



Curso de Posgrado: Ciencia de datos aplicada: Fundamentos, técnicas y herramientas

Profesor Responsable: Dr. Facundo A. Lucianna

Expositores: Dr. Facundo A. Lucianna

Mg. Cecilia E. Saavedra Fresia

Dr. Alvaro Gabriel Pizá

Facultad: FaCET - UNT

Resumen

La capacidad de recolectar y convertir los datos en información y conocimiento se está volviendo una habilidad competitiva que muchas organizaciones requieren para desempeñarse con rapidez, certeza y eficiencia.

Esta "ciencia de los datos", es una ciencia multidisciplinar que requiere habilidades en informática, matemáticas/estadística, aprendizaje automático, entre otras. Además exige saber manejar grandes volúmenes de datos, capacidad de extraer y transmitir el conocimiento que se extraen de los datos, visión de negocio, etc.

Estos conjuntos de datos pueden provenir de los datos generados por todo tipo de dispositivos electrónicos (como un móvil, todo tipo de sensores, secuenciadores de genoma), redes sociales, datos médicos, páginas web; y afectan de manera muy significativa la investigación actual en muchos campos como las ciencias biológicas, la informática médica, la salud, las ciencias sociales, por citar sólo algunos.

El data science es la evolución de lo que hasta ahora se conocía como analista de datos, pero a diferencia de éste que sólo se dedica a analizar fuentes de datos de una única fuente, explora y analiza datos de múltiples fuentes que pueden tener formatos muy diferentes.

El presente curso está concebido para que los alumnos se introduzcan en el flujo de trabajo típico en Data Science, entre ello:

- el concepto de análisis exploratorio de datos,
- a extraer los datos, independientemente de su fuente (webs, csv, logs, APIs, etc.) y de su volumen (Big Data o Small Data),
- a limpiar los datos, para eliminar lo que distorsiona las mismas;
- a procesar los datos usando diferentes métodos estadísticos (inferencia estadística, pruebas de hipótesis, etc.);
- a diseñar nuevos tests o experimentos en caso necesario;



- qué son y cómo utilizar librerías de código para realizar operaciones de análisis y
- cómo realizar distintos análisis mediante gráficos y visualizaciones.

Docentes

El Dr. Ing. Facundo Adrián Lucianna es becario Postdoctoral del CONICET, jefe de trabajos prácticos del Departamento de Bioingeniería de la FaCET y docente de la asignatura Análisis Matemático II de la Universidad Tecnológica Nacional - Facultad Regional Tucumán. Su área de trabajo es la biomimética y bio-robótica, y tiene gran experiencia en el uso de algoritmos de Machine Learning, particularmente redes neuronales, con trabajos publicados en revistas indexadas en el tema.

La Mg. Cecilia Saavedra Fresia, es Profesor Adjunto del Área de Matemática Aplicada de la FaCET y alumna del doctorado en Ciencias Biológicas.

El Dr. Ing. Álvaro Gabriel Pizá es becario posdoctoral del CONICET y jefe de trabajos prácticos en el Departamento de Bioingeniería de la FaCET. Se especializa en la adquisición, procesamiento y análisis de señales electrofisiológicas, orientado a la rehabilitación y el deporte.

Programa

Unidad I: Definición de Data Science. Descripción del proceso de trabajo en Data Science: Obtener, limpiar, explorar, visualizar, modelar e interpretar los datos.

Unidad II: Introducción al lenguaje de programación Python, junto con las librerías propias del análisis de datos. Introducción a SQL.

Unidad III: Obtención y limpieza de datos. Tipos de datos. Fuente de datos Identificación y limpieza de errores, valores faltantes y corruptos. Casos de uso en el manejo de pequeños volúmenes y grandes cantidades de datos.

Unidad IV: Análisis Exploratorio de Datos. Análisis y estructuración de los campos de los datos. Creación de información de datos crudos. Extracción de características mediante el uso de inferencia estadística. Test de variables. Reducción de dimensionalidad.

Unidad V: Visualización de datos. Técnicas de visualización de datos. Búsqueda de patrones de comportamiento de los datos a través de visualizaciones o gráficos. Buenas prácticas de uso de herramientas de visualización.



Unidad VI: Modelando los datos con Machine Learning. Tipos de algoritmos de aprendizaje automático. Clasificación, regresión, predicción y clusterización. Tipos de modelos, estudio de casos de uso. Interpretación de resultados y el problema del sobreajuste.

Bibliografía:

Bishop, C. M. (2016). Pattern Recognition and Machine Learning (Information Science and Statistics) (Softcover reprint of the original 1st ed. 2006 ed.). Springer.

Bruce, P., Bruce, A., & Gedeck, P. (2020). Practical Statistics for Data Scientists: 50+ Essential Concepts Using R and Python (2nd ed.). O'Reilly Media.

Bundău, O., Juratoni, A., Chevereșan, A., Simos, T. E., Psihoyios, G., & Tsitouras, C. (2010).

Burkov, A. (2019). The Hundred-Page Machine Learning Book. Andriy Burkov.

Burkov, A. (2020). Machine Learning Engineering. True Positive Inc.

Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems (2nd ed.). O'Reilly Media.

Grus, J. (2019). Data Science from Scratch: First Principles with Python (2nd ed.). O'Reilly Media.

Hertz, J. A., Krogh, A. S., & Palmer, R. G. (1991). Introduction To The Theory Of Neural Computation (Santa Fe Institute Series) (1st ed.). CRC Press.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics) (2nd ed. 2021 ed.). Springer.

Prerrequisitos:

- Manejo fluido de Álgebra Lineal.
- Conocimientos básicos de Probabilidad y Estadística.
- Conocimientos básicos de algún lenguaje de programación (Python, Matlab, Fortran, C, etc.)

Duración: 70 hs.



Sistema de Evaluación:

Presentación de Trabajos Prácticos. Elaboración de una monografía.

Asistencia requerida: 80%

Tribunal examinador:

Dr. Facundo A. Lucianna, Mg. Cecilia E. Saavedra Fresia, Dr. A. Gabriel Pizá

Número Mínimo de Inscritos: 3 alumnos.

Fecha de Inicio: Setiembre de 2022, en horario a convenir con los alumnos.

Fecha estimada de finalización: Noviembre de 2022.

Financiamiento: No requiere de financiamiento externo de la Maestría.