



Ingeniería de Datos

Agosto – Diciembre de 2025

Facultad de Ingeniería y Tecnologías

Inteligencia Artificial y Ciencia de Datos

DESARROLLO DEL CURSO	3
OBJETIVOS	3
RESULTADOS ESPERADOS DEL APRENDIZAJE	3
ESTRUCTURA GENERAL DEL CURSO Y METODOLOGÍA.....	5
Framework pedagógico integrado:	5
Estructura típica de clase	5
Unidades temáticas	5
EVALUACIÓN DEL CURSO	6
ESCALA DE NOTAS	7
COMPONENTES DE LA CALIFICACIÓN DEL CURSO	7
CRONOGRAMA.....	8
INSTRUCCIONES GENERALES DE LA MATERIA.....	9
HORARIOS Y ASISTENCIA.....	9
USO DEL COMPUTADOR PORTÁTIL	10
PRUEBAS INDIVIDUALES Y EN EQUIPO.....	10
TRABAJOS DE APLICACIÓN, PORTAFOLIOS Y EJERCICIOS DOMICILIARIOS.....	10
CONSIDERACIONES SOBRE EL USO DE LA WEB ASIGNATURA.....	11
HONESTIDAD ACADÉMICA.....	12
BIBLIOGRAFÍA	13
BIBLIOGRAFÍA BÁSICA.....	13

DESARROLLO DEL CURSO

Los datos son esenciales en cualquier proyecto. Este curso se centra en los aspectos fundamentales de su preparación, tratamiento y representación, tanto en forma individual como en forma integral para diferentes aplicaciones de Inteligencia Artificial. Se recomienda haber aprobado Probabilidad y Estadística para un mejor aprovechamiento de esta actividad formativa.

OBJETIVOS

Al finalizar el curso, el estudiante será capaz de dominar las técnicas de preparación y limpieza de datos para su uso en modelos de Inteligencia Artificial, explorar diversas formas de representar datos y determinar las más convenientes para aplicaciones prácticas en IA.

RESULTADOS ESPERADOS DEL APRENDIZAJE

- Emplear herramientas y métodos para representar datos de manera efectiva en proyectos de Inteligencia Artificial
- Analizar conjuntos de datos especiales y determinar las estrategias más adecuadas para su tratamiento y modelado
- Diseñar un proceso completo de preparación y tratamiento de datos para un proyecto de ciencia de datos
- Evaluar la calidad de los datos en conjuntos de datos reales y proponer mejoras
- Desarrollar modelos de datos especializados para aplicaciones de Inteligencia Artificial específicas.

ESTRUCTURA GENERAL DEL CURSO Y METODOLOGÍA

El curso de Ingeniería de Datos tiene un enfoque aplicado y por competencias, abordando el tratamiento de datos desde problemas organizacionales reales y midiendo el éxito por el impacto y la calidad de las soluciones, no solo por métricas técnicas.

Framework pedagógico integrado:

- Clase invertida + lecturas previas: El estudiante llega preparado para aplicar la teoría en clase.
- ≠ Team-Based Learning (TBL): Equipos estables de 3–4 estudiantes, con IRAT (Individual Readiness Assurance Test) y TRAT (Team Readiness Assurance Test) al inicio de cada unidad.
- ≠ Problem-Based Learning (PBL): Cada unidad se articula en torno a un caso organizacional (retail analytics, fraude, control de calidad, IoT, streaming, etc.).
- ≠ Portfolio profesional: Documentación continua en GitHub con notebooks, código, documentación técnica en MkDocs y reflexiones.
- ≠ Hands-on continuo: Más del 60% de cada clase se dedica a implementación práctica y desarrollo de pipelines reales.

Estructura típica de clase

1. Al inicio de cada unidad
 - a. IRAT
 - b. TRAT
2. Mini-lecture
3. Prácticas
 - a. Práctica guiada
 - b. Práctica autónoma
 - c. Desafío con continuación fuera del aula

Unidades temáticas

- UT1 – EDA & Fuentes de Datos (Clases 1–3)
- UT2 – Calidad de Datos & Ética (Clases 4–6)
- ≠ UT3 – Feature Engineering (Clases 7–9)
- ≠ UT4 – Datos Especiales (Clases 10–12).
- ≠ UT5 – Pipelines ETL & DataOps (Clases 13–15)
- ≠ Defensa Final (Clase 16)

EVALUACIÓN DEL CURSO

Filosofía. Evaluación continua que combina preparación individual, colaboración efectiva, aplicación práctica y comunicación profesional.

Distribución total (100%)

- **IRAT (quizzes individuales): 10%**
- **TRAT (quizzes en equipo): 5%**
- **Prácticas estructuradas en clase: 10%**
- **Participación + Peer evaluation: 5%**
- **Portfolio de proyectos: 60%**
- **Defensa oral: 10%**

Formativa vs Sumativa

- **Formativa (30%):** IRAT + TRAT + Prácticas + Participación/Peer.
- **Sumativa (70%):** Portfolio + Defensa oral.

