

Práctico Integrador – Predicción de Diabetes con Machine Learning

En este TP debes aplicar conocimientos sobre exploración de datos, modelos supervisados y evaluación para identificar el mejor modelo predictivo. Luego, exportar el modelo y utilizarlo en una aplicación gráfica (Tkinter) para realizar predicciones con nuevos datos.

1- Dataset

Utiliza el siguiente dataset: [diabetes.csv](#)

Subí manualmente el archivo diabetes.csv al entorno de Google Colab.

Actividades

1. Carga y exploración de datos

- Cargar el dataset con Pandas.
- Analizar valores faltantes y tipos de datos.
- Realizar visualizaciones básicas.

2. Preprocesamiento

- Tratar valores faltantes o nulos.
- Sumar columnas con 0 y reemplazar por la mediana (`median()`)
- Normalizar el df

```
from sklearn.preprocessing import StandardScaler  
  
escala = StandardScaler()  
  
df_normado = escala.fit_transform(df)
```

- Dividir en conjuntos de entrenamiento y prueba.

3. Entrenamiento de modelos

- Entrenar 3 modelos y hacer predecir:
 - Regresión Logística
 - Árbol de Decisión
 - Bosque Aleatorio

4. Evaluación y comparación

- Calcular métricas para cada modelo:
 - Accuracy (precision)
 - Score (puntaje)
 - Matriz de Confusión
- Comparar métricas y justificar el mejor modelo.

5. Exportación del modelo

- Usar joblib para guardar el modelo.
- Descargar el archivo desde Colab.

6. Interfaz gráfica con Tkinter

- Crear una GUI que permita ingresar datos nuevos.
- Cargar el modelo exportado.
- Mostrar predicción (Diabetes: Sí / No).

Entregables

- Notebook de Colab completo
- Modelo exportado (modelo_diabetes.pkl)
- Captura o video de la app Tkinter (opcional)

Criterios de Evaluación

- Exploración y limpieza de datos 2 pts
- Implementación de los 3 modelos 3 pts
- Evaluación y justificación del mejor modelo 2 pts
- Exportación del modelo para reutilización 1 pt
- Interfaz gráfica funcional con Tkinter 2 pts