

---

## ECONOMETRÍA II MACHINE LEARNING

---

### **Descripción del curso:**

En este curso introduciremos los principios fundamentales de ciencia de datos a través del pensamiento análisis de datos y de la construcción de modelos de machine learning con el fin de extraer conocimiento útil y de valor para el negocio. Se utilizarán herramientas de programación basada en Python para desarrollar algoritmos de clasificación y regresión. Adicionalmente se explorará a profundidad cada uno de los pasos del flujo de trabajo para la construcción e implementación de modelos de machine learning con el fin de generar conocimiento en diferentes situaciones y giros de negocio.

### **Catedrático:**

Ing. Preng Biba: prengsen@galileo.edu  
Lic. Marvin Lopez: marvin.lopezdubon@galileo.edu

### **Objetivos del Curso:**

- Introducir modelos de toma de decisiones basadas en Datos (Data Drive Decision-Making).
- Entender y practicar las distintas técnicas de minería de datos actuales utilizando Python.
- Caracterizar algoritmos de aprendizaje supervisado y su implementación en pipelines de Machine Learning.
- Practicar en casos teóricos y casos reales la aplicación de algoritmos de Machine Learning
- Conoce y aplica los fundamentos para la implementación y despliegue de modelos de machine learning.
- Utiliza mlFlow como herramienta para la gestión del ciclo de vida de MLOps.

### **Metodología:**

- El curso se impartirá de forma sincrónica en el horario de clase y asincrónica mediante vídeos de clase.
  - Es su propia responsabilidad descargar/ver los vídeos correspondientes cada semana.
  - Las dudas serán resueltas por medio de correo electrónico.
- Los exámenes serán presenciales y tendrán 3 horas para resolverlo. Se recomienda ver el calendario de clase y estar enterado de las semanas en las que se realizarán. No hay cambios de fechas ni reposiciones de exámenes

**Calendario Tentativo:**

Sesión	Tema
Semana 1	<b>Machine Learning Workflow:</b> Ingestión de Datos. Ingeniería de Características. Selección de Variables. Entrenamiento y Selección de Modelos. Optimización de Hyper-parámetros. Implementación y Despliegue de Modelos. Monitoreo y Control de Modelos. Git & Github para Machine Learning
Semana 2	<b>Ingeniería de Características - Parte 1:</b> Data Imputation. Categorical Encoding.
Semana 3	<b>Ingeniería de Características - Parte 2:</b> Manejo de Outliers. Transformación de Variables. Feature Scaling.
Semana 4	<b>Pipelines para Regresión - Parte 1:</b> Regresión Lineal Múltiple. Métricas para Regresión: $R^2$ , RMSE, MAE. Gráficos de resultados. Stepwise Regression.

Sesión	Tema
Semana 5	<b>Pipelines para Regresión - Parte 2:</b> SVR, KNN, Árboles de Decisión, Radom Forest. Regularización de Regresiones. Optimización de Hyper-Parámetros para Regresión. sklearn pipelines para regresiones.
Semana 6	<b>Pipelines para Clasificación - Parte 1:</b> Regresión Logística, Matriz de Confusión. Curva ROC-AUC. Features Selection para Clasificación.
Semana 7	<b>Pipelines para Clasificación - Parte 2:</b> SVM, Naive Bayes, KNN, Árboles de Decisión, Random Forest. Optimización de Hyper-Parámetros para Clasificación. Pipelines de sklearn para clasificación
Semana 8	<b>Machine Learning Operations - :</b> Definición de MLOps Ciclo de Vida de MLOps. Introducción a mlFlow.
Semana 9	<b>mlFlow - Parte 1</b> Tracking and logging Components en mlFlow.
Semana 10	<b>mlFlow - Parte 2</b> Models and Experiments en mlFlow.

**Bibliografía:**

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112, p. 18). New York: springer..
- John Lu, Z. Q. (2010). *The elements of statistical learning: data mining, inference, and prediction*. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(3), 693-694
- Cichosz, P. (2015). *Data Mining Algorithms: Explained Using R* (1 edition). Wiley.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about*.
- *Data Mining and Data-Analytic Thinking* (1 edition). Sebastopol, Calif.: O'Reilly Media.

**Evaluación:**

Actividades a desarrollar	Puntuación Asignada
Laboratorios.	25 pts.
Proyecto I.	25 pts.
Proyecto Final.	25 pts.
Examen Final.	25 pts.
<b>Total</b>	<b>100 pts.</b>