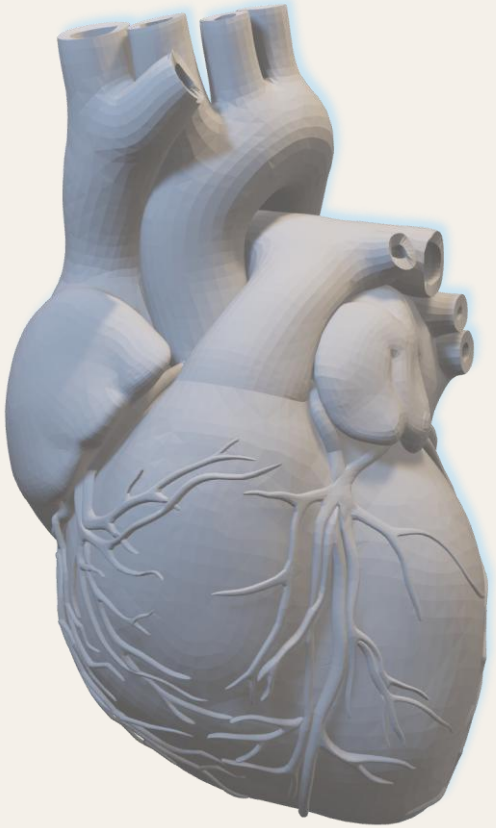
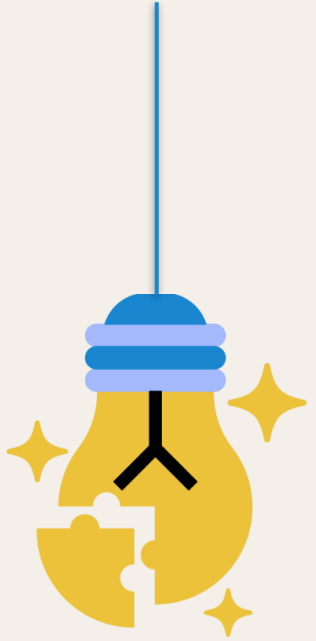
جامعة الملك سعود
King Saud University

Cardiovascular Disease Prediction

Group 15



PROBLEM



Cardiovascular diseases are a major cause of death worldwide. With more people at risk due to unhealthy lifestyles and genetics, it's important to find ways to identify those at high risk. This project focuses on using data mining techniques to analyze a dataset related to cardiovascular diseases. The goal is to classify individuals based on their risk factors, helping to predict who may develop CVDs.

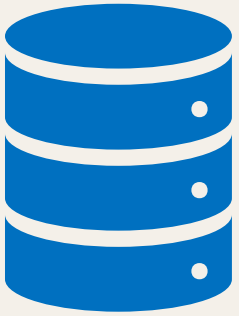
DATASET

- Dataset Overview:

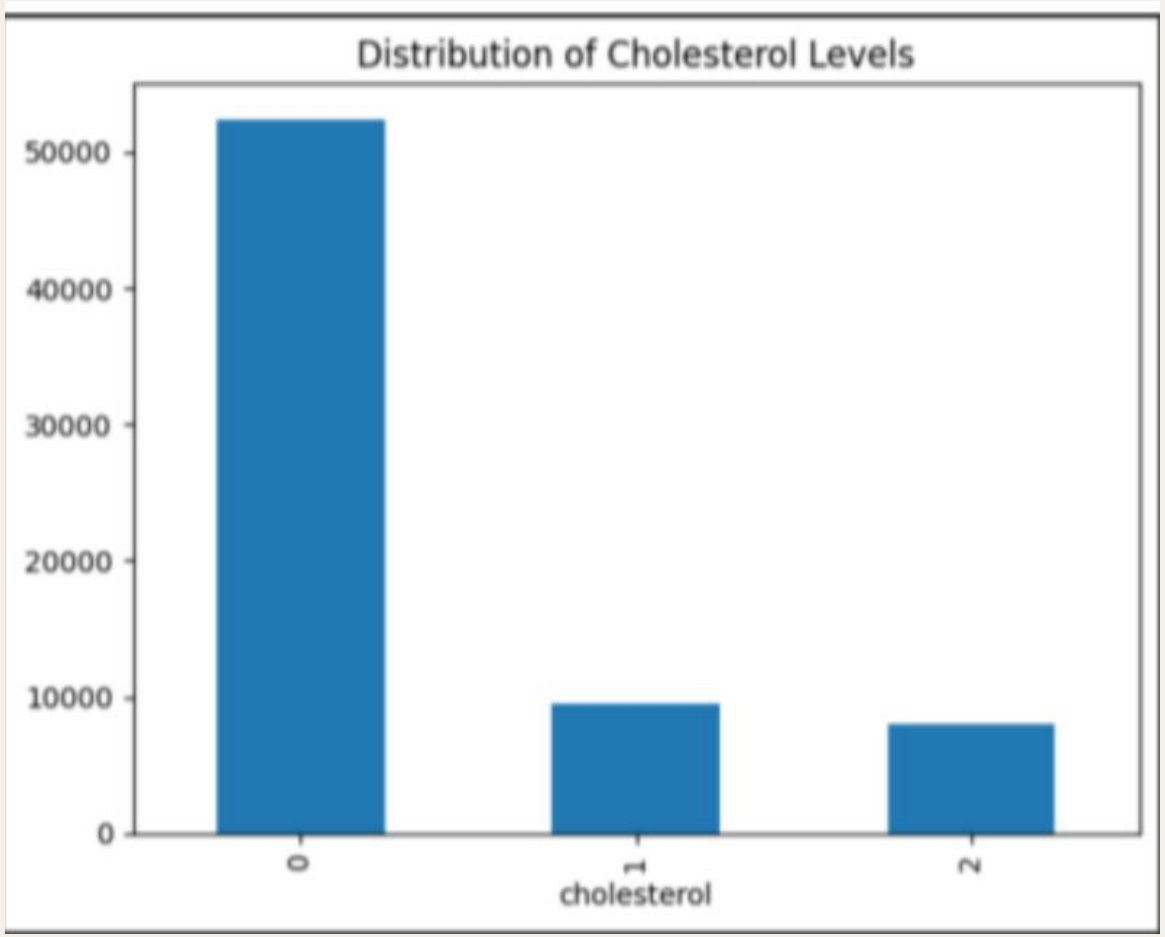
Number of Objects: 70000
Number of Attributes: 14·

