

CARRERA: Ciencias de Datos	CURSO LECTIVO: 2025
CÁTEDRA: Inteligencia Artificial y Aprendizaje Automático I	CURSO: 3º año - 1º semestre
DURACIÓN: Semestral	Hs. TOTALES: 64
SEMANAS: 15	Hs. TEÓRICAS: 32 Hs. PRÁCTICAS: 32

PROFESOR ASISTENTE: Rodriguez, Federico Martín

PROFESOR Pro-Titular: Di Pasquale, Ricardo Héctor

1. OBJETIVOS DE LA ASIGNATURA

Que los alumnos logren:

- Dominar los elementos fundamentales de las técnicas de Machine Learning.
- Comprender los alcances y límites de las técnicas de Machine Learning para aplicarlas de manera adecuada en el caso adecuado.
- Enmarcar en el ciclo de vida de Machine Learning los proyectos pertenecientes a las disciplinas de AI/ML.
- Dominar la implementación del Pipeline de ML.
- Comprender la dependencia y relevancia de los Datos en este tipo de paradigmas.
- Conocer la complejidad de los sistemas de Machine Learning.
- Estimar el esfuerzo necesario para implementar en producción un producto basado en AI/ML.
- Distinguir AI/ML en la Investigación Académica de AI/ML en Sistemas Productivos.
- Dominar las técnicas de preparación de datos para el uso en AI/ML.
- Dominar la disciplina de Ingeniería de Features y las distintas técnicas aplicadas.
- Dominar el proceso de entrenamientos de Modelos de ML para clasificación, regresión (supervisados) y clusterización (no supervisados).
- Conocer las métricas de evaluación de los modelos y saber elegir la técnica indicada.
- Conocer el uso de Feature Stores para el escalamiento de los procesos de ML.
- Conocer la importancia de la disciplina de MLOps.

2. UNIDADES TEMÁTICAS

I. Introducción.

Repaso de Machine Learning. Definición y conceptos fundamentales. Ejemplos de Aplicación. Taxonomía de métodos: Supervisado: regresión y clasificación; No supervisado: clustering y reducción de dimensionalidad; Aprendizaje por refuerzo: toma de decisiones en entornos dinámicos; Métodos híbridos y avanzados. Principales Desafíos en Machine Learning: Cantidad insuficiente de datos de entrenamiento; Datos de entrenamiento no representativos; Calidad pobre de los datos; Características (Features) irrelevantes; Sobreajuste (overfitting); Subajuste (underfitting). Principales frameworks: scikit-learn, pytorch, Tensorflow+Keras, XGBoost, LightGBM. Principales Plataformas y Nubes: Kubernetes+Kubeflow, AWS Sagemaker, GCP Vertex.AI, Azure ML, Databricks, Weight & Biases, H2O.AI.

II. Machine Learning Production Systems

¿Qué es el Machine Learning en Producción? Definición y objetivos del Machine Learning en entornos productivos. Diferencias entre investigación y producción en Machine Learning. Retos comunes en la implementación de modelos. Beneficios del Uso de ML Pipelines. Enfoque en el desarrollo de nuevos modelos y Minimización del mantenimiento de modelos existentes. Prevención de errores y depuración. Creación de registros: Trazabilidad para depuración y reproducción de resultados. Estandarización. El caso de negocio para los ML Pipelines. Cuando Utilizar ML Pipelines. Pasos de un pipeline: Ingesta y versionado de datos, Validación de datos, Verificación de calidad, consistencia y conformidad de los datos, Ingeniería de características (Features), Selección, entrenamiento y ajuste del modelo, Análisis del modelo, Despliegue del modelo.

III. Ciclo de vida de Machine Learning

ML vs software tradicional. Ciclo de vida de ML: Comprensión del problema de negocio y formulación del problema de ML; Comprensión y preparación de los datos; Entrenamiento y evaluación del modelo; Despliegue del modelo; Monitoreo del modelo; Seguimiento de métricas de negocio. Arquitectura de soluciones de ML: Comprensión del negocio y transformación mediante ML. Identificación y verificación de técnicas de ML. Diseño e implementación de la arquitectura del sistema. Automatización del flujo de trabajo en plataformas de ML. Seguridad y cumplimiento normativo.

