de las redes neuronales profundas y sus aplicaciones actuales. Se buscará que los estudiantes desarrollen la capacidad de seleccionar, implementar, evaluar e interpretar modelos adecuados para distintos problemas reales de análisis de datos.



UNIVERSIDAD
NACIONAL
DE LA PLATA

RESULTADOS DE APRENDIZAJE

Al finalizar el curso, los estudiantes serán capaces de:

- Comprender los fundamentos de la Minería de Datos y del Aprendizaje Automático.
- Preparar conjuntos de datos mediante técnicas de limpieza, transformación y normalización.
- Aplicar técnicas supervisadas y no supervisadas para resolver problemas reales.
- Construir y evaluar modelos basados en árboles de decisión, reglas, agrupamiento y redes neuronales.
- Analizar críticamente los resultados obtenidos utilizando métricas apropiadas.

CONTENIDOS MINIMOS

- Introducción. Minería de Datos y Aprendizaje Automático. Definiciones básicas y conceptos fundamentales. Preparación de Datos. Operaciones de limpieza y transformación.
- Tareas y Métodos. Tareas predictivas y descriptivas. Aprendizaje supervisado y aprendizaje no supervisado.
- Árboles de decisión. Métricas de selección de atributos basadas en entropía. Poda y Sobreajuste. Construcción de árboles para grandes volúmenes de datos.
- Evaluación de modelos predictivos
- Reglas de clasificación. Reglas de asociación. Concepto de ítem frecuente.
- Métricas de calidad: soporte, cobertura, confianza, interés y convicción.
- Técnicas de Agrupamiento. Métricas de calidad del agrupamiento. Tipos de agrupamiento: Jerárquico, partitivo y probabilista.
- Aprendizaje Automático Profundo (Deep Learning). Redes Neuronales Artificiales.
- Arquitecturas de redes neuronales convolucionales (CNN) y recurrentes (RNN).
- Aplicaciones de redes neuronales profundas en visión por computadora y procesamiento del lenguaje natural.

PROGRAMA ANALÍTICO

- Introducción a la Inteligencia Artificial. Aprendizaje Automático y Minería de Datos. Definiciones y conceptos fundamentales. Metodología Crisp-DM y proceso KDD. Tipos de tareas. Aprendizaje supervisado y no supervisado.
- Preparación de Datos. Limpieza y transformación. Tratamiento de valores faltantes. Normalización, discretización y selección de atributos.
- Árboles de decisión. Entropía, ganancia de información, tasa de ganancia e índice Gini. Algoritmos ID3, C4.5 y CART. Poda y sobreajuste.



UNIVERSIDAD
NACIONAL
DE LA PLATA

-
- Ensamblados de clasificadores. Bagging, Random Forest y Boosting.
 - Evaluación de modelos predictivos. Conjuntos de entrenamiento y prueba. Validación cruzada. Matriz de confusión. Precisión, recall, sensibilidad, especificidad, F-measure y curvas ROC.
 - Técnicas de agrupamiento. Agrupamiento jerárquico, partición y probabilístico. Algoritmo k-means. Métricas de calidad.
 - Redes neuronales artificiales. Perceptrón multicapa. Algoritmo Backpropagation.
 - Aprendizaje Profundo. Redes neuronales profundas. Técnicas de regularización y optimización.
 - Redes Neuronales Convolucionales (CNN). Aplicaciones en visión por computadora y procesamiento de lenguaje natural.

BIBLIOGRAFÍA

- Ian H. Witten, Eibe Frank, Mark A. Hall. Data Mining: Practical Machine Learning Tools and Techniques, (Fourth Edition). Morgan Kaufmann. 2017. ISBN 978-0-12-804291-5.
- Mehmed Kantardzic. Data Mining. Concepts, Models, Methods, and Algorithms. (Third Edition). Wiley. IEEE Press. 2020. ISBN: 978-1-119-51604-0.
- François Chollet. Deep Learning with Python (Second Edition). Manning Publications. 2021.
- Andreas C. Müller, Sarah Guido. Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media; 1er edición. 2017.
- Mohammed Zaki, Wagner Meira, Jr. Data Mining and Machine Learning. Fundamental Concepts and Algorithms (Second Edition). Cambridge University Press. 2020. ISBN: 978-1-108-47398-9.
- Jiawei Han, Jian Pei, Hanghang Tong. Data Mining: Concepts and Techniques, (Fourth Edition). Morgan Kaufmann. 2022. ISBN- 978-0-128-11761-3.

