



INSTITUTO
Data Science

Diplomatura en **Python orientado a Científico de Datos**

Acuerdo Institucional con la Universidad Tecnológica Nacional.

La calidad académica de la Diplomatura en Python orientado a Científico de Datos ha sido avalada por la Resolución Nro 290/20 de la Universidad Tecnológica Nacional sede Reconquista.

Información destacada.

Certificado en Python orientado a Científico de Datos: Modalidad Online. Una clase semanal en vivo, con un fuerte componente práctico orientado a proyectos y experiencias reales para que aprendas haciendo.

Duración: 15 semanas. Dedicación recomendada: 10 Hs semanales – Carga horaria total 150 Hs. Todas las clases quedan grabadas y no se exige asistencia. La duración de cada clase es de 30 a 60 minutos dependiendo del tema, la cantidad de participantes y el nivel de intervenciones conseguido. El estudiante tiene actividad complementaria con material de soporte de las clases, material teórico para estudiar, ejercicios resueltos y para resolver. Se realizan correcciones de ejercicios y autoevaluaciones. Los grupos de participantes son reducidos y de varios países hispanoparlantes. Lo que aporta diferentes visiones y la posibilidad de networking internacional. Soporte docente por Grupo de Telegram, email y foro.

Objetivos

La Diplomatura en Python orientada a Científico de Datos está enfocada a aprender a programar en Python y su aplicación orientada al mundo de la ciencia de datos.

Aprenderás el poderoso lenguaje Python al tiempo que incorporas las técnicas de análisis exploratorio, y los paquetes que permiten crear modelos como vecinos cercanos, reglas de asociación, redes neuronales, máquinas de soporte vectorial, árboles de decisión, bayes ingenuo y discriminante tanto lineal como cuadrático.

Que los estudiantes:

- a. Entiendan los conceptos del campo.
- b. Manejen los principales algoritmos para crear modelos predictivos.
- c. Utilicen las herramientas de aprendizaje no supervisado.
- d. Sepan medir con criterio del negocio los resultados de sus modelos.
- e. Participen de las experiencias aportadas por los instructores.

Plan académico de la Diplomatura en Python orientado a Científico de Datos

Definiciones de Ciencia de Datos

- 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 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

Arboles 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

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

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

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

Instancias de evaluación

- Examen final de la diplomatura (Mínimo 60 % de respuestas correctas).
- Presentación de proyecto final con tutoría docente.
- Autoevaluación recursos de nivelación positiva (opcional).

Equipo docente

Ignacio Urteaga

Laura Siri

Susana Sarmiento

Guillermo Garofalo



Instituto Data Science SA

Contacto: info@institutodatascience.org Tel: +54 9 11 3914 8021

CUIT: 30 71719900 2

Adm. Académica: Riobamba 1059 C1116ABE

Domicilio Legal: Uriburu 660 P 2 C1027AAN

Buenos Aires - Argentina