

BOOTCAMP

# DATA SCIENTIST



# ¿Por qué debo llevar este Bootcamp?

La demanda de profesionales en torno a la ciencia de datos y al machine learning viene en aumento y continuará con esta tendencia en los próximos años, ya que quedó demostrado que los datos son un gran aliado empresarial como soporte al diseño de estrategias de negocios y a la toma de decisiones en entornos de alta incertidumbre.

Por ello DMC Perú, presenta el "Bootcamp Data Scientist" que te permitirá incursar e iniciar tu carrera en este fascinante mundo, a través del desarrollo de casos prácticos sobre las diferentes técnicas de regresión de datos, de clasificación binaria y de clústerización, y con ello iniciar tu portafolio de proyectos.



## Información académica

 **Horas:** 52 horas académicas

 **Modalidad:** Online en vivo



**Certificación:**

Por DMC:

- Por haber aprobado el Bootcamp Data Scientist



## ¿A quién está dirigido?

- Egresados de las carreras de sistemas, informática, desarrollo de software y afines que deseen incursionar en el campo de la ciencia de datos y el machine learning.

## Requisitos

- Contar con conocimientos en programación de preferencia en lenguaje Python.
- Contar con una laptop o computadora de escritorio con disponibilidad de micrófono y cámara web.
- Tener instalado los softwares y herramientas señalados en la sección Malla Curricular.

## ¿Qué aprenderás en el Bootcamp?

- A aplicar la metodología SCRUM a través de sus principales roles y artefactos a proyectos relacionados a datos.
- A emplear el lenguaje de programación Python, con énfasis en la librería Pandas, para la manipulación de datos a través de Dataframes; así como para realizar análisis exploratorio básico de datos incluyendo su representación tabular, su representación gráfica, su selección manual y su muestreo (división train y test).
- A implementar modelos de clasificación binaria de datos mediante las técnicas árboles de decisión (Decision-trees), vecinos cercanos (KNN), máquinas de soporte vectorial (SVM y Naive Bayes, así como a evaluar su rendimiento y capacidad de predicción mediante matriz de confusión y sus principales métricas asociadas. Conocerás cómo operan las técnicas de aprendizaje no supervisado para clusterización como K-Means y DBScan, mediante simuladores y su implementación básica.

# Malla Curricular

## TALLER DE PROJECT & AGILE

### 1. Project & Agile

- Dinámica de sociabilización.
- Formación de equipos.
- SCRUM Immersion. Definición, principales roles y artefactos.
- Agile Team Simulation.

## DATA ANALYSIS & PRE-PROCESSING

### 2. Introducción al Machine Learning y sus herramientas

- Conceptos generales: Inteligencia artificial, Ciencia de datos y Aprendizaje de máquinas (machine learning).
- Herramientas para implementación de proyectos de machine learning. IDEs vs. Notebooks.
- **Taller:** Exploración del entorno de trabajo de Google Colab. Carga de archivos (datasets).
- **Taller:** Estructura típica de un Notebook para proyectos de machine learning.
- **Taller:** Configurar Google Colab para trabajo colaborativo.

### 3. Fundamentos de Python aplicado a datos

- Python. Definición, justificación de su uso en ciencia de datos.
- **Taller:** Importación y uso de librerías en Python.
- **Taller:** Declaración, escritura y lectura de variables en Python.
- **Taller:** Operaciones aritméticas básicas con variables y constantes en Python.
- **Taller:** Declaración, escritura y lectura de Dataframes en Python.

### 4. Análisis exploratorio de datos

- **Taller:** Uso de Python para carga de un dataset. Descripción estadística e identificación de sus variables y tipos de datos.
- **Taller:** Uso de Python para la representación tabular y gráfica de variables (frecuencias).
- **Taller:** Uso de Python para la selección de variables de un dataset. Uso de corchetes, loc, iloc.
- **Taller:** Uso de Python para muestreo básico de datos. Criterios para dividir el dataset en data-training y data-testing.

# Malla Curricular

## MACHINE LEARNING FUNDAMENTALS

### 5. Introducción a las técnicas de regresión

- Revisión de conceptos: Linealidad de datos y la función lineal, variable dependiente e independiente, tipos de datos asociados a la regresión.
- Tipos de regresión: Simple, múltiple y polinomial. Formulación matemática y representación visual.

• **Taller:** Ajuste de un modelo de regresión lineal simple con Python. Representación visual, análisis e interpretación de las métricas del modelo.

• **Taller:** Ajuste de un modelo de regresión lineal múltiple con Python. Análisis e interpretación de las métricas del modelo.

