

> CURSO DESARROLLADOR WEB SCRAPING CON PYTHON

DURACIÓN: **48 HORAS ACADÉMICAS**
MODALIDAD: **ONLINE**

En este curso, entenderás los algoritmos de Machine Learning que permiten simular la navegación de un ser humano en una página web con el objetivo extraer información pública relevante para su posterior análisis.

I. DIRIGIDO A

- Profesionales, egresados y estudiantes, interesados en aprender técnicas Machine Learning para la extracción de información web pública.

II. METODOLOGÍA

- El enfoque práctico-reflexivo de cada una de las sesiones y la participación a lo largo de estas favorecerá no solo el aprendizaje de contenidos, sino también, su aplicación en la resolución de situaciones reales en diferentes contextos que se planteen. Por ello, el curso considera el desarrollo de laboratorios guiados e individuales, permitiendo la consolidación del aprendizaje del tema.

III. BENEFICIOS



- Acceso a la plataforma de Microsoft Azure Lab Services, desde cualquier PC o laptop con una conexión de Internet de 2 Mbps como mínimo
- La máquina virtual estará configurada con el software y hardware necesario, con velocidad 2.1 Gbps de navegación disponible para las sesiones de clases
- Correo institucional de Cibertec
- Office 365 Web (Word, Excel, Power Point, etc.)
- Acceso a Microsoft OneDrive y Microsoft Teams

IV. LOGROS DEL CURSO



Al finalizar el curso, el alumno será capaz de utilizar algoritmos de Machine Learning en la recopilación de información de páginas web. **Adicionalmente, estará preparado para:**

- Aplicar las herramientas de Web Scraping con Python.
- Identificar los elementos claves en una página web.
- Analizar la información web como tablas con Python.

V. CERTIFICACIONES



- Al aprobar la capacitación, obtendrás un certificado en **Desarrollador Web Scraoing con Python** a nombre de Cibertec.

CERTIFICACIONES ASOCIADAS

- Especialización en Ciencia de Datos con Python
- Actualización Desarrollo de Aplicaciones de Analítica de Datos con Python

VI. PRERREQUISITOS



- Conocimiento de lenguaje Python (nivel intermedio)
- Conocimiento de HTML (nivel básico)

CAPACITACIÓN RECOMENDADA

- Python Datascience Fundamentals

VIII. CONCEPTOS Y TERMINOLOGÍA

Data Scientist
Zen de Python
Python 3.7
Programación
Funciones
Argumentos

Operadores
Módulos
Interprete
Ámbito
Pandas
Matplotlib

Machine Learning
Random forest
scikit-learn
Algoritmos supervisados
Algoritmos no supervisados
Regresión logística

Document Similarity
Django
TF-IDF
Arboles de decisión
Distancia coseno
Clasificación

CONTENIDO TEMÁTICO

1 Fundamentos Web Scraping

Definición de Web Scraping
Fundamentos de Web Scraping
Definiciones de API
Buenas prácticas
HTML

2 Beautiful Soup

Definición
Ventajas y desventajas
Tipos de objetos
Navegando árboles
Web Crawler

3 Selenium / Scrapy

Selenium
XPath
CSS Selector
Scrapy
Splash
Playwright