Mahavir Education Trust's
SHAH & ANCHOR KUTCHHI ENGINEERING COLLEGE
Chembur, Mumbai - 400 088
UG Program in Information Technology

- **Student Name: Purva Mahamunkar**
- **Roll Number: 12/28**
- **PRN : 22UF17661IT090**
- **Department: IT**
- **Batch/Division: BTECH-1**
- **Experiment Title: Heart Disease Prediction**
- **Date: 27/09/2025**

Experiment No. 4				
Date of Performance:	**27/09/2025**			
Date of Submission:	**04/10/2025**			
Program Execution/ formation / correction/ ethical practices	Timely Submission	Viva	Experiment Total	Sign with Date

Experiment No 4

4.1 Aim: Heart Disease Prediction

4.2 Course Outcome: Students will be able to use different machine learning algorithms and compare its accuracy.

4.3 Learning Outcome: Students will be able to use different machine learning algorithms and compare its accuracy.

4.4 Requirement: Python (with TensorFlow/PyTorch), Jupyter/Colab, OpenCV, NumPy, Matplotlib, and a labeled Heart Diseases Prediction dataset.

4.5 Related Theory:

- **Introduction to Problem**
  - Heart disease is one of the leading causes of death globally.
  - Early prediction can help in prevention, timely diagnosis, and treatment.
  - AIML provides predictive models based on medical data.
- **Objective**
  - To analyze patient data (age, gender, cholesterol, blood pressure, etc.).
    To build a model that predicts the likelihood of heart disease.
- **Data Collection**
  - Public datasets (e.g., UCI Heart Disease Dataset, Kaggle datasets).
  - Features include:
    - Age, Gender
    - Chest pain type
    - Resting blood pressure
    - Serum cholesterol
    - Fasting blood sugar

    - Resting ECG results
    - Maximum heart rate achieved
    - Exercise-induced angina
    - Oldpeak (ST depression)
    - Slope, number of major vessels, thalassemia, etc.
- **Data Preprocessing**
    - Handle missing values.
    - Normalize/standardize values (e.g., cholesterol, blood pressure).
    - Convert categorical features (e.g., chest pain type, thalassemia) into numerical using encoding.
    - Split the dataset into training and testing sets.

- **Exploratory Data Analysis (EDA)**
    - Statistical analysis and visualization (histograms, correlation heatmaps).
    - Identify which factors strongly influence heart disease.
    - Example: Higher cholesterol, blood pressure,smoking often show strong correlation.
- **Feature Selection**
    - Select the most important features that impact prediction.
    - Techniques: Correlation, Chi-square test, PCA, Feature importance from models.
- **Model Selection**
    - Common ML algorithms used:
        - **Logistic Regression** → good for binary classification (disease or no disease).
        - **Decision Trees & Random Forests** → handle non-linear data well.
        - **Support Vector Machine (SVM)** → effective classification clear margins.
        - **K-Nearest Neighbors (KNN)** → simple and effective for small datasets.
        - **Neural Networks (Deep Learning)** → captures complex patterns.
- **Model Training & Testing**
    - Train model using training data.
    - Test accuracy, precision, recall, F1-score, ROC curve on testing data.
- **Evaluation Metrics**
    - **Accuracy** → % correctly predicted cases.
    - **Precision** → How many predicted positives are actually positive.
    - **Recall (Sensitivity)** → Ability to detect actual patients with heart disease.
    - **F1-score** → Balance between precision & recall.
    - **ROC-AUC** → Overall performance of classifier.
- **Deployment**
- The model can be deployed as a **web app** or **mobile app**.
- Example: Flask/Django for web deployment.
- Doctors or patients can input health data → system predicts risk level.

11. **Applications**

- Early detection of heart disease risk.
- Helps doctors in decision-making.
- Can be integrated into wearable devices & health monitoring systems.

12. **Challenges**

- Data imbalance (fewer positive cases than negative).
- Privacy & security of medical data.
- Overfitting if the dataset is small.
- Model interpretability (black-box problem in deep learning).

