

ADEN UNIVERSITY
FACULTAD DE CIENCIAS EMPRESARIALES
PLANIFICACIÓN ACADÉMICA - PROGRAMA DE GRADO

Docente:	<u>Ulises Gonzalez</u>	Identificación Personal:	<u>CI 80760404</u>
Período Lectivo:	<u>2025</u>	Modalidad:	<u>online</u>

DATOS GENERALES DE LA ASIGNATURA

Nombre de Asignatura:	<u>Machine Learning</u>	Horas Teóricas:	<u>3</u>
Código de Asignatura	<u>CT TD 100</u>	Horas Prácticas:	<u>0</u>
Total, Créditos:	<u>3</u>	Total, Horas	<u>3</u>

DESCRIPCIÓN

Esta asignatura introduce los fundamentos del Machine Learning, explorando su concepto, las necesidades que aborda, las soluciones que ofrece y las plataformas más relevantes utilizadas en este campo. El enfoque principal del curso radica en el desarrollo de técnicas que permitan a los sistemas informáticos aprender de datos para identificar patrones y mejorar su rendimiento a través de la experiencia acumulada. Los estudiantes trabajarán con escenarios y bases de datos reales, lo que les permitirá aplicar los conceptos aprendidos en proyectos de gran relevancia práctica y estratégica.

A lo largo del curso, se enfatizará la aplicación de estos conocimientos en proyectos que resuelvan problemas del mundo real. Los participantes no sólo comprenderán la teoría detrás del Machine Learning, sino que también adquirirán las habilidades prácticas necesarias para implementar soluciones efectivas. La integración de escenarios y bases de datos auténticas asegura que los estudiantes desarrollen una comprensión profunda de cómo el Machine Learning puede ser utilizado para generar valor y tomar decisiones informadas en diversos contextos empresariales y tecnológicos.

OBJETIVO GENERAL:

Diseñar soluciones de Machine Learning para resolver escenarios reales, considerando las limitaciones del problema y los campos adyacentes.

OBJETIVOS ESPECÍFICOS:

- **Identificar** los principios del Machine Learning, considerando herramientas, limitaciones y fuentes de información asociadas.
- **Analizar** las etapas para la creación de modelos, el rol del Deep Learning y los métodos de trabajo en proyectos de IA.
- **Utilizar** plataformas y procesos de Machine Learning como Python y Scikit-Learn para la implementación de algoritmos.
- **Evaluar** los fundamentos de distintos clasificadores y modelos de regresión para el análisis y la minería de datos.
- **Crear** un proyecto de Machine Learning de principio a fin para la toma automatizada de decisiones, encontrando patrones en los datos y prediciendo situaciones

CONTENIDOS DE LA ASIGNATURA:

MÓDULO I. Principios del Machine Learning

1. Desafíos para resolver.
2. Herramientas y conceptos.
3. Delimitaciones.
4. Fuentes de información.

MÓDULO II. Características del Machine Learning y de la (IA)

1. Herramientas estadísticas para el análisis de datos
2. Supervisado vs. No Supervisado.
3. Etapas para la creación de modelos.
4. El lenguaje Natural (NLP).

MÓDULO III. Plataformas y procesos

1. Introducción a Python.
2. Introducción a Rstudio
3. Introducción a Matlab.
4. Introducción a RapidMiner

MÓDULO IV. Fundamentos de clasificadores en RapidMiner

1. Clasificador probabilístico Naive Bayes.
2. Regresión logística.
3. Árboles de decisión.

4. Deep Learning.

MÓDULO V. Productos

1. Planificación de un proyecto de Machine Learning o de Ciencia de datos
2. Implementación de un proyecto de Machine Learning o Ciencia de datos
3. Evaluación de un proyecto de Machine Learning o de Ciencia de datos.
4. Rol de capital Humano en un proyecto de Ciencia de datos o en la transformación digital

METODOLOGÍA

Para alcanzar los objetivos propuestos, nos apoyaremos en una metodología de "aula invertida" y aprendizaje práctico (hands-on) que motive la interacción e integración del grupo. Las actividades incluyen:

- Exposiciones dialogadas y guiadas del docente.
- Análisis y comprensión de lecturas complementarias (papers).
- Análisis y resolución de casos prácticos y de estudio.
- Asignaciones asíncronas, soportadas en la plataforma de aprendizaje (Canvas).
- Realización de talleres grupales y laboratorios de codificación.

RECURSOS

- **Plataforma de aprendizaje:** Canvas
- **Material de Referencia:** Papers fundamentales, Módulos de Google Developers, Notebooks de GitHub.
- **Herramientas tecnológicas:** Python, Scikit-learn, Google Colab.
- **Tablero y Marcadores:** Para discusiones y diagramas en clase.

CRITERIOS DE EVALUACIÓN

De acuerdo con los requerimientos del Programa, la evaluación será integral y permanente; su fin primordial será el detectar e impulsar el progreso individual del estudiante y, por tanto, su dinamismo, su sentido de responsabilidad y compromiso en el desarrollo de actividades propuestas.

La evaluación estará compuesta de la siguiente forma:

–

Sesión	Criterio de Evaluación	Ponderación (%)
2	Análisis de Caso: Definición del pipeline para un proyecto End-to-End.	10%

3	Cuestionario: Diferencias entre aprendizaje supervisado y no supervisado.	10%
4	Laboratorio 1: Entrega de notebook con preprocesamiento de datos.	10%
5	Taller Práctico: Resolución de ejercicios de introducción a Scikit-Learn.	10%
6	Laboratorio 2: Implementación y análisis de modelos de regresión.	10%
7	Laboratorio 3: Implementación y comparación de modelos de clasificación.	10%
8	Laboratorio 4: Entrega de notebook con árboles de decisión y ensambles.	10%
9	Taller Práctico: Cálculo e interpretación de métricas de evaluación.	10%
10	Análisis de Resultados: Presentación de impacto de la validación cruzada.	10%
11	Proyecto Final: Entrega y sustentación del modelo optimizado.	10%
Total		100%

BIBLIOGRAFÍA

- Géron, A. (2022). ***Hands-on machine learning with Scikit-Learn, Keras & TensorFlow*** (3rd ed.). O'Reilly Media.
- Chollet, F. (2021). ***Deep learning with Python*** (2nd ed.). Manning Publications.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). ***Deep learning***. MIT Press.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). ***An introduction to statistical learning: With applications in R*** (2nd ed.). Springer.
- Bishop, C. M. (2006). ***Pattern recognition and machine learning***. Springer.
- Molnar, C. (2022). ***Interpretable machine learning: A guide for making black box models explainable***. Lulu.com.
- Burkov, A. (2019). ***The hundred-page machine learning book***. Andriy Burkov.
- Kleppmann, M. (2017). ***Designing data-intensive applications: The big ideas behind reliable, scalable, and maintainable systems***. O'Reilly Media.
- Sutton, R. S., & Barto, A. G. (2018). ***Reinforcement learning: An introduction*** (2nd ed.). MIT Press.
- Zheng, A., & Casari, A. (2018). ***Feature engineering for machine learning: Principles and techniques for data scientists***. O'Reilly Media.
- Management, Vol. 39, No. 6, pp. 491–502.
[https://doi.org/10.1016/S0378-7206\(01\)00120-3](https://doi.org/10.1016/S0378-7206(01)00120-3)

BIBLIOGRAFÍA AJUSTADA PARA CADA SESIÓN:

SESIÓN 1: ¿Qué es el ML? y el Pipeline del Proyecto

- **Lectura Principal (Paper):**
 - Sutton, R. S. (2019). **The Bitter Lesson**. *Incomplete Ideas Blog*.
(<http://www.incompleteideas.net/InclIdeas/BitterLesson.html>)
 - **Recurso de Google:**
 - Introducción al enmarcado de problemas de AA
(<https://developers.google.com/machine-learning/problem-framing/problem?hl=es-419>)
 - **Laboratorio Práctico (GitHub):**
 - Cap. 1: El panorama del Machine Learning
(https://github.com/gonzalezulises/handson-ml3/blob/main/01_the_machine_learning_landscape.ipynb)
 - **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulo 1).
-

