

### Descriptor Taller

<b>Nombre taller</b>	Fundamentos de <i>Data Science</i> con Python
<b>Modalidad</b>	<input type="checkbox"/> Presencial <input type="checkbox"/> Semipresencial <input checked="" type="checkbox"/> A distancia
<b>Perfil de los Asistentes</b>	El Taller de Fundamentos de <i>Data Science</i> con Python está dirigido a docentes del Instituto de Matemática Física y Estadística que deseen adquirir conocimientos y habilidades en Ciencia de Datos, con aplicación en análisis de información y modelos de aprendizaje automático.
<b>Tipos de Perfiles</b>	<p><b>P1.</b> Principiantes:</p> <ul style="list-style-type: none"> <li>• No tienen experiencia en <i>Python</i> ni en programación.</li> <li>• Necesitan una introducción al lenguaje antes de abordar Ciencia de Datos.</li> </ul> <p><b>P2.</b> Intermedios:</p> <ul style="list-style-type: none"> <li>• Tienen conocimientos básicos de <i>Python</i> (estructuras de datos, funciones, manipulación de datos).</li> <li>• Pueden iniciar directamente con análisis exploratorio y <i>Machine Learning</i>.</li> </ul> <p><b>P3.</b> Con experiencia en <i>Python</i>:</p> <ul style="list-style-type: none"> <li>• Manejan <i>Python</i> y bibliotecas como <i>pandas</i> y <i>numpy</i>, pero pueden no haber trabajado con <i>Machine Learning</i>.</li> <li>• Buscan aplicar herramientas de Ciencia de Datos en sus áreas de enseñanza o investigación.</li> </ul>
<b>Total Horas</b>	21 horas, distribuidas en seis sesiones virtuales
<b>Objetivos de Aprendizaje</b>	<p><b>OA1.</b> Cognitivos (Conocimientos)</p> <ul style="list-style-type: none"> <li>• Comprender los conceptos fundamentales de <i>Data Science</i> y su aplicación en ingeniería y negocios.</li> <li>• Identificar el rol del <i>Data Scientist</i> y las principales herramientas utilizadas en el análisis de datos.</li> <li>• Conocer los principios básicos del <i>Machine Learning</i> y su aplicación en problemas reales.</li> </ul> <p><b>OA2.</b> Procedimentales (Habilidades)</p> <ul style="list-style-type: none"> <li>• Manipular datos con <i>Python</i> utilizando bibliotecas como <i>pandas</i> y <i>numpy</i>.</li> <li>• Realizar análisis exploratorio de datos (EDA) y generar visualizaciones con <i>matplotlib</i> y <i>seaborn</i>.</li> <li>• Implementar modelos básicos de <i>Machine Learning</i> con <i>scikit-learn</i>.</li> <li>• Aplicar técnicas de preprocessamiento y evaluación de modelos en conjuntos de datos reales.</li> </ul> <p><b>OA3.</b> Actitudinales (Valores y Actitudes)</p> <ul style="list-style-type: none"> <li>• Fomenten el trabajo colaborativo en el análisis y resolución de problemas.</li> <li>• Desarrollen una actitud crítica y reflexiva sobre la interpretación de datos.</li> <li>• Valoren la importancia de la ética en el uso de la Ciencia de Datos.</li> </ul>