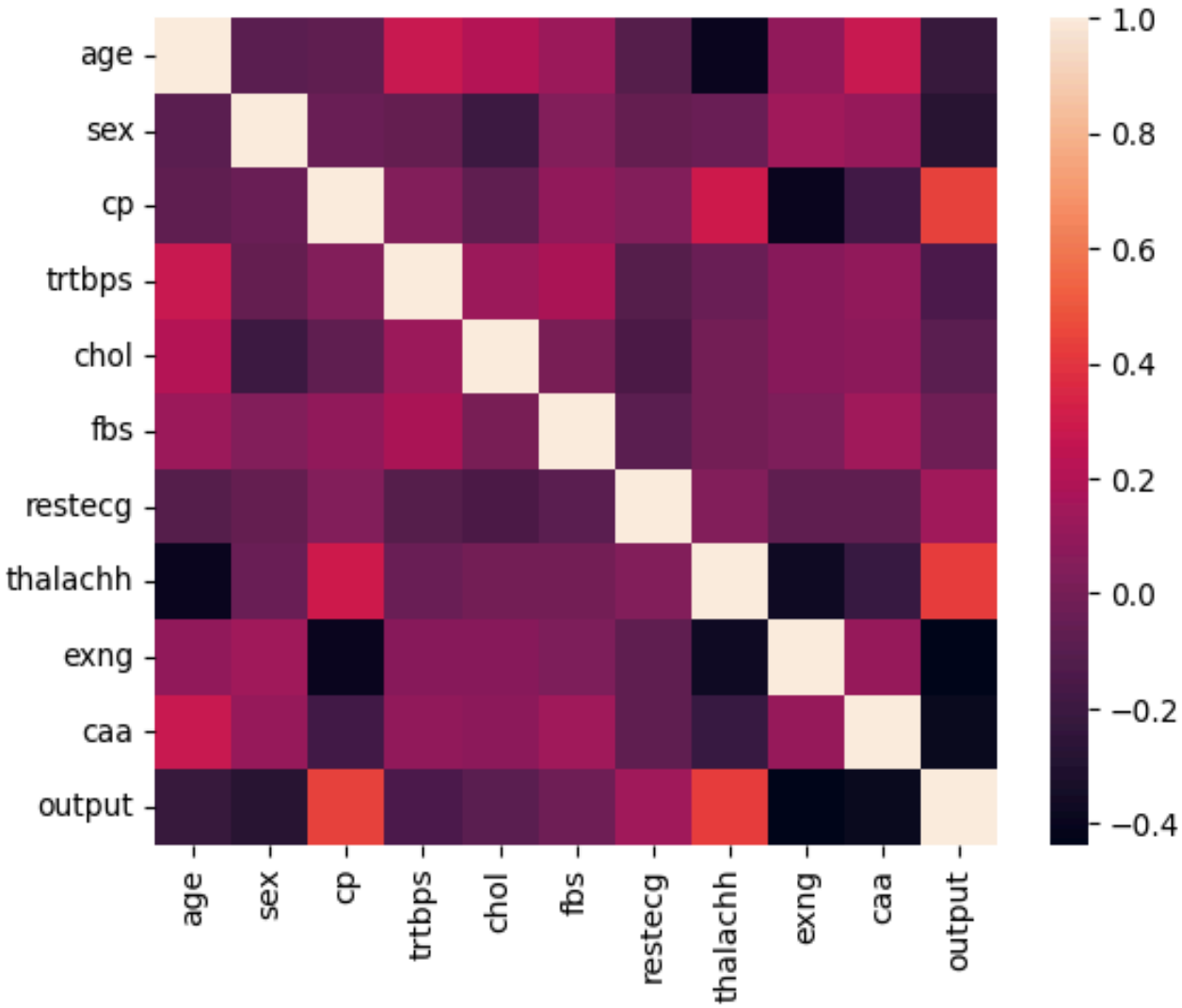
13. **Future Scope**

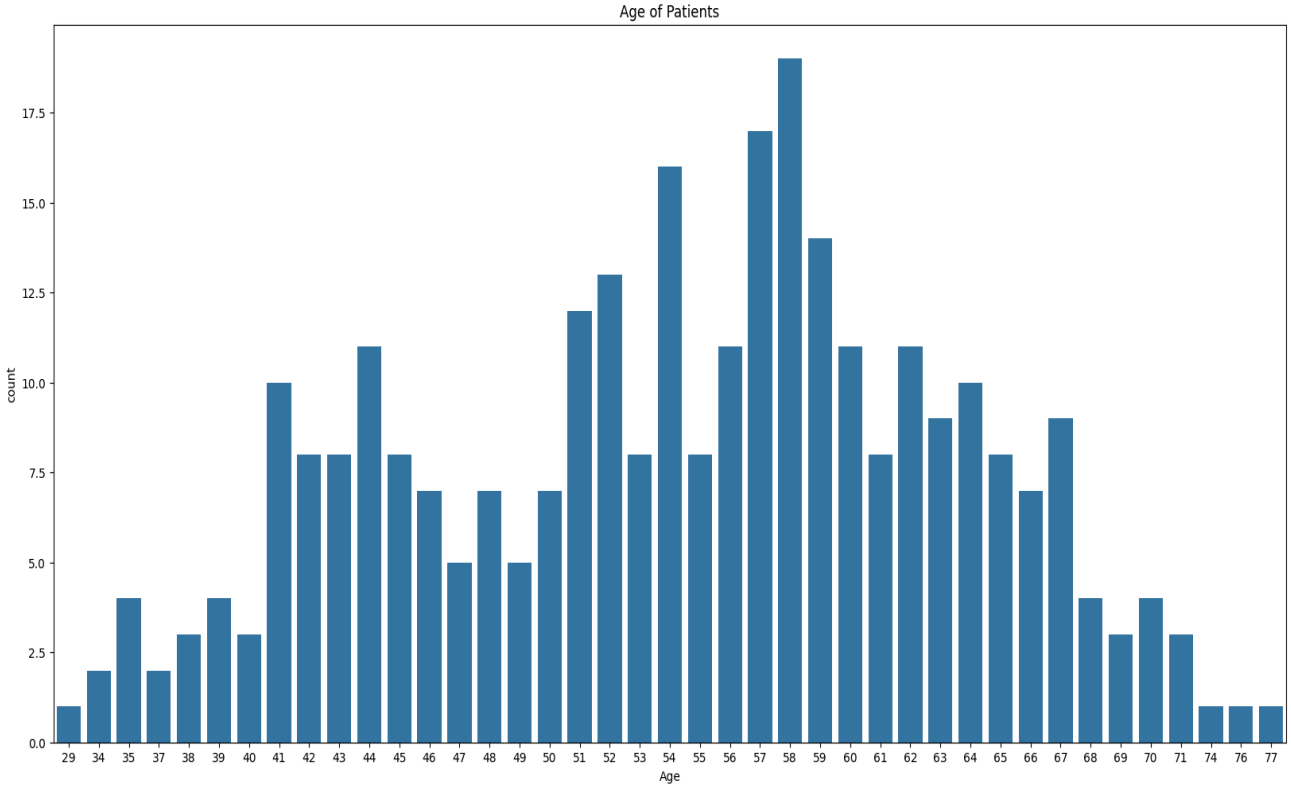
- Use **real-time IoT data** from wearables (heart rate, BP).
- Apply **Deep Learning** for better accuracy.
- Integration with **cloud healthcare systems**.
- Personalized health recommendations.

