

Fundamentos de Data Science con Python

Sesión 6: Jornada Integradora

Dr. Julio Lopez-Nunez

Diciembre, 2025



Objetivos de la sesión.



Aplicar técnicas aprendidas (EDA, limpieza, visualización, ML).



Resolver en equipo un problema real de Data Science.



Presentar resultados del proyecto final.



Reflexionar sobre aprendizajes y dificultades.

¿Qué revisamos en este taller?

A lo largo de las sesiones pasadas del curso trabajamos el proceso completo de Ciencia de Datos utilizando Python, integrando conceptos teóricos y actividades prácticas. Durante este recorrido abordamos:



Fundamentos de Data Science y herramientas de trabajo.



Limpieza, preparación y exploración de datos reales.



Visualización e interpretación de patrones y tendencias.



Modelos supervisados de regresión y clasificación.

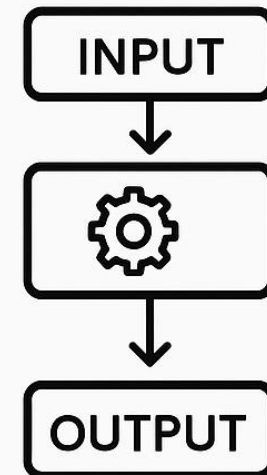
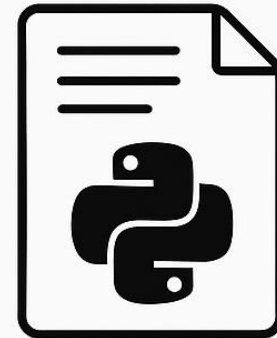
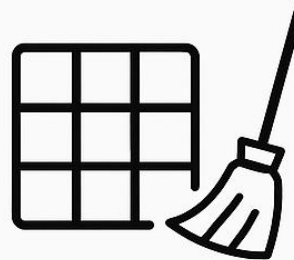
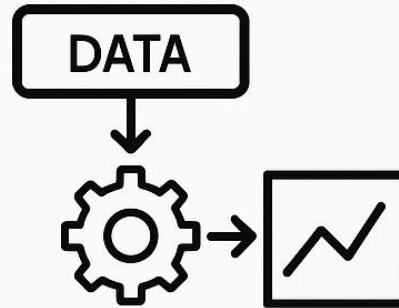


Principios de redes neuronales y Deep Learning (MLP, CNN, Transfer Learning).

Con ello, cada participante obtuvo una visión integral del análisis de datos y el uso de la IA en contextos académicos, profesionales y de investigación.

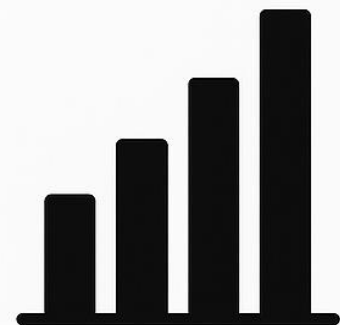
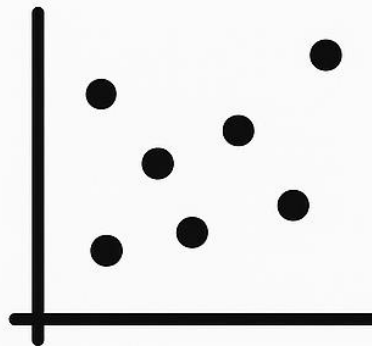
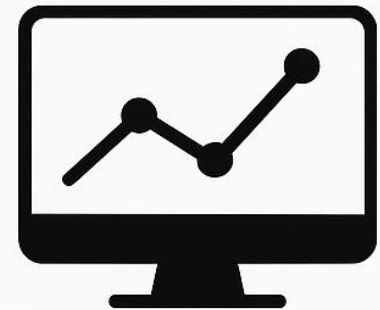
Sesión 1

Introducción



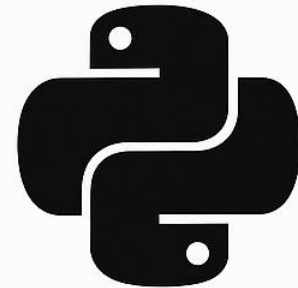
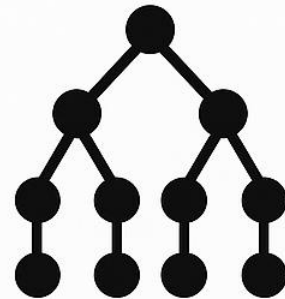
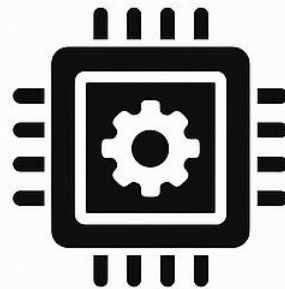
Sesión 2

