

Negocios Digitales

Certificación Avanzada en Data Science

Big data se refiere a las grandes colecciones de datos, estructurados, o no estructurados, que pueden crecer a volúmenes enormes y a un ritmo tan alto que es complejo manejarlos con las técnicas habituales de los sistemas de bases de datos y las herramientas de análisis existentes hasta hace un tiempo.

La gestión y el análisis de estas cantidades masivas de datos exigen nuevas soluciones que van más allá de los procesos tradicionales o herramientas de software habituales en la práctica diaria, lo que ha llevado a acuñar el término "big data analytics".

La Certificación Avanzada en Data Science está orientada a facilitar el conocimiento de las herramientas que permiten diseñar, preparar, analizar y manejar información, estructurada y no-estructurada.

OBJETIVOS DEL PROGRAMA

1. Generar comprensión de los pasos centrales del workflow de un proyecto de data science, desde la comprensión del problema hasta la construcción de modelos y visualización de los resultados.
2. Presentar las herramientas y habilidades necesarias para formar parte de un equipo de data science: modelado, diseño y consultas (utilizando SQL) de bases de datos y data warehouses, construcción de modelos de machine learning utilizando R (y opcionalmente Python) y manejo de otros tipos de datos más allá de las bases de datos relacionales, como datos geográficos y bases de datos de grafos.



OBJETIVOS DE FORMACIÓN

Que los participantes puedan:

1. Crear y desarrollar modelos de datos para el análisis de diversos escenarios que requieran el uso de datos para la toma de decisiones.
2. Integrar datos provenientes de diversas fuentes y técnicas para asegurar su consistencia.
3. Modelar e implementar data warehouses.
4. Manejar diferentes tipos de datos, en particular datos geográficos y bases de datos de grafos (para el análisis de redes).
5. Crear visualizaciones que ayuden a entender los data sets.
6. Construir y analizar modelos de machine learning
7. Presentar y comunicar resultados y hallazgos (incluso para una audiencia inexperta).

PERFIL DEL PARTICIPANTE

Profesionales de cualquier área o especialidad, que quieran adquirir conocimientos sobre técnicas y herramientas que les permitan aplicar el análisis de datos en sus actividades.

METODOLOGÍA

Durante la certificación se realizarán encuentros sincrónicos de desarrollo de contenido, presentación de materiales y trabajos grupales. De forma asincrónica se compartirán materiales que los y las estudiantes deberán revisar y actividades que refuercen lo trabajado en los encuentros sincrónicos.

Cada módulo se desarrolla según la dinámica que mejor se adecue a las características de los temas a impartir. En algunos módulos se desarrolla un



proyecto de principio a fin, y en otros cursos el sistema consiste en explicar cada tema y resolver un trabajo práctico relacionado.

CONOCIMIENTOS MÍNIMOS

Conocimientos de Bases de Datos y SQL. Se tomará una evaluación en base a los contenidos del curso de nivelación. El curso de Nivelación es un curso de conceptos básicos de Base de Datos y SQL. El curso es opcional, el examen es obligatorio. Se recomienda ver y hacer el curso a todos/as los participantes.

Todo el software utilizado en el curso es *open source*. Se sugiere una computadora con un mínimo de 16GB de RAM y procesador mínimo i5 o equivalente.

PROGRAMA

CURSO DE NIVELACIÓN

- Conceptos básicos de Base de datos y SQL.

MÓDULO 1: DATA WAREHOUSING & OLAP

- Modelos de datos.
- Arquitecturas.
- Análisis de requerimientos.
- Diseño conceptual, lógico y físico.
- El modelo multidimensional: estrella, snowflake y constellation.
- Manejo de cambios: Slowly changing dimensions.
- Diseño físico.
- Online Analytical Processing: OLAP. OLAP vs OLTP.
- SQL/OLAP: funciones de ventana.

MÓDULO 2: INTELIGENCIA ARTIFICIAL Y MACHINE LEARNING

- Conceptos básicos de Machine Learning.



- Modelos descriptivos y predictivos.
- Técnicas fundamentales: clasificación, clustering, patrones.
- Árboles de decisión.
- Aplicación a problemas concretos de predicción.
- Casos de estudio.
- Redes Neuronales.

MÓDULO 3: SISTEMAS DE INFORMACIÓN GEOGRÁFICA Y BASES DE DATOS DE GRAFOS

Parte 1:

- Bases de datos espaciales.
- Open Geospatial Consortium (OGC): objetos geométricos y geográficos.
- Lenguaje de Consulta: funciones y predicados para el análisis espacial.
- Concepto de índice espacial.
- Visualización de resultados en QGIS y OpenJump.