- Attribute Details:

Data types: Integer, Decimal data
Missing Values: No missing values



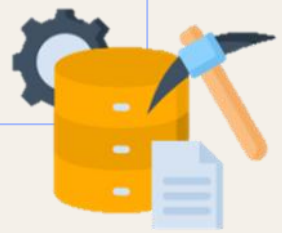
DATA VISUALIZATION



DATA MINING TECHNIQUES

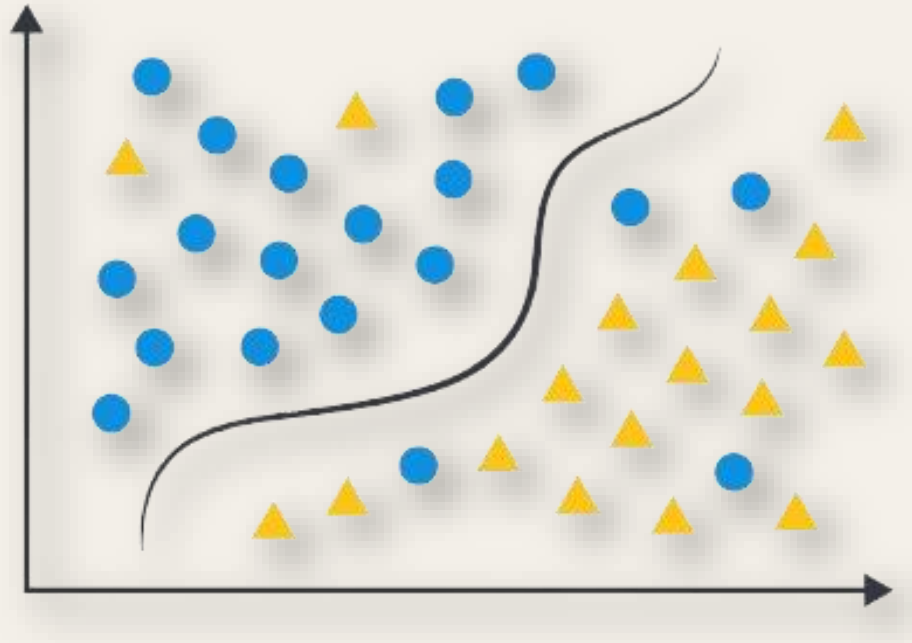
Classification

Clustering



CLASSIFICATION

- **Supervised Learning.**
- **Selection Measures Used:** Gini Index and Entropy.
- **Data Splits Tested:**90%-10%, 80%-20%, and 70%-30%.

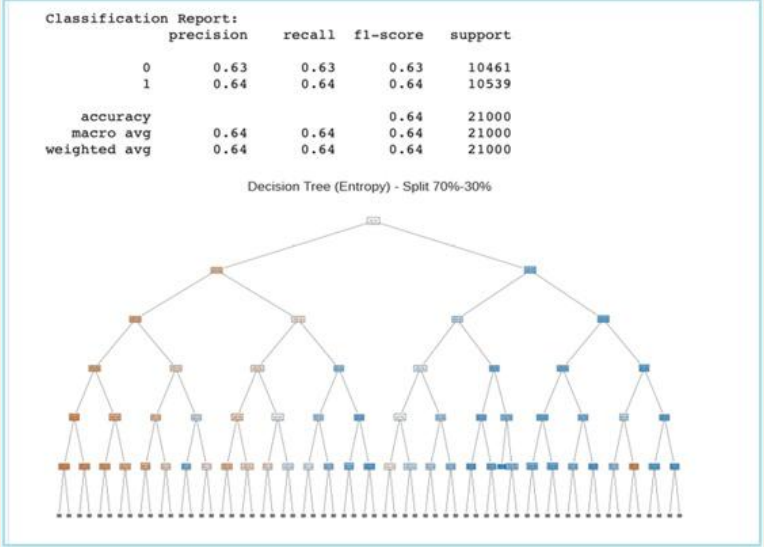


CLASSIFICATION FINDINGS

- Higher training data ratio improved precision, recall, and F1-score.
- The best accuracy was achieved with a data split of 70% training data and 30% test data.
- Confusion Matrix showed clear accuracy in predicting target classes.
- Gini criterion showed balanced splits.
- Entropy criterion focused more on reducing uncertainty in final nodes.

