

CERTIFICACIÓN PROFESIONAL

+ POWER BI

CIENCIA DE DATOS CON R Y PYTHON

MODALIDAD	DURACIÓN
ONLINE, EN VIVO	8 MESES

3 CLASES EN VIVO POR SEMANA



INSTITUTO

Data Science

Aval académico: Universidad Tecnológica Nacional Sede Reconquista
Resolución Nro 290/20.



Objetivos

Convertirte en un Experto en Ciencia de Datos: Adquirirás las habilidades esenciales en R y Python para convertirte en un experto capaz de analizar, interpretar y utilizar datos para tomar decisiones informadas y estratégicas en cualquier industria.

Impulsar tu Carrera Profesional: El curso está diseñado para abrirte las puertas a nuevas y mejores oportunidades laborales en el campo de la ciencia de datos, un área con alta demanda y excelentes perspectivas de crecimiento.

Desarrollar Proyectos Reales: Aprenderás a aplicar tus conocimientos a proyectos reales, preparando informes y modelos que tengan un impacto directo y positivo en el entorno laboral.

Contactos Profesionales: Formarás parte de una comunidad de aprendizaje vibrante, donde podrás conectar con otros profesionales y expandir tu red de contactos en el mundo de la ciencia de datos.

Estructura de la cursada

La estructura del curso ha sido cuidadosamente diseñada para maximizar el aprendizaje y garantizar que todos los estudiantes, independientemente de su nivel previo de conocimientos, puedan avanzar de manera efectiva.

La **Etapa de Nivelación** permite a los estudiantes reforzar sus fundamentos en programación lógica, bases de datos y estadísticas, asegurando una base sólida para el aprendizaje avanzado.

La **Etapa de Introducción** proporciona un marco conceptual integral sobre la ciencia de datos y las herramientas esenciales, estableciendo un contexto claro y relevante.

Finalmente, la **Etapa de Especialización** se enfoca en técnicas avanzadas, librerías y aplicaciones prácticas, permitiendo a los estudiantes aplicar sus conocimientos en proyectos del mundo real y prepararse para oportunidades laborales de alto nivel.

Etapa de Nivelación (optativa)	Programación Lógica	Bases de Datos	Probabilidad y Estadística
Etapa de Introducción	Introducción Conceptual	Introducción a Python	Introducción a R
Etapa de Especialización (circular)	Especialización en R clase teórica	Especialización en R clase práctica	Especialización en Python clase teórica-práctica

Programa académico

Nivelación (Optativa)

- Programación lógica.
- Bases de datos relacionales y no relacionales.
- Probabilidad y Estadística.

Introducción Conceptual

- Introducción a la Ciencia de Datos.
- Niveles a los que opera la ciencia de datos.
- Introducción a Data Warehouse.
- Introducción a Data Mining.
- Introducción a Knowledge Discovery.
- Introducción a Herramientas OLAP y Tableros de comando.
- Repaso de herramientas disponibles.
- Taxonomía de las competencias de un científico de datos.

Introducción a R

- Variables, vectores y matrices.
- Dataframes.
- Paquetes y librerías.
- Manejo de archivos csv.
- Conversiones de tipos.
- Conexión a bases de datos.
- Ejecución condicional.
- Bucles.

- Funciones en R.
- Medidas estadísticas.
- Correlaciones.
- Funciones estadísticas.

Introducción a Python

- Descarga e instalación.
- Principales librerías.
- Variables y tipos de datos.
- Listas, tuplas y diccionarios.
- Ejecución condicional.
- Ciclos definidos e indefinidos.
- Manejo de Excepciones.
- Funciones y Generadores.
- Clases y objetos.
- Manejo de archivos y directorios.
- Conexión a archivos planos.
- Conexión a Excel.
- Conexión a Bases de Datos.

Tests básicos

- Repaso de conceptos de probabilidad y estadística.
- Test de Hipótesis.
- Correlaciones.
- AB Test.
- Cálculo de correlaciones en R con cor.
- Funciones de distribución en R.
- Histogramas en R.
- Gráficos de líneas en R.
- Gráficos de áreas en R.
- Uso de NumPy.
- Uso de Pandas.
- Uso de Matplotlib.
- Módulo matplotlib.
- Módulo math.
- Módulo numpy.
- Módulo yt.
- Módulo mayavi.

