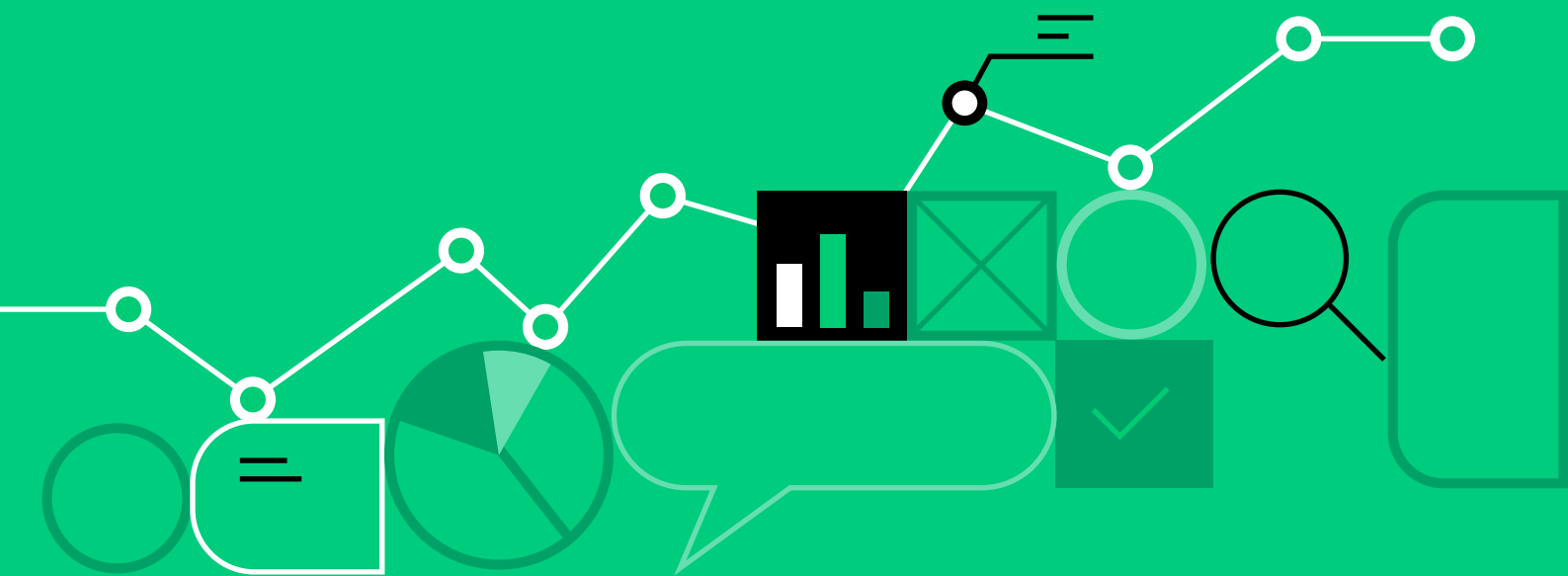


DIGITALHOUSE

# Data Science

Programa del curso



Crea modelos de aprendizaje automático **para descubrir** nuevos insights y resolver problemas de negocio complejos.

# ¿Por qué estudiar Data Science hoy?

El crecimiento tanto en la disponibilidad de datos como en las capacidades de procesamiento ha creado un panorama competitivo para las organizaciones que desean aprovechar los datos como un recurso estratégico. En ese panorama, la capacidad de analizar profundamente los datos y dar sentido a la nueva información rápidamente se ha convertido en un factor diferenciador para las organizaciones modernas.

Este programa enseña algunos modelos básicos de aprendizaje automático (Machine Learning) y la metodología de trabajo de un científico de datos para enfrentar el análisis profundo de datos. Estas herramientas son clave para poder encontrar soluciones a diversos desafíos comerciales complejos y dinámicos. El programa tiene la idea de preparar al alumnado para un rol de data scientist jr. como para complementar perfiles de análisis, investigación y profesionales en general.



#### **Modalidad**

100% a distancia.



#### **Duración**

28 semanas.