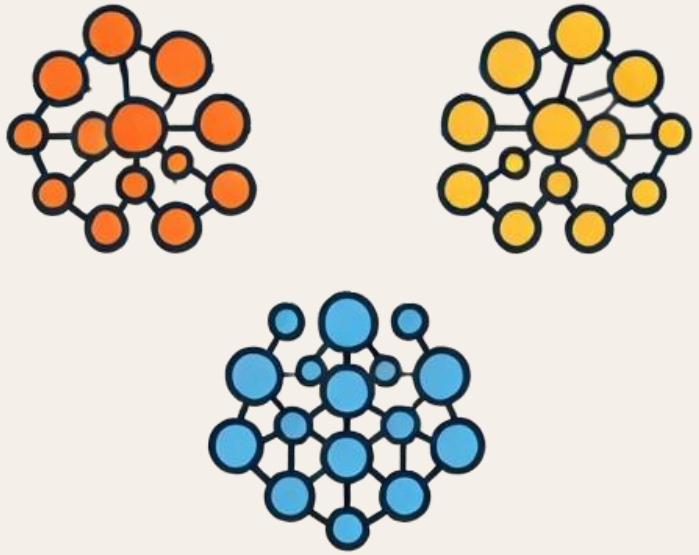
Classification Report:

	precision	recall	f1-score	support
0	0.63	0.63	0.63	10461
1	0.64	0.64	0.64	10539
accuracy			0.64	21000
macro avg	0.64	0.64	0.64	21000
weighted avg	0.64	0.64	0.64	21000



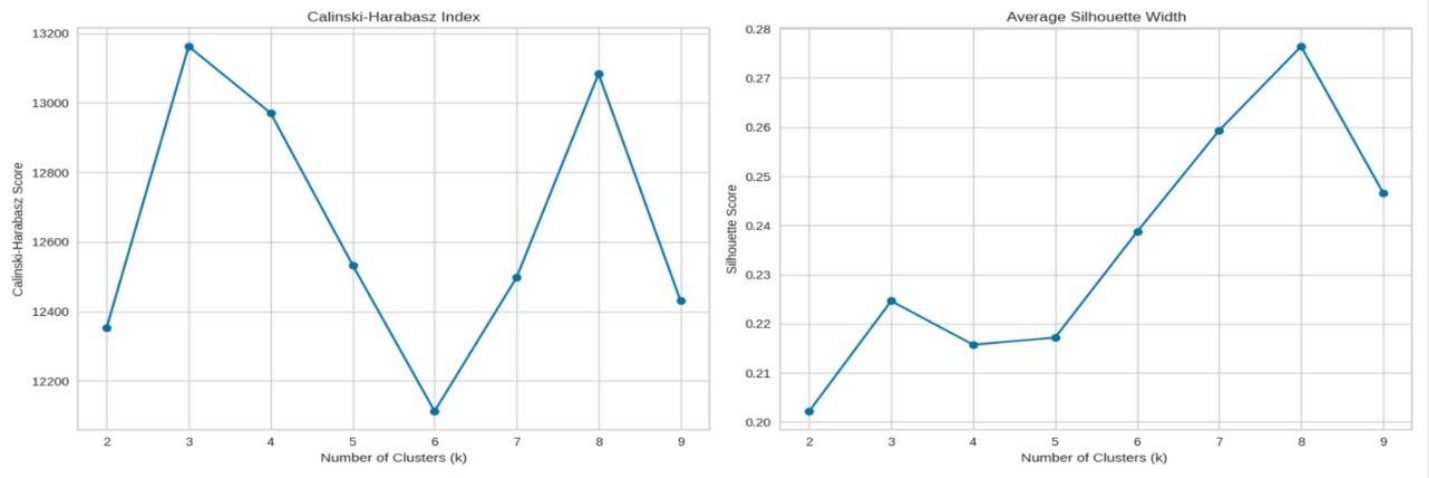
CLUSTERING

- **Unsupervised Learning.**
- **Method Used:** K-Means clustering.
- **Evaluation:** Silhouette width for K=2,4,8.



CLUSTERING FINDINGS

- K-Means Clustering was evaluated to determine the optimal number of clusters.
- Silhouette Analysis: Suggested k=4 for balanced cluster quality.
- Calinski-Harabasz Index: Highlighted distinct clusters with a peak at k=4.
- Elbow Method: Optimal clusters identified at k=8 based on reduced improvement in WSS.



THANK YOU!