EXPERIMENT 5

Aim of the Experiment

To implement Support Vector Machine (SVM) for classification on the Heart Disease dataset and improve model performance using Hyperparameter Tuning (GridSearchCV).

Theory

Support Vector Machine (SVM) is a supervised machine learning algorithm used for classification and regression. In classification, SVM finds an optimal hyperplane that separates data points of different classes with maximum margin. The closest data points to the hyperplane are called Support Vectors.

Kernel Trick

For non-linear datasets, SVM uses kernel functions such as:

- Linear Kernel
- Radial Basis Function (RBF)
- Polynomial Kernel

Important Hyperparameters

1. C (Regularization Parameter)

Controls trade-off between margin size and classification error.

- Small C → wider margin, more tolerance to misclassification
- Large C → smaller margin, strict classification

---

2. Kernel

Determines type of decision boundary.

---

3. Gamma

Controls influence of individual training samples.

- High Gamma → complex boundary
- Low Gamma → smoother boundary

Dataset Description

Dataset Name

Heart Disease Dataset

Dataset Type

Medical classification dataset

Dataset Size

- ~303 records
- 13 input features
- 1 target variable

Target Variable

Column	Description
target	0 = No heart disease, 1 = Heart disease

Methodology / Workflow

Step 1: Data Collection

- Load heart.csv dataset

Step 2: Data Preprocessing

- Check missing values
- Separate features and target

Step 3: Train-Test Split

- 80% training
- 20% testing

Step 4: Feature Scaling

- StandardScaler applied
- Required because SVM depends on distance

Step 5: Baseline Model

- Train SVM with default parameters

- Calculate baseline accuracy

Step 6: Hyperparameter Tuning

- GridSearchCV used
- Tested combinations of:
    - C
    - Kernel
    - Gamma
- 5-fold cross-validation used

Step 7: Model Evaluation

Evaluate using:

- Accuracy
- Confusion Matrix
- Precision
- Recall
- F1 Score

Results Summary

After training the SVM model on the Heart Disease dataset, the following results were obtained:

Baseline Model

Baseline Accuracy = **88.78%**

The default SVM model achieved good classification performance but left room for improvement.

Hyperparameter Tuning Results

Using GridSearchCV with 5-fold cross-validation, the best combination of parameters was selected.

Tuned Model Accuracy = **98.54%**

This shows a significant improvement in classification performance.

Confusion Matrix

$$\begin{bmatrix} 102 & 0 \\ 3 & 100 \end{bmatrix}$$

Interpretation:

- 102 patients correctly classified as No Heart Disease
- 100 patients correctly classified as Heart Disease
- 3 patients misclassified
- 0 false positives in class 0

The model demonstrates excellent classification ability with very low misclassification rate.

Classification Report Summary

Class	Precision	Recall	F1-Score
0	0.97	1.00	0.99
1	1.00	0.97	0.99

Overall Accuracy = **99% (≈ 98.54%)**

Macro Average F1 Score = **0.99**

This indicates:

- Very high precision (few false positives)
- Very high recall (few false negatives)
- Excellent model balance

Results Summary

- Baseline Accuracy: **88.78%**
- Tuned Accuracy: **98.54%**
- Hyperparameter tuning improved performance significantly.
- Only 3 misclassifications out of 205 test samples.

CONCLUSION

In this experiment, Support Vector Machine (SVM) was successfully implemented on the Heart Disease dataset for binary classification. Initially, the baseline SVM model achieved an accuracy of 88.78%. After applying hyperparameter tuning using GridSearchCV, the model accuracy improved significantly to 98.54%. The confusion matrix and classification report show that the tuned model achieved very high precision, recall, and F1-score, indicating strong predictive performance and minimal classification errors. This experiment demonstrates that:

- SVM is a powerful margin-based classification algorithm.
- Feature scaling is essential for SVM performance.
- Hyperparameter tuning plays a crucial role in improving model accuracy.
- Proper selection of C, kernel, and gamma can significantly enhance model generalization.

Overall, the tuned SVM model provided excellent classification results and outperformed the baseline model by a large margin.