

Train–Test Split & Evaluation Metrics

Project Title

Heart Disease Prediction – Train–Test Split & Model Evaluation

Project Description

This project focuses on building and evaluating a simple machine learning model using the Heart Disease dataset. The objective is to understand how to split data into training and testing sets, train a Logistic Regression model, and evaluate its performance using different evaluation metrics.

This task helps in understanding the fundamentals of model validation and performance measurement.

Objectives

- Split the dataset into training and testing sets
- Understand the purpose of training vs testing data
- Train a Logistic Regression model
- Make predictions on test data
- Evaluate the model using accuracy, precision, and recall
- Understand and interpret the confusion matrix

Dataset Used

Heart Disease Dataset

Target Variable:

- Presence of heart disease (0 = No, 1 = Yes)

Features include:

- Age, Sex, Chest Pain Type, Cholesterol, Blood Pressure, etc.

Tools & Libraries Used

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

Steps Performed

1. Loaded the Heart Disease dataset
2. Identified input features (X) and target variable (y)
3. Encoded categorical features into numerical values
4. Split the dataset into training and testing sets (80/20)
5. Trained a Logistic Regression model
6. Made predictions on the test dataset
7. Calculated accuracy, precision, and recall
8. Generated and interpreted the confusion matrix

Evaluation Metrics Used

- **Accuracy:** Measures overall correctness of the model
- **Precision:** Measures how many predicted positives are correct
- **Recall:** Measures how many actual positives are correctly predicted
- **Confusion Matrix:** Shows correct and incorrect predictions

Key Learnings

- Train-test split helps evaluate model performance on unseen data
- Logistic Regression requires numerical input features
- Encoding categorical data is necessary before training
- Accuracy alone is not sufficient; precision and recall are important
- Confusion matrix provides detailed performance insight

Files in this Repository

- heart.csv – Dataset
- Heart_Disease_Model.ipynb – Jupyter Notebook
- README.md – Project documentation

How to Run

1. Clone this repository
2. Open the Jupyter Notebook
3. Run all cells step by step
4. View evaluation metrics and outputs

Conclusion

This task provided hands-on experience in training and evaluating a machine learning model. Understanding train-test split and evaluation metrics is essential for building reliable ML models.