

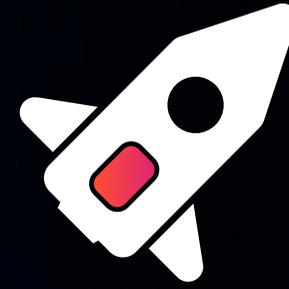
Curso TRATAMIENTO DE DATOS E INTRODUCCIÓN A MACHINE LEARNING CON PYTHON

40 horas



Universidad de
América

Código SNIES 1715



Este curso tiene como objetivo brindar herramientas avanzadas de tratamiento de datos que permitan su posterior análisis mediante herramientas básicas de machine learning.

El enfoque será práctico y con ejercicios de programación utilizando el lenguaje python mediante los frameworks pandas y sklearn; cada clase se tendrá una explicación teórica del tema a tratar y se solucionará un ejercicio utilizando como herramienta lo visto en la explicación teórica.

PRESENTACIÓN

JUSTIFICACIÓN

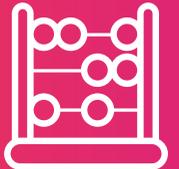


Una de las principales actividades de los científicos de datos e ingenieros es la preparación de los datos para posteriores análisis con herramientas de machine learning.



Este curso busca brindar herramientas para preparar cualquier conjunto de datos estructurados para su exploración y posterior análisis mediante herramientas básicas de machine learning. Se verán los algoritmos clásicos de machine learning con un enfoque más práctico que teórico, que permita generar resultados de una manera rápida para de esta forma iterar con otros modelos y escoger el que brinde los mejores resultados, para determinar el comportamiento del fenómeno a analizar.

En esta actividad se invierte más del 70% del tiempo en los proyectos y los principales problemas que generan errores de código se derivan del tratamiento inadecuado de los datos. Además, dentro del análisis exploratorio se pueden determinar comportamientos y características naturales de los datos antes de un análisis más avanzado.





GENERAL

Mostrar la preparación de datos como la etapa más importante antes de realizar cualquier análisis o plantear modelos de machine learning.



ESPECÍFICOS

Hacer una introducción a los algoritmos de machine learning mediante el uso de la librería sklearn.

Enseñar a los estudiantes a realizar un análisis exploratorio EDA y poder preparar los datos para algoritmos de machine learning mediante el framework pandas de Python.

OBJETIVOS



PERFIL DEL ASPIRANTE

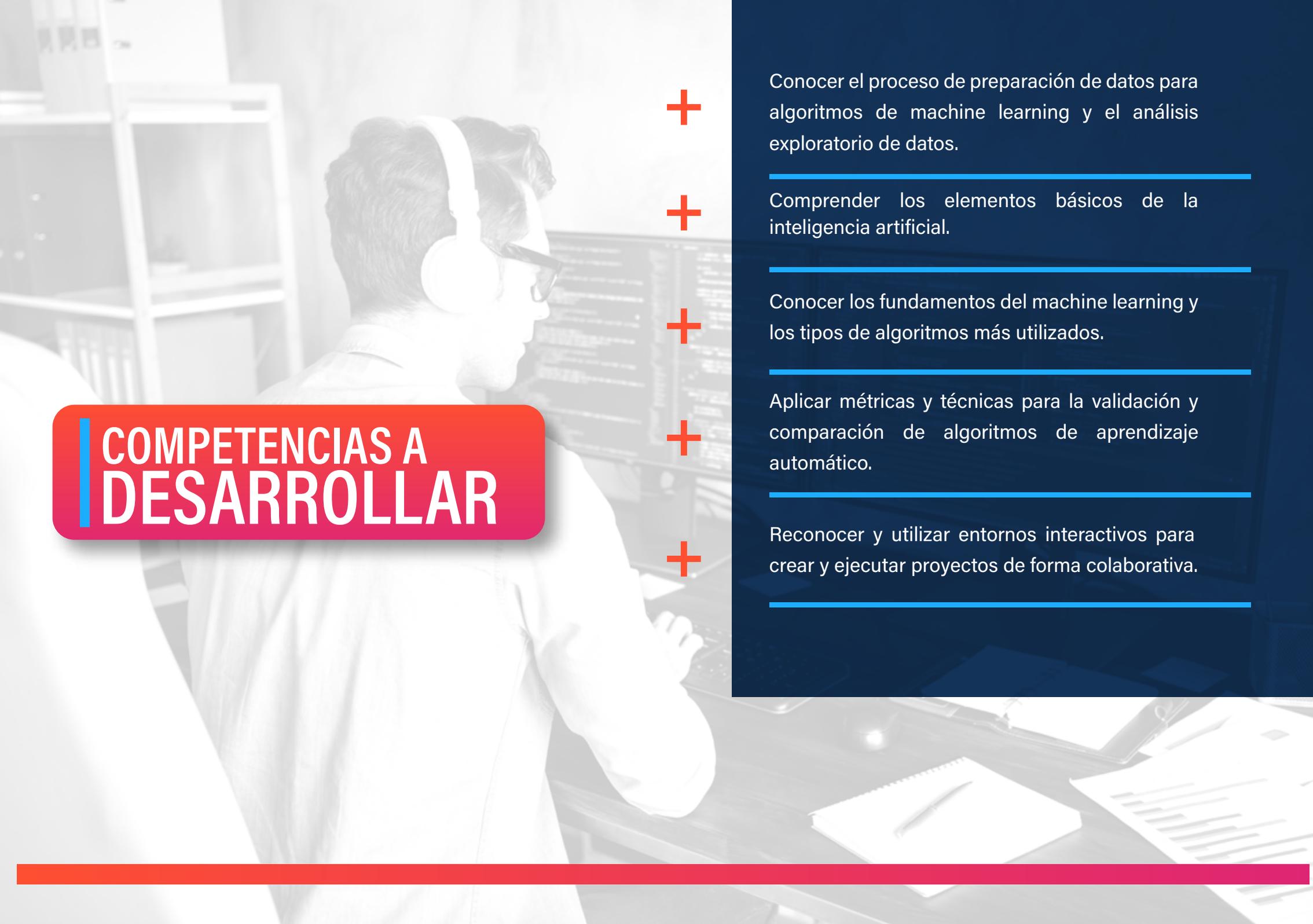
El aspirante al curso debe tener conocimientos básicos de Python y del manejo de entornos virtuales con anaconda, virtualenv o poetry; además, debe tener un conocimiento básico en estadística descriptiva para entender las métricas de evaluación de modelos, contar con 8 horas semanales: 6 de clase y dos horas para resolver los ejercicios propuestos, aproximadamente.