Rubricas y criterios

- **IRAT/TRAT:**
 - comprensión conceptual
 - aplicación técnica
 - relevancia organizacional.
- **Prácticas:**
 - funcionalidad
 - comprensión
 - colaboración.
- **Portfolio:**
 - implementación técnica
 - calidad de documentación
 - relevancia organizacional
 - crecimiento progresivo durante el curso
- **Defensa oral:**
 - claridad técnica
 - conexión con negocio
 - demo efectiva
 - manejo de preguntas.

ESCALA DE NOTAS

La escala de notas a usarse en cada instancia de calificación será la oficial usada en la Universidad Católica:

- a) Deficiente (1)
- b) Regular (2)
- c) Bueno (3)
- d) Bueno Muy Bueno (4)
- e) Muy Bueno (5)
- f) Sobresaliente (6)

Para la correspondencia entre puntajes y notas se utilizarán los siguientes rangos:

%	Nota
0-30	D
31-60	R
61-74	B
75-86	BMB
87-93	MB
94-100	S

COMPONENTES DE LA CALIFICACIÓN DEL CURSO

Para aprobar en primera instancia se requiere:

1. **Calificación final \geq B** (Bueno).
2. **Asistencia \geq 75%** de las clases efectivamente dictadas.
3. **Portfolio completo** (contenido de 5 unidades + proyecto integrador).
 - a. El contenido de cada unidad sera evaluado al final de la misma.
4. **Defensa oral realizada exitosamente.**
5. **Peer evaluation sin problemas severos.**

CRONOGRAMA

Clases	Unidad	Resultados de Aprendizaje Esperados
13-ago 20-ago 27-ago	UT1: EDA & Fuentes	<p>Al completar UT1, el estudiante será capaz de:</p> <p>Cargar y explorar datasets de diferentes formatos (CSV, JSON, SQLite)</p> <p>Aplicar técnicas básicas de EDA con pandas y PySpark</p> <p>Crear visualizaciones informativas con matplotlib/seaborn</p> <p>Documentar hallazgos usando MkDocs (para portafolio) y mejores prácticas</p> <p>Interpretar resultados de análisis exploratorio</p> <p>Configurar entornos de desarrollo colaborativo con GitHub</p> <p>Comparar performance entre Pandas y PySpark para datasets grandes</p>
03-sep 10-sep 17-sep	UT2: Calidad & Ética	<p>Al completar UT2, el estudiante será capaz de:</p> <p>Distinguir entre MCAR, MAR y MNAR en datasets reales</p> <p>Detectar patrones de missing data y outliers</p> <p>Aplicar estrategias de imputación apropiadas según el contexto</p> <p>Implementar pipelines de limpieza reproducibles</p> <p>Prevenir data leakage usando validación cruzada apropiada</p> <p>Identificar y mitigar sesgo en datasets históricos</p> <p>Evaluar fairness usando métricas estándar</p> <p>Documentar decisiones éticas en el tratamiento de datos</p>
24-sep 01-oct 15-oct	UT3: Feature Engineering	<p>Al completar UT3, el estudiante será capaz de:</p> <p>Crear features derivadas relevantes según el dominio</p> <p>Aplicar técnicas avanzadas de encoding categórico</p> <p>Manejar variables de alta cardinalidad efectivamente</p> <p>Implementar PCA para reducción dimensional</p> <p>Interpretar componentes principales y varianza explicada</p> <p>Fusionar features temporales usando ventanas deslizantes</p>

		Versionar datasets y cambios con DVC Construir pipelines de feature engineering escalables
22-oct 29-oct 05-nov	UT4: Datos especiales	Al completar UT4, el estudiante será capaz de: Manipular datos geoespaciales con GeoPandas Crear y analizar grafos con NetworkX Extraer features de audio (MFCC) e imagen (keypoints) Procesar datos de sensores IoT Diseñar esquemas de data warehouse Implementar agregaciones temporales eficientes Construir pipelines para datos no estructurados
12-nov 19-nov 26-nov	UT5: Pipelines ETL & Spark	Al completar UT5, el estudiante será capaz de: Diseñar arquitecturas ETL escalables Implementar pipelines con Airflow y PySpark Manejar streaming de datos en tiempo real Aplicar principios de DataOps y CI/CD Escribir tests de calidad de datos Dockerizar pipelines para reproducibilidad Orquestar workflows complejos de datos
03-dic	Defensa Final	

INSTRUCCIONES GENERALES DE LA MATERIA

HORARIOS Y ASISTENCIA.

- Miercoles 08:00–10:50.
- Se exige **75% de asistencia** mínima.
 - Se pasa lista al 08.00

USO DEL COMPUTADOR PORTÁTIL

- Cada **equipo** debe contar con **al menos 2 laptops** configuradas.

- Requisitos mínimos: Python 3.10+, VS Code, Git/GitHub, Docker (para UT5), acceso a Google Colab u otros recursos de cómputo cuando aplique.
- Traer el equipo **siempre** con el entorno listo para ejecutar notebooks y scripts del curso.

PRUEBAS INDIVIDUALES Y EN EQUIPO.