SESIÓN 2: Sistemas y Datos (Enfoque Data-Centric)

- **Lectura Principal (Paper):**
 - Artículo conceptual de Andrew Ng sobre **Data-Centric AI**
(<https://www.deeplearning.ai/the-batch/a-chat-with-andrew-ng-about-data-centric-ai/>)
- **Recurso de Google:**
 - Preparación de datos e ingeniería de atributos
(<https://developers.google.com/machine-learning/data-prep>)
- **Laboratorio Práctico (GitHub):**
 - Cap. 2: Proyecto de ML de principio a fin
(https://github.com/gonzalezulises/handson-ml3/blob/main/02_end_to_end_machine_learning_project.ipynb)

- **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulo 2).
-

SESIÓN 3: Supervisado vs. No Supervisado

- **Lectura Principal (Paper):**
 - LeCun, Y. (2022). **A Path Towards Autonomous Machine Intelligence.**
OpenReview. (<https://openreview.net/pdf?id=BZ5a1r-kVsf>)
 - **Recurso de Google:**
 - Introducción al agrupamiento en clústeres
(<https://developers.google.com/machine-learning/clustering/overview?hl=es-419>)
 - **Laboratorio Práctico (GitHub):**
 - Cap. 9: Técnicas de aprendizaje no supervisado
(https://github.com/gonzalezulises/handson-ml3/blob/main/09_unsupervised_learning.ipynb)
 - **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulo 9).
-

SESIÓN 4: Ingeniería de Características

- **Lectura Principal (Paper):**
 - Mikolov, T., et al. (2013). **Efficient Estimation of Word Representations in Vector Space.** *arXiv preprint arXiv:1301.3781*.
(<https://arxiv.org/abs/1301.3781>)
- **Recurso de Google:**
 - Preparación de datos e ingeniería de atributos
(<https://developers.google.com/machine-learning/data-prep>)
- **Laboratorio Práctico (GitHub):**

- Cap. 13: Carga y preprocesamiento de datos
(https://github.com/gonzalezulises/handson-ml3/blob/main/13_loading_and_preprocessing_data.ipynb)
 - **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulo 13).
-

SESIÓN 5: Introducción a Scikit-Learn

- **Lectura Principal (Paper):**
 - Pedregosa, F., et al. (2011). **Scikit-learn: Machine Learning in Python**. *Journal of Machine Learning Research*, 12, 2825-2830.
(<https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>)
 - **Recurso de Google:**
 - Curso intensivo de aprendizaje automático
(<https://developers.google.com/machine-learning/crash-course?hl=es-419>)
 - **Laboratorio Práctico (GitHub):**
 - Cap. 2: Proyecto de ML de principio a fin
(https://github.com/gonzalezulises/handson-ml3/blob/main/02_end_to_end_machine_learning_project.ipynb)
 - **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulo 2).
-

SESIÓN 6: Modelos de Regresión

- **Lectura Principal (Libro):**
 - James, G., et al. (2021). **An introduction to statistical learning...** (Capítulo 3).
(<https://www.statlearning.com/>)
- **Recurso de Google:**
 - Curso intensivo de aprendizaje automático (Secciones de Regresión)
(<https://developers.google.com/machine-learning/crash-course?hl=es-419>)

- **Laboratorio Práctico (GitHub):**
 - Cap. 4: Entrenamiento de modelos lineales
(https://github.com/gonzalezulises/handson-ml3/blob/main/04_training_linear_models.ipynb)
 - **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulo 4).
-

SESIÓN 7: Algoritmos de Clasificación

- **Lectura Principal (Paper):**
 - Cover, T., & Hart, P. (1967). **Nearest neighbor pattern classification**. *IEEE Transactions on Information Theory*, 13(1), 21-27.
(<https://ieeexplore.ieee.org/document/4037264>)
 - **Recurso de Google:**
 - Clasificación de imágenes
(<https://developers.google.com/machine-learning/image-classification>)
 - **Laboratorio Práctico (GitHub):**
 - Cap. 3: Clasificación
(https://github.com/gonzalezulises/handson-ml3/blob/main/03_classification.ipynb)
 - **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulo 3).
-