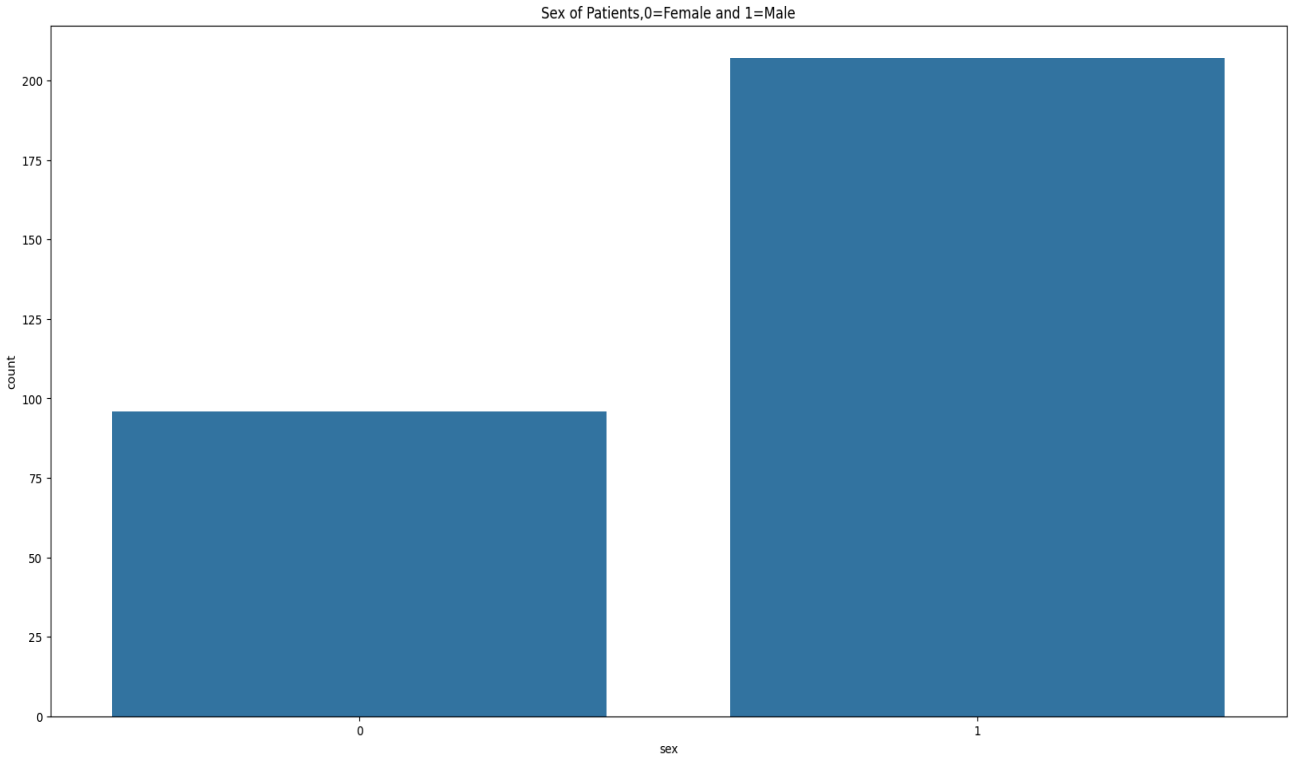
4.6 Program and Output:

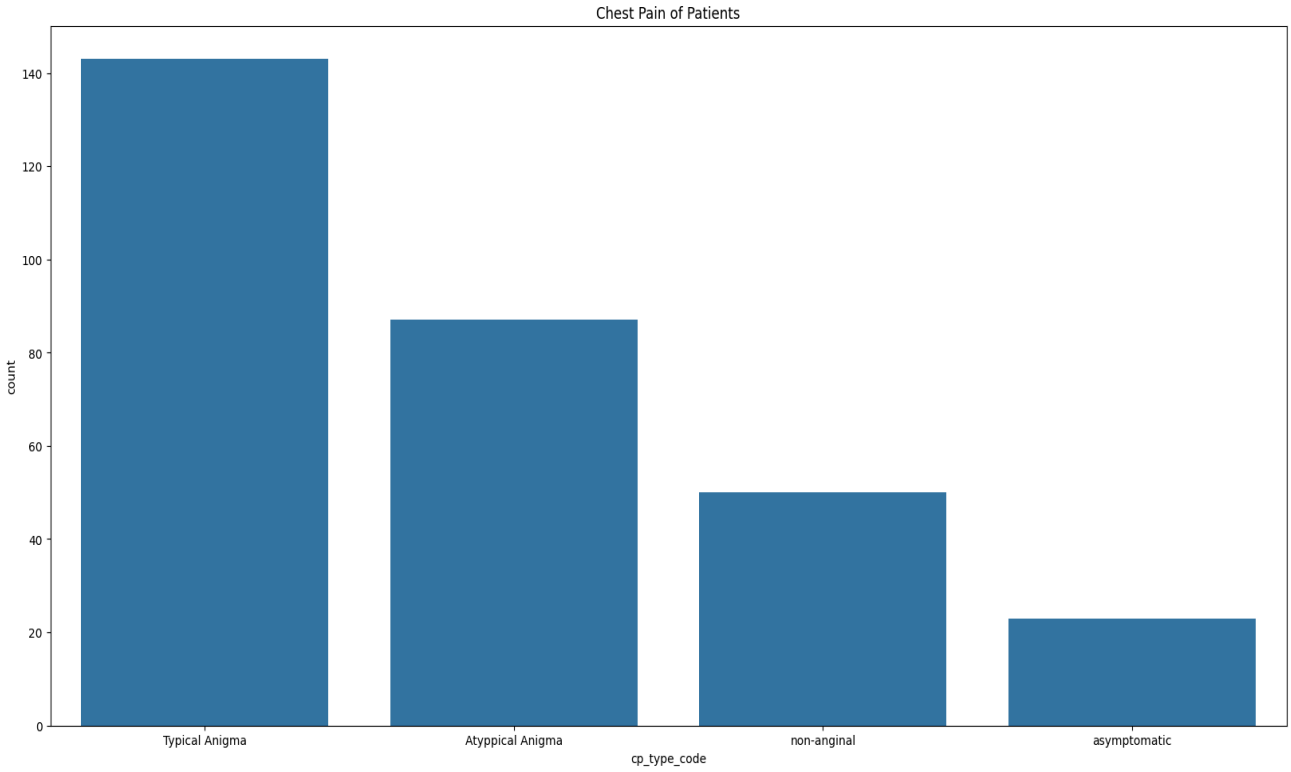
Completed Google Colab AIML4 Link

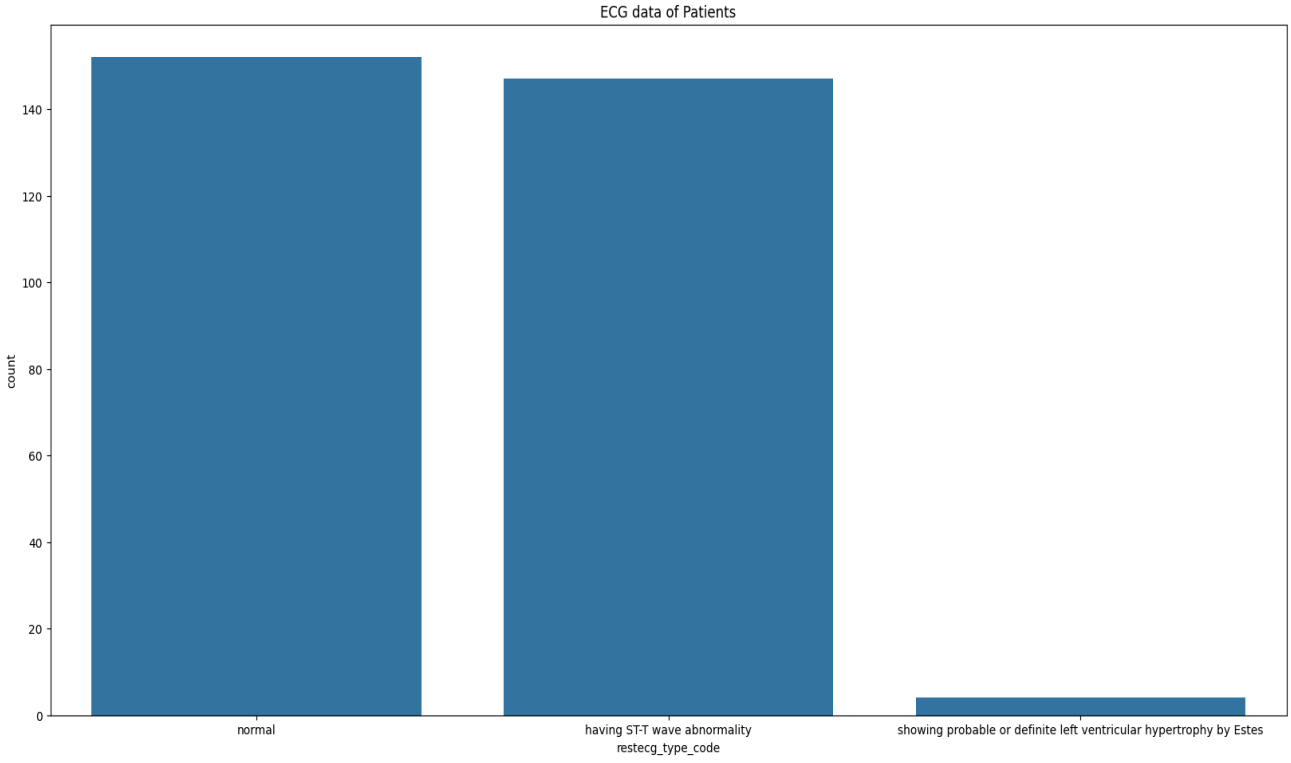
4.7 Result SnapShots:

