

## “Ciencia de Datos y Big Data”

### Presentación y objetivos:

La ciencia de datos trata sobre la explotación de datos con el fin de realizar análisis descriptivos y/o predictivos a partir de ella. En el mundo de hoy, donde se generan una enorme cantidad de datos en forma continua (*Big Data*), el análisis de esta información mediante técnicas de Minería de Datos (*Data Mining*) y Aprendizaje Automático (*Machine Learning*) juega un papel crucial en los más variados campos y disciplinas: investigación médica y bioquímica, análisis de redes sociales, finanzas, operaciones comerciales y de producción, pronósticos meteorológicos, reconocimiento de imágenes, etc.

El objetivo de este curso es presentar los fundamentos de Data Mining y de Machine Learning, sus técnicas y algoritmos, utilizando para ello el lenguaje de programación Python.

A su vez se presentarán conceptos de arquitecturas de Big Data como sistemas distribuidos y escalables de almacenamiento y procesamiento, frameworks Hadoop y Spark.

**Docente a Cargo:** Ing. Carlos Arana

## Contenido:

### *Revisión de Conceptos de Estadística y de Programación en Python*

- Preliminares de Probabilidad/Estadística y Álgebra Lineal
- Conceptos de Programación en Python y sus Librerías Especializadas

### *Modelos de Regresión*

- Regresión Lineal Simple y Múltiple
- Regresión Polinomial, No Lineal y por Splines
- Selección del mejor modelo (Stepwise): Base, Forward, Backwards
- Concepto de Sobreajuste (Overfitting) Regulación: Ridge y Lasso (L1 y L2)

### *Validación y Prueba de Modelos*

- Definición de Aprendizaje Supervisado
- Sets de Entrenamiento, Validación y Prueba
- Compensación Sesgo-Varianza (*Bias-Variance Tradeoff*)
- Validación Cruzada (*Cross Validation*)
- k-folds Cross Validation

### *Modelos de Clasificación*

- K-Vecinos más Cercanos (KNN),
- Regresión Logística
- Matriz de Confusión
- Árboles de Decisión para Clasificación y Regresión (CART)
- Matriz de Costo-Beneficio
- Ensamblados de Árboles de Decisión (Bagging y Random Forests)
- Gradient Boosting Machines (Adaboost y XGBoost)

### *Aprendizaje No Supervisado*

- Análisis de Componentes Principales
- Clustering K-Means y Jerarquizado

### *Redes Neuronales Artificiales y Aprendizaje Profundo*

- Perceptrón Multicapa
- Parámetros e Hiperparámetros, Funciones de Pérdida y BackPropagation
- Aprendizaje Profundo (Deep Learning). Aprendizaje de Representaciones de Datos
- Redes Neuronales Convolucionales
- Práctica en Framework de Deep Learning Tensorflow

### *Big Data*

- Cloud Services
- Arquitecturas de Big Data

- GFS y Map Reduce
- Hadoop y Spark

**Condiciones De Ingreso:**

- Ser egresado de una carrera universitaria en cualquier especialidad.
- Conocimientos básicos de Probabilidad y Estadística, Álgebra Lineal y Cálculo Diferencial
- Fundamentos de programación

**Duración del Curso:**

12 clases de 3.5 hs (42 hs)

**Fecha de Inicio:**

Lunes 18 de abril de 2022

**Día y Horario:**

Lunes de 18 a 21.30

**Formato y Condiciones de Cursada:**

Virtual. Asistencia mínima: 75% de las clases (obligatoriedad de cámara encendida durante todas las clases)

**Aprobación:**

Trabajo práctico integrador

**Arancel:**

Importe Total del Curso: \$ 80.000.-

Financiamiento: Matrícula \$20.000 y 3 cuotas de \$20.000

**Apto Crédito Fiscal Para Capacitación: Somos UCAP**

La "Escuela de Graduados Ingeniería de Dirección Empresaria" de la Facultad de Ingeniería (UBA) es una Unidad Capacitadora, de manera que los cursos que dictamos pueden canalizarse a través del Régimen de Crédito Fiscal para Capacitación.

**Informes e Inscripción:**

Secretaría de la Escuela de Graduados en Ingeniería de Dirección Empresaria –

Av. Paseo Colón 850 3<sup>er</sup> Piso - Buenos Aires

Consultas e inscripción a través del e-mail: [egide@fi.uba.ar](mailto:egide@fi.uba.ar)