Desafíos de Ciencia de Datos - Regresión y Clasificación con Datos Reales

Contexto: Mortalidad en centros de cultivo y rendimiento de cosechas.

🕵 RETO 1 (Regresión): Predecir cantidad de mortalidades semanales

Archivo: Mortalidades Centro de cultivos.csv

Objetivo: Construir un modelo de regresión para predecir la cantidad de **mortalidades** semanales usando edad, especie, centro, densidad, alimentación, etc.

Pasos a seguir:

- 1. Importar bibliotecas
- 2. Cargar el archivo CSV
- 3. Mostrar primeras filas y .info()
- 4. Mostrar estadísticas descriptivas
- 5. Detectar y tratar valores faltantes con la mediana
- 6. Convertir tipos de datos apropiadamente
- 7. Verificar que no queden valores nulos
- 8. Generar gráfico interactivo: Edad vs Mortalidades
- 9. Generar gráfico: Centro vs Mortalidades
- 10. Detectar outliers (rango intercuartílico IQR)
- 11. Codificar Centro, Especie, Alimentación CON OneHotEncoder
- 12. Separar X (variables predictoras) e y (mortalidades)
- 13. Dividir en datos de entrenamiento y prueba (80/20)
- 14. Agregar constante para modelo
- 15. Entrenar modelo de regresión con statsmodels.ols
- 16. Mostrar resumen del modelo
- 17. Mostrar coeficientes de variables
- 18. Predecir mortalidades en el conjunto de prueba
- 19. Evaluar con MSE y R²
- 20. Graficar reales vs predichos + guardar imagen

🥄 RETO 2 (Clasificación): ¿Fue la cosecha de alto rendimiento?

Archivo: Cosechas_Cosechas_2023.csv

Objetivo: Predecir si una cosecha fue de **alto rendimiento** o **bajo rendimiento**, basado en el peso total, número de cajas, tipo de cultivo y semana.

Pasos a seguir:

- 1. Importar bibliotecas
- 2. Cargar el archivo CSV
- 3. Mostrar primeras filas e información
- 4. Estadísticas descriptivas
- 5. Imputar valores faltantes con mediana
- 6. Convertir tipos (por ejemplo, semana como int)
- 7. Confirmar ausencia de valores nulos
- 8. Visualizar: Peso total vs Semana
- 9. Visualizar: Cultivo vs Peso total
- 10. Detectar outliers con IQR
- 11. Codificar Cultivo, Región, etc.
- 12. Crear columna rendimiento = 'alto' si peso total > Q75, 'bajo' si < Q25
- 13. Definir X e y (rendimiento_binario)
- 14. Dividir datos en entrenamiento y prueba
- 15. Agregar constante
- 16. Entrenar modelo de regresión logística
- 17. Mostrar coeficientes

- 18. Predecir si la cosecha fue alto o bajo rendimiento
- 19. Evaluar con accuracy y matriz de confusión
- 20. Graficar resultados y guardar imagen

🔍 RETO 3 (Clasificación): ¿Hubo evento crítico de mortalidad?

Archivo: Mortalidades_Centro de cultivos.csv

Objetivo: Clasificar si una semana tuvo un evento **crítico de mortalidad,** es decir, si las mortalidades superaron el percentil 90.

Pasos a seguir:

- 1. Importar bibliotecas
- 2. Cargar el archivo CSV
- 3. Visualizar estructura del dataset
- 4. Ver estadísticas de mortalidad
- 5. Imputar nulos con mediana
- 6. Convertir tipos de datos
- 7. Confirmar limpieza del dataset
- 8. Gráfico: Edad vs Mortalidad
- 9. Gráfico: Centro vs Mortalidad
- 10. Detectar outliers con IQR
- 11. Codificar variables categóricas
- 12. Crear columna evento critico = 1 si mortalidades > Q90, 0 en caso contrario
- 13. Definir X e y
- 14. Separar entrenamiento y prueba
- 15. Agregar constante
- 16. Entrenar modelo de clasificación
- 17. Mostrar coeficientes e interpretación
- 18. Realizar predicciones
- 19. Evaluar con accuracy y matriz de confusión
- 20. Graficar predicciones y guardar imagen

← RETO 4 (Regresión): ¿Cuánto tarda una cosecha en completarse?

Archivo: Cosechas Cosechas 2023.csv

Objetivo: Predecir la semana de cosecha en función de variables como el tipo de cultivo, cajas, peso total, y región.

Pasos a seguir:

- 1. Importar bibliotecas
- 2. Cargar el archivo CSV
- 3. Mostrar primeras filas y resumen
- 4. Ver estadísticas de duración
- 5. Imputar valores faltantes con mediana
- 6. Convertir tipos
- 7. Confirmar que no hay nulos
- 8. Gráfico: N cajas vs Semana
- 9. Gráfico: Peso total vs Semana
- 10. Detectar valores atípicos (semana o peso)
- 11. Codificar cultivo y región
- 12. Definir X e y (semana)
- 13. Dividir en conjuntos de entrenamiento y prueba
- 14. Agregar constante
- 15. Entrenar modelo de regresión lineal
- 16. Mostrar resumen del modelo
- 17. Mostrar coeficientes
- 18. Predecir semanas de cosecha
- 19. Evaluar con MSE y R²
- 20. Graficar semana_real vs semana_predicha y guardar

Materiales necesarios:

Archivo CSV Descripción

Mortalidades_Centro de cultivos.csv Datos semanales por centro: especie, edad, mortalidad, etc.

Cosechas_Cosechas_2023.csv
Datos de cada cosecha: semana, peso, cultivo, cajas