METODOLOGÍA DE ENSEÑANZA

La asignatura tiene como objetivo central capacitar al alumno en el uso de técnicas de Minería de Datos y Aprendizaje Automático para la construcción, evaluación e interpretación de modelos aplicados a problemas reales.

Esta articulación se llevará a cabo mediante la presentación de conceptos teóricos acompañados por la resolución de ejercicios prácticos y el desarrollo de aplicaciones concretas utilizando herramientas actuales de aprendizaje automático.

Durante el curso se promoverá el uso responsable y crítico de modelos de Inteligencia Artificial Generativa como apoyo al proceso de enseñanza y aprendizaje. Los estudiantes podrán utilizar estas herramientas para explorar alternativas de solución, analizar resultados, documentar procesos y asistir el desarrollo de implementaciones, comprendiendo sus alcances y limitaciones.



UNIVERSIDAD
NACIONAL
DE LA PLATA

ACTIVIDADES PRÁCTICAS

Durante el desarrollo del curso se propondrán ejercicios prácticos orientados a afianzar los conceptos abordados en clase y la elaboración de un trabajo práctico integrador grupal, cuyo objetivo será aplicar de manera conjunta las técnicas estudiadas a un problema concreto de Minería de Datos y Aprendizaje Automático.

METODOLOGÍA DE EVALUACIÓN

Cada alumno podrá aprobar la asignatura cumpliendo los siguientes requisitos:

- Entregar y aprobar con una nota mayor o igual a 6 (seis) puntos un trabajo práctico integrador grupal.
- Aprobar con una nota mayor o igual a 6 (seis) puntos un examen individual escrito.

El trabajo práctico integrador deberá incluir la preparación de los datos, construcción de modelos, evaluación de resultados e interpretación de las soluciones obtenidas. Asimismo, se valorará la correcta utilización de herramientas modernas de Inteligencia Artificial como apoyo al proceso de desarrollo.

El examen individual escrito permitirá verificar la comprensión conceptual de los temas desarrollados durante el curso y la capacidad de análisis y resolución de problemas por parte de cada estudiante. El examen contará con dos (2) instancias de recuperación.

Se otorgará el final de la asignatura a aquellos alumnos que hayan aprobado las dos instancias de evaluación, siendo el promedio de ambas la nota final de promoción.

Recordar que para poder pasar en SIU-GUARANI el final de la asignatura el alumno debe tener aprobado el examen final de las materias correlativas.

CRONOGRAMA

Semana		Contenidos/Actividades
1	26-ago	Introducción a la IA. Tipos de IA. Historia.
2	02-sep	Aprendizaje Automático y Minería de Datos. Metodologías. Evaluación de modelos. Preparación de los datos.
3	09-sep	Árboles de decisión.
4	16-sep	Ensamblados.
5	23-sep	Técnicas de Machine Learning para extracción de características, balance de clases y modelos de recomendación.
6	30-sep	Aprendizaje No Supervisado. K-means. Coeficiente Silhouette. DBSCAN. Reducción de dimensionalidad. PCA y t-SNE.
7	07-oct	Redes neuronales. Regresión lineal, Regresión logística.



UNIVERSIDAD
NACIONAL
DE LA PLATA

8	14-oct	Redes neuronales. Segunda parte. Métricas de clasificación y regresión.
9	21-oct	Deep Learning. Redes Convolucionales. Visión por computadora.
10	28-oct	Deep Learning. Segunda parte. Autoencoders. Transformers. Problemas de NLP. Modelos Generativos.
11	04-nov	1ra. Fecha de examen
12	11-nov	Consultas y muestra de exámenes de la 1ra. Fecha
13	18-nov	2da. Fecha de examen
14	25-nov	Consultas y muestra de exámenes de la 2da. Fecha
15	2-dic	3ra. Fecha de examen
16	9-dic	Muestra de exámenes de la 3ra. Fecha

Contacto

Toda la información de la cátedra, durante la cursada, se refleja en el entorno virtual de enseñanza y aprendizaje Moodle.

Los docentes de la cátedra atenderán consultas durante las clases.

Email de contacto: fronchetti@lidi.info.unlp.edu.ar

Franco Ronchetti

Waldo Hasperué