



ACM PYGROUP – Data Science I

Versión 2025-1

Daniela Cubillos – Hamilton Espitia

Sesión 1: Introducción a Data Science y Numpy

Temas:

- ¿Qué es Data Science?
- Instalación y configuración del entorno (Deepnote, Jupyter Notebooks)
- Introducción Numpy (con repaso de Python)

Puntos clave:

- Roles y habilidades en Data Science
- Ecosistema de herramientas para Data Science
- ¿Por qué se utiliza Numpy?
- Funciones fundamentales Numpy

Sesión 2: Manipulación de Datos con Pandas y Numpy

Temas:

- Introducción a Numpy y Pandas (Series y DataFrames)
- Carga de datos (CSV, API's)
- Operaciones básicas con DataFrames

Puntos clave:

- Operaciones eficientes con Numpy
- Manejo de datos con Pandas
- Funciones de agregación y groupby

Sesión 3: Visualización de Datos

Temas:

- Tipos de gráficos (básicos y avanzados)
- Histogramas
- Kernel de estimación de densidad

Puntos clave:

- Matplotlib, Plotly
- Personalización de gráficos (colores, estilos, anotaciones)



Sesión 4: Bases matemáticas de inteligencia artificial

Temas:

- Estadística descriptiva
- Distribuciones de probabilidad
- Medidas de tendencia central y dispersión
- Redes Neuronales y Perceptrones

Puntos clave:

- Aplicaciones de diferentes distribuciones en Data Science
- Detección de outliers con estadística
- Media, mediana, moda, rango y desviación estándar
- Funciones de Activación
- Importancia del cálculo en las redes neuronales

Sesión 5: Procesamiento de Datos

Temas:

- Limpieza de datos
- Codificación de variables categóricas
- Normalización y estandarización

Puntos clave:

- Tratar faltantes, outliers y duplicados
- Label encoding, ordinal encoding, one-hot encoding
- Pipelines de preprocesamiento

Sesión 7: Introducción al machine learning - Regresión

Temas:

- Regresión lineal y linealización
- Coeficientes de determinación e incertidumbre

Puntos clave:

- Supuestos de la regresión lineal
- Interpretación de coeficientes y su significancia



Sesión 8: Clasificación - Agrupamiento

Temas:

- Introducción a la clasificación
- Implementación de modelo KNN
- Evaluación de modelos (precisión, recall, F1-score)
- Introducción al agrupamiento
- Algoritmos de agrupamiento (K-means)

Puntos clave:

- Hiperparámetros en KNN y su optimización
- Curvas ROC y AUC
- Matriz de confusión y su interpretación
- Validación cruzada
- Determinación del número óptimo de clusters
- Otras técnicas de clustering (DBSCAN, clustering jerárquico)

Sesión 10: MongoDB y Docker

Temas:

- Introducción a los contenedores
- Manejo de Docker
- Introducción a mongoDB
- CRUD mongoDB

Puntos clave:

- Ventajas de la contenerización en Data Science
- Dockerfiles y docker-compose
- Modelo de datos en MongoDB (documentos y colecciones)

Sesión 6 y 9: Competencias

- Competencias en Kaggle