

Aprendizaje de máquinas: aplicaciones en el sector público

Bogotá Summer School in Economics — 2026

Profesor: Jorge Gallego | Banco Interamericano de Desarrollo | jandresg@iadb.org

1. Introducción

Las técnicas de aprendizaje automático (*machine learning*) se han consolidado como herramientas fundamentales para el análisis de datos en la era del *big data*. Su impacto trasciende el sector tecnológico y se aplica en múltiples ámbitos. Desde la detección temprana de deserción estudiantil y la focalización de programas sociales, hasta la identificación de patrones de corrupción y la optimización de servicios ciudadanos, el aprendizaje automático permite a las organizaciones públicas anticiparse a los problemas, asignar recursos con mayor precisión y diseñar políticas basadas en evidencia.

En la primera parte del curso, exploraremos las principales técnicas de aprendizaje supervisado, aplicando modelos predictivos a diversos problemas empíricos. En la segunda parte, abordaremos el aprendizaje no supervisado, analizando métodos para la segmentación de datos, detección de anomalías y reducción de dimensionalidad. Finalmente, profundizaremos en el aprendizaje profundo (*deep learning*), explorando arquitecturas de redes neuronales y sus aplicaciones en problemas complejos de análisis de datos.

2. Metodología

Este curso de 18 horas será teórico-práctico, con presentaciones magistrales en las que el profesor expondrá los principales conceptos estadísticos y matemáticos subyacentes a estas técnicas, seguidas de talleres prácticos en los que se enseñará a los estudiantes a resolver problemas de política pública concretos. Para tal propósito, en el curso usaremos el paquete estadístico Python en la plataforma *Google Colab*, e introduciremos los conocimientos básicos necesarios para aplicar las técnicas que estudiaremos en el curso. Todos los materiales del curso se encontrarán en un repositorio de GitHub <https://github.com/jagallegod/>.

3. Contenido

I. Aprendizaje supervisado: clasificación

- Introducción
- Clasificación usando vecinos más cercanos
- Aprendizaje probabilístico: Naive Bayes
- Support Vector Machines
- Árboles de decisión

II. Aprendizaje supervisado: predicción numérica

- Regresión
- Ridge
- Lasso

III. Desempeño de modelos

- Evaluación y mejoramiento del desempeño
- Feature extraction, selection y engineering
- Importancia de variables
- Desbalance de clase
- Ensamblaje de modelos y overfitting

IV. Aprendizaje no supervisado

- Introducción
- Clustering
- Detección de anomalías
- Aprendizaje de representación y embeddings

V. Aprendizaje profundo

- Introducción
- Redes neuronales artificiales (ANNs)
- Redes neuronales convolucionales (CNNs)
- Redes neuronales recurrentes (RNNs)

4. Políticas del curso

- Se espera que los estudiantes realicen las lecturas asignadas antes de clase.
- Se espera participación activa durante las clases.

5. Bibliografía

- Abadi, M. et al. (2016). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. <https://www.tensorflow.org/>
- Bisong, E. (2019). *Building Machine Learning and Deep Learning Models on Google Cloud Platform*. Apress.
- Chollet, F. (2021). *Deep Learning with Python* (2.^a ed.). Manning Publications.
- Conway, D. y J. Myles (2012). *Machine Learning for Hackers*. O'Reilly Media.
- Hastie, T., R. Tibshirani y J. Friedman (2009). *The Elements of Statistical Learning*. Springer.
- James, G., D. Witten, T. Hastie y R. Tibshirani (2013). *An Introduction to Statistical Learning*. Springer.
- Lantz, B. (2015). *Machine Learning with R*. Packt Publishing.
- Matloff, N. (2011). *The Art of R Programming*. No Starch Press.
- Müller, A. C. y S. Guido (2016). *Introduction to Machine Learning with Python*. O'Reilly Media.
- Raschka, S., J. Patterson y C. Nolet (2020). Machine Learning in Python. *Information*, 11(4), 193. <https://doi.org/10.3390/info11040193>
- Russell, S. y P. Norvig (2020). *Inteligencia Artificial: Un Enfoque Moderno*. Pearson Educación.
- Siegel, E. (2013). *Predictive Analytics*. John Wiley & Sons.
- Torres, J. (2020). *Python Deep Learning: Introducción práctica con Keras y TensorFlow 2*. Marcombo.
- Zhang, A., Z. C. Lipton, M. Li y A. J. Smola (2021). *Dive into Deep Learning*. <https://d2l.ai/>
- Zumel, N. y J. Mount (2014). *Practical Data Science with R*. Manning Publications.