Heart Disease Prediction - ML Model Performance Analysis Report

Objective

To evaluate and compare the performance of different machine learning models (Logistic Regression and SVM) on heart disease prediction, with and without feature reduction techniques like feature selection and PCA.

1. Logistic Regression Model

Feature Reduction Techniques Used:

Method	Parameters	Train Accuracy	Test Accuracy
Feature Selection	k = 0	84.70%	83.60%
PCA	**n_components = 9**	**85.53%**	**85.26%**
Simple (No Reduction)	-	85.53%	80.32%

Analysis:

- Best Test Accuracy: Achieved using PCA (85.26%), closely followed by Feature Selection (83.60%).
- Observation: Feature reduction via PCA maintained model generalization better than no reduction (which overfits slightly).

2. SVM Model

A. With PCA-Based Feature Extraction

Kernel	n_components	Train Accuracy	Test Accuracy
Linear	7	83.47%	83.60%
Polynomial	5	84.71%	75.40%
Gaussian (RBF)	5	83.47%	77.04%

B. With Feature Selection

Kernel	k	Train Accuracy	Test Accuracy

Linear	10	86.76%	81.96%
Polynomial	7	91.32%	80.32%
Gaussian (RBF)	8	91.22%	82.60%

C. Without Feature Reduction (All Features)

Kernel	Train Accuracy	Test Accuracy
Linear	85.55%	81.96%
Polynomial	97.10%	75.40%
Gaussian (RBF)	97.10%	70.60%

SVM Model Analysis

- With PCA:
  • Best Test Accuracy: Linear kernel (83.60%) with n_components=7.
  • Polynomial and Gaussian show signs of overfitting due to lower test accuracy despite reasonable training scores.

- With Feature Selection:
  • Best Test Accuracy: Gaussian (82.60%) with k=8.
  • Polynomial kernel gives very high train accuracy but test accuracy drops—suggesting overfitting.

- Without Feature Reduction:
  • Overfitting observed in Polynomial and Gaussian kernels (97.10% train but <76% test).
  • Linear kernel shows balance: 85.55% train, 81.96% test.

Overall Conclusion

Model	Technique	Best Kernel	Test Accuracy
Logistic Regression	**PCA (n=9)**	**N/A**	**85.26%**
SVM	Feature Selection (k=8)	Gaussian	82.60%
SVM	PCA (n=7)	Linear	83.60%
SVM	All Features	Linear	81.96%

Diabetes Prediction Model Analysis

1. Logistic Regression Model

Method	Train Accuracy	Test Accuracy
Feature Selection	Train: 70.66%	Test: 76.62%
Feature Extraction	Train: 70.33%	Test: 75.97%
Normal (No Feature Reduction)	Train: 70.82%	Test: 75.97%

2. SVM Model

Feature Strategy	Kernel	Train Accuracy	Test Accuracy
Feature Extraction (n_comp=3)	Linear	Train: 73.15%	Test: 76.62%
Feature Extraction (n_comp=3)	Poly	Train: 76.7%	Test: 70.0%
Feature Extraction (n_comp=3)	**Gaussian**	**Train: 76.05%**	**Test: 76.62%**
Feature Selection (k=6)	Linear	Train: 70.50%	Test: 77.27%
Feature Selection (k=2)	Poly	Train: 76.22%	Test: 76.62%
Feature Selection (k=3)	Gaussian	Train: 70.66%	Test: 75.32%
Simple	Linear	Train: 70.66%	Test: 77.27%
Simple	Poly	Train: 84.52%	Test: 72.72%
Simple	Gaussian	Train: 85.66%	Test: 72.07%

3. Neural Network

Method	Train Accuracy	Test Accuracy

Simple	Train: 81.56%	Test: 76.40%
Feature Extraction	**Train: 79.23%**	**Test: 72.07%**
Feature Selection	Train: 76.71%	Test: 72.72%