Gender-Based Anthropometric Analysis

# Project Overview

As part of my continuous learning in data science, I **have taken a deep dive into practical machine learning using Python, Pandas, and Scikit-Learn**, all executed within Google Colab. The objective has been to build and evaluate a classification model step by step—with a core focus on **Logistic Regression**, one of the foundational algorithms in supervised learning.

This is about transforming concepts into practical skills, understanding how machine learning operates beneath the surface, and building real confidence in data preprocessing, model training, and evaluation.

**Key Steps and Learnings**

# Data Loading and Exploration

* The dataset was loaded using pandas.read\_csv() to work directly in Google Colab.
* Performed initial exploration (head(), info(), describe()) to understand the structure, types, and distributions.

**Tools used:** pandas, matplotlib, seaborn

# Preprocessing and Label Encoding

* Categorical variables were converted into numerical values using LabelEncoder from sklearn.preprocessing.
* Cleaned the data to ensure it's suitable for training the model.

This step is essential to prepare raw data for ML models, especially for algorithms like logistic regression that can't handle string values.

# Splitting the Dataset

* Used train\_test\_split() to divide the data into training and testing sets (usually 70-30 or 80-20).
* This allows evaluation of model performance on unseen data and prevents overfitting.

# Model Building with Logistic Regression

* Implemented the **Logistic Regression** model from sklearn.linear\_model.
* Trained the model on the training dataset using .fit().

**Evaluation & Insights**

After model training, the following metrics were used to evaluate performance:

# Accuracy Score

* Computed using accuracy\_score() from sklearn.metrics
* Helped measure the proportion of correct predictions.

# Confusion Matrix

* Displayed as a heatmap using seaborn.heatmap() for better visualization.
* Helped understand **True Positives, False Positives, False Negatives**, and **True Negatives**.

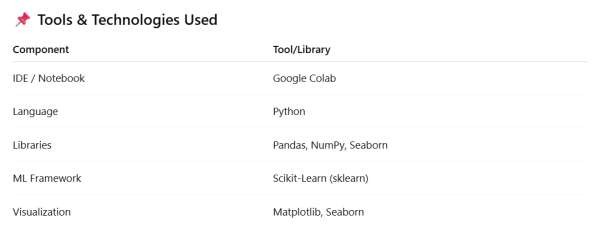
**Classification Report**

* Generated detailed metrics like **precision**, **recall**, and **F1-score**.

# Reflections and Learnings

This hands-on notebook taught me:

* How **data preprocessing** is the foundation of successful model building.
* How to choose the right **metrics** depending on the nature of the classification problem.
* The importance of **data splitting**, **encoding**, and **model interpretation**.
* How to interpret outputs like **coefficients** and **confusion matrix** to draw conclusions.



Practiced code link : [https://colab.research.google.com/drive/1doYklGo-3C\_TShAPOak2-](https://colab.research.google.com/drive/1doYklGo-3C_TShAPOak2-Kfcr4GudNUM)

[Kfcr4GudNUM](https://colab.research.google.com/drive/1doYklGo-3C_TShAPOak2-Kfcr4GudNUM)