SESIÓN 8: Árboles de Decisión y Ensamblados

- **Lecturas Principales (Papers):**
 - Breiman, L. (2001). **Random Forests**. *Machine Learning*, 45(1), 5-32.
(<https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>)
 - Chen, T., & Guestrin, C. (2016). **XGBoost: A Scalable Tree Boosting System**. *KDD*. (<https://arxiv.org/abs/1603.02754>)

- **Recurso de Google:**
 - Introducción a los bosques de decisión
(<https://developers.google.com/machine-learning/decision-forests?hl=es-419>)
 - **Laboratorio Práctico (GitHub):**
 - Cap. 6: Árboles de Decisión
(https://github.com/gonzalezulises/handson-ml3/blob/main/06_decision_trees.ipynb)
 - Cap. 7: Ensamblados y Random Forests
(https://github.com/gonzalezulises/handson-ml3/blob/main/07_ensemble_learning_and_random_forests.ipynb)
 - **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulos 6 y 7).
-

SESIÓN 9: Métricas de Evaluación

- **Lectura Principal (Paper):**
 - Davis, J., & Goadrich, M. (2006). **The Relationship Between Precision-Recall and ROC Curves.** *ICML*.
(https://www.researchgate.net/publication/220387544_The_Relationship_Between_Precision-Recall_and_ROC_Curves)
- **Recurso de Google:**
 - Pruebas y depuración en el aprendizaje automático
(<https://developers.google.com/machine-learning/testing-debugging>)
- **Laboratorio Práctico (GitHub):**
 - Cap. 3: Clasificación
(https://github.com/gonzalezulises/handson-ml3/blob/main/03_classification.ipynb)
- **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulo 3).

SESIÓN 10: Overfitting y Validación Cruzada

- **Lectura Principal (Paper):**
 - Srivastava, N., et al. (2014). **Dropout: A Simple Way to Prevent Neural Networks from Overfitting**. *JMLR*, 15(1), 1929-1958.
(<https://jmlr.org/papers/v15/srivastava14a.html>)
- **Recurso de Google:**
 - Curso intensivo de aprendizaje automático (Sección de Generalización)
(<https://developers.google.com/machine-learning/crash-course?hl=es-419>)
- **Laboratorio Práctico (GitHub):**
 - Cap. 2: Proyecto de ML de principio a fin
(https://github.com/gonzalezulises/handson-ml3/blob/main/02_end_to_end_machine_learning_project.ipynb)
- **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulo 2).

SESIÓN 11: Optimización y Proyecto Final

- **Lecturas Principales (Papers):**
 - Kingma, D. P., & Ba, J. (2014). **Adam: A Method for Stochastic Optimization**. *arXiv preprint arXiv:1412.6980*. (<https://arxiv.org/abs/1412.6980>)
 - Snoek, J., et al. (2012). **Practical Bayesian Optimization of Machine Learning Algorithms**. *NeurIPS*. (<https://arxiv.org/abs/1206.2944>)
- **Recurso de Google:**
 - Curso intensivo de aprendizaje automático (Sección de Producción)
(<https://developers.google.com/machine-learning/crash-course?hl=es-419>)
- **Laboratorio Práctico (GitHub):**
 - Cap. 2: Proyecto de ML de principio a fin
(https://github.com/gonzalezulises/handson-ml3/blob/main/02_end_to_end_machine_learning_project.ipynb)

- **Lectura de Apoyo (Libro):**
 - Géron, A. (2022). **Hands-on machine learning...** (Capítulo 2).