<b>Estructura de Contenidos y Actividades por Clase (Número de clase y perfil asistente)</b>	
<b>Clase 1 (P1)</b>	<p><b>Contenidos:</b></p> <ul style="list-style-type: none"> <li>• Qué es <i>Data Science</i>: Conceptos básicos, aplicaciones en negocios e ingeniería.</li> <li>• El rol del <i>Data Scientist</i>: Funciones y habilidades necesarias.</li> <li>• Introducción a <i>Python</i>: Uso de <i>notebooks</i> interactivos (<i>Google Colab, CoCalc</i>) para ejecutar código en <i>Python</i> y trabajar con bibliotecas esenciales de Ciencia de Datos.</li> <li>• Manipulación básica de datos con <i>pandas</i> y <i>numpy</i>: Cargar, explorar y limpiar datos.</li> </ul> <p><b>Actividades:</b></p> <ul style="list-style-type: none"> <li>• Ejercicio práctico: Importar y explorar un conjunto de datos simple (archivo CSV).</li> <li>• Discusión grupal sobre cómo la Ciencia de Datos puede aplicarse en sus áreas de enseñanza.</li> </ul>
<b>Clase 2 (P1 y P2)</b>	<p><b>Contenidos:</b></p> <ul style="list-style-type: none"> <li>• Análisis Exploratorio de Datos (EDA): Descripción y visualización de datos.</li> <li>• Limpieza de datos: Tratamiento de valores faltantes, duplicados y tipos de datos.</li> <li>• Visualización con <i>matplotlib</i> y <i>seaborn</i>: Creación de gráficos básicos (barras, líneas, histogramas).</li> </ul> <p><b>Actividades:</b></p> <ul style="list-style-type: none"> <li>• Ejercicio práctico: Realizar un análisis exploratorio de un conjunto de datos real, identificando patrones y creando visualizaciones.</li> <li>• Discusión: Análisis de los resultados obtenidos y cómo interpretar las visualizaciones.</li> </ul>
<b>Clase 3 (P1 y P2)</b>	<p><b>Contenidos:</b></p> <ul style="list-style-type: none"> <li>• Qué es <i>Machine Learning</i>: Definición y tipos de aprendizaje (supervisado, no supervisado).</li> <li>• Regresión Lineal: Concepto y aplicación para predicción de valores continuos.</li> <li>• Clasificación: Introducción a modelos básicos de clasificación.</li> <li>• Uso de <i>scikit-learn</i>: Entrenamiento, validación y evaluación de modelos.</li> </ul> <p><b>Actividades:</b></p> <ul style="list-style-type: none"> <li>• Ejercicio práctico: Implementar un modelo de regresión lineal para predecir una variable.</li> <li>• Evaluación del modelo: Cálculo de métricas como el RMSE (<i>Root Mean Squared Error</i>).</li> </ul>
<b>Clase 4 (P1, P2 y P3)</b>	<p><b>Contenidos:</b></p> <ul style="list-style-type: none"> <li>• Qué es <i>Deep Learning</i>: Conceptos básicos y diferencias con <i>Machine Learning</i>.</li> <li>• Redes neuronales artificiales: Estructura básica y cómo funcionan.</li> <li>• Uso de <i>TensorFlow/Keras</i>: Implementación básica de una red neuronal.</li> </ul> <p><b>Actividades:</b></p> <ul style="list-style-type: none"> <li>• Ejercicio práctico: Crear un modelo simple de red neuronal para clasificar imágenes (clasificación de dígitos manuscritos usando el <i>dataset MNIST</i>).</li> <li>• Discusión: Cómo elegir el modelo adecuado y qué factores afectan su rendimiento.</li> </ul>
<b>Clase 5 (P1, P2 y P3)</b>	<p><b>Contenidos:</b></p> <ul style="list-style-type: none"> <li>• Redes Neuronales Convolucionales (CNN)</li> <li>• Arquitecturas modernas</li> <li>• Transfer Learning</li> <li>• Ética, sesgos y riesgos en modelos de visión computacional</li> </ul>

	<p><b>Actividades:</b></p> <ul style="list-style-type: none"> <li>• Ejercicio práctico guiado: Clasificar imágenes reales mediante un modelo preentrenado (MobileNetV2).</li> <li>• Actividad práctica autónoma: Experimentación con EfficientNetB0 modificando hiperparámetros seleccionados (transfer learning)</li> </ul>
<p><b>Clase 6 (P1, P2 y P3)</b></p>	<p><b>Contenidos:</b></p> <ul style="list-style-type: none"> <li>• Revisión de conceptos clave: Repaso de herramientas y técnicas aprendidas.</li> <li>• Desarrollo de un proyecto final: Aplicación de todo lo aprendido en un caso real.</li> </ul> <p><b>Actividades:</b></p> <ul style="list-style-type: none"> <li>• Trabajo en grupo: Los participantes trabajarán en grupos para desarrollar un proyecto de análisis de datos que incluya preprocesamiento, análisis exploratorio, visualización y un modelo básico de <i>Machine Learning</i>.</li> <li>• Presentación de proyectos: Cada grupo presentará su solución, destacando los resultados obtenidos y las herramientas utilizadas.</li> </ul>
<p><b>Descripción General</b></p>	<p>El Taller de Fundamentos de <i>Data Science</i> con <i>Python</i> es un curso introductorio diseñado para docentes del Instituto de Matemática Física y Estadística, con el objetivo de proporcionar conocimientos esenciales en Ciencia de Datos y su aplicación a través de <i>Python</i>. A lo largo de seis sesiones virtuales, los participantes desarrollarán habilidades en manipulación, análisis y visualización de datos, así como en la implementación de modelos básicos de <i>Machine Learning</i>. El propósito del taller es que los docentes comprendan los fundamentos del análisis de datos y su relevancia en la toma de decisiones, adquieran destrezas en el uso de herramientas clave de <i>Python</i> y fomenten un enfoque crítico y colaborativo en la resolución de problemas basados en datos.</p> <p>Desde una perspectiva conceptual, los participantes explorarán los principios fundamentales de la Ciencia de Datos, las técnicas esenciales de análisis y modelado, y el uso de bibliotecas como <i>pandas</i>, <i>numpy</i> y <i>scikit-learn</i>. En el ámbito procedural, desarrollarán habilidades para importar, limpiar y analizar datos, generar visualizaciones efectivas e implementar algoritmos básicos de <i>Machine Learning</i>. En cuanto a los aspectos actitudinales, se promoverá el trabajo en equipo, el pensamiento analítico basado en evidencia y la reflexión sobre la ética en el uso de datos.</p> <p>El curso se desarrollará en modalidad virtual mediante clases expositivas interactivas, ejercicios prácticos con <i>Python</i> aplicados a casos reales, trabajo colaborativo en la resolución de problemas y acceso a materiales complementarios.</p> <p>La evaluación del aprendizaje estará orientada a la aplicación práctica de los conocimientos adquiridos, a través de ejercicios individuales en cada sesión, discusiones sobre el análisis de datos y el desarrollo de un ejercicio final integrador, donde los docentes trabajarán en equipo para resolver un problema de Ciencia de Datos y presentar sus hallazgos en la última sesión.</p>
<p><b>Prerrequisitos</b></p>	<p><b>Conocimientos básicos de lógica y matemáticas:</b>  Se recomienda familiaridad con conceptos matemáticos fundamentales como álgebra, estadística básica y funciones.</p> <p><b>Competencia digital y manejo de herramientas informáticas:</b>  Se requiere que los participantes tengan habilidades básicas en el uso de software, instalación de programas y navegación en entornos virtuales de aprendizaje.</p>

**Bibliografía sugerida**

- What is CoCalc? : <https://doc.cocalc.com/>
- Google Colaboratory: <https://research.google.com/colaboratory/faq.html>
- NumPy Documentation: <https://numpy.org/doc/>
- Pandas Documentation: <https://pandas.pydata.org/docs/>
- scikit-learn Documentation: <https://scikit-learn.org/stable/>
- TensorFlow Documentation: [https://www.tensorflow.org/api\\_docs](https://www.tensorflow.org/api_docs)
- Keras Documentation: <https://keras.io/>
- Matplotlib Documentation: <https://matplotlib.org/stable/index.html>
- Seaborn Documentation: <https://seaborn.pydata.org/>
- Vujošić, Ž. Đ. (2021). Classification model evaluation metrics. International Journal of Advanced Computer Science and Applications (IJACSA), 12(6). <https://doi.org/10.14569/IJACSA.2021.0120670>
- Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education, 20, 38. <https://doi.org/10.1186/s41239-023-00408-3>
- Wang, G., Zhao, B., Wu, B., Zhang, C., & Liu, W. (2023). Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases. International Journal of Mining Science and Technology. 33(1), 47-59. <https://doi.org/10.1016/j.ijmst.2022.07.002>
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN Computer Science, 2, 160. <https://doi.org/10.1007/s42979-021-00592-x>