IV. Ingeniería de Features

Introducción a la Ingeniería de Features. Relevancia. Qué NO puede hacer la ingeniería de Features. Importancia de los datos para el rendimiento de los modelos. El pipeline de la ingeniería de features. Validación de Datos: Detección de Data Issues. TensorFlow Data Validation. Manejo de Datasets desbalanceados. Data Imputation: Eliminación de observaciones con datos faltantes. Imputación utilizando la media o la mediana. Imputación de variables categóricas. Reemplazo de valores faltantes con un número arbitrario. Identificación de valores extremos para imputación. Marcado de valores imputados. Imputación multivariada mediante ecuaciones encadenadas. Estimación de datos faltantes utilizando vecinos más cercanos.

V. Ingeniería de Features: Codificación y Transformación

Codificación de variables categóricas: Creación de variables binarias mediante codificación one-hot; Codificación one-hot de categorías frecuentes; Reemplazo de categorías por conteos o frecuencias de observaciones; Reemplazo de categorías por números ordinales; Codificación ordinal basada en el valor objetivo; Target mean encoding; Codificación con el peso de la evidencia; Agrupación de categorías raras o poco frecuentes; Codificación binaria. Variable numérica: función logarítmica; función recíproca; raíz cuadrada; transformaciones de potencias; Box-Cox; Yeo-Johnson. Discretización de variables: equal-width; equal-frequency; en intervalos arbitrarios; con k-means clustering; binarización; con árboles de decisión. Features de Fecha y

Tiempo. Feature Scaling. Creación de nuevos features. Manejo de Outliers. Ingeniería de Features en Series de Tiempo: Generación de Features y Extracción automática.

VI. Reducción de la dimensionalidad en Ingeniería de Features

Reducción de dimensionalidad usando PCA. Reducción de la dimensionalidad utilizando Feature Extraction: Reducción con Principal Components; Reducción cuando los datos son linealmente inseparables; por Maximización de Class Separability; mediante el uso de factorización de matrices; Reducción en datos raros (sparse data). Reducción de la dimensionalidad utilizando Feature Selection: Numerical Feature Variance; Binary Feature Variance; manejando Features altamente co-relatados; Removiendo Features irrelevantes; Eliminando Features recursivamente.

VII. Introducción a MLOps

Machine Learning Engineering (MLE) y MLOps. Integración y Despliegue continuos (CI/CD). DevOps vs MLOps. DataOps. Automatización de la plataforma de ML. Casos de Estudio.

VIII. Feature Stores

Introducción a los Feature Stores. Problemas que atacan los Feature Stores. Necesidad y Filosofía de los Feature Stores. Integración de los Feature Stores con el Pipeline de ML. Variantes de Feature Stores. Caso práctico.

IX. Entrenamiento de Modelos (Supervisados)

Entrenamiento de modelos de Clasificación: Binarios; Multiclass; Multilabel; Multioutput. Métricas de Performance: Cross-Validation; Matrices de confusión; Precision y Recall; ROC Curve. Ajuste Precision/Recall. Análisis de Errores. Entrenamiento de modelos de Regresión: Lineal; Polinomial; Ridge; Lasso; Elastic Net; Regresión Logística. Gradient-Descent. Entrenamiento de SVM: Linear SVM Classification; Nonlinear SVM Classification; SVM Regression; Problema primal y dual; Kernelized SVM. Entrenamiento de Árboles de Decisión: CART; Gini impurity y entropía; Sensibilidad de la regresión a la orientación del eje; Varianza en Árboles de Decisión.

X. Ensamblados, Random Forest y Gradient-Boosted Decision Trees

Voting Classifiers. Bagging & Pasting; Random Forests; Boosting; AdaBoost; Gradient Boosting; Histogram-Based Gradient Boosting. Stacking. Caso práctico con LightGBM. Comparativa entre LightGBM, XGBoost y enfoques de Deep Learning.