Análisis
Exploratorio



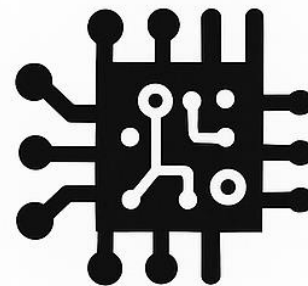
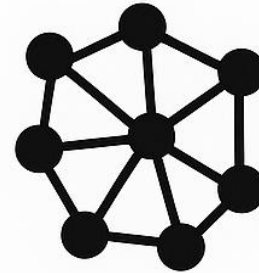
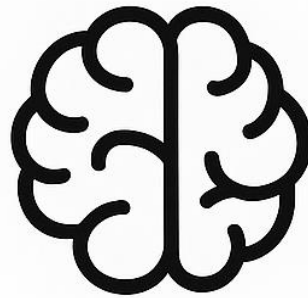
Sesión 3

Machine Learning



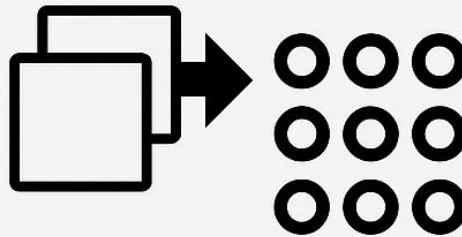
Sesión 4

Deep Learning

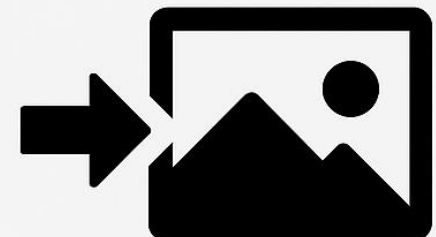


Sesión 5

Redes Neuronales



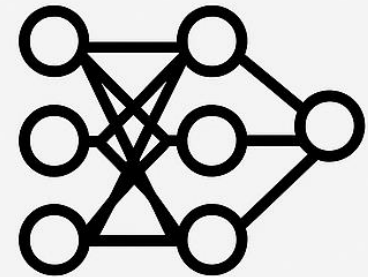
CNN



Transfer Learning



**Riesgos, sesgos
y ética de la IA**



**Conceptos avanzados
de redes neuronales**

Objetivos del Taller Integrador

- Aplicar el flujo completo de un proyecto de Ciencia de Datos (preprocesamiento, EDA, visualización y modelado).
- Seleccionar y justificar las técnicas utilizadas (pandas, numpy, visualización, ML, redes neuronales).
- Trabajar en equipo para resolver un problema práctico basado en datos reales.
- Comunicar resultados mediante una presentación clara y estructurada.
- Reflexionar sobre riesgos, sesgos y consideraciones éticas en el uso de IA

Análisis de Factores Asociados al Rendimiento Académico en Cursos STEM

1. Contexto General

El Instituto de Física y Matemáticas, al igual que otras unidades STEM, enfrenta el desafío de comprender qué factores influyen en el rendimiento académico de los estudiantes. El análisis sistemático de datos educativos permite identificar patrones, anticipar dificultades y fortalecer la toma de decisiones pedagógicas basadas en evidencia.

El Taller Integrador de la sesión 6 busca que los participantes apliquen de manera articulada los contenidos revisados en las cinco sesiones previas del curso Fundamentos de Data Science con Python, utilizando un dataset ficticio pero verosímil sobre desempeño estudiantil en cursos STEM. Este ejercicio permitirá recorrer el ciclo completo de un proyecto de Ciencia de Datos, desde el preprocesamiento hasta la construcción de un modelo predictivo basado en redes neuronales.

Análisis de Factores Asociados al Rendimiento Académico en Cursos STEM

2. Descripción del Dataset Ficticio

Se proporcionará un dataset simulado con información de aproximadamente 500 estudiantes de cursos iniciales de matemáticas, estadística y física. Cada registro representa a un estudiante e incluye:

Características generales: Sexo, Edad y Dependencia del establecimiento escolar de origen.

Indicadores académicos iniciales: Nota de diagnóstico en matemáticas,

Nivel de álgebra (autoevaluado 1–4), Promedio de actividades de laboratorio (física) y Resultados en quizzes de estadística.

Variables de comportamiento académico: Horas de estudio semanal,

Asistencia, Participación en clases, Número de entregas atrasadas, Uso de plataformas de aprendizaje y Asistencia a tutorías.