Regresiones

- Regresión Lineal.
- Regresión polinómica.
- Regresión exponencial y logarítmica.
- Regresión de dos variables.
- Cálculo de regresiones en R con lm.

- Cálculo de regresiones logísticas en R con glm.
- Regresiones en Python.
- Regresión logística en Python.

Árboles de decisión

- Algoritmo básico en Excel.
- División en entrenamiento y prueba.
- Uso de Rpart y cp.
- Predicción y valoración de la solución.
- Uso de Party.
- Aplicación al problema del call center.
- Instalación y uso del paquete en Python.
- Ejemplo en Python.
- Principales parámetros de ajuste y control.

Clusters

- Algoritmo básico en Excel.
- Uso de kmeans en R.
- Ejemplo de aplicación real en R.
- Otros algoritmos de agrupamiento en R.
- Ejercicio de aplicación de agrupamiento de mascotas.
- Uso de Scikit-Learn en Python.
- Análisis de componentes principales en Python.
- K-Medias en Python.
- Clustering jerárquico en Python.

Reglas de Asociación

- Algoritmo básico en Access.
- División en entrenamiento y prueba en R.
- Uso de arules en R.
- Ejemplo de una aplicación real votación de reglas en R.
- Ejemplo de una aplicación real a datos de ventas en supermercados en R.
- Instalación y uso del paquete en Python.
- Ejemplo en Python.
- Principales parámetros de ajuste y control.
- Esquema de votación de reglas en Python.

Redes Neuronales

- Algoritmo básico en Excel.
- División en entrenamiento y prueba en R.
- Uso de neuralnet.
- Predicción y valoración de la solución en R.
- Discusión del problema de legibilidad de los resultados desde el negocio.
- Competencia entre árboles y redes en un caso concreto en R.

- Instalación y uso del paquete en Python.
- Ejemplo en Python.
- Principales parámetros de ajuste y control en Python.
- Problema concreto en Python.

Algoritmos Genéticos

- Algoritmo básico en Excel.
- Discusión del tipo de problemas en los que se aplica.
- Discusión de otros mecanismos de optimización.
- Implementación en R del uso de algoritmos genéticos.
- Valores reales.
- Binarios.
- Permutaciones.
- Implementación en Python.

Series Temporales

- Taxonomía.
- Separación de componentes.
- Predicciones.
- ARIMA implementado en R.
- Predicción en series con un único período.
- Predicción en series con múltiples períodos.
- Predicción en series con períodos variables: renormalización.
- Implementación en Python.

Método de Simulación de Montecarlo

- Algoritmo básico en Excel.
- Discusión de la utilidad del método.
- Comparación con el análisis de escenarios.
- Ejemplo de una aplicación real.
- Ejemplo simple implementado en R.
- Ejemplo simple implementación en Python.

Minería de Textos

- Clasificación supervisada de piezas de texto.
- Construcción de una red semántica.
- Discusión de las posibilidades de reconocimiento de voz.
- Ejemplos de aplicación real del algoritmo de clasificación.
- Aplicación a la detección de sentimientos.
- Ejemplo de implementación en Python.

Vecinos Cercanos (Knn)

- Instalación y uso del paquete en R.
- Ejemplo conceptual en Excel.

- Ejemplo en R.
- Instalación y uso del paquete en Python.
- Ejemplo en Python.
- Principales parámetros de ajuste y control.

Bayes Ingenuo

- Implementación en Excel.
- Ejemplo de aplicación en R.
- Comparación de 4 algoritmos predictivos (Bayes ingenuo, árboles, redes neuronales y regression logística).
- Ejemplo de aplicación en Python.

Random Forest

- Bootstrap y bagging.
- Descripción conceptual del método random forest.
- Paquete randomForest para R.
- Ejemplo de aplicación.
- Comparación con otras técnicas.
- Implementación en Python.

Métodos Bayesianos Avanzados

- Descripción conceptual del método.
- Regresión lineal bayesiana.
- Regresión logística bayesiana.
- Inferencia bayesiana.
- Red bayesiana.
- Ejemplos de aplicación en R.
- Paquete BAS.
- Paquete brms.
- Paquete arm.
- Paquete bnlearn.
- Comparación con otras técnicas.
- Ejemplos de aplicación en Python.

Máquina de Soporte Vectorial

- Descripción conceptual del método.
- Paquete e1071 para R.
- Ejemplo de aplicación en R.
- Ejemplo de aplicación en Python.
- Comparación con otras técnicas.

Discriminante Lineal y Cuadrático

- Instalación y uso del paquete.
- Ejemplo conceptual en Excel.

- Ejemplo en R.
- Ejemplo en Python.
- Principales parámetros de ajuste y control en R.
- Principales parámetros de ajuste y control en Python.
- Problema concreto.

Análisis de Fourier

- Descripción conceptual.
- Ejemplo conceptual en Excel.
- Instalación y uso del paquete en R.
- Instalación y uso del paquete en Python.
- Ejemplo en R.
- Ejemplo en Python.

Herramientas Geográficas

- Distancias.
- Implementación en R.
- Paquete sf.
- Paquete ngeo.
- Aplicación práctica en R.
- Implementación en Python.
- Aplicación práctica en Python.

Bases de Datos Documentales

- Instalación de MongoDB.
- Conexión a R.
- Conexión a Python.
- Aplicación práctica en R.
- Aplicación práctica en Python.

Diseño de Datawarehouses

- Diferencias entre los DW y los OLTP.
- Tipos de datos y soportes.
- Dimensiones y jerarquías.
- Estimación de recursos y tiempos según tipos de datos.
- Ejemplos de staging.
- Interacción entre equipos (DB-ETL-Soporte-FrontEnd-Usuarios).
- Ejercicios de diseño de DW.

Diseño y Construcción de ETL

- Procesos ETL: generalidades, buenas prácticas, esquemas de ETL.
- Detalles de cada etapa: extracción, limpieza, normalización, transformación, carga.
- Estrategias de update.

- Ejercicios de diseño de ETL para la carga inicial.
- Ejercicios de diseño de ETL para la actualización.
- Licencia, descarga e instalación de Open Refine.
- Uso general como herramienta de limpieza de datos.
- Pre-procesado de los datos.
- Casos de interés y ejemplos.
- Licencias, descarga e instalación de Pentaho-Kettle.
- Pentaho Kettle: funciones, características, utilización.
- Auditoría y documentación.
- Ejemplos de uso.

Big Data

- ¿Qué es Big Data?
- ¿Cuándo usar Big Data?
- ¿Cómo armar una infraestructura para Hadoop?
- Ejemplo de una aplicación de Map Reduce: wordcount.
- Cómo evitar Hadoop.

Se incluye el Curso Experto en Power BI

100% BONIFICADO CON TU INSCRIPCIÓN

Experto en Power BI (On Demand)

- Datos dentro de la administración empresarial con BI
- Introducción al Power BI como herramienta clave.
- Cargar datos a Power BI.
- Utilizando Power Query
- Generación de columnas y tabla calendario para lograr una buena gestión de los datos.
- Utilizar Dax para gestionar datos.
- Diseño de informes / gráficos.
- Proyecto final.

Evaluaciones

- **Por Unidades:** Son opcionales y se corrigen.
- **Examen Final:** Mínimo 60% de respuestas correctas.
- **Proyecto Final:** Presentación y defensa.

Soporte y Recursos Adicionales

ChatBot Híbrido: Podes consultarle sobre contenidos, horarios, grabaciones y links de acceso a las clases en vivo.

Consultas con docentes: En la clase en vivo o través del grupo de Telegram que comparten docentes y estudiantes. También se pueden coordinar videollamadas.

Clases Grabadas: Disponibles para repaso. No se exige asistencia a las clases en vivo.

Grupos de Estudio: Coordinados por tutores

Más información en www.institutodatascience.org

Consultas: info@institutodatascience.org - +549 11 3914 8021