A man wearing a blue cap and a grey sweater is sitting at a desk in a dimly lit room. He is looking at a laptop screen which displays a grid of images under the heading "Search Photos". He has his hand to his chin, appearing to be in deep thought. The scene is overlaid with a dark blue tint.

METODOLOGÍA

La metodología será muy práctica y enfocada en el desarrollo de software donde cada clase se verá una explicación teórica del tema a tratar con un ejemplo y su solución en Python.

De igual manera, se propondrá un ejercicio corto para aplicar lo aprendido en la parte teórica: el ejercicio se resolverá de manera participativa por los estudiantes y se resolverán todas las dudas en el mismo espacio.



COMPETENCIAS A DESARROLLAR



Conocer el proceso de preparación de datos para algoritmos de machine learning y el análisis exploratorio de datos.



Comprender los elementos básicos de la inteligencia artificial.



Conocer los fundamentos del machine learning y los tipos de algoritmos más utilizados.



Aplicar métricas y técnicas para la validación y comparación de algoritmos de aprendizaje automático.



Reconocer y utilizar entornos interactivos para crear y ejecutar proyectos de forma colaborativa.



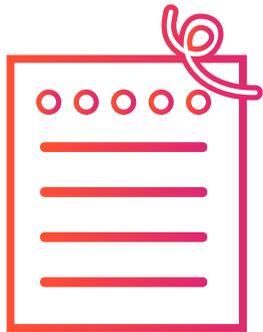
CONTENIDO TEMÁTICO

1) Módulo 1 Repaso de conceptos básicos de python (6 horas)

- Instalación de entorno Anaconda.
- Creación de entornos virtuales y el uso de los mismos.
- Programación en notebooks.
- Programación de scripts clases y métodos (conceptos básicos de programación orientada a objetos).
- Tipos de datos en Python.

2) Módulo 2 Manipulación de datos con pandas (20 horas)

- Pandas como librería para manipulación de datos.
- Lectura de archivos desde diferentes fuentes, csv, xls, SQL, txt, json y lectura desde url.
- Manipulación básica para poder entender los datos que se leyeron en el apartado anterior.
- Visualizaciones básicas para entender el comportamiento estadístico básico de los datos.
- Reemplazo de valores faltantes o fuera de rango: entender el concepto de outlier.
- Creación de conjuntos y subconjuntos condicionados.
- Agrupaciones de datos y operaciones con datos agrupados.
- Agrupaciones de varias fuentes de datos para creación de una única fuente consolidada (en esta parte se verán las operaciones de conjuntos aplicadas a la manipulación de datos).
- Correlación estadística.



CONTENIDO TEMÁTICO

3) Módulo 3 Aprendizaje supervisado (7 horas)

- Regresión lineal.
- Regresión lineal múltiple.
- Validación de la regresión lineal.
- Regresión logística.
- Árboles de decisión.

4) Módulo 4 Aprendizaje no supervisado (7 horas)

- Clustering.
- Análisis de componentes principales.
- Kmeans.



Curso

TRATAMIENTO DE DATOS E INTRODUCCIÓN A MACHINE LEARNING CON PYTHON

40 horas

www.uamerica.edu.co
www.uniamerica.edu.co

Oficina de Mercadeo y Admisiones
educacion.continua@uamerica.edu.co
PBX: (60 1) 3376680 Ext. 249, 149 y 156
PBX: (60 1) 6580658 Ext. 670, 677, 656 y 650

 (+57) 3108680919  (+57) 3108699046
 (+57) 3147269388  (+57) 3108809757

EcoCampus de Los Cerros: Avenida Circunvalar No 20 -53
Tel: (60 1) 3376680 / Fax. (60 1) 3362941
Sede Norte: Calle 106 No. 19-18
Tel: (60 1) 6580658
Bogotá D.C., Colombia.



Universidad de
América

Código SNIES 1715