#### **Encuentros**

2 clases online  
semanales de 2:30hs  
cada uno



#### **Dedicación estimada**

Hasta 15 horas semanales  
de dedicación (incluye  
horas de cursada en vivo  
y prácticas obligatorias)

# ¿Para qué te capacitarás a lo largo del curso?

El curso está estructurado en base al ciclo de vida completo de los datos: desde su recolección y preparación, hasta su análisis y visualización efectiva. Además, hace énfasis en la utilización de los algoritmos de Machine Learning más comunes para la resolución de distintos tipos de problemas de análisis. Como estudiante, vas a aprender las habilidades y conocimientos fundamentales para desempeñarse en tu primer trabajo como Data Scientist Jr. De esta manera, al finalizar el curso tendrás la capacidad de:

- Comprender diferentes problemas de clasificación y regresión junto con los algoritmos clásicos utilizados para resolverlos.
- Trabajar con criterio a la hora de entrenar, evaluar y optimizar modelos de Machine Learning.
- Identificar problemas comerciales modernos y aplicar soluciones basadas en algoritmos para resolverlos.



**El alumnado va a aprender las habilidades y conocimientos fundamentales para desempeñarse en su primer trabajo como Data Scientist Jr.**

# Plan de estudios

---

El programa de Data Science de Digital House ofrece conocimientos sólidos en la metodología de trabajo con modelos de aprendizaje automático (Machine Learning), la base sobre la que se construyen las implementaciones tecnológicas más modernas (detección de fraude, sistemas de recomendación, clasificación de imágenes/texto, entre muchas otras). Además, el curso se enfoca en la utilización de distintas librerías de Python para la extracción, procesamiento y análisis de grandes cantidades de datos.

Trabajaremos en profundidad distintas herramientas necesarias para entrenar modelos de aprendizaje automático (Machine Learning), optimizar y evaluar sus resultados. También, desarrollaremos un proyecto de Data Science desde cero con el objetivo de aprender las habilidades y conocimientos fundamentales para que el alumnado se pueda desempeñar de forma óptima en el mercado.

El curso se compone de seis módulos de formación que agrupan los distintos temas. Cada uno de ellos está organizado para que aprendas a:

- Analizar, extraer, limpiar, combinar y visualizar datos para su análisis.
- Identificar, desarrollar y evaluar diferentes problemáticas de Machine Learning utilizando los algoritmos apropiados para buscar resolverlas.
- Introducir algunas implementaciones específicas como análisis de series temporales o clasificación de texto para sentar las bases de problemas más complejos.
- Aplicar las técnicas aprendidas a diferentes proyectos y datasets, para mostrar experiencias en distintos tipos de problemas y soluciones del mundo real.

# 01

## Fundamentos: Python, Numpy, Pandas y probabilidad

Se introduce a la dinámica de trabajo con Python utilizada en el curso. Además, se aprende a usar las librerías Pandas y Numpy para la **manipulación, limpieza y cálculos numéricos con datos.**

# 02

## Limpieza, manipulación y visualización de datos Estadística inferencial

En este módulo, se profundizan las prácticas de limpieza y manipulación de datos. Se introducen las buenas prácticas de análisis exploratorio de **datos, librerías de visualización y de estadística inferencial.**

# 03

## Introducción a Machine Learning y Scikit-Learn. Reducción de la dimensionalidad. Problemas de Regresión, Regularización y optimización de modelos

Con las bases de análisis exploratorio y limpieza de datos más consolidadas, se introduce a **la metodología de trabajo con modelos de Machine Learning**, junto a una de sus librerías más importantes, Scikit-Learn. Además, se empieza a profundizar en el primer tipo de problema a resolver del curso: regresión.

# 04

## Clasificación y Clustering. Evaluación de modelos y Pipelines

Luego de comprender los problemas básicos de regresión, **se profundiza en el segundo tipo de problema: clasificación.** Además, se hará foco en las distintas métricas de evaluación de modelos y el armado de procesos automáticos en Pipelines.