MÓDULO	OBJETIVOS GENERALES DE APRENDIZAJE	ACTIVIDADES DEL DOCENTE	ACTIVIDADES DEL ESTUDIANTE	RECURSOS	EVALUACIÓN Y CRITERIOS
I. Principios del Machine Learning ¿Qué es el ML? Pipeline del Proyecto, Paradigmas de IA. Sistemas de ML y Enfoque en Datos, estructura End-to-End	<ul style="list-style-type: none"> • Identificar la definición y componentes de un proyecto de ML. • Distinguir entre enfoque centrado en modelo y en datos. • Comprender el impacto de los paradigmas de IA. OBJETIVOS ESPECÍFICOS DE LA ACTIVIDAD INTEGRADORA <ul style="list-style-type: none"> • Identificar una oportunidad de negocio para aplicar ML. • Describir el pipeline preliminar para la solución. 	<ul style="list-style-type: none"> • Diseña y publica lecturas. • Elabora material audiovisual (videos, PPTs). • Diseña la actividad integradora. • Interactúa en la plataforma. 	<ul style="list-style-type: none"> • Lee el material asignado. • Visualiza el video y PPT. • Consulta en la plataforma. • Elabora el trabajo integrador. 	<ul style="list-style-type: none"> • Canvas: Aula Virtual, videos, papers, etc. • Lecturas: "The Bitter Lesson", "Data-Centric AI". • Videos: Módulo "Enmarcado de problemas de AA". • Lab: Notebooks Cap. 1 y 2. 	Análisis de Caso: El estudiante aplica los conceptos del Módulo I. <ul style="list-style-type: none"> • Identifica un problema resoluble con ML. • Distingue las fases del pipeline. • Comprende la importancia del enfoque en datos.
II. Características del ML y la IA Aprendizaje Supervisado vs.	OBJETIVOS GENERALES DE APRENDIZAJE <ul style="list-style-type: none"> • Reconocer las 	<ul style="list-style-type: none"> • Desarrolla lecturas y material audiovisual. • Confecciona el cuestionario. 	<ul style="list-style-type: none"> • Lee el material. • Visualiza video y PPT. • Responde el cuestionario. 	<ul style="list-style-type: none"> • Canvas: Papers, enlaces, notebooks. • Lecturas: "A Path 	<ul style="list-style-type: none"> • Cuestionario: Distingue tipos de aprendizaje • Laboratorio 1

<p>No Supervisado.</p> <p>Ingeniería de Características (limpieza, preprocesamiento).</p>	<p>diferencias entre aprendizaje supervisado y no supervisado. • Comprender las fases de preparación de datos. • Reconocer la importancia de la ingeniería de características.</p> <p>OBJETIVOS ESPECÍFICOS DE LA ACTIVIDAD INTEGRADORA</p> <p>• Aplicar técnicas de limpieza y preprocesamiento a un dataset. • Justificar las decisiones tomadas.</p>	<ul style="list-style-type: none"> • Diseña el Laboratorio 1. • Da feedback. 	<p>Elabora el Laboratorio 1.</p>	<p>Towards...", "Word2Vec". • Videos: Módulos "Agrupamiento", "Preparación de datos". • Lab: Notebooks Cap. 9 y 13.</p>	<p>Aplica preprocesamiento.</p> <ul style="list-style-type: none"> • Aplica funciones de limpieza. • Transforma variables. • Justifica sus decisiones.
<p>III. Plataformas y procesos</p> <p>Introducción a Scikit-Learn (API, flujo de trabajo).</p> <p>Modelos de Regresión (Lineal y Logística).</p>	<p>OBJETIVOS GENERALES DE APRENDIZAJE</p> <p>• Utilizar Scikit-Learn para construir un flujo de trabajo de ML. • Implementar y evaluar modelos de regresión.</p> <p>OBJETIVOS ESPECÍFICOS DE LA ACTIVIDAD</p>	<ul style="list-style-type: none"> • Elabora demos en vivo. • Diseña el Taller Práctico y Laboratorio 2. • Provee retroalimentación. 	<ul style="list-style-type: none"> • Lee la documentación. • Replica demos y completa el Taller. • Elabora el Laboratorio 2. 	<ul style="list-style-type: none"> • Canvas: Papers, enlaces, notebooks. • Lecturas: "Scikit-learn", Cap. 3 de "ISL". • Videos: Módulo "Curso intensivo de AA". • Lab: Notebooks Cap. 2 y 4. 	<ul style="list-style-type: none"> • Taller Práctico (10%): Resuelve ejercicios en Scikit-Learn. • Laboratorio 2): Implementa un modelo de regresión. • Utiliza .fit() y .predict() • Implementa regresión lineal y