- **IRAT**: quiz individual al inicio.
- **TRAT**: el mismo quiz resuelto en **equipo**, con discusión y justificación.
- Las pruebas comienzan **puntualmente**. Si ingresas tarde, tendrás el tiempo restante hasta el cierre.
- Plataforma: **WebSignatura**. Trae tu laptop configurada.

TRABAJOS DE APLICACIÓN, PORTAFOLIOS Y EJERCICIOS DOMICILIARIOS

Los trabajos de aplicación en Equipos son la parte más importante del proceso de aprendizaje de cada Unidad Temática. Los mismos serán de diferentes índoles, procurando alcanzar mediante ellos los Resultados Esperados del Aprendizaje (REA) definidos para el Curso y para cada Unidad Temática.

El objetivo del Portafolios es demostrar, que has alcanzado las capacidades indicadas en los objetivos del curso, es decir, que has logrado los Resultados Esperados del Aprendizaje – REA -del curso y de cada Unidad Temática.

Tanto los Trabajos de Aplicación como los Ejercicios Domiciliarios están diseñados alineados con esos REA, y por tanto también constituyen una evidencia de haberlos alcanzado.

Es por tanto de la máxima importancia que tú y cada miembro de tu equipo se prepare de la mejor manera posible, antes de concurrir a la clase, a efectos de lograr el mayor éxito y aprendizaje en cada Trabajo de Aplicación.

Todas las semanas te proveeremos de diferentes ejercicios domiciliarios y lecturas o fuentes de información adicionales, que te serán útiles para prepararte para trabajar en equipo con tus compañeros de la forma más efectiva y lograr mayores niveles en las destrezas objetivo.

Una parte significativa de la evaluación del Portafolios está relacionada a las evidencias presentadas que permitan apreciar que has desarrollado esas capacidades. Entre estas evidencias, los ejercicios en clase como practicas domiciliarias, si bien no son exclusivos, sí los consideramos muy importantes.

En todo momento podrás realizar consultas sobre todos los trabajos prácticos, utilizando a tales efectos los foros correspondientes de cada Unidad Temática.

CONSIDERACIONES SOBRE EL USO DE LA WEB ASIGNATURA

La Web asignatura de nuestro curso contiene todos los materiales puestos a tu disposición por los docentes, y también es la plataforma donde hemos de realizar muchas de las pruebas de preparación, parciales y entregas de trabajos.

El foro de anuncios generales de la web asignatura es el medio de comunicación básico que tenemos para relacionarnos contigo continuamente.

NOTA: si tienes consultas por temas administrativos, de agenda, gestión o personales, preséntalos personalmente a los docentes (o a las autoridades correspondientes de la Carrera).

HONESTIDAD ACADÉMICA

Para la realización de tu portafolio, proyecto y presentaciones, tanto en equipo como individuales, debes tener especialmente en cuenta que se espera un producto desarrollado exclusivamente por ti y no algo copiado o desarrollado por otra persona.

Si presentaras como propio un trabajo copiado o realizado por otra persona, además de no servirte en absoluto para alcanzar los resultados esperados del aprendizaje, estarías cometiendo una deshonestidad académica grave.

Vale la pena transcribir algunos párrafos de lo que a ese respecto expresa el folleto “HONESTIDAD ACADÉMICA: RESPONSABILIDAD DE TODOS” elaborado por la Universidad Católica:

“Una Universidad es por esencia un ámbito de libertad y búsqueda de la verdad, por esto el respeto a la honestidad académica es un principio y un valor que debe ser defendido en toda circunstancia. Los universitarios continuamente leen, analizan, debaten ideas de varios autores que luego toman en cuenta al elaborar trabajos académicos.”

“...Reconocer y valorar el trabajo de otros, no apropiarse indebidamente del pensamiento o la creación de terceros, es la base que permite producir y compartir conocimientos en una comunidad académica...”

“...Más allá de las disposiciones reglamentarias, compete a todos preservar estos valores y hacer posible que nuestra Universidad forme profesionales y personas éticamente insobornables y socialmente responsables...”

Recuerda: puedes consultar a los profesores en cualquier momento sobre dudas acerca de cómo citar, transcribir o parafrasear correctamente trabajos, textos o imágenes de terceros.

BIBLIOGRAFÍA

Toda la bibliografía y recursos necesarios (lecturas, videos, ejemplos, ejercicios, recursos web, etc.) para la preparación de cada Unidad Temática será indicada en la sección correspondiente de cada tema en la Webasignatura.

En términos generales, se utilizarán los siguientes materiales:

BIBLIOGRAFÍA BÁSICA

- Brust, A. V. (s. f.). Ciencia de Datos para Gente Sociable. (2023), de https://bitsandbricks.github.io/ciencia_de_datos_gente_sociable/
- Leskovec, J., Rajaraman, A., Ullman, J. (2023). Mining of Massive Datasets. <http://www.mmds.org/>
- Murty, M. N., Avinash, M. (2023). Representation in Machine Learning. Springer Nature. <https://doi.org/10.1007/978-981-19-7908-8>
- Zheng, A., Casari, A. (2018). Feature Engineering for Machine Learning [Book]. O'Reilly Media