# 05

## Árboles, ensambles y Boosting

A esta altura, tendrás conocimiento de la metodología de trabajo de Machine Learning para poder resolver distintos tipos de problemas, evaluar los modelos utilizados y optimizarlos. En este módulo, se abordará algunas **técnicas más avanzadas como árboles de decisión, modelos de ensamble y boosting.**

# 06

## Series de tiempo y Text Mining

Con la metodología de trabajo de distintos modelos y diferentes técnicas útiles incorporadas, introduciremos las problemáticas encontradas en dos de las **implementaciones más comunes cuando se trabaja con datos: series temporales y texto.** Si bien estas implementaciones tienen mucha más profundidad que la abarcada por el curso, se aprenderá a incorporarlas en los análisis con distintos modelos y los problemas a tener en cuenta en estas implementaciones.

# 07

## Despliegue de modelos y APIs

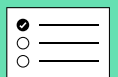
En el último módulo del curso verás cómo disponibilizar **los modelos y cómo trabajar con APIs.** A esta altura ya estarás haciendo foco fuertemente en el proyecto integrador final.

# ¿Cómo es el proceso de admisión?



## **Prework**

Material de nivelación y cursos recomendados sobre programación en Python y conceptos de estadística para reforzar conceptos necesarios para el desarrollo de la cursada.



## **Desafío**

Prueba online para evaluar las habilidades necesarias para el curso, realizada en una plataforma online (Mumuki). Los ejercicios de programación y estadística no tienen límite de tiempo ni de intentos. La aprobación de esta prueba es necesaria para la admisión final al curso de Data Science.



## **(Opcional) Curso de nivelación en Python**

Existe un curso de nivelación intensivo (2 semanas) de Python, que tiene el objetivo de nivelar las bases mínimas necesarias para aprobar la sección de programación del desafío de admisión.

# Requisitos previos

- Nociones de programación imperativa, preferentemente en Python 3. Por ejemplo: estructuras de datos, control de flujo, definición de funciones, etc.
- Conocimientos sólidos de estadística descriptiva. Por ejemplo: medidas de tendencia central (como media y mediana), desvío estándar, correlación, comprensión de histogramas y diagramas de caja, etc.
- Conocimientos de estadística inferencial. Por ejemplo: muestreo, intervalos de confianza, test de hipótesis y distribuciones de probabilidad, etc.
- Nociones de probabilidad clásica. Por ejemplo: calcular la probabilidad de eventos aleatorios simples y combinar probabilidades.
- Lecto-comprensión media del idioma Inglés. Si bien el curso de Data Science no tiene clases dictadas en inglés, existe mucho material adicional que puede ser usado de referencia o como complemento, además de estar mejor desarrollado que sus traducciones en muchos casos. Esto también tiene como objetivo presentar algunos conceptos con su nombre en inglés para facilitar que el alumnado pueda encontrar material de referencia sobre todos los conceptos abordados en el curso.

## Cursos recomendados:

Recomendamos algunos cursos para asegurarse de tener las bases necesarias tanto de experiencia en programación como de conocimiento en estadística para realizar el curso y sacarle el mayor provecho posible.

**NOTA: ¡Todos estos materiales son de acceso gratuito!** Únicamente en el caso de Educative.io tienes que registrarte y luego seleccionar "Audit this course" para acceder al curso de forma gratuita. En ningún caso deberías pagar ni tener que poner datos de tarjeta de crédito.

■ Conceptos de estadística (en Inglés):  
<https://www.udacity.com/course/intro-to-descriptive-statistics--ud827>  
<https://www.udacity.com/course/intro-to-inferential-statistics--ud201>

■ Conceptos de estadística (en Español):  
<https://es.khanacademy.org/math/probability>  
<https://es.khanacademy.org/math/statistics-probability>

■ Python 3 (en Inglés):  
<https://www.educative.io/courses/learn-python-3-from-scratch>



# Programa de desarrollo profesional

Desde el equipo de desarrollo profesional ofrecemos un programa que te ayudará a posicionar tu perfil digital en el mercado y a desarrollar las habilidades blandas necesarias para potenciar tu inserción en el mundo laboral.