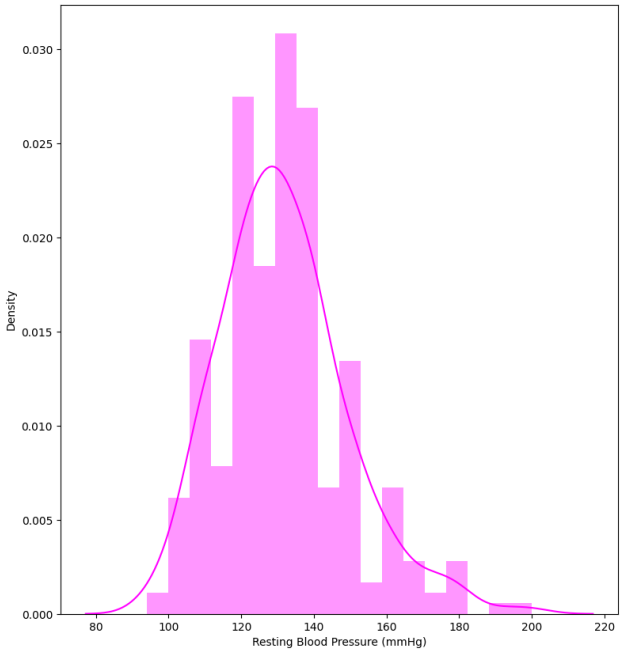


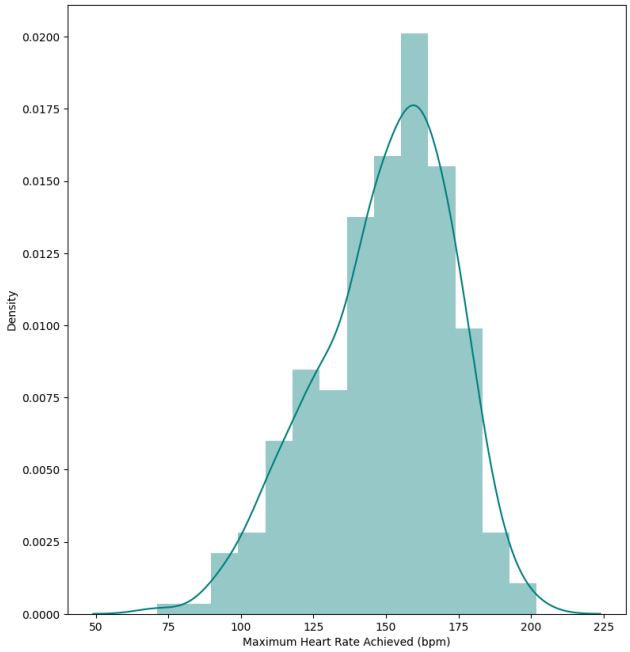

Age of Patients
count
Age
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
29 34 35 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 74 76 77