XI. Entrenamiento de Modelos (No Supervisados)

Algoritmos de Clustering. k-means. Límites de k-means. Segmentación de imágenes con Clustering. Semi-Supervised Learning. DBSCAN. Gaussian Mixtures: Detección de anomalías. Selección del número de Clusters. Bayesian Gaussian Mixture Models. Otros algoritmos para la detección de anomalías y novedades.

XII. AutoML como Baseline de performance y Tuning de Hiperparámetros

AutoML: Caso práctico con FLAML y LightGBM. Introducción al tuning de hiperparámetros. Parámetros vs Hiper-Parámetros. Espacio de hiperparámetros y distribuciones. Búsqueda exhaustiva: búsqueda manual; grid y random. Optimización Bayesiana: BO; BO GP; SMAC; TPE; Metis. Búsquedas heurísticas: recocido simulado, algoritmos genéticos, particle swarm optimization. Caso práctico con scikit-learn.

3. BIBLIOGRAFÍA

3.1 BIBLIOGRAFÍA GENERAL OBLIGATORIA

- “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow” 3rd Edition - Aurélien Géron - O'Reilly Media, Inc. - 2022 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/hands-on-machine-learning/9781098125967/>
- “The Machine Learning Solutions Architect Handbook” Second Edition - David Ping - Packt Publishing - 2024 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/the-machine-learning/9781805122500/>
- “Machine Learning Production Systems” - Robert Crowe, Hannes Hapke, Emily Caveness, Di Zhu - O'Reilly Media, Inc. - 2024 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/machine-learning-production/9781098156008/>
- “Feature Engineering Bookcamp” - Sinan Ozdemir - Manning Publications - 2022 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/feature-engineering-bookcamp/9781617299797/>
- “Machine Learning with Python Cookbook” - 2nd Edition - Kyle Gallatin, Chris Albon - O'Reilly Media, Inc. - 2023 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/machine-learning-with/9781098135713/>
- “Python Feature Engineering Cookbook” Second Edition - Soledad Galli - Packt Publishing - 2022 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/python-feature-engineering/9781804611302/>
- “Practical MLOps” - Noah Gift, Alfredo Deza - O'Reilly Media, Inc. - 2021 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/practical-mlops/9781098103002/>
- “Feature Store for Machine Learning” - Jayanth Kumar M J - Packt Publishing - 2022 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/feature-store-for/9781803230061/>
- “Machine Learning with LightGBM and Python” - Andrich van Wyk - Packt Publishing - 2023 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/machine-learning-with/9781800564749/>
- “Hyperparameter Tuning with Python” - Louis Owen - Packt Publishing - 2022 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/hyperparameter-tuning-with/9781803235875/>
- “Machine Learning at Scale with H2O” - Gregory Keys, David Whiting - Packt Publishing - 2022 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/machine-learning-at/9781800566019/>
- “Practical Machine Learning on Databricks” - Debu Sinha - Packt Publishing - 2023 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/practical-machine-learning/9781801812030/>

- “Azure Machine Learning Engineering” - Sina Fakhraee, Balamurugan Balakreshnan, Megan Masanz - Packt Publishing – 2023 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/azure-machine-learning/9781803239309/>

3.2 BIBLIOGRAFÍA ADICIONAL

- "Feature Engineering for Machine Learning" - Alice Zheng, Amanda Casari - O'Reilly Media, Inc. - 2018 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/feature-engineering-for/9781491953235/>
- “Machine Learning: Make Your Own Recommender System” - Oliver Theobald - Packt Publishing – 2024 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/machine-learning-make/9781835882061/>
- “Architecting Data and Machine Learning Platforms” - Marco Tranquillin, Valliappa Lakshmanan, Firat Tekiner - O'Reilly Media, Inc. - 2023 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/architecting-data-and/9781098151607/>
- “MLOps Engineering at Scale” - Carl Osipov - Manning Publications – 2022 - Disponible en Biblioteca: <https://learning.oreilly.com/videos/mlops-engineering-at/9781617297762AU/>
- “Building Machine Learning Systems with a Feature Store” - Jim Dowling - O'Reilly Media, Inc. - 2025 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/building-machine-learning/9781098165222/>
- “Designing Machine Learning Systems” - Chip Huyen - O'Reilly Media, Inc. - 2022 - Disponible en Biblioteca: <https://learning.oreilly.com/library/view/designing-machine-learning/9781098107956/>

