

## Fundamento Cuaderno 1: Conceptos Básicos de EDA

<https://colab.research.google.com/drive/1QzzyFXlbG9IZv4oA0WS1hUNYVM6wzWY1?usp=sharing>

*El código realiza un Análisis Exploratorio de Datos (EDA) sobre un dataset simulado de estudiantes, utilizando Pandas, Seaborn y Matplotlib para visualizar distribuciones, relaciones y estadísticas descriptivas. No entrena un modelo predictivo, sino que analiza patrones y características de los datos.*

## Fundamento Cuaderno 2: Manejo de Valores Nulos

[https://colab.research.google.com/drive/1\\_CtHfMwSGUwfwxbEj50MANY8cwqN71t6?usp=sharing](https://colab.research.google.com/drive/1_CtHfMwSGUwfwxbEj50MANY8cwqN71t6?usp=sharing)

*Este código detecta y trata valores nulos en un conjunto de datos, utilizando técnicas como eliminación, imputación con estadísticas básicas (media, mediana, moda) y métodos avanzados como KNN o MICE para garantizar la calidad del análisis.*

## Fundamento Cuaderno 3: Manejo de Datos Categóricos

<https://colab.research.google.com/drive/1YgePba-1IXAJDfrhxgXFyLMDtCw8W7Ac?usp=sharing>

*El modelo entrenado es un **Random Forest Classifier** que predice el **género** de una persona en función de sus características físicas, como edad, peso, altura, tipo de cabello y talla de calzado.*

## Fundamento Cuaderno 4: Tratamiento de Valores Atípicos (Outliers)

<https://colab.research.google.com/drive/1WBLiAyPqTokobhW0X8DwLSGCeUPhbeSn?usp=sharing>

*El modelo entrenado identifica y trata valores atípicos en un conjunto de datos utilizando métodos estadísticos como Z-Score e IQR, para mejorar la calidad del análisis y la robustez de los modelos de machine learning.*

#### Fundamento Cuaderno 5: Creación de Nuevas Características

[https://colab.research.google.com/drive/1iXfE34LTMiSkSWOPUCdfo\\_TfiU-bidLV?usp=sharing](https://colab.research.google.com/drive/1iXfE34LTMiSkSWOPUCdfo_TfiU-bidLV?usp=sharing)

*El modelo entrenado analiza datos y crea nuevas características (feature engineering) para mejorar la calidad y precisión de futuros modelos predictivos.*

#### Fundamento Cuaderno 6: Selección de Características Fundamento Cuaderno 6: Selección de Características

<https://colab.research.google.com/drive/1F-XNQnTO7Fk5n9fSiEvT1jvXMWbCG0xy?usp=sharing>

*El modelo entrenado con Lasso (L1 Regularization) selecciona las características más relevantes para predecir la variable objetivo, eliminando aquellas con coeficientes cero para reducir la dimensionalidad y mejorar la interpretabilidad.*

#### Fundamento Cuaderno 7: Dividir los Datos de Prueba y Entrenamiento y el de Balanceo

<https://colab.research.google.com/drive/120N4u0NyP8GpEZ9khc8-Phy61fiKuTIJ?usp=sharing>

*El modelo entrenado sirve para predecir la clase de un conjunto de datos desbalanceado, aplicando técnicas como el sobremuestreo (SMOTE) o el submuestreo (RandomUnderSampler) para equilibrar las clases antes de la predicción. Esto mejora su capacidad para clasificar correctamente las clases minoritarias y evitar que se sesgue hacia las clases mayoritarias.*

#### Fundamento Cuaderno 8: Modelo Naive Bayes

[https://colab.research.google.com/drive/1h\\_obGlai\\_2s8K\\_8ZpTU5C54autjnu6V3?usp=sharing](https://colab.research.google.com/drive/1h_obGlai_2s8K_8ZpTU5C54autjnu6V3?usp=sharing)

*El modelo Naive Bayes entrenado sirve para predecir si un paciente tiene riesgo de problemas cardíacos basándose en su edad y nivel de colesterol, utilizando un*