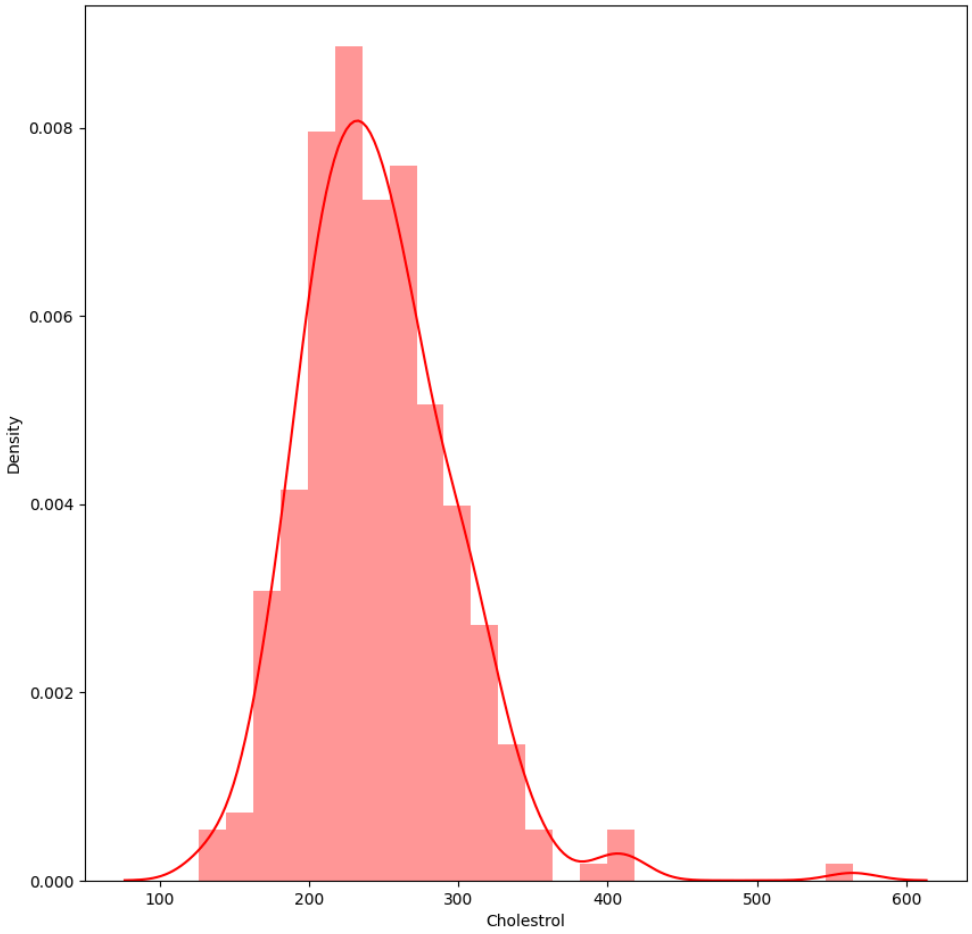

Sex of Patients,0=Female and 1=Male
count
sex
0
25
50
75
100
125
150
175
200
0
1

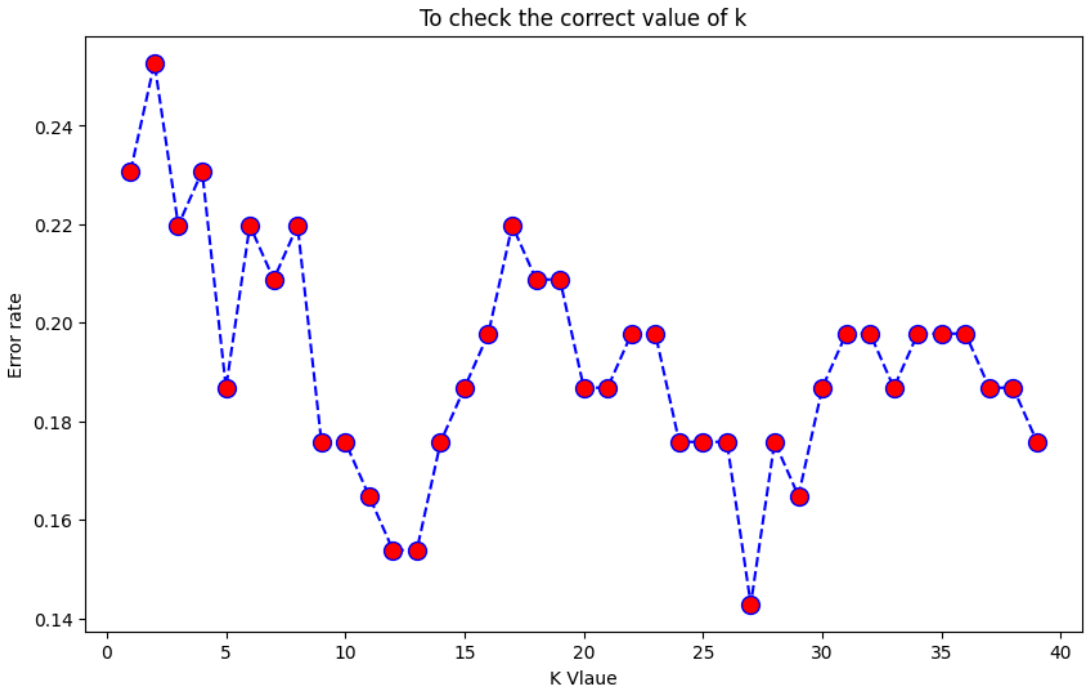

Chest Pain of Patients
count
0
20
40
60
80
100
120
140
Typical Anigma
Atyppical Anigma
non-anginal
asymptomatic
cp_type_code


ECG data of Patients
count
0
20
40
60
80
100
120
140
normal
having ST-T wave abnormality
showing probable or definite left ventricular hypertrophy by Estes
restecg_type_code


Density
Resting Blood Pressure (mmHg)


Density
Maximum Heart Rate Achieved (bpm)


Density
Cholestrol



4.8 Conclusion:

The experiment on **Heart Disease Prediction using AIML** demonstrates how machine learning models can effectively analyze patient health data and predict the risk of heart disease. By applying preprocessing, feature selection, and classification algorithms such as Logistic Regression, Random Forest, and Neural Networks, we can achieve reliable accuracy in prediction. This helps in **early diagnosis, better decision-making, and preventive healthcare**. Although challenges like data privacy, imbalance, and interpretability exist, AIML-based prediction systems have great potential in supporting doctors, reducing mortality, and advancing smart healthcare solutions.

4.9 Questions:

1. **What are the main steps involved in this experiment?**
   A. Data collection
   B. Data preprocessing (cleaning, encoding, normalization)
   C. Exploratory Data Analysis (EDA)
   D. Feature selection
