> Especialización en

MACHINE LEARNING

DMC | INSTITUTE

DMC | INSTITUTE

FORMACIÓN PARA LOS NUEVOS LÍDERES DE DATA & AI

DMC INSTITUTE, según el Estudio de Adopción de la Ciencia de Datos y Brecha de Talento 2025 de Apoyo Consultoría, es la institución no universitaria

Nº1

más valorada por

Chief Data Officers y empleadores del Perú

para la formación en

CIENCIA DE DATOS

Fuente: Estudio de Adopción de la Ciencia de Datos y Brecha de Talento 2025. Apoyo Consultoría

Presentación

¿Sabías que puedes aprovechar tus conocimientos en programación y/o estadística para el desarrollo de modelos de machine learning?

En virtud a ello, DMC INSTITUTE ha diseñado la Especialización en Machine Learning que te preparará en el diseño e implementación de modelos de machine learning destinados tanto a la predicción binaria como al agrupamiento de datos (clusterización), para que puedas emplearlos como soporte a la toma de decisiones anticipada en toda industria.

Sobre esta Especialización

10
sesiones

40
horas académicas

24
talleres prácticos

¿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.**
- **Mentoría especializada** con docentes expertos.
- Acompañamiento **constante.**

¿Porqué estudiar esta especialización?

Amplía tu potencial a través de la implementación de modelos de machine learning.

Prepárate con una capacitación enfocada en el core de la ciencia de datos: los modelos predictivos y los modelos de clusterización.

Aprende con la guía y acompañamiento de un experto.

Objetivos de la especialización

- Entrena y evalúa modelos de clasificación binaria empleando el lenguaje Python y sus librerías asociadas a este propósito, considerando algoritmos como regresión lineal, regresión logística, árboles de decisión y técnicas ensemble.

- Entrena y evalúa modelos de clusterización de datos empleando el lenguaje Python y sus librerías asociadas a este propósito, considerando algoritmos como K-Means y DBScan. Así como la aplicación de la técnica de reducción dimensional.

- Entrena y evalúa modelos de machine learning (supervisados y no supervisados) sobre un data masiva o Big Data, empleando PySpark como interfaz de implementación.

Objetivo Final
Implementa proyectos de machine learning de tipos supervisados y no supervisados, acorde a las necesidades y casos de uso de cada negocio.

¿A quién está dirigido?

1. Egresados de las carreras de ingeniería, estadística y afines
Personas que cuenten con conocimientos de programación, que actualmente estén buscando:

- Aplicar sus habilidades técnicas en una disciplina en auge relacionada al entrenamiento y evaluación de modelos predictivos y de clusterización.

¿Cuáles son los requisitos?

Conocimientos / Habilidades	Experiencia Laboral	Tecnológicos
• Conocimientos básicos de programación, preferentemente Python. • Conocimientos básicos de funciones matemáticas y estadísticas.	• No requerido.	• 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.

Perfil del egresado

El egresado de la Especialización en Machine Learning estará en la capacidad de:

- Entrenar y evaluar modelos de clasificación binarios con los principales algoritmos disponibles, como regresión lineal, regresión logística, árboles de decisión, y técnicas avanzadas como los ensembles.
- Entrenar y evaluar modelos de agrupamiento o clústering de datos con técnicas como K-Means y DBScan.
- Dominar los fundamentos de PySpark.

Campo Laboral

Podrá laboral en puestos relacionados a:

- Científico de datos Jr.
- Analista de ciencia de datos.
- Machine Learning Developer.

Herramientas

Google Colab

Python

PySpark

Databricks Community

TensorFlow

Malla Curricular

I. Aprendizaje Supervisado y modelos binarios

1. 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.

2. 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.

3. Evaluación 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.

4. 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.

5. 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.

II. Aprendizaje No Supervisado y algoritmos de clusterización

6. 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).

7. Evaluación de técnicas de clusterización

· Evaluación de modelos de clusterización. Evaluación externa vs. evaluación interna.
• **Taller:** Evaluación externa de modelos de clusterización, método matricial, índice 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

8. 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).

III. Fundamentos de machine learning sobre Big Data

9. 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.

10. 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.

Certificación DMC INSTITUTE

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

DMC | INSTITUTE

METODOLOGÍA
LEARNING AGILE

Prepárate para nuevos retos, aprende con nuestro método de primer nivel y lánzate a conquistar el mundo.

DESCUBRE
Tiempo de la sesión
20% aprox.

Revisa el marco conceptual de la sesión.

Trabaja con presentaciones PPT, prototipos, material de análisis.

EXPLORA
Tiempo de la sesión
40% aprox.

Aprende con el acompañamiento del docente.

Resuelve retos, proyectos y casos.

APLICA
Tiempo de la sesión
40% aprox.

Aplica lo aprendido.

Te proveemos de material extra para completar tu aprendizaje.

DMC | INSTITUTE

¿Por qué elegirnos?

+16 Más de 16 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.

Nuestros Partners

CertiProf® | Partner

Google Partners

aws partner network

PYTHON INSTITUTE
Open Education & Development Group

Microsoft Partner

Estas empresas confían en nosotros

BBVA

ANTAMINA

PROM PERÚ

SCOTIABANK

PACÍFICO SEGUROS

SUNAT

CAJA HUANCAYO

BUENAVENTURA

PRONABEC

CAJA AREQUIPA

RIMAC

BCRP

MIBANCO

MAPFRE

ONCOSALUD

LOS ANDES

Métodos de pago

J&J DATA MINING CONSULTING S.A.C.
RUC: 20520972740

1. Depósito en cuenta BCP

- Corriente soles BCP: **193-225-1181-0-01**
- CCI BCP: **00219300225118100116**
- Corriente BCP dólares: **193-2318515-1-52**
- CCI BCP dólares: **002-193-002318515152-11**

2. Depósito en cuenta BBVA

- Ahorros BBVA soles: **0011-0177-02-00180473**
- CCI BBVA: **011-177-000200180473-37**

2. Pago Online

Generamos un link de pago online donde se acepta todas las tarjetas.

3. Pago con Yape

A nombre de J J Data Mining Consulting Sac

4. Pago online por PayPal

06 CUOTAS SIN INTERESES pagando con:

DMC | INSTITUTE

*Analiza, aplica, **transforma***

Visita nuestra web

www.dmc.pe

Síguenos en: