TEMARIO DEL CURSO

Python aplicado a Machine Learning

Python · Machine Learning · FORMACIÓN ESPECIALIZADA

25 horas | Intermedio

Formación práctica orientada al desarrollo de modelos de Machine Learning con Python. Trabaja preparación de datos, feature engineering, clasificación, regresión, clustering, evaluación y optimización de modelos mediante Pandas, NumPy y scikit-learn.

OBJETIVOS DEL CURSO

- Preparar y transformar conjuntos de datos para proyectos de Machine Learning.
- Realizar análisis exploratorio de datos mediante Python.
- Construir modelos de aprendizaje supervisado y no supervisado.
- Aplicar algoritmos de regresión y clasificación.
- Trabajar con árboles de decisión y métodos ensemble.
- Implementar técnicas de clustering y reducción de dimensionalidad.
- Crear pipelines reproducibles de procesamiento y modelado.
- Evaluar modelos utilizando métricas adecuadas para cada problema.
- Detectar problemas de overfitting y underfitting.
- Ajustar hiperparámetros y comparar diferentes modelos.
- Interpretar los resultados obtenidos por modelos de Machine Learning.

A QUIÉN VA DIRIGIDO

- Desarrolladores Python
- Analistas de datos
- Data Analysts
- Data Scientists
- Ingenieros de datos
- Ingenieros de software
- Profesionales de Business Intelligence
- Consultores tecnológicos
- Profesionales que trabajan con análisis predictivo
- Profesionales que necesitan incorporar Machine Learning a proyectos basados en datos

EL CURSO INCLUYE

- Manuales digitales
- Diploma de asistencia
- Tramitación de FUNDAE incluida

TEMARIO

MÓDULO 01.

Introducción al Machine Learning con Python

- Conceptos fundamentales de Machine Learning
- Aprendizaje supervisado y no supervisado
- Problemas de regresión, clasificación y clustering
- Flujo de trabajo de un proyecto de Machine Learning
- Ecosistema Python para ciencia de datos
- NumPy
- Pandas
- Matplotlib
- scikit-learn
- Preparación del entorno de trabajo

MÓDULO 02.

Preparación y exploración de datos

- Carga de conjuntos de datos
- DataFrames con Pandas
- Exploración de variables
- Tipos de datos
- Estadística descriptiva
- Identificación de valores ausentes
- Tratamiento de valores nulos
- Detección de valores atípicos
- Duplicados
- Distribuciones
- Correlaciones
- Visualización exploratoria de datos

MÓDULO 03.

Preprocesamiento y feature engineering

- Separación entre variables predictoras y variable objetivo
- Variables numéricas
- Variables categóricas
- Codificación de variables categóricas
- Escalado y normalización
- StandardScaler
- MinMaxScaler
- Transformación de variables
- Selección de características
- Creación de nuevas variables
- ColumnTransformer
- Preparación reproducible de datos

Fecha de descarga: 26/11/2026

MÓDULO 04.

Preparación de datos para entrenamiento

- División train/test
- Train-test split
- Conjunto de entrenamiento
- Conjunto de validación
- Conjunto de prueba
- Data leakage
- Sesgos en los datos
- Balanceo de clases
- Reproducibilidad
- Random state
- Estrategias de muestreo

MÓDULO 05.

Modelos de regresión

- Fundamentos de regresión
- Regresión lineal
- Regresión múltiple
- Entrenamiento del modelo
- Predicciones
- Evaluación de regresión
- MAE
- MSE
- RMSE
- Coeficiente R^2
- Interpretación de resultados
- Limitaciones de los modelos lineales

MÓDULO 06.

Modelos de clasificación

- Fundamentos de clasificación
- Clasificación binaria
- Clasificación multiclase
- Regresión logística
- K-Nearest Neighbors
- Entrenamiento de clasificadores
- Probabilidades de predicción
- Umbrales de clasificación

Fecha de descarga: 26/11/2026

MÓDULO 07.

Evaluación de modelos de clasificación

- Matriz de confusión
- Accuracy
- Precision
- Recall
- F1-score
- Sensibilidad y especificidad
- Curva ROC
- AUC
- Selección de métricas según el problema
- Clases desbalanceadas
- Comparación entre modelos

MÓDULO 08.

Árboles de decisión y métodos ensemble

- Árboles de decisión
- División de nodos
- Profundidad del árbol
- Criterios de división
- Importancia de características
- Overfitting en árboles
- Random Forest
- Bagging
- Métodos ensemble
- Gradient Boosting
- Comparación de modelos basados en árboles

MÓDULO 09.

Aprendizaje no supervisado

- Fundamentos del aprendizaje no supervisado
- Clustering
- K-Means
- Selección del número de clusters
- Método del codo
- Silhouette Score
- Interpretación de clusters
- Casos de uso de segmentación
- Escalado previo al clustering

MÓDULO 10.

Reducción de dimensionalidad

- Problemas de alta dimensionalidad
- Selección frente a extracción de características
- Principal Component Analysis
- PCA
- Varianza explicada
- Componentes principales
- Visualización de datos multidimensionales
- Aplicación de PCA antes del modelado

MÓDULO 11.

Pipelines de Machine Learning

- Concepto de pipeline
- Pipeline de scikit-learn
- Integración de preprocesamiento y modelos
- ColumnTransformer
- Prevención de data leakage
- Automatización del entrenamiento
- Reutilización de procesos
- Comparación reproducible de modelos

MÓDULO 12.

Validación y optimización de modelos

- Overfitting
- Underfitting
- Bias y variance
- Validación cruzada
- K-Fold Cross Validation
- Stratified K-Fold
- Ajuste de hiperparámetros
- GridSearchCV
- RandomizedSearchCV
- Selección del mejor modelo
- Evaluación final sobre datos no utilizados

MÓDULO 13.

Interpretación y puesta en práctica

- Interpretación de predicciones
- Importancia de variables
- Coeficientes de modelos
- Análisis de errores
- Comparación entre algoritmos
- Selección del modelo adecuado
- Persistencia de modelos
- Guardado y reutilización del modelo
- Flujo de trabajo de un proyecto completo de Machine Learning