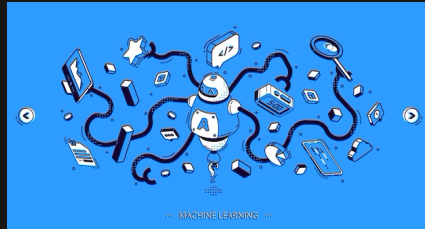


Syllabus

Aprendizaje de Máquina aplicado



Marco Teran
EAFIT

2025

Contenido

1 Presentación

2 Información del curso

- Objetivos
- Horarios
- Calificación y expectativas del curso
- Proyecto aplicado
- Presentación de tópicos avanzados

3 Bibliografía

4 Contactos

Presentación

Información del curso

Información del curso

Aprendizaje de Máquina aplicado

Este curso de **Aprendizaje de Máquina aplicado** combina fundamentos teóricos con implementación práctica y toma de decisiones basada en evidencia. Aborda el ciclo completo de un proyecto de *Machine Learning* (definición del problema, datos, modelado, validación, despliegue ligero) con énfasis en buenas prácticas, interpretabilidad y comunicación de resultados.

Enfoque del curso

- **Teoría + práctica:** principios estadísticos y de generalización → implementación reproducible en *notebooks* y repositorios.
- **Cobertura equilibrada:** modelos lineales, SVM, ensambles, no supervisado, *intro a deep learning* y series de tiempo.
- **Herramientas estándar:** Python, scikit-learn, pandas, matplotlib; vista corta a TensorFlow/Keras cuando sea pertinente.
- **Énfasis aplicado:** tratamiento de clases desbalanceadas, selección de características, *regularización*, *cross-validation*, *hyper-parameter tuning*.
- **Comunicación:** reportes claros, tableros de métricas y argumentos sólidos para decisiones técnicas.

Objetivo general del curso

Aplicar técnicas y algoritmos de *Machine Learning* para resolver problemas de clasificación y regresión con datos reales, garantizando validación honesta, interpretabilidad y comunicación efectiva de resultados.

Objetivos específicos

Al finalizar, el estudiante será capaz de:

- Diseñar el flujo de trabajo completo de un proyecto ML (**CRISP-DM**) con control de calidad de datos.
- Preprocesar, explorar y visualizar datos; definir métricas apropiadas al problema y al *stakeholder*.
- Entrenar y comparar modelos base: regresión/CLF lineales, SVM, árboles y *ensembles* (RF, Gradient Boosting, XGBoost).
- Implementar estrategias para *clases desbalanceadas* (re-muestreo, costes, umbrales).
- Reducir dimensionalidad (PCA) y seleccionar características con criterio estadístico.
- Introducir arquitecturas de *deep learning* cuando aporten valor (MLP, CNN/RNN para casos acotados).
- Modelar series de tiempo con enfoques clásicos y *baselines* robustos.
- Evaluar, interpretar y comunicar con rigor (curvas ROC/PR, *calibration*, SHAP/permutation).

Contenido

- 1 Introducción y **CRISP-DM**: problema, métricas, *baselines*, división honesta de datos.
- 2 **Preprocesamiento y EDA**: limpieza, *pipelines*, *feature engineering*.
- 3 **Clases desbalanceadas**: re-muestreo, costes, umbral, métricas adecuadas.
- 4 **Modelos lineales y regularización**: Ridge/Lasso/Elastic Net; regresión logística.
- 5 **SVM**: núcleos, márgenes y *tuning*.
- 6 **Ensamble**: Árboles, Random Forest, Gradient Boosting, XGBoost.
- 7 **Reducción de dimensionalidad**: PCA & proyección; selección de variables.
- 8 **No supervisado**: k -means, clustering jerárquico, validación.
- 9 **Intro Deep Learning**: MLP; CNN/RNN breves para casos acotados.

Horario fijo

Día	Hora inicio	Hora fin	Salón
Jueves	18:00	21:00	B14 201

Cuadro 1: Sesiones semanales.

Nota: Los jueves comprendidos para este curso son las fechas listadas en el cronograma.

Cronograma tentativo

Sesión	Fecha	Tema	Evaluación
1	11/09/2025	Intro + CRISP-DM; métricas + baselines	—
2	18/09/2025	EDA, preprocesamiento + pipelines	Se asigna Entrega 1
3	25/09/2025	Modelos lineales + regularización	Entrega 1 (semanal)
4	02/10/2025	SVM: núcleos, márgenes, CV	—
5	09/10/2025	Ensamblados: RF, GBM, XGBoost	Entrega 2 (mitad del curso)
6	16/10/2025	PCA + selección de features	—
7	23/10/2025	No supervisado: k-means, jerárquico	—
8	30/10/2025	Intro DL (MLP) + nociones de series de tiempo	—
9	05/11/2025	Cierre: interpretabilidad, calibración, comunicación	Examen final + Entrega 3 + Presentaciones

Cuadro 2: Sesiones (jueves 18:00–21:00; cierre el 05/11).

Métodos de evaluación

(Sujeto a ajustes)

Componente	Porcentaje
Proyecto aplicado (3 entregas : semanal, mitad, final)	35 %
Entrega 1 (2 semana)	5 %
Entrega 2 (mitad del curso)	10 %
Entrega 3 (día final)	20 %
Examen final (teórico–práctico)	20 %
Presentación de tópicos avanzados	20 %
Talleres, tareas cortas y <i>quices</i>	25 %
Las actividades de taller se asignarán de forma continua para reforzar contenidos.	

Cuadro 3: Esquema de evaluación.

Proyecto aplicado

Proyecto aplicado: objetivo

Ejecutar un proyecto de *Machine Learning* de principio a fin con **CRISP-DM**, garantizando: datos de calidad, validación honesta, selección de modelos apropiados, interpretación de resultados y comunicación clara orientada a decisiones.

Entregas del proyecto

Entrega 1 (semanal)

- Problema y métrica; EDA y *data card*; *baseline* reproducible.
- **ZIP** con *notebook*, reporte breve y repo Git actualizado.

Entrega 2 (mitad del curso)

- Comparación de 2–3 familias (lineales/SVM/árboles/ensambles); *tuning* y validación.
- Análisis de umbral/desbalance y PCA/selección de *features*.

Entrega 3 (día final)

- Modelo final + *interpretabilidad* (SHAP/permutation), *calibration* y recomendaciones.
- **ZIP** con *notebook*, reporte ejecutivo y póster.

Conjuntos de datos sugeridos

Conjuntos de datos

- Google Play Store Apps
- Trip Advisor Hotel Reviews
- Netflix Movies and TV Shows
- Avocado Prices
- Fashion MNIST
- Students Performance in Exams
- Credit Card Fraud Detection
- Melbourne Housing Market
- IBM HR Analytics Employee Attrition & Performance
- UJIIndoorLoc
- COVID19 Global Forecasting (Week 5)

Presentación de tópicos avanzados

Presentación de tópicos avanzados

Selecciona uno (individual o en parejas) para una charla corta y un demo:

- **Optimización y *tuning*** avanzados: Bayesian/Optuna, *early stopping*.
- **Interpretabilidad** moderna: SHAP de interacciones, *counterfactuals*.
- **Desbalance extremo**: focal loss, *cost-sensitive* pipelines.
- **AutoML** responsable: límites, fugas, gobernanza.
- **Graph ML** (nociones): *node/edge* tasks, *message passing*.
- **Fairness** en ML: métricas y *trade-offs*.
- **MLOps ligero**: *pipelines*, *model cards*, versiones y *drift*.

Bibliografía

Bibliografía esencial

- **Hastie, Tibshirani, Friedman** (2009). *The Elements of Statistical Learning*. Springer.
- **Murphy, K. P.** (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.
- **Bishop, C. M.** (2006). *Pattern Recognition and Machine Learning*. Springer.
- **Goodfellow, Bengio, Courville** (2016). *Deep Learning*. MIT Press.
- **scikit-learn documentation**: <https://scikit-learn.org/>

Contactos

Repositorio de GitHub



► Repositorio Aprendizaje de Máquina aplicado
url: github.com/marcoteran/machinelearning

¡Muchas gracias por su atención!

¿Preguntas?



Contacto: Marco Teran
webpage: marcoteran.github.io/
e-mail: mtteranl@eafit.edu.co