

Week 2 – Heart Disease Prediction Report

This report summarizes the Week 2 Machine Learning task, where multiple classification models (Logistic Regression, Decision Tree, Random Forest) were applied to the heart disease dataset. PCA (Principal Component Analysis) and Standard Scaling were used to analyze performance changes.

Dataset Overview

The dataset contains 13 input features and 1 target column indicating the presence (1) or absence (0) of heart disease. No missing values were found, and data was ready for processing.

Model Accuracy Before PCA

Model	Accuracy
Logistic Regression	0.8525
Decision Tree	0.8361
Random Forest	0.8525

Model Accuracy After PCA (n_components = 5)

Model	Accuracy
Logistic Regression	0.8852
Decision Tree	0.7541
Random Forest	0.8033

Analysis & Observations

- Logistic Regression improved after PCA because PCA reduces noise and enhances linear separability.
- Decision Tree performance dropped after PCA since tree-based models do not benefit from PCA-reduced features.
- Random Forest saw a slight decline for the same reason.
- Logistic Regression with PCA gave the highest accuracy (approx. 89%).

Conclusion

PCA combined with standardized scaling significantly improved the performance of linear models such as Logistic Regression. This task demonstrates strong understanding of preprocessing, dimensionality reduction, and classification model comparison. The best performing model in this task was Logistic Regression with PCA.