Variables objetivo: Nota final del curso (variable continua), Estado de aprobación (0 = reprobado / 1 = aprobado) y Riesgo de deserción estimado (variable ficticia 0/1).

Nota: El dataset contiene variabilidad natural, posibles valores faltantes y diferencias de escala, permitiendo aplicar técnicas de preprocesamiento, análisis exploratorio y modelado predictivo.

Análisis de
Factores
Asociados al
Rendimiento
Académico
en Cursos
STEM

3. Objetivo General del Taller

Aplicar el flujo completo de un proyecto de Ciencia de Datos utilizando Python, integrando preprocesamiento, análisis exploratorio, visualización, implementación obligatoria de una red neuronal MLP y reflexión ética sobre los modelos predictivos.

Análisis de Factores Asociados al Rendimiento Académico en Cursos STEM

4. Actividades del Taller (1)

Cada grupo deberá desarrollar las siguientes tareas en su notebook:

4.1 Preprocesamiento: Identificación y tratamiento de valores faltantes. Revisión de outliers y datos inconsistentes. Normalización o escalamiento cuando corresponda. Codificación de variables categóricas.

4.2 Análisis Exploratorio (EDA): Estadísticos descriptivos por variable. Análisis de correlación. Comparaciones entre grupos (sexo, dependencia escolar, etc.). Identificación de patrones relevantes y posibles predictores.

4.3 Visualizaciones: Generar al menos tres visualizaciones significativas, del tipo Histogramas, Scatter plots, Boxplots y Matriz de correlación.

Las visualizaciones deben ayudar a responder preguntas del EDA.

Análisis de Factores Asociados al Rendimiento Académico en Cursos STEM

4. Actividades del Taller (2)

4.4 Modelo Predictivo: Implementación Obligatoria de una MLP:

Cada grupo deberá construir y entrenar una red neuronal multicapa (MLP) utilizando TensorFlow/Keras.

Requisitos mínimos de la MLP: Una capa de entrada según las variables seleccionadas, Al menos una capa oculta, Entre 8 y 32 neuronas en la(s) capa(s) oculta(s), Activación ReLU en la capa oculta.

Activación: sigmoid o softmax si el modelo es de clasificación, linear si es de regresión.

Entrenar con: Optimizador adam.

Pérdida: mse para regresión, binary_crossentropy o categorical_crossentropy para clasificación

Entrenamiento: Entrenar entre 20 y 50 epochs

Graficas: Graficar la curva de pérdida (loss)

Tarea predictiva a elección: Regresión en caso de predecir nota final. Clasificación en caso de predecir aprobación o riesgo de deserción.

Los equipos deberán justificar: Arquitectura utilizada, Selección de hiperparámetros, Rendimiento obtenido, Limitaciones del modelo.

Análisis de Factores Asociados al Rendimiento Académico en Cursos STEM

4. Actividades del Taller (3)

4.5 Evaluación del Modelo: Presentar las métricas principales según sea el caso.

Regresión: RMSE, MAE, Clasificación: accuracy, precision, recall, F1.

Interpretar los resultados respondiendo:

¿Qué tan bien predice el modelo? ¿Dónde se equivoca? ¿Cómo se compararía con un modelo más simple?

4.6 Reflexión Ética y Sesgos: Incorporar un breve análisis sobre **a)** Sesgos potenciales en las variables (por sexo, dependencia escolar, etc.). **b)** Riesgos de usar modelos predictivos en educación. **c)** Posibles decisiones erróneas o injustas basadas en predicciones automáticas.

Análisis de Factores Asociados al Rendimiento Académico en Cursos STEM

5. Productos Finales

a) Notebook completo en Python (Google Colab)

Debe incluir: Código limpio y comentado, Gráficos EDA, MLP implementada y entrenada Métricas, Análisis final y Reflexión ética.

b) Presentación de 4 a 6 diapositivas

Debe contener: Descripción del dataset, Principales descubrimientos del EDA, Arquitectura de la MLP, Resultados y métricas, Conclusiones, Comentario ético

Análisis de Factores Asociados al Rendimiento Académico en Cursos STEM

6. Normativa y Metodología del Trabajo

Trabajo en grupos de 3 a 5 personas.

Tiempo para desarrollar la actividad: 90–120 minutos.

Tiempo de presentación: 5 minutos + 2 de preguntas.

Se permite el uso de documentación oficial de Python, pandas, numpy, sklearn y Keras.

Cada grupo debe asegurarse de que todos los integrantes participen en alguna parte del análisis.