Diploma

DATA SCIENTIST

DMC | INSTITUTE

Presentación

¿Sabías que es posible anticipar (o predecir) con certeza matemática el comportamiento futuro de un indicador como las ventas, o de una persona como un cliente?

En virtud a esta realidad, DMC INSTITUTE ha diseñado el Diploma Data Scientist destinado a brindarte una capacitación integral en el campo de la ciencia de datos, desde los primeros pasos en Python, pasando por el entrenamiento y evaluación de modelos de machine learning, hasta la selección y adopción de la metodología idónea acorde a los requerimientos y recursos de cada proyecto.

Sobre este Diploma

25	96	47	01
sesiones	horas académicas	talleres prácticos	proyecto para tu portafolio

¿Cómo impulsamos tu carrera?

- Sesiones 80% **enfocadas en la práctica.**
- Enfoque en **Casos Aplicados a Negocio,** enfrentando los retos del mercado.
- Énfasis en **habilidades técnicas y blandas.**
- **Mentoría especializada** con docentes expertos.
- Acompañamiento **constante.**

¿Porqué estudiar este diploma?

¿Listo para iniciar tu carrera en ciencia de datos?

Prepárate con una capacitación que lleva desde los fundamentos del lenguaje Python hasta su aplicación en el diseño e implementación de soluciones machine learning, tanto para la predicción binaria como para la clusterización de datos.

Comprende las diferentes metodologías aplicables a la ciencia de datos y como aplicarlas a cada proyecto al que te enfrentes.

Desarrolla un proyecto integrador con el acompañamiento y asesoramiento de un experto a partir de una casuística de negocio.

DMC | INSTITUTE

¿A quién está dirigido?

• **Egresados de las carreras de estadística, matemática e ingenierías relacionadas a informática**, que actualmente estén buscando:
- Incursionar en la ciencia de datos y el machine learning como alternativa laboral o especialización profesional.
- Aplicar sus conocimientos en un campo profesional en crecimiento mediante herramientas orientadas al dato.

• **Analistas de datos y puestos afines al Dashboarding y Reporting**, que estén buscando:
- Incorporar a sus soluciones resultados de predicción o agrupamiento de datos.
- Dar el siguiente paso en su profesionalización en el manejo de los datos.

¿Cuáles son los requisitos?

Conocimientos / Habilidades	Experiencia Laboral	Tecnológicos
· Predisposición a aprender un lenguaje de programación (Python). · Capacidad de plantear problemas en función a procesos y algoritmos. · Conocimientos básicos de funciones matemáticas y estadísticas.	· Mínimo un año de experiencia trabajando con informes, reportes o bases de datos (deseable).	· 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 Contenidos.

Malla Curricular

1. Taller de marca personal y empleabilidad

- Actividad de sociabilización y contacto.
- Marca personal ¿Qué es y cómo desarrollarla?
- Empleabilidad y ser empleable. Diferencias clave.
- ¿Cómo hacer más atractivo el curriculum?
- ¿Cómo afrontar una entrevista de trabajo?

PYTHON FOR ANALYTICS

2. Introducción a Python

- Python. Definición, proceso de instalación (solo descriptivo).
- Software y editores de código para ciencia de datos.
- **Taller:** Exploración del entorno de trabajo de Google Colab.

3. Variables y estructuras de datos en Python

- Variables. Definición y tipos de datos.
- **Taller:** Declaración, escritura y lectura de variables de diferentes tipos. Operaciones aritméticas básicas.
- **Taller:** Declaración, escritura y lectura de listas. Casos de uso.
- **Taller:** Declaración, escritura y lectura de matrices. Casos de uso.
- **Taller:** Declaración, escritura y lectura de diccionarios. Casos de uso.

4. Estructuras de control en Python

- **Taller:** Uso de la condicional (IF) y sus variantes en Python.
- **Taller:** Uso de bucles (WHILE, FOR) en Python.

5. Manejo de Dataframes en Python (Pandas)

- **Taller:** Declaración, escritura y lectura de dataframes.
- **Taller:** Agregar, cambiar y eliminar elementos de dataframe.
- **Taller:** Técnicas de selección de filas y columnas en un dataframe.

ESTADÍSTICA DESCRIPTIVA PARA CIENCIA DE DATOS

6. Introducción a la estadística descriptiva

- Estadística y Data Science. Aplicaciones de estadística descriptiva y de estadística inferencial.
- Población y muestra. Técnicas de muestreo.
- Tabulación y visualización de frecuencias categóricas y numéricas.
- Introducción a las medidas de tendencia central y dispersión.
- Cuantiles.
- La correlación estadística.

Malla Curricular

7. Análisis exploratorio de datos (EDA)

- **Taller:** Uso de Python para obtener los principales estadísticos de un dataset: tendencia central, dispersión, cuantiles.
- **Taller:** Uso de Python para visualización de distribución de frecuencias categóricas y numéricas.
- **Taller:** Uso de Python para detección y recuento de valores perdidos en un dataset.
- **Taller:** Uso de Python para detección, recuento y visualización de valores atípicos en un dataset.
- **Taller:** Uso de Python para analizar la correlación de variables numéricas de un dataset.

