



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Instituto de
Investigación en
Ciencias
Básicas y
Aplicadas



Centro de
Investigación en
Ciencias

INSTITUTO DE INVESTIGACIÓN EN CIENCIAS BÁSICAS Y APLICADAS
CENTRO DE INVESTIGACIÓN EN CIENCIAS

"2019, a 100 años del asesinato del General Emiliano Zapata Salazar"

DIPLOMADO EN CIENCIA DE DATOS CON PYTHON

(CInC - UAEM 2019)

I. Información general

- **Dependencia que lo propone:** Centro de Investigación en Ciencias, Universidad Autónoma del Estado de Morelos.
- **Área Académica:** Departamento de Ciencias Computacionales
- **Coordinador académico:** Dr. Jorge Hermosillo Valadez
- **Modalidad:** Presencial.
- **Duración:** 143 horas.
- **Periodo:** Del 4 de octubre al 7 de marzo.
- **Horario:** Viernes de 17:00 a 20:00 y sábados de 9:00 a 14:00.
- **Dirigido a:** Profesionales, egresados y estudiantes de últimos semestres de áreas afines a las Ciencias Computacionales y al reconocimiento de patrones, como ingenierías, informática, economía, matemáticas aplicadas, ciencias de la salud, etc.



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Instituto de
Investigación en
Ciencias
Básicas y
Aplicadas



Centro de
Investigación en
Ciencias

"2019, a 100 años del asesinato del General Emiliano Zapata Salazar"

II. Objetivos del diplomado

El **Diplomado en Ciencia de Datos con Python** que ofrece el Centro de Investigación en Ciencias tiene como objetivo presentar los fundamentos, herramientas y métodos para realizar procesos de análisis de datos y así extraer información práctica. El desarrollo de esta tarea tiene aplicaciones en tanto en los sectores académicos como en los empresariales e industriales.

Este diplomado se divide en 6 módulos: Introducción a Python; Análisis estadístico de datos; Pre-procesamiento y Normalización; Clasificación y Agrupamiento; Aprendizaje profundo y Taller experimental. Durante los 6 módulos en la resolución de ejemplos y ejercicios se hará uso de librerías específicas que tiene Python para la exploración de datos y el aprendizaje de máquina.

III. Requisitos

- Se recomienda a los interesados tener habilidades y conocimientos básicos en matemáticas y programación.
- Traer su propia laptop con carga al 100%.
- Tener una **cuenta activa en Google**.

IV. Módulos

1. **Programación en Python (24 horas)**
 - 1.1. Instalación de Python y librerías. (1 hora)
 - 1.2. Variables y tipos de datos. (1 hora)
 - 1.2.1. Tipos de datos
 - 1.2.2. Listas
 - 1.2.3. Diccionarios
 - 1.3. Control de flujo de programa. (2 horas)
 - 1.4. Funciones. (2 horas)
 - 1.5. Clases, objetos y módulos en Python (2 horas)
 - 1.6. NumPy (8 horas)
 - 1.7. Pandas (6 horas)
 - 1.8. Matplotlib (2 horas)



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Instituto de
Investigación en
Ciencias
Básicas y
Aplicadas



Centro de
Investigación en
Ciencias

"2019, a 100 años del asesinato del General Emiliano Zapata Salazar"

2. **Análisis estadístico de datos. (32 horas)**

- 4.1. Estadística Descriptiva. (4 horas)
- 4.2. Teoría muestral. (4 horas)
- 4.3. Estimación estadística. (8 horas)
- 4.4. Decisión estadística. (8 horas)
- 4.5. Aprendizaje estadístico. (8 horas)

3. **Pre-procesamiento y Normalización. (24 horas)**

- 3.1. Introducción al Pre-procesamiento. (30 min)
- 3.2. Visualización. (2horas)
 - 3.2.1. Gráficas de Dispersión.
 - 3.2.2. Histogramas.
 - 3.2.3. Diagramas de Caja.
- 3.3. Normalización de Variables Independientes. (8horas)
 - 3.3.1. Variables Categóricas a Numéricas.
 - 3.3.2. Clasificación.
 - 3.3.2.1. Weight of Evidence.
 - 3.3.3. Regresión.
 - 3.3.3.1. z-score.
 - 3.3.3.2. Normalización Logarítmica.
 - 3.3.4. Pre-procesamiento en Texto.
 - 3.3.4.1. Matriz de Frecuencia y de Frecuencia Inversa.
 - 3.3.4.2. Visualización: Word Clouds.
- 3.4. Identificación y tratamiento de outliers. (2 horas)
- 3.5. Tratamiento de datos faltantes. (4 horas)
- 3.6. Selección de Variables. (4:30 horas)

