

## PROGRAMA DE CURSO

### EAE3706 MACHINE LEARNING Y BIG DATA

#### IDENTIFICACIÓN

CURSO	:	MACHINE LEARNING Y BIG DATA
TRADUCCIÓN	:	MACHINE LEARNING AND BIG DATA
SIGLA	:	EAE3706
CRÉDITOS	:	10
MÓDULOS	:	3
REQUISITOS	:	SIN REQUISITOS
RESTRICCIONES	:	SIN RESTRICCIONES
CONECTOR	:	NO APLICA
CARÁCTER	:	OPTATIVO
TIPO	:	CÁTEDRA
CALIFICACIÓN	:	ESTÁNDAR
PALABRAS CLAVE	:	DATA SCIENCE, DATA MINING, ANÁLISIS DE DATOS, PREDICCIÓN, CAUSALIDAD.
NIVEL FORMATIVO	:	MAGÍSTER

#### I. DESCRIPCIÓN DEL CURSO

En este curso, las y los estudiantes aplicarán los principales conceptos y metodologías empleadas en el análisis de big data y machine learning. Además, podrán evaluar y analizar, desde una visión integral, los principales usos, fortalezas y debilidades de estas herramientas, así como la capacidad de aplicación práctica en ámbitos de la macroeconomía, economía financiera y políticas públicas. La modalidad del curso consiste en clases expositivas y laboratorios prácticos. El aprendizaje será evaluado mediante tareas prácticas, presentaciones y un trabajo aplicado.

#### II. RESULTADOS DE APRENDIZAJE

1. Sintetizar los principales conceptos de big data y machine learning para aplicarlos a datos y situaciones reales de análisis.
2. Planificar una línea de trabajo de análisis de datos considerando objetivos claros de trabajo, además de los desafíos y potenciales resultados en cada escenario.
3. Aplicar distintos modelos de machine learning para realizar predicciones considerando diferentes escenarios, objetivos y disponibilidad de datos.
4. Interpretar predicciones de los modelos aplicados buscando una utilidad práctica.
5. Evaluar y analizar el desempeño de modelos aplicados incorporando posibles ideas de mejora y determinando condiciones suficientes para su uso.
6. Aplicar ML para análisis causal considerando los aportes y limitaciones que herramientas de data science pueden aportar a la econometría.

#### III. CONTENIDOS

1. Data, big data y machine learning
  - 1.1. Análisis de datos:
    - 1.1.1. Tipos de datos y análisis
    - 1.1.2. Bases de datos relacionales
    - 1.1.3. Data science workflow

## PROGRAMA DE CURSO

### EAE3706 MACHINE LEARNING Y BIG DATA

#### 1.2. Big data:

- 1.2.1. Los datos en el actual “escenario digital”: fuentes, volumen, capacidad de almacenamiento y computación.
- 1.2.2. Nuevo paradigma en el análisis de datos: data mining
- 1.2.3. Modelo conceptual 5Vs: volumen, velocidad, variedad, veracidad y valor.
- 1.2.4. Map-reduce y tecnologías de big data
- 1.2.5. Aspectos éticos relacionados con big data

#### 1.3. Machine Learning:

- 1.3.1. Qué es ML y para qué se usa (ML vs. Econometría: predicción vs. causalidad)
- 1.3.2. Tipos de aprendizaje: supervisado, no supervisado y reinforcement learning
- 1.3.3. El proceso de aprendizaje: entrenamiento – testeo – predicción
- 1.3.4. Generalizando el aprendizaje (predicción)
- 1.3.5. Cross-validation, overfitting y performance de los modelos (learning curve, precision, recall, F1 score)

#### 2. Modelos de machine learning

- 2.1 Árboles de decisión, random forest y regression y model trees
- 2.2 Clustering – K-Means
- 2.4 Naive Bayes – predicción probabilística
- 2.5 Support Vector Machines
- 2.6 Perceptrones, redes neuronales y deep learning
- 2.7 Text mining

#### 3. Usando ML para análisis de inferencia causal (econometría)

- 3.1 ¿Basta con predecir? La economía requiere del análisis de causalidad
- 3.2 Estimación de parámetros
- 3.3 Análisis de heterogeneidad

#### 4. Casos aplicados

## IV. ESTRATEGIAS METODOLÓGICAS

- Clases expositivas
- Laboratorios prácticos
- Relatores externos con experiencia aplicada
- Presentaciones de los estudiantes

## V. ESTRATEGIAS EVALUATIVAS

- Tareas prácticas (30%): entregas de reportes individuales con el desarrollo, resultados y comentarios sobre un set de ejercicios prácticos propuestos.
- Presentación paper/tema relevante (20%): exposición grupal al resto de los estudiantes sobre un tema relevante, guiando la clase y generando discusión sobre los temas asignados.
- Trabajo grupal (50%): Entrega de un reporte y código con el desarrollo y resultados de un análisis práctico escogido y diseñado por cada grupo en el cual se empleen los conceptos propios del contenido del curso.

## VI. BIBLIOGRAFÍA

Mínima:

## PROGRAMA DE CURSO

### EAE3706 MACHINE LEARNING Y BIG DATA

- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. W. McKinney. O'Reilly Media, 1 edition, (Feb 5, 2013).
- J. Leskovec, A. Rajaraman, and J. Ullman. Mining of Massive Datasets. Cambridge University Press, Second edition, (2014).
- Bishop, Christopher M. Pattern Recognition and Machine Learning. New York: Springer, 2006.
- Grégoire Montavon, Geneviève B. Orr, Klaus-Robert Müller. Neural Networks: Tricks of the Trade. Springer, 2012.

#### Complementaria:

- Athey (2017). Beyond prediction: Using big data for policy problems. Science
- Bardoscia, Bianconi, Ferrara (2019). Multiplex network analysis of the UK over-the-counter derivatives market. International Journal of Finance and Economics
- Blumenstock, Cadamuro, On (2015). Predicting poverty and wealth from mobile phone metadata. Science
- Caballo and Rigobon (2016). The Billion Prices Project: Using Online Prices for Measurement and Research. Journal of Economic Perspectives
- Donaldson and Storeygard (2016). The View from Above: Applications of Satellite Data in Economics. Journal of Economic Perspectives
- Galbraith and Tkacz (2015). *Nowcasting GDP with electronic payments data*. ECB's Statistics Paper Series
- Henderson, Storeygard and Weil (2009). Measuring Economic Growth from Outer Space. American Economic Review
- McBride and Nichols (2018). Retooling Poverty Targeting Using Out-of-Sample Validation and Machine Learning. World Bank Economic Review.
- Mullainathan and Spiess (2017). Machine Learning: An Applied Econometric Approach. Journal of Economic Perspectives
- Murphy, K. Machine Learning: A Probabilistic Perspective. MIT Press.
- Witten, Frank, Hall. Data Mining: Practical Machine Learning Tools and Techniques. MK Series.