*enfoque probabilístico rápido y eficiente. Es especialmente útil en entornos clínicos para evaluaciones preliminares o cuando se requieren predicciones rápidas con datos limitados.*

#### Fundamento Cuaderno 9: Modelo de Máquinas de Vectores de Soporte

<https://colab.research.google.com/drive/144-rIoAMPtq5BFjXRYeXPgKKmaTtv3R?usp=sharing>

*El modelo SVC entrenado **\*\*predice problemas cardíacos\*\*** basado en edad y colesterol, clasificando pacientes como **\*\*"Sí"\*\*** (riesgo alto) o **\*\*"No"\*\*** (riesgo bajo). Sirve para **\*\*diagnóstico asistido\*\*** y **\*\*evaluación preventiva\*\*** en entornos médicos.*

#### Validación Cruzada en Machine Learning Fundamento Cuaderno 9A: Cross\_Validation\_Tutorial.ipynb

[https://colab.research.google.com/drive/11TOI3Ubo\\_H3AhVz9\\_1mvyhiR\\_US7h-64?usp=sharing](https://colab.research.google.com/drive/11TOI3Ubo_H3AhVz9_1mvyhiR_US7h-64?usp=sharing)

*Este notebook cubre los conceptos fundamentales de **Cross Validation** (validación cruzada), incluyendo métodos aplicados a datos comunes y series temporales. También se incluyen ejemplos prácticos con scikit-learn.*

#### Fundamento Cuaderno 10: K-Nearest Neighbors KNN

[https://colab.research.google.com/drive/1ruRjR7ErjbaEz\\_AJTAxWQ5nsEZS\\_Mrnn?usp=sharing](https://colab.research.google.com/drive/1ruRjR7ErjbaEz_AJTAxWQ5nsEZS_Mrnn?usp=sharing)

*El modelo KNN entrenado predice problemas cardíacos basado en edad y colesterol normalizados, clasificando pacientes como "1" (con problema) o "0" (sin problema). Es útil para evaluación de riesgo clínico rápido y clasificación automatizada en entornos médicos.*

#### Fundamento Cuaderno 11: Árboles de Decisión para Clasificación

<https://colab.research.google.com/drive/1nk1qyWx1wIWhM2kKGrcxnjIsymqS1Dm9?usp=sharing>

*El modelo de Árbol de Decisión entrenado clasifica pacientes según su riesgo cardíaco (edad, colesterol, etc.) en "Sí" (problema) o "No" (sin problema). Sirve para diagnóstico asistido rápido y evaluación preventiva en entornos clínicos.*

## Fundamento Cuaderno 12: Random Forest (Bosques Aleatorios)

<https://colab.research.google.com/drive/1OjZ9-x8gG1CYITVQCGB14kYYdgb-RHm?usp=sharing>

*El modelo de Random Forest entrenado clasifica pacientes según su riesgo de diabetes (basado en glucosa, presión arterial, IMC, etc.) en "1" (diabético) o "0" (no diabético). Sirve para detección temprana y evaluación clínica automatizada en entornos médicos.*

## Fundamento Cuaderno 13: Regresión lineal múltiple

[https://colab.research.google.com/drive/1\\_gOxlmpf\\_Gvhb6VaglwjKda10GP6omb2?usp=sharing](https://colab.research.google.com/drive/1_gOxlmpf_Gvhb6VaglwjKda10GP6omb2?usp=sharing)

*El modelo de regresión lineal múltiple entrenado predice el precio de viviendas basado en características como distancia, habitaciones, baños y áreas. Sirve para valoración inmobiliaria automatizada y análisis de mercado, proporcionando estimaciones rápidas y objetivas.*

## Fundamento Cuaderno 14: Regresión logística

<https://colab.research.google.com/drive/1pHUsJW1EsH1HIy5FRgDqCP7zIcqxMg4A?usp=sharing>

*El modelo de regresión logística entrenado clasifica tumores mamarios como benignos (0) o malignos (1) basado en 10 características clave como radio, textura y concavidad. Sirve para diagnóstico médico asistido y detección temprana de cáncer, con un 95% de precisión.*

## Fundamento Cuaderno 15: XGBOOST

<https://colab.research.google.com/drive/17DIYmL-7z8OkaYEHQTfw9uqeWcoz7Hb2?usp=sharing>

*El modelo XGBoost entrenado predice diabetes basado en 8 variables clínicas como glucosa e IMC, con alta precisión (78-85%). Sirve para detección temprana y evaluación médica asistida en entornos clínicos.*

Fundamento Cuaderno 16: Kmeans

<https://colab.research.google.com/drive/1qKCjhVgfDY3IEFsUBgcnxNMgGLAYkY00?usp=sharing>

*El modelo K-means entrenado segmenta clientes en 11 grupos basados en edad, ingresos y hábitos de gasto, identificando patrones de comportamiento. Sirve para marketing dirigido y personalización de estrategias comerciales según perfiles de consumo.*

Fundamento Cuaderno 17: k-medoids

<https://colab.research.google.com/drive/1FvRrxop8tH7yUHUqkA34ThqgX2IGxSd1?usp=sharing>

*El modelo K-Medoids entrenado agrupa clientes en 14 segmentos basados en edad, ingresos y hábitos de compra, usando puntos reales (medoids) como centros. Es ideal para marketing personalizado y detección de patrones robusta, especialmente con datos ruidosos o outliers.*

Fundamento Cuaderno 18: Modelo Jerárquico Fundamento Cuaderno 18: Modelo Jerárquico

<https://colab.research.google.com/drive/1VCsmGuSz8hMKEyBc4x8EOv3JQ5BzIFIN?usp=sharing>

*El modelo jerárquico entrenado segmenta clientes en 12 grupos basados en patrones de edad, ingresos y gastos, revelando relaciones jerárquicas entre ellos. Es ideal para análisis exploratorio y estrategias de marketing personalizado, especialmente cuando se necesita visualizar la estructura de agrupamiento.*

Fundamento Cuaderno 19: DBSCAN Fundamento Cuaderno 19: DBSCAN

[https://colab.research.google.com/drive/1qISKylyPn6h4JFyLx399\\_qi98jLw-akx?usp=sharing](https://colab.research.google.com/drive/1qISKylyPn6h4JFyLx399_qi98jLw-akx?usp=sharing)

*El modelo DBSCAN entrenado identifica 16 grupos naturales de clientes basados en densidad de datos, detectando automáticamente outliers (clientes atípicos). Es ideal para segmentación de mercados complejos y detección de patrones ocultos, especialmente cuando los clusters tienen formas irregulares o densidades variables.*

Fundamento Cuaderno 20: PCA

<https://colab.research.google.com/drive/1TlfgL6aJGHFp0SrzeVGG6-dvC2i33IXX?usp=sharing>

*El modelo PCA entrenado reduce las 23 variables originales a 13 componentes principales, capturando el 81% de la varianza total. Sirve para simplificar análisis complejos y eliminar redundancias en datos multivariados, mejorando la eficiencia de modelos posteriores.*

Fundamento Cuaderno 21: LDA

[https://colab.research.google.com/drive/1WcqzDVKH1j\\_mniZET0qbqVjnm6VlVwbp?usp=sharing](https://colab.research.google.com/drive/1WcqzDVKH1j_mniZET0qbqVjnm6VlVwbp?usp=sharing)

*El modelo entrenado con LDA sirve para clasificar clientes bancarios en dos categorías ("Exited" o "No Exited") basándose en sus características financieras y demográficas, optimizando la separación entre clases para mejorar la precisión predictiva. Además, reduce la dimensionalidad del conjunto de datos conservando solo las componentes más discriminantes.*

Fundamento Cuaderno 22: t-SNE

<https://colab.research.google.com/drive/1e7nwBatACHftKxMiD0vfNIwF5Qw4-0LP?usp=sharing>

*El modelo t-SNE sirve para visualizar datos de alta dimensión (como tablas con muchas columnas) en 2D o 3D, manteniendo las relaciones cercanas entre puntos y permitiendo identificar patrones o agrupamientos. No se usa para predicción, sino para exploración y análisis visual de estructuras complejas en los datos.*