Parte 2:

- Bases de datos de grafos NoSQL.
- Modelado de Property Graphs.
- Lenguaje de Consulta: búsqueda de patrones y uso de agregaciones.
Soporte para el análisis de datos espaciales.

MÓDULO 4: VISUALIZACIÓN DE LA INFORMACIÓN

- Introducción, definiciones, antecedentes, gráficos notables.
- Principios de la excelencia gráfica. Observaciones y variables. Tipo de variables.
- Visualizando tablas, jerarquías y redes. Utilización del color. Representación eficiente de la información, sumarización y visualización de grandes volúmenes de datos.
- Prácticas con Tableau, Flourish, Datawrapper, RAWGraphs, Google Colab (R y Python) y Observable (JavaScript).



REQUISITOS DE APROBACIÓN

Aprobación examen de nivelación. Asistencia igual o superior al 75% de cada módulo. Aprobación de trabajo en cada módulo.

CUERPO DOCENTE

ALEJANDRO VAISMAN

Doctor en Ciencias de la Computación por la Universidad de Buenos Aires. Postdoctorado en la Universidad de Toronto, Canadá. Es especialista en Data Warehousing y Business Intelligence (BI), Data Mining, Sistemas de Información Geográfica (GIS), y en publicación de datos en la web, siguiendo los lineamientos de Linked Open Data. Co-autor, junto con el Prof. Esteban Zimányi (Universidad Libre de Bruselas) del libro "Data Warehouse Systems: Design and Implementation". Ha publicado, además, 5 capítulos en libros internacionales y ha publicado artículos en las más prestigiosas revistas y conferencias de la especialidad.

MARCELA RICCILLO

- Doctora en Ciencias de la Computación de la Universidad de Buenos Aires. Cuenta con +15 años de experiencia en empresas e instituciones en proyectos de tecnología de información e Inteligencia Artificial aplicada a industrias.
- Dictado de conferencias en eventos nacionales e internacionales (como TEDxRosario, y TEDxBarcelona). Profesora universitaria. Dictado de charlas y cursos de Robótica e IA (IA, Robótica, Data Mining, Machine Learning, Redes Neuronales). Publicaciones en congresos y revistas de divulgación científica (como Muy Interesante).
- Profesora en la Maestría de Ciencia de Datos (ITBA) y en la Certificación Avanzada en Data Science (ITBA).
- Fue columnista de Robótica en Radio Palermo y jurado en el programa de ciencia de TV Eureka de Canal Encuentro.
- Estudios de idiomas Inglés, Chino, Japonés, Coreano, Alemán, Latín.



LETICIA GÓMEZ

Licenciada en Ciencias de la Computación de la UBA y Doctora en Ingeniería en Informática del ITBA. Ha participado en diferentes proyectos en el área de Data Warehousing como “Diseño e Implementación de Herramientas para la Administración y Análisis de la Evolución de Datos Complejos en Big Data Utilizando Bases de Datos de Grafos” y “Big Data Cubes: Publicación y Análisis de Datos Multidimensionales en la Web Semántica en Tiempo Casi-Real”. Directora del Proyecto FITAP 008 “Desarrollo de Laboratorio de Seguimiento de Ojos para Medición de Usabilidad en SaaS”, MINCYT, Argentina. Actualmente es miembro del comité de UXQCC (User Experience Quality Certification Center) en UXPA, UXQCC. Es investigadora en el área de Bases de Datos, Business Intelligence (Online Analytical Processing y Data Mining) y Grafos. Ha realizado publicaciones en diferentes revistas científicas internacionales. Es docente titular en el Departamento de Ingeniería Informática del ITBA.

ARIEL AIZEMBERG

Licenciado en Ciencias de la Computación, Departamento de Computación, FCEyN, UBA. 25 años de experiencia docente (Inteligencia Artificial, Sistemas Expertos, Data Mining y Visualización de Información), 10 años como consultor independiente en las siguientes áreas: Spatial Databases, GIS, Data Quality, InfoVis y BI. Se desempeñó también como responsable de desarrollos GIS en diferentes organizaciones.



↗ Título: **Certificación Avanzada en Data Science**

↗ Modalidad: **Virtual**

↗ Inicio mail de bienvenida: **31/MAR**

↗ Examen nivelatorio: **23/ABR**

↗ Primera clase sincrónica: **28/ABR**

↗ Duración: **119 hs**

↗ Cursada: **martes y jueves de 18 a 21.30 hs**

↗ Tutorías optativas: **miércoles de 18 a 19.00 hs**