• **Taller:** Ajuste de un modelo de regresión polinomial con Python. Análisis e interpretación de las métricas del modelo.

• **Taller:** Predicción empleando modelos de regresión con Python.

### 6. Introducción a las técnicas de clasificación binaria

- Decision-Tree. Componentes gráficos de un árbol de decisión. Descripción del índice de Gini, de entropía y de ganancia de información. Algoritmos para la implementación de árboles, tipos de datos asociados.

• **Taller:** Ajuste de un modelo de clasificación binario basado en Decision-Tree con Sklearn.tree (Python). Representación visual e interpretación.

- KNN. Modo de operación, descripción de la distancia euclidiana. Tipos de datos asociados.

• **Taller:** Ajuste de un modelo de clasificación binario basado en KNN con Sklearn.neighbors (Python). Representación visual e interpretación.

- Support Vector Machine (SVM). Modo de operación, descripción de vector, plano e hiperplano.

• **Taller:** Ajuste de un modelo clasificación binario basado en SVM con Sklearn.svm (Python). Representación visual e interpretación.

- Naive Bayes. Modo de operación, descripción del teorema de Bayes.

• **Taller:** Ajuste de un modelo de clasificación binario basado en Naive Bayes con Sklearn.naive\_bayes (Python). Análisis e interpretación de resultados.

### 7. Evaluación básica de modelos de clasificación binario

• **Taller:** Predicción empleando modelos de clasificación binarios con Python.

• **Taller:** Evaluación de modelos de clasificación binario mediante matriz de confusión y métricas básicas: Accuracy, Precision, Recall, F1-score.

# Malla Curricular

## 8. Introducción a las técnicas de clusterización

- Enfoque basado en particionamiento. Definición y algoritmos.
- K-Means. Definición, descripción del algoritmo de cálculo, casos de uso. Análisis del algoritmo K-Means mediante un simulador on-line.

• **Taller:** Agrupamiento de datos con Sklearn.cluster.KMeans (Python). Representación visual e interpretación de resultados.

- Enfoque basado en densidad. Definición y algoritmos.
- DBScan. Definición, descripción del algoritmo de cálculo, casos de uso. Análisis del algoritmo DBScan mediante un simulador on-line.

• **Taller:** Agrupamiento de datos con Sklearn.cluster.DBScan (Python). Representación visual e interpretación de resultados.

## DATATON

## 9. Dataton

- Sesión 1. Planteamiento del caso a resolver, asesoría y acompañamiento.

- Sesión 2. Presentación y sustentación de entregables.

## Metodología DMC



### Aprende Haciendo

Desarrolla casos con datos reales, incluso puedes proponer casos de tu propio sector.



### Clases en Vivo

El 100% de las clases que se desarrollan en el programa son en vivo.



### Asesoría Académica

Resuelve tus dudas con el asistente académico en línea.



### Plataforma E-learning

Accede en cualquier momento a materiales complementarios: videos, clases grabadas, etc.

# Docentes Expertos

Aprende con los líderes de las mejores empresas de Latam.



**Arnaldo  
Alvarado** ■■

Jefe de Gestión  
de Portafolio

**INTERBANK**

\*En caso de contingencias podría cambiar alguno de los docentes por otro profesional de similar perfil.



# Certificación

## Certificado del Bootcamp

Por aprobación del Bootcamp Data Scientist, por un total de 52 horas académicas.



## Constancias por Primer y Segundo Puesto en la Dataton

Por haber ocupado el primer / segundo puesto en la Dataton Data Scientist.



## Constancias de participación en el Bootcamp

Por haber participado en el Bootcamp Data Scientist, registrando asistencia a diversos módulos específicos.



# ¿Por qué elegirnos?

## Docentes líderes



Nuestros docentes son destacados expertos en data & analytics que lideran equipos de alto rendimiento en las empresas más grandes de Latam.

## Portafolio especializado



Tenemos el portafolio más completo con +150 capacitaciones sincrónicas y asincrónicas que se ajustan a diferentes perfiles y niveles de conocimiento.

## Metodología innovadora



Nuestra metodología "Aprende haciendo" ha logrado que nuestra comunidad de +25K profesionales en todo Latam mejoren su situación laboral.

## Reconocimiento empresarial



Las empresas worldclass de Latam confían en nosotros para acompañarlas en su transformación hacia el enfoque data driven desde hace 15 años.

## Atención personalizada



Desde que te matriculas recibirás atención y asesorías en todo lo que necesites para que aproveches al máximo tu inversión.





[www.dmc.pe](http://www.dmc.pe)