El programa comienza con un curso online 100% a distancia que te permitirá conocer aquellos temas que promuevan tu desarrollo profesional, tales como la creación de un CV tech, la construcción de tu marca personal, estrategias de posicionamiento en LinkedIn, entrevistas laborales, portafolios, trabajo freelance, entre otros. Además, podrás acceder a nuestra bolsa de trabajo, participar de webinars con las principales empresas referentes de la industria digital y recibir el asesoramiento de nuestro equipo de profesionales.

Este servicio es opcional y gratuito. Será brindado a quienes cumplan con todos los requisitos de aprobación por un plazo de hasta 6 meses desde la finalización del curso.

Digital House, creamos talento digital. Te invitamos a formar parte de nuestra comunidad y a desarrollar las competencias digitales claves para liderar tu transformación profesional.

## ¡El cambio comienza hoy!

DIGITALHOUSE

# Requerimientos técnicos

A continuación vas a encontrar el listado de los requerimientos técnicos necesarios para poder utilizar los programas usados en el curso:

## ➤ Anaconda

- Computadora con sistema operativo de 64 bits (Windows 8 o mayor, macOS High Sierra 10.13+, o Linux, incluyendo Ubuntu, RedHat, CentOS, etc. 6+).
- Al menos 5 GB de espacio libre en disco local para la instalación.
- Al menos 6 GB de RAM.

Fuente: <https://docs.anaconda.com/anaconda/install/>

Los cambios en los requerimientos técnicos de las herramientas y/o los sistemas operativos soportados no son responsabilidad de Digital House y quedan a discreción de los desarrolladores de las mismas.

# Equipamiento recomendado

---

A continuación vas a encontrar el listado de requerimientos y recomendaciones para que puedas acceder a las salas virtuales y aprovechar al máximo las clases a distancia.

## ➤ **Conexión a internet**

De ser posible mediante cable de red, sino inalámbrica (Wi-Fi). Ancho de banda mínimo, 25Mbps (carga/descarga), recomendado 50Mbps en adelante. Se puede verificar la velocidad de tu conexión mediante FAST o Speedtest. Un micrófono / headset. Integrado en tu equipo, USB o inalámbrico Bluetooth.

## ➤ **Auriculares**

Aunque tu equipo tenga parlantes es importante que utilices auriculares para no generar feedback dentro de la sala virtual. Para evitar esto, otra opción es que silencies el micrófono (sin transmitir sonido) en todo momento, pero esto prevendría que puedas interactuar con la clase. Los auriculares pueden ser cableados al equipo (por Jack 3.5mm, o USB) o inalámbricos bluetooth. Recomendaciones de micrófonos USB de mejor calidad: Jabra, Plantronics, Logitech. Webcam / Cámara Integrada en tu equipo, o una que se conecta a tu equipo por USB. Recomendaciones de Webcam USB de mejor calidad: Logitech y Microsoft HD.

Sistemas operativos compatibles con Zoom:  
Mac OS X with MacOS 10.6.8 //(Snow Leopard) en adelante/ Windows 10 / Windows 8 or 8.1 Windows 7 /Windows Vista with SP1 en adelante Windows XP with SP3 en adelante  
Ubuntu 12.04 en adelante / Mint 17.1 en adelante  
Red Hat Enterprise Linux 6.4 en adelante  
Oracle Linux 6.4 en adelante/ CentOS 6.4 en adelante  
Fedora 21 en adelante/ OpenSUSE 13.2 en adelante  
ArchLinux (64-bit only)

Tablets y dispositivos móviles compatibles con Zoom:  
Equipos con iOS o Android.  
Windows Surface PRO 2 corriendo Win 8.1  
Windows Surface PRO 3 corriendo Win 10  
Equipos Blackberry

Browsers compatibles con Zoom:  
Windows: IE7+, Firefox, Chrome, Safari5+  
Mac: Safari5+, Firefox, Chrome  
Linux: Firefox, Chrome

# DIGITALHOUSE



Digital House se reserva el derecho de modificar este programa o cualquiera de los componentes del curso, sin previo aviso, cuando lo considere necesario.