



**INSTITUTO**  
Data Science

# **Diplomatura en Ciencia de Datos con R y Python**

---

## **Acuerdo Institucional con la Universidad Tecnológica Nacional.**

La calidad académica de la Diplomatura en Ciencia de Datos con R y Python ha sido avalada por la Resolución Nro 290/20 de la Universidad Tecnológica Nacional sede Reconquista.

## **Información destacada.**

Certificado en Ciencia de Datos con R y Python: Modalidad Online. Tres clases semanales en vivo, un fuerte componente práctico orientado a proyectos y experiencias reales para que aprendas haciendo. Duración: 25 semanas. Dedicación recomendada: 10 Hs semanales – Carga horaria total 250 Hs. Todas las clases quedan grabadas y no se exige asistencia. La duración de cada clase es de 30 a 75 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.

Aprenderás los principales algoritmos de aprendizaje supervisado y no supervisado. Entre los algoritmos supervisados verás árboles de decisión, redes neuronales, reglas de asociación, bayes ingenuo, bosques aleatorios. Y no supervisado como agrupamiento.

Descubrirás además cómo usar series temporales, el método de montecarlo, los algoritmos genéticos, y las técnicas de regresión, minería de textos y big data. Estas aplicaciones podrás utilizarlas en múltiples campos. Como negocios, salud, recursos humanos, cobranzas, finanzas, publicidad, marketing, procesos de servicios y de producción, urbanismo, políticas públicas, etc.

## 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.
- f. Realicen proyectos enfocados a sus áreas de interés.

# Plan académico de la Diplomatura en Ciencia de Datos con R y Python.

## 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
- Primeros pasos en R
- Instalando y cargando paquetes en R

## 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

- 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 como 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 RPaquete 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

## 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

Nicolás del Frate



### Contacto

[www.institutodatascience.org](http://www.institutodatascience.org)

info@institutodatascience.org

Teléfono: +54 9 11 39148021

### Información institucional

Instituto Data Science SA

CUIT: 30 71719900 2

Adm. Académica: Riobamba 1059 C1116ABE

Domicilio Legal: Uriburu 660 P 2 C1027AAN

Buenos Aires - Argentina