4. **Clasificación y Agrupamiento. (24 horas)**

- 4.1. Fundamentos de la clasificación. (3 horas).
 - 4.1.1. El problema de la clasificación.
 - 4.1.2. Modelos lineales de clasificación.
 - 4.1.3. Modelo Perceptrón.
 - 4.1.4. Exactitud, Precisión y Sensibilidad.
- 4.2. Regresión logística (3 horas)
 - 4.2.1. Regresión logística
 - 4.2.2. Descenso de gradiente



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Instituto de
Investigación en
Ciencias
Básicas y
Aplicadas



Centro de
Investigación en
Ciencias

"2019, a 100 años del asesinato del General Emiliano Zapata Salazar"

4.2.3. Función de costo regularizada

4.3. Máquinas de soporte vectorial. (2 horas)

4.3.1. Vectores de soporte y márgenes.

4.3.2. Clasificación con Kernels.

4.4. Árboles de decisión. (3 horas)

4.4.1. Medidas de información.

4.4.2. Algoritmo de árbol de clasificación.

4.5. Clasificador Bayesiano. (5 horas)

4.5.1. Teorema de Bayes.

4.5.2. Método de programación probabilista.

4.5.3. Bayes ingenuo.

4.5.4. Modelos de lenguaje.

4.5.5. Aplicación a clasificación de texto.

4.6. Agrupamiento (clustering). (8 horas)

4.6.1. El problema del agrupamiento.

4.6.2. Métricas de agrupamiento.

4.6.3. K-medias (K-means).

4.6.4. Clustering jerárquico.

4.6.5. Mapas auto-organizados.

5. Aprendizaje profundo. (24 horas)

5.1. Introducción al aprendizaje profundo. (3 horas)

5.1.1. Frameworks para deep learning.

5.1.2. El modelo tensorflow y keras.

5.2. Redes neuronales (MLP). (5 horas)

5.2.1. Perceptron Multi-Capa.

5.2.2. Backpropagation.

5.2.3. Redes neuronales from scratch.

5.2.4. MLP con tensorflow.

5.3. Redes neuronales de convolución. (CNN). (3 horas)

5.3.1. Arquitectura de una CNN.

5.3.2. Entrenamiento.

5.4. Secuencias y redes neuronales recurrentes (RNN). (5 horas)

5.4.1. Redes recurrentes.

5.4.2. Long-Short Term Memories.

5.4.3. Tópicos avanzados en Redes Recurrentes.



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Instituto de
Investigación en
Ciencias
Básicas y
Aplicadas



Centro de
Investigación en
Ciencias

"2019, a 100 años del asesinato del General Emiliano Zapata Salazar"

6. **Taller experimental. (15 horas)**
Resolución de problemas.

V. Proceso de Admisión

Entregar en la Oficina del Servicios de Extensión del CInC:

- 1 Fotocopia de identificación oficial.
- Pagar al menos el 50% del costo, y cubrir el resto antes del 1 de octubre de 2019.
- Original y 1 fotocopia de la ficha de depósito.
- Llenar y firmar la ficha de registro de participante.

Restricciones:

- Fecha límite para cubrir el 100% del costo: **martes 1 de octubre de 2019.**
- Cupo mínimo: 15 estudiantes.
- Cupo máximo: 40 estudiantes.
- El CInC se reserva el derecho de cancelar el diplomado si no se logra el cupo mínimo de inscritos, haciendo los reembolsos correspondientes.

Costos del diplomado:

	Externos UAEM	Comunidad UAEM
Académicos e investigadores	\$8,500	\$7,000
Estudiantes	\$6,500	\$5,000
Público general	\$9,500	NA

Datos de cuenta para realizar el pago:

Banco: Santander

No. de Cuenta: 65504841845

CLABE: 014540655048418458

Nombre de la Cuenta: UAEM CENTRO DE INVESTIGACION EN CIENCIAS



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Instituto de
Investigación en
Ciencias
Básicas y
Aplicadas



Centro de
Investigación en
Ciencias

"2019, a 100 años del asesinato del General Emiliano Zapata Salazar"

V. Criterios de acreditación

Para acreditar *cada módulo* el participante deberá tener una calificación aprobatoria de parte del instructor.

