EVALUATION REPORT

Heart Disease Dataset – Logistic Regression

Dataset:

The Heart Disease dataset was used, obtained from the UCI repository via sklearn OpenML.

It contains medical attributes such as age, sex, chest pain, cholesterol, blood pressure,

and a target variable indicating presence of heart disease.

Model Used:

Logistic Regression

Train-Test Split:

The dataset was split into 80% training data and 20% testing data.

Training data was used to fit the model.

Testing data was used to evaluate performance.

Evaluation Metrics:

1. Accuracy:

Accuracy represents the proportion of correct predictions made by the model.

The model achieved high accuracy, showing good overall performance.

2. Precision:

Precision indicates how many predicted heart disease cases were actually correct.

High precision means fewer false positives.

3. Recall:

Recall shows how many actual heart disease cases were successfully identified.

High recall means fewer false negatives.

4. Confusion Matrix:

The confusion matrix shows:

True Positives (TP) – correctly predicted disease cases

True Negatives (TN) – correctly predicted non-disease cases

False Positives (FP) – incorrectly predicted disease

False Negatives (FN) – missed disease cases

Interpretation:

The Logistic Regression model generalizes well to unseen data.

It shows balanced precision and recall, meaning reliable medical predictions.

Conclusion:

The model effectively predicts heart disease using basic clinical features.

Evaluation metrics confirm strong and stable classification performance.