8. Tratamiento y procesamiento de datos

- **Taller:** Uso de Python para el tratamiento de valores perdidos en un dataset.
- **Taller:** Uso de Python para el tratamiento de valores atípicos en un dataset.
- **Taller:** Uso de Python para el tratamiento de registros duplicados.
- **Taller:** Limpieza de un dataset y selección de variables basada en análisis exploratorio y correlación (con Python).
- **Taller:** Muestreo de datasets (con Python).

APRENDIZAJE SUPERVISADO Y MODELOS BINARIOS

9. Técnicas de clasificación

- El principio de linealidad en modelamiento de datos.
- El modelo más sencillo de todos: La regresión Lineal.
- **Taller:** Entrenamiento y análisis de un modelo de regresión lineal con Python.
- Regresión Logística y la función sigmoide.
- **Taller:** Entrenamiento y análisis de un modelo de clasificación binaria basado en regresión logística con Python.
- Árboles de decisión. Algoritmos para su estimación CART, C45.
- **Taller:** Entrenamiento y análisis de árboles de clasificación binaria basados en CART y C45 con Python.
- Fundamentos de evaluación de modelos de clasificación binaria. La matriz de confusión y errores tipo I y tipo II.
- **Taller:** Implementación y análisis de matrices de confusión para modelos entrenados.

10. Técnicas de clasificación combinadas (Ensembles)

- Técnicas Ensembles. Definición, algoritmos homogéneos y heterogéneos, descripción de Bagging y Boosting.
- **Taller:** Entrenamiento y análisis de un modelo estándar de clasificación binaria basado en RandomForest con Python.
- **Taller:** Entrenamiento y análisis de un modelo estándar de clasificación binaria basado en AdaBoost con Python.
- **Taller:** Entrenamiento y análisis de un modelo estándar de clasificación binaria basado en GBM.
- **Taller:** Entrenamiento y análisis de un modelo estándar de clasificación binaria basado en XGBoost con Python.

Malla Curricular

11. Evaluación avanzada de modelos de clasificación

- Métricas de evaluación de modelos: Método de cálculo e interpretación de F1-Score, Sensibilidad y Especificidad.
- Gráficas de evaluación de modelos: Interpretación de curva Precision-Recall y de curva ROC.
- Matriz de confusión y métricas de evaluación de clasificadores multiclase.

12. Técnicas de optimización de modelos de clasificación

- Sobreajuste y subajuste de modelos de clasificación. Definición, métodos para su identificación.

• **Taller:** Tratamiento de datos desbalanceados. Impacto en el modelo.
• **Taller:** Muestreo basado en Cross-validation. Impacto en el modelo.

13. Fundamentos de redes neuronales

- Principio de No linealidad en modelamiento de datos.
- De la neurona biológica a la neurona artificial: Arquitectura y componentes del perceptrón simple.
- Del perceptrón a la red neuronal. Arquitecturas ("The neural network Zoo"), y aplicaciones.

• **Taller:** Implementación de un modelo de regresión basado en perceptrón simple con TensorFlow. Caso conversión de unidades.
• **Taller:** Implementación de un modelo de clasificación binaria construido con TensorFlow y Python.

APRENDIZAJE NO SUPERVISADO Y ALGORITMOS DE CLUSTERIZACIÓN

14. Técnicas de clusterización

- Clústering de datos. Definición, objetivo, técnicas y algoritmos para su implementación.
- Exclusive clustering vs. Non-Exclusive clustering. Definiciones, casos de uso. Data regular e irregular.
- K-Means. Definición, descripción de la distancia euclidiana, métodos de selección del mejor número de clústeres (K). Análisis del algoritmo mediante un simulador on-line.

• **Taller:** Clusterización de datos empleando K-Means con Python. Representación visual, interpretación de resultados.
• **Taller:** Estimación del mejor número de clústeres para K-Means mediante los métodos visual y matemático con Python.

- DBScan. Definición, métodos de selección del mejor número de clústeres (epsilon). Análisis del algoritmo mediante un simulador on-line.

• **Taller:** Clusterización de datos empleando DBScan. Representación visual, interpretación de resultados (Python).
• **Taller:** Estimación del mejor número de clústeres para DBScan (Python).

15. Evaluación de técnicas de clusterización

- Evaluación de modelos de clusterización. Evaluación externa vs. evaluación interna.

• **Taller:** Evaluación externa de modelos de clusterización, método matricial, indice de rand e índice de Fowlkes-Mallows (Python).
• **Taller:** Evaluación interna de modelos de clusterización, coeficiente de silueta y coeficiente de Calinski-Harabaz (Python).

Malla Curricular

16. Introducción a la reducción dimensional

- Análisis de componentes principales (PCA). Definición, casos de uso.
- **Taller:** Reducción dimensional a partir de una nube de puntos 2D. Representación visual, análisis e interpretación de resultados. (Python).
- **Taller:** Reducción dimensional a partir de una nube de puntos 3D. Representación visual, análisis e interpretación de resultados. (Python).