Para acreditar *el diplomado*:

- El participante deberá cursar y aprobar todos los módulos.
- Tener al menos el 80% de asistencia.
- Haber cubierto el 100% del costo del Diplomado

Si se satisfacen estos criterios, al finalizar el diplomado se otorgará un diploma por parte del Centro de Investigación en Ciencias y de la Universidad Autónoma del Estado de Morelos. El costo del diploma está incluido en la inscripción.

VI. Instructores

- **Dr. Jorge Hermosillo Valadez.** Es doctor en Robótica por el IPN Grenoble, Francia. Fue investigador en el Instituto de Investigaciones Eléctricas y director del Centro Morelense de Innovación y Transferencia Tecnológica. Actualmente es profesor investigador Titular del Centro de Investigación en Ciencias de la UAEM. Imparte los cursos de Lógica, Inteligencia Artificial, y Machine Learning en la Licenciatura y el posgrado en Ciencias. Es miembro fundador de la Academia Mexicana de Tecnologías del Lenguaje, y miembro del Sistema Nacional de Investigadores y de la Academia Mexicana de Ciencias Computacionales.



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Instituto de
Investigación en
Ciencias
Básicas y
Aplicadas



Centro de
Investigación en
Ciencias

"2019, a 100 años del asesinato del General Emiliano Zapata Salazar"

- **Dr. Dan Sidney Díaz Guerrero.** Es Licenciado en Ciencias (Matemáticas) y Doctor en Ciencias (Modelado Computacional y Cómputo Científico) por la UAEM, México. Fue investigador posdoctoral en la Unidad Académica de Física de la Universidad Autónoma de Zacatecas. Actualmente es profesor por horas en la Licenciatura en Ciencias de la UAEM, donde ha impartido los cursos de Introducción a la Computación, Ciencia y Sociedad, Matemáticas Discretas 1 y 2, Bioestadística, Probabilidad, Métodos Numéricos, Compiladores, e Introducción a la Programación Científica. También fue coordinador del Curso Propedéutico de la Licenciatura en Ciencias . Es miembro del Sistema Nacional de Investigadores.
- **Lic. Javier Alberto Pérez Garza.** Es Licenciado en Ciencias (Computación) y egresado de la Maestría en Optimización y Cómputo Aplicado por la UAEM, México. Fue investigador en el Instituto Nacional de Electricidad y Energías Limpias y ha participado como consultor en múltiples proyectos relacionados con el análisis de datos. Actualmente está involucrado en proyectos de investigación en la industria eléctrica y en el área de genómica. Como profesor, ha impartido cursos de Big Data, Deep Learning, y aprendizaje de máquinas para el Tecnológico de Monterrey y UAEM. Se especializa en estudios relacionados con sistemas distribuidos, optimización, y algoritmos de Machine Learning y Deep Learning.
- **Dr. Guillermo Santamaría Bonfil.** Es Doctor en Ciencias Computacionales por el Tecnológico de Monterrey (ITESM). Actualmente es catedrático CONACYT en el Instituto Nacional de Electricidad y Energías Limpias. Su investigación incluye enfoques de ciencia de datos para la resolución de problemas, aprendizaje automático (aprendizaje supervisado / no supervisado), modelado y pronóstico de series de tiempo, pronóstico de energía eólica, optimización metaheurística, complejidad y análisis no lineal. Además de aplicaciones de Realidad Virtual a la capacitación del personal del sector energético. Pertenece al Sistema Nacional de Investigadores.



UNIVERSIDAD AUTÓNOMA DEL
ESTADO DE MORELOS



Instituto de
Investigación en
Ciencias
Básicas y
Aplicadas



Centro de
Investigación en
Ciencias

"2019, a 100 años del asesinato del General Emiliano Zapata Salazar"

VII. Bibliografía

- Morris H. DeGroot, Mark J. Schervish. Probability and statistics. 2012, 4ta edición. Pearson.
- George Casella, Roger L. Berger. Statistical Inference. 2001, 2da edición. Duxbury Press.
- Montgomery, Peck, Vining. Introducción al Análisis de Regresión Lineal. 2002. 3a edición Compañía Editorial Continental.
- James G., Witten D., Hastie T. and Tibshirani R. An Introduction to Statistical Learning. 2013. Springer.
- Trevor Hastie, Robert Tibshirani, Jerome Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. 2017. 2nd edition. Springe
- Ian Goodfellow, Yoshua Bengio y Aaron Courville . *Deep Learning*. MIT Press. 2016.