Homework Assignment
Diabetes Risk Prediction

Summary

The goal of this assignment is to use the *load_diabetes* dataset from *sklearn.datasets* to create a binary classification model that predicts whether a patient is at risk of diabetes based on their target value. The task involves:

- Preprocessing the dataset and defining two thresholds (150 and 250) to classify patients as "Endangered" or "Not Endangered."

- Training and evaluating a machine learning model for binary classification.

- Creating visualizations to evaluate the data distribution and the model's performance.

Note: The focus is on building a robust model and providing clear visualizations to interpret the results.

Task

- Load the dataset using *diabetes = load_diabetes(scaled=False)* - [link].

- Explore the dataset and provide a summary of its features and target variable.

- Define two scenarios with different threshold values applied to the target variable (last column):

- **Scenario 1**: Patients with a target value above 150 are classified as "Endangered," and those below are "Not Endangered."

- **Scenario 2**: Patients with a target value above 250 are classified as "Endangered," and those below are "Not Endangered."

- Build a binary classification model for each scenario and discuss the results, including model performance, models, evaluation methods, and insights.

Deliverables

- A Python script or Jupyter Notebook containing the code for data preparation, model training, evaluation, and visualization.

- A brief report summarizing the findings, including:

  - Data exploration results.
  - Model performance metrics.
  - Visualizations and their interpretations.