Obesity Classification

Group Members: Dileep Malothu, Mukhesh Ravi & Sai Kiran

1.0 - A brief description of the dataset

The dataset that we have chosen for this class activity indulges in Obesity Classification.
The dataset is titled "Obesity Classification" and contains information about individuals' physical attributes and their corresponding obesity classification. Here is a brief description of the dataset:

- **ID**: A unique identifier for each individual.
- **Age**: The age of the individual.
- **Gender**: The gender of the individual (Male or Female).
- **Height**: The height of the individual in centimeters.
- **Weight**: The weight of the individual in kilograms.
- **BMI**: The Body Mass Index, calculated as weight in kilograms divided by the square of height in meters.
- **Label**: The obesity classification label, which can be either Underweight, Normal Weight, Overweight and Obese

2.0 – Model Comparison

Results Without PCA

Method	Metrics	Values
Decision Tree	Accuracy	1.0000
	Precision	1.0000
	Recall	1.0000
	F1 Score	1.0000
K-Nearest Neighbours	Accuracy	0.8636
	Precision	0.8818
	Recall	0.8636
	F1 Score	0.8680
Random Forest	Accuracy	1.0000
	Precision	1.0000
	Recall	1.0000
	F1 Score	1.0000
Naïve Bayes (GaussianNB)	Accuracy	0.8636
	Precision	0.8727

	Recall	0.8636
	F1 Score	0.8650

Results With PCA

Method	Metrics	Values
Decision Tree	Accuracy	0.5455
	Precision	0.6040
	Recall	0.5455
	F1 Score	0.5562
K-Nearest Neighbours	Accuracy	0.8636
	Precision	0.8712
	Recall	0.8636
	F1 Score	0.8571
Random Forest	Accuracy	0.6818
	Precision	0.7136
	Recall	0.6818
	F1 Score	0.6831
Naïve Bayes (GaussianNB)	Accuracy	0.7727
	Precision	0.7857
	Recall	0.7727
	F1 Score	0.7643

3.0 - Analysis

Decision Tree and Random Forest Classifiers performed perfectly without PCA, achieving 100% accuracy, precision, recall, and F1 score. However, their performance significantly dropped after applying PCA, especially for the Decision Tree.

K-Nearest Neighbors (KNN) Classifier maintained consistent performance with and without PCA, showing high accuracy and balanced precision, recall, and F1 scores.

Naive Bayes (GaussianNB) also showed consistent performance, though slightly lower than KNN, with and without PCA.

4.0- Observations on dimensionality reduction and its impact.

PCA Impact

In this analysis, PCA affected the models differently. For K-Nearest Neighbors (KNN) and Naive Bayes (GaussianNB), it maintained relatively stable performance, indicating that these models are less dependent on the original feature space and can effectively utilize the reduced dimensions. On the other hand, Decision Tree and Random Forest classifiers experienced a decline in performance after PCA was applied. This suggests that these models rely more heavily on the original features to make accurate predictions and may struggle when the data is transformed and reduced, potentially leading to loss of important information and increased risk of underfitting.

5.0 – Best Model

Based on the results, K-Nearest Neighbors (KNN) and Naive Bayes (GaussianNB) classifiers demonstrate consistent performance with and without the application of Principal Component Analysis (PCA), indicating their robustness to dimensionality reduction. This makes them potentially better choices for the "Obesity Classification" dataset when dimensionality reduction is applied, as they maintain high accuracy, precision, recall, and F1 scores even after the feature space is reduced. In contrast, models like Decision Tree and Random Forest show significant performance drops with PCA, suggesting they rely more heavily on the original feature space and may be prone to overfitting.