4. METODOLOGÍA

El curso está organizado en 12 unidades temáticas divididas en 2 encuentros de 3hs reloj semanales, a realizarse en formato presencial. La modalidad adoptada para el dictado será **teórico-práctica**. En las clases se presentarán los temas de cada unidad, proponiendo espacios de intercambio con el docente y entre los estudiantes a partir de consignas específicas. Se facilitará material de lectura obligatoria y complementaria para complementar la comprensión de las unidades.

5. EVALUACIONES Y CRITERIOS PARA LA APROBACIÓN

La aprobación de la materia estará supeditada al cumplimiento de la condición de asistencia exigida por la Universidad, la aprobación de todas las actividades prácticas, la aprobación de la instancia de Coloquio y la aprobación del examen integrador.

Los trabajos prácticos podrán ser individuales o grupales, debiéndose cargar a través de la plataforma de Entornos Virtuales de Aprendizaje en tiempo y forma, otorgándole una única

instancia de revisión y recuperación. Las actividades prácticas deberán contar con su aprobación para acceder a la instancia de evaluación final.

Para los trabajos prácticos y la evaluación final se realizarán sesiones de consultas individuales y grupales, haciendo además puesta en común general si el caso lo requiera. A los estudiantes que presenten dificultades se les observará y se los guiará para resolver el conflicto.

La instancia del coloquio es una modalidad tomada de materias de maestría en donde el profesor ofrece a los alumnos una serie de papers, capítulos de libros y otros recursos. Estos recursos son asignados entre los alumnos (individualmente o en grupos de a pares). El objetivo es que los alumnos puedan hacer suyo, entender, expandir, y comprender todas las aristas del artículo seleccionado. Esta elaboración se hace en colaboración con los docentes. El objetivo final es que los alumnos presenten en coloquio oral su síntesis. Esta síntesis debe ser comunicacionalmente efectiva (se espera que el resto de los alumnos use los apuntes obtenidos como material de estudio) y académicamente completa. Vale aclarar que los docentes van a evaluar tanto a los alumnos comprometidos con el coloquio, como a los alumnos asistentes al coloquio en función de sus aportes y preguntas.

La instancia de recuperación está prevista para estudiantes que no hayan aprobado el examen integrador, que hayan reprobado el coloquio, o que hayan estado ausentes.

Criterios de Evaluación:

- Respeto de las consignas presentadas.
- Resolución correcta de los problemas planteados.
- Adecuada respuesta a los contenidos teóricos.
- Participación en los coloquios (tanto expositiva como inquisitiva)

6. CRITERIOS y MODALIDAD PARA LA EVALUACIÓN DEL EXAMEN FINAL

El examen final consiste en una evaluación oral y escrita, presencial e individual, donde el alumno deberá demostrar conocimientos teóricos y prácticos. El examen final se diferencia en que abarca todos los temas del programa. Los ejercicios prácticos tendrán un carácter integrador, articulando los distintos contenidos vistos en la materia. Finalmente, en las preguntas teóricas se pretende que el alumno demuestre un conocimiento profundo de los temas, relacionando conceptos entre sí.

Criterios de Evaluación:

- Adecuada respuesta a los contenidos teóricos.
- Relación de conceptos pertinente.
- Resolución correcta de los problemas planteados.
- Fundamentación bibliográfica de los temas.

- Respeto de las consignas presentadas.

Se agrega la posibilidad de defender un trabajo final en la instancia del examen final. Esta opción se habilitará únicamente a los alumnos que hayan aprobado la cursada con una nota de 7 (siete) o mayor.