



Diplomado en Ciencia de Datos

<u>Duración:</u>	<u>Horario:</u>	<u>Modalidad:</u>	<u>Fecha de inicio:</u>
96 horas	<ul style="list-style-type: none">Viernes de 6:00 pm a 9:00 pmSábado de 9:00 am a 1:00 pm	<ul style="list-style-type: none">Online7 horas a la semana síncrono1 hora de trabajo asíncronoPlataformas: Microsoft teams<ul style="list-style-type: none">UPY-virtualClassroom	17 junio 2023 <u>Fecha de término:</u> 2 de septiembre del 2023

Sobre el Diplomado

Se le proveerá entrenamiento práctico y teórico en los métodos de procesamiento y análisis de datos, haciendo énfasis en un entendimiento del funcionamiento de los algoritmos de aprendizaje. El entrenamiento práctico se realizará utilizando el lenguaje de programación Python y sus librerías: pandas, scipy, scikit-learn y tensorflow.

Objetivo General: Aprender qué es la ciencia de datos, así como el flujo de trabajo a desarrollar como científico de datos.

¿A quien va dirigido?: Profesionistas con interés para el descubrimiento de patrones a partir de los datos para tener un impacto de negocio.

Perfil de ingreso y prerrequisitos:

- Conocimiento básico de programación.



Competencias a desarrollar:

- Capacidad de aplicar las metodologías para el desarrollo de proyectos de ciencia de datos.
- Aplicar los conceptos básicos de programación en ciencia de datos.
- Realizar el análisis exploratorio de datos, incluyendo la selección pre-procesamiento y transformación de los mismos.
- Capacidad de aplicar regresión lineal y árboles de regresión a problemas de datos.
- Capacidad de aplicar diferentes técnicas de clasificación a problemas de datos que requieran la predicción de pertenencia a una o múltiples clases.
- Capacidad de diseñar y aplicar redes neuronales profundas para problemas de clasificación y regresión.
- Capacidad de realizar agrupamiento sobre conjuntos de datos y evaluar la calidad del mismo.
- Capacidad de reconocer aspectos prácticos relevantes al implementar proyectos de ciencia de datos.

Plan de estudios

Ø Módulo 1: Introducción a la ciencia de Datos (19 hrs).

Descripción: En este módulo se abordarán los conceptos previos para abordar las técnicas de ciencia de datos, incluyendo las perspectivas de negocios, introducción a Python para ciencia de datos, pandas y conceptos de estadística y visualización de datos.

Introducción a la Ciencia de Datos (4 hrs.)

Python para ciencia de datos (8 hrs.)

Estadística y visualización 1 (5 hrs.)

Estadística y visualización 2 (5 hrs.)



Ø Módulo 2: Preprocesamiento y extracción de Características (15 hrs.)

Descripción: En este módulo se abordarán las técnicas para el conocimiento y consolidación de los datos, así como su limpieza y preprocesamiento.

Análisis exploratorio de datos (5 hrs.)

Limpieza y preprocesamiento de datos (5 hrs.)

Selección y extracción de atributos (5 hrs.)

Ø Módulo 3: Regresión (11 hrs.)

Descripción: En este módulo se estudiarán las técnicas para realizar regresión lineal simple y múltiple en conjuntos de datos que necesiten la predicción de variables reales.

Regresión Lineal (8 hrs.)

Regresión por árboles (3 hrs.)

Ø Módulo 4: Clasificación (11 hrs.)

Descripción: En este módulo se abordarán diferentes técnicas de clasificación, para realizar la predicción de datos que pertenezcan a un grupo de clases. Se analizarán las técnicas de K-vecinos más cercanos, algoritmos bayesianos, máquinas de soporte vectorial y bosques aleatorios.

K-nearest neighbors (3 hrs.)

Máquinas de soporte vectorial (5 hrs.)

Random forest (3 hrs.)



Ø Módulo 5: Deep Learning (13 hrs.)

Descripción: En este módulo se estudiarán los algoritmos de redes neuronales, su funcionamiento y optimización por medio de descenso de gradiente, así como su aplicación a clasificación de imágenes por medio de redes convolucionales.

Redes Neuronales (5 hrs.)

Deep Learning (5 hrs.)

Redes Convolucionales (3 hrs.)

Ø Módulo 6: Clustering (10 hrs.)

Descripción: Se analizarán diferentes técnicas de aprendizaje no supervisado y su aplicación en problemas de agrupamiento, incluyendo K-means, clustering jerárquico y DB-Scan. De igual manera se analizarán métricas para la medición de la calidad del clustering.

K-means (2 hrs.)

Clustering jerárquico (3 hrs.)

DB-Scan (3 hrs.)

Calidad de clustering (2 hrs.)

Ø Módulo 7: Aspectos prácticos de Machine Learning (17 hrs.)

Descripción: Se abordarán aspectos prácticos del ML, incluyendo la gestión de proyectos y de datos, la comparación de infraestructura en la nube para proyectos de ML y consejos sobre regularización y despliegue de los modelos. De igual manera los participantes desarrollarán un proyecto integrador



aplicado a su propia información, utilizando los conceptos y técnicas estudiadas a lo largo del diplomado.

Aspectos prácticos de ML (5 hrs.)

Proyecto Integrador (12 hrs.)

Sobre el instructor

Nombre del instructor: *Mario Alejandro Campos Soberanis*

Reseña: Licenciado y Maestro en Ciencias de la Computación por la Universidad Autónoma de Yucatán, se ha desempeñado como desarrollador de software en la industria pública y privada especializándose en desarrollo web y de productos usando aprendizaje automático. Fungió como director de tecnología de SoldAI. Cuenta con diplomados en Diseño e Implementación de Intranets y Seguridad Informática, así como diferentes cursos y certificaciones en inteligencia Artificial incluyendo especializaciones en Aprendizaje Profundo, Procesamiento de Lenguaje Natural, Redes Neuronales Adversarias Generativas, entre otras. Tiene certificaciones como desarrollador profesional de Tensorflow por coursera, Ingeniero de aprendizaje profundo por Workera y DeepLearning practitioner por IBM. Ha participado como ponente en eventos nacionales e internacionales de emprendimiento e Inteligencia Artificial. Ha ejercido como docente a nivel universitario en instituciones públicas y privadas. Cuenta con publicaciones científicas en los ramos de Procesamiento de Lenguaje Natural y Tutores Inteligentes Conversacionales. Actualmente funge como Director de Investigación en la empresa SoldAI, así como docente en la Universidad Politécnica de Yucatán enseñando asignaturas de Sistemas predictivos y Procesamiento de lenguaje natural.

Datos para la admisión



(999) 316 71 53

upy@yucatan.gob.mx

www.upy.edu.mx

Carr. Mérida-Tetiz. Km.4.5, Ucu, Yucatán, México



- **General:** 9,500
- **Canieti/UNAM:** 7,600
- **Comunidad UPY:** 6650

Contacto:

Correo: genesis.amaya@upy.edu.mx

Teléfono: (999) 916 71 53 Ext. 108



☎ (999) 316 71 53

✉ upy@yucatan.gob.mx

www.upy.edu.mx

Carr. Mérida-Tetiz. Km.4.5, Ucu, Yucatán, México