	INTEGRADORA <ul style="list-style-type: none"> • Resolver ejercicios prácticos con la API de Scikit-Learn. • Construir un modelo de regresión completo. 				logística. • Interpreta coeficientes.
IV. Fundamentos de clasificadores Algoritmos de Clasificación (K-NN, Naive Bayes). Árboles de Decisión y Ensamblados (Random Forest, XGBoost)	OBJETIVOS GENERALES DE APRENDIZAJE <ul style="list-style-type: none"> • Reconocer el funcionamiento de algoritmos de clasificación. • Comprender el poder de los métodos de ensamble. OBJETIVOS ESPECÍFICOS DE LA ACTIVIDAD INTEGRADORA <ul style="list-style-type: none"> • Aplicar y comparar el rendimiento de múltiples clasificadores. • Implementar un ensamble y analizar su superioridad. 	<ul style="list-style-type: none"> • Elabora material audiovisual. • Diseña los laboratorios 3 y 4. • Modera un debate en clase. 	<ul style="list-style-type: none"> • Lee los papers. • Completa el Laboratorio 3. • Completa el Laboratorio 4. 	<ul style="list-style-type: none"> • Canvas: Papers, enlaces, notebooks. • Lecturas: "Nearest neighbor", "Random Forests", "XGBoost". • Videos: Módulos "Clasificación de imágenes", "Bosques de decisión". • Lab: Notebooks Cap. 3, 6 y 7. 	<ul style="list-style-type: none"> • Laboratorio 3: Compara el rendimiento de clasificadores. • Laboratorio 4: Implementa un Random Forest. • Construye árbol y ensamble. • Compara métricas. • Explica por qué el ensamble es superior.
V. Productos (Ciclo de Vida de un Proyecto)	OBJETIVOS GENERALES	<ul style="list-style-type: none"> • Diseña talleres y la rúbrica final. 	<ul style="list-style-type: none"> • Completa el taller de métricas. 	<ul style="list-style-type: none"> • Canvas: Papers, enlaces, 	<ul style="list-style-type: none"> • Taller Práctico: Calcula e interpreta

<p>Métricas de Evaluación.</p> <p>Overfitting y Validación Cruzada.</p> <p>Optimización (HPO) y Proyecto Final.</p>	<p>DE APRENDIZAJE</p> <ul style="list-style-type: none"> • Evaluar modelos de forma robusta con métricas adecuadas. • Diagnosticar y mitigar el sobreajuste. • Optimizar el rendimiento de un modelo. <p>OBJETIVOS ESPECÍFICOS DE LA ACTIVIDAD INTEGRADORA</p> <ul style="list-style-type: none"> • Seleccionar la métrica más apropiada. • Aplicar validación cruzada. • Presentar un modelo final optimizado. 	<ul style="list-style-type: none"> • Guía en la selección de métricas. • Organiza la presentación de proyectos. 	<p>Presenta análisis de validación.</p> <ul style="list-style-type: none"> • Entrega y sustenta su proyecto final. 	<p>notebooks.</p> <ul style="list-style-type: none"> • Lecturas: "Precision-Recall", "Dropout", "Adam", "Bayesian Optimization". • Videos: Módulos "Pruebas y depuración", "Curso Intensivo". • Lab: Notebooks Cap. 2 y 3. 	<p>métricas.</p> <ul style="list-style-type: none"> • Análisis de Resultados): Presenta impacto de la validación. • Proyecto Final : Entrega y sustenta el modelo. • Aplica el pipeline completo. • Justifica sus elecciones.
---	--	---	---	--	---