Train-Test Split & Evaluation Metrics

Objective

The objective of this task is to split the dataset into training and testing data, train a machine learning model, and evaluate its performance using different evaluation metrics.

Dataset Used

- Heart Disease Dataset

Tools Used

- Python
- Pandas
- NumPy
- Scikit-learn
- VS Code
- Jupyter Notebook

Task Performed

1. Data Splitting

- Loaded the Heart Disease dataset using Pandas.
- Separated input features and target variable.
- Split the dataset into training data (80%) and testing data (20%) using train_test_split.
- Training data was used to train the model and testing data was used to evaluate performance.

2. Model Training

- Trained a **Logistic Regression** model using the training dataset.

- Logistic Regression was chosen as a simple and effective classification algorithm.

3. Model Prediction

- Used the trained model to predict outcomes on the test dataset.

4. Model Evaluation

- Calculated **Accuracy**, **Precision**, and **Recall**.
- Generated a **Confusion Matrix** to understand prediction results.
- Visualized confusion matrix using a heatmap.

Conclusion

This task helped in understanding how machine learning models are trained and evaluated. Using proper evaluation metrics ensures that the model performs well on unseen data and helps in making reliable predictions.