   E. Model selection (Logistic Regression, Decision Tree, Random Forest, etc.)
   F. Training & testing the model
   G. Evaluation (accuracy, precision, recall, F1-score, ROC-AUC)
   H. Deployment for practical use.

2. **Why is data preprocessing important?**
   Medical data often has missing values, inconsistent units, or categorical features.
   Preprocessing ensures:

- Missing values are handled.
- Features are normalized/scaled.
- Categorical data is encoded into numeric form.
- The dataset is split into training and testing sets.
  This improves model accuracy and reliability.

3. **Which machine learning algorithms are suitable for heart disease prediction?**
   Commonly used algorithms include:
- Logistic Regression
- Decision Tree
- Random Forest
- K-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)
- Neural Networks (Deep Learning)

4. **What evaluation metrics are used for prediction models?**
   Accuracy – Overall correct predictions.
   Precision – Correctly predicted positive cases out of all predicted positives.
   Recall (Sensitivity) – Correctly identified actual positive cases.
   F1-score – Balance between precision and recall.
   ROC-AUC – Measures performance across different thresholds.

5. **What are the challenges in heart disease prediction?**
   - ➢ Limited and imbalanced datasets.
   - ➢ Privacy and security of medical data.
   - ➢ Overfitting in small datasets.
   - ➢ Interpretability of complex models like deep learning