FUNDAMENTOS DE MACHINE LEARNING SOBRE BIG DATA

17. Introducción a Big Data y Databricks

- Big Data como concepto. Definición, características.
- Big Data como tecnología. Arquitectura típica de un clúster Big Data. Tipos de archivos asociados (RDD, Parquet, JSON).
- Introducción a PySpark. Definición y su relación con Big Data.
- **Taller:** Exploración del entorno de Databricks Community.

18. Aprendizaje Supervisado y No Supervisado sobre Big Data

- **Taller:** Implementación de un pipeline de procesamiento de datos con PySpark.
- **Taller:** Entrenamiento y evaluación de modelos de clasificación binaria empleando regresión logística y bosques aleatorios, con PySpark.
- **Taller:** Clusterización de datos con PySpark.

METODOLOGÍAS PARA CIENCIA DE DATOS

19. Metodologías para ciencia de datos (marcos de trabajo)

- Metodología KDD. Fases y casos de uso.
- Metodología SEMMA. Fases y casos de uso.
- Metodología CRISP-DM. Fases y casos de uso.
- Metodología TDSP. Fases y casos de uso.
- Metodología ASUM-DM. Fases y casos de uso.
- **Taller:** Identificación de la metodología idónea para proyectos propuestos.

20. Data Science Team

- Business analyst (procesos). Funciones típicas.
- Data Scientist. Funciones típicas.
- Data Engineer. Funciones típicas.
- Stakeholders / Product Owners. Functions típicas.
- **Taller:** Elección de los miembros del Data Science Team para proyectos propuestos.

Malla Curricular

21. Agilidad en proyectos de ciencia de datos

- Integración de SCRUM y CRISP-DM.

- **Taller:** Elaboración de un Plan de Desarrollo y Despliegue de un Proyecto de Data Science.

PROYECTO INTEGRADOR

22. Proyecto integrador

- Sesión 1. Lineamientos generales y presentación de formatos para los entregables. Exposición por equipos de la idea del proyecto.

- Sesión 2. Seguimiento de avance y feedback.
- Sesión 3. Exposición y sustentación de proyectos.

Nuestra Propuesta de Capacitación

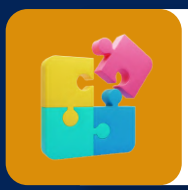
Las metodologías que aplicamos

Desarrollo de competencias clave en el mundo de los datos

Analiza · Innova · Transforma

Aprendizaje Secuencial

- Descubre conocimiento de vanguardia
- Explora con la guía del experto
- Aplica lo aprendido



Aprendizaje basado en práctica (Learning by Doing)

- Resuelve retos
- Aprende en base a proyectos
- Analiza casos

Docentes Expertos

Aprende con los líderes de las mejores empresas de Latam.

César Quezada

Científico de Datos - Seventec
en KONECTA

Luis Chacón

Sr Data Scientist Celula de Analitica Avanzada Lingo 4.0
en MARCOBRE

Nilton Yanac

Gerente de Desarrollo de Negocios
en BUSINESS ANALYTICS SAC

Claudia Aguilar

Managing Director
en OREKAN

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

Certificación DMC INSTITUTE

Por aprobación del Diploma Data Scientist, por un total de 96 horas académicas.

Por aprobación de la Especialización en Machine Learning, por un total de 40 horas académicas.

Certificación Internacional

Al finalizar la capacitación, tendrás la oportunidad de acceder al **Artificial Intelligence Professional Certification (CAIPC)** de CertiProf con un descuento especial.

Costo del examen		Precio con descuento
150 dólares	**67.33%**	**49 dólares**

IMPORTANTE:

• No es una certificación automática de CertiProf
Completar la capacitación no garantiza automáticamente el certificado de CertiProf, es un paso previo, pero no sustituye el proceso oficial de certificación.

• Aprueba la capacitación y activa tu beneficio
El primer paso para acceder al beneficio es aprobar la capacitación. ¡Superarlo te abre la puerta!

• Tú eliges si aplicar o no
El beneficio es opcional. Solo tú puedes decidir aprovecharlo

• Más de una oportunidad para alcanzar tu meta
Con el beneficio, tendrás hasta 2 intentos para rendir el examen y asegurar que tu esfuerzo sea recompensado con la certificación.

¿Por qué elegirnos?

+15 Más de 15 años de experiencia.

+300 Más de 300 empresas asesoradas en Perú, Ecuador y Bolivia.

35k 35 mil profesionales capacitados en más de 20 países de América Latina.

Propuesta integra en formación en Data & AI.

+150 Más de 150 docentes expertos de Latinoamérica, España y Estados Unidos.

Comunidad más grande en Data & AI con beneficios exlusivos: Networking, empleabilidad, habilidades blandas.

Excelente nivel de servicio.

DMC | INSTITUTE

Analiza, aplica, ***transforma***

Visita nuestra web

www.dmc.pe

Síguenos en: