Suresh Angadi Education Foundation's
ANGADI INSTITUTE OF TECHNOLOGY AND MANAGEMENT
Savagaon Road, Belagavi-590009
Approved by AICTE, New Delhi
Affiliated to Visvesvaraya Technological University, Belagavi
(ACCREDITED BY NAAC)

HEART DISEASE PREDICTION

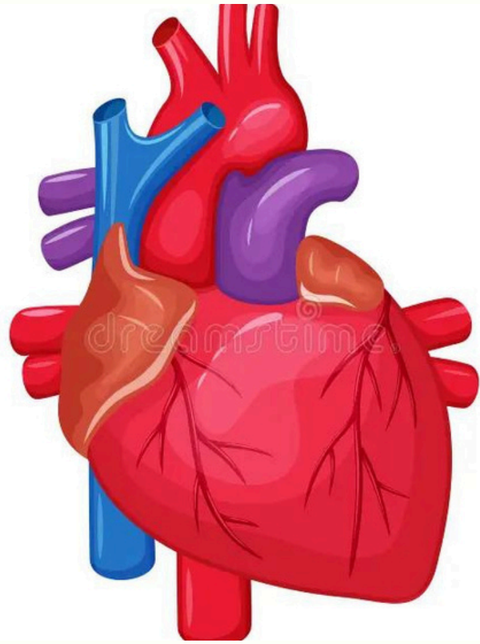
USING LOGISTIC REGRESSION

Name :Pooja Koparde (2AG21CS062)

Name :Madhura S Shenolkar (2AG21CS103)

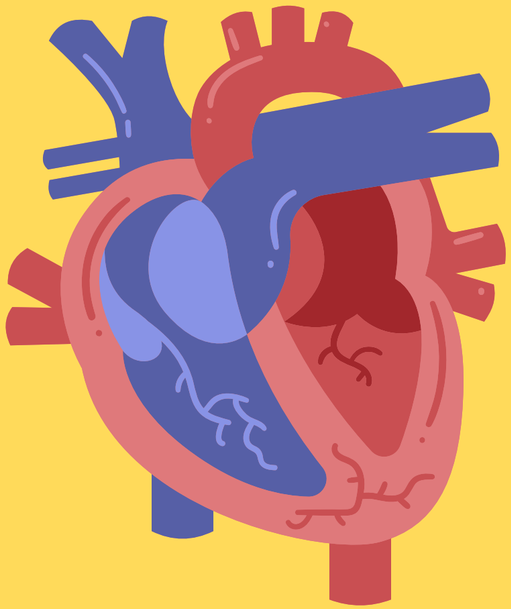
Name : Khushi Lad (2AG21CS043)

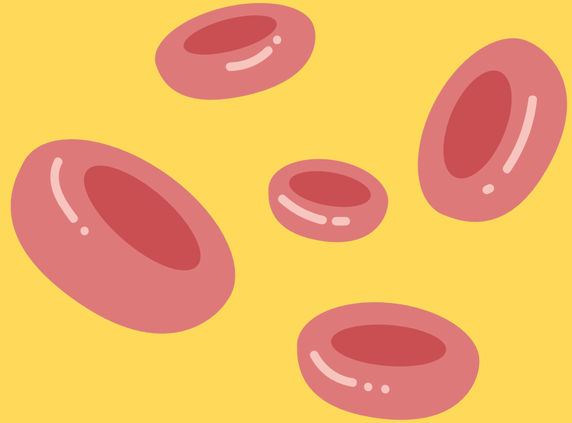
Name : Nagaraj Guledagudda (2AG21CS057)



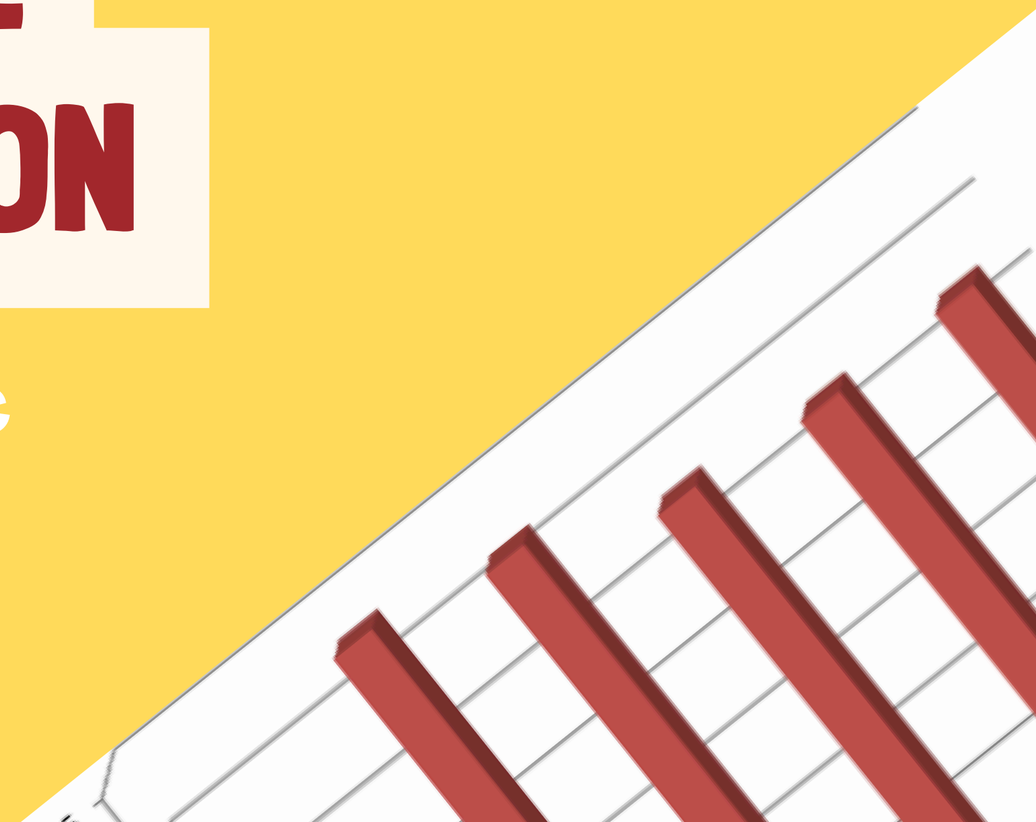
Guided by:
Mr. Gautam Dematti
Asst. Prof. CSE Department, AITM Belagavi

HEART DISEASE PREDICTION

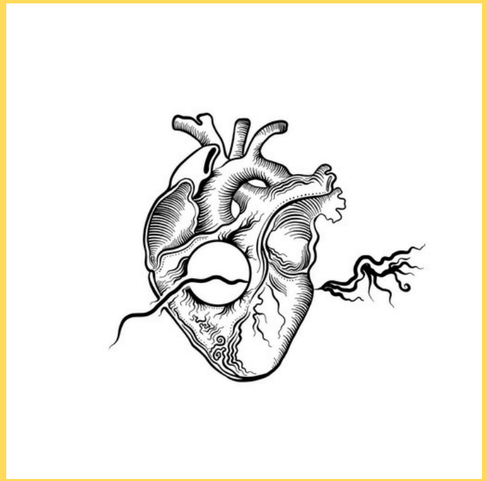




USING LOGISTIC REGRESSION



CONTENT

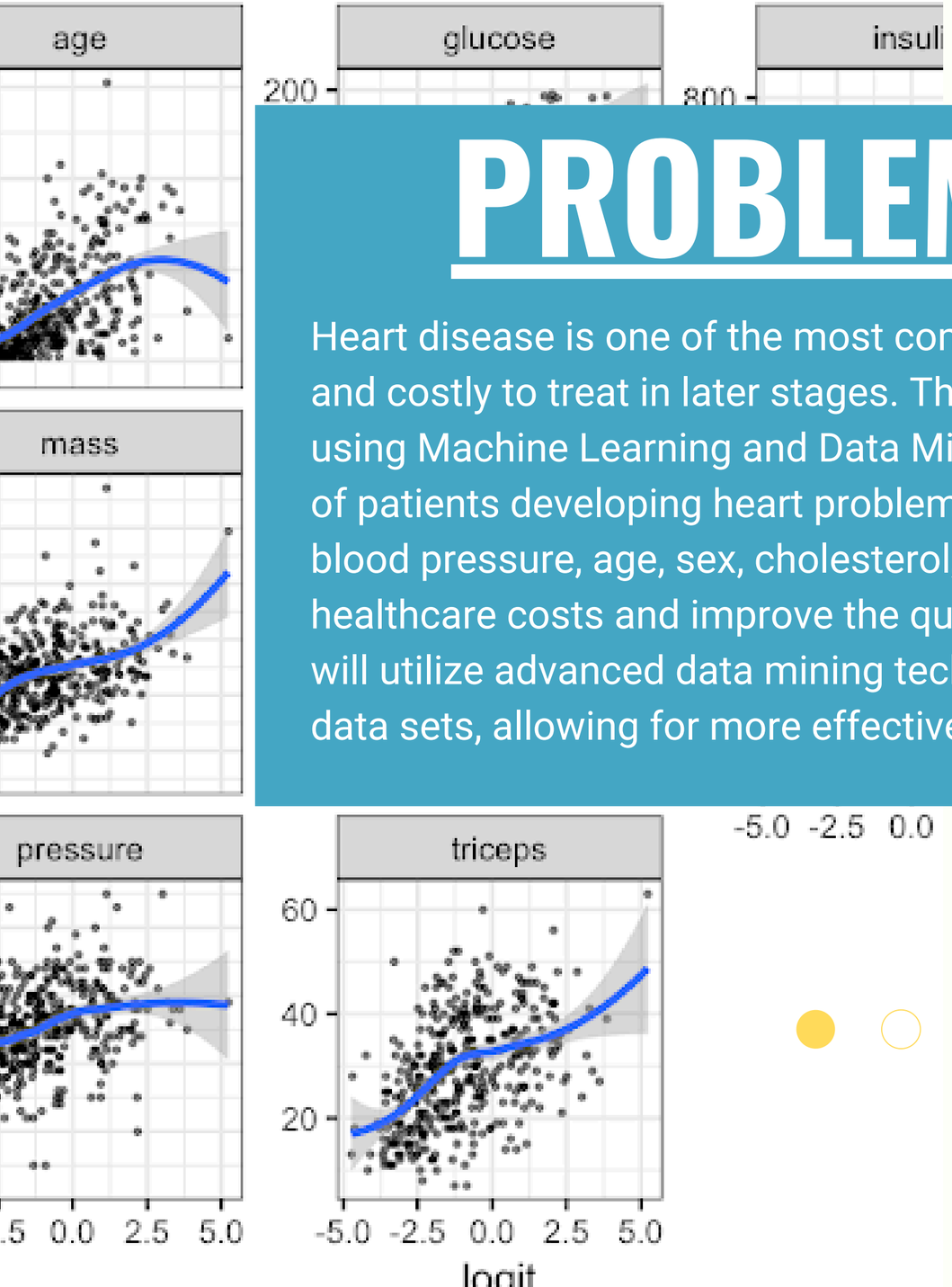


ABSTRACT

In contemporary times, Heart disease is the most common disease. But, unfortunately the treatment of heart disease is somewhat costly that is not affordable by common man. Hence, we can reduce this problem in some amount just by predicting heart disease before it becomes dangerous using Heart Disease Prediction System Using Machine Learning and Data mining. If we can find out heart disease problem in early stages then it becomes very helpful for treatment. Machine Learning and Data Mining techniques are used for the construction of Heart Disease Prediction System. In healthcare biomedical field, there is large use of heath care data in the form of text, images, etc but, that data is hardly visited and is not mined. So, we can avoid this problem by introducing Heart Disease Prediction System. This system will help us reduce the costs and to enhance the quality treatment of heart patients. This system can able to identify complex problems and can able to take intelligent medical decisions. The system can predict likelihood of patients of getting heart problems by their profiles such as blood pressure, age, sex, cholesterol and blood sugar. Also, the performance will be compared by calculation of confusion matrix. This can help to calculate accuracy, precision, and recall. The overall system provides high performance and better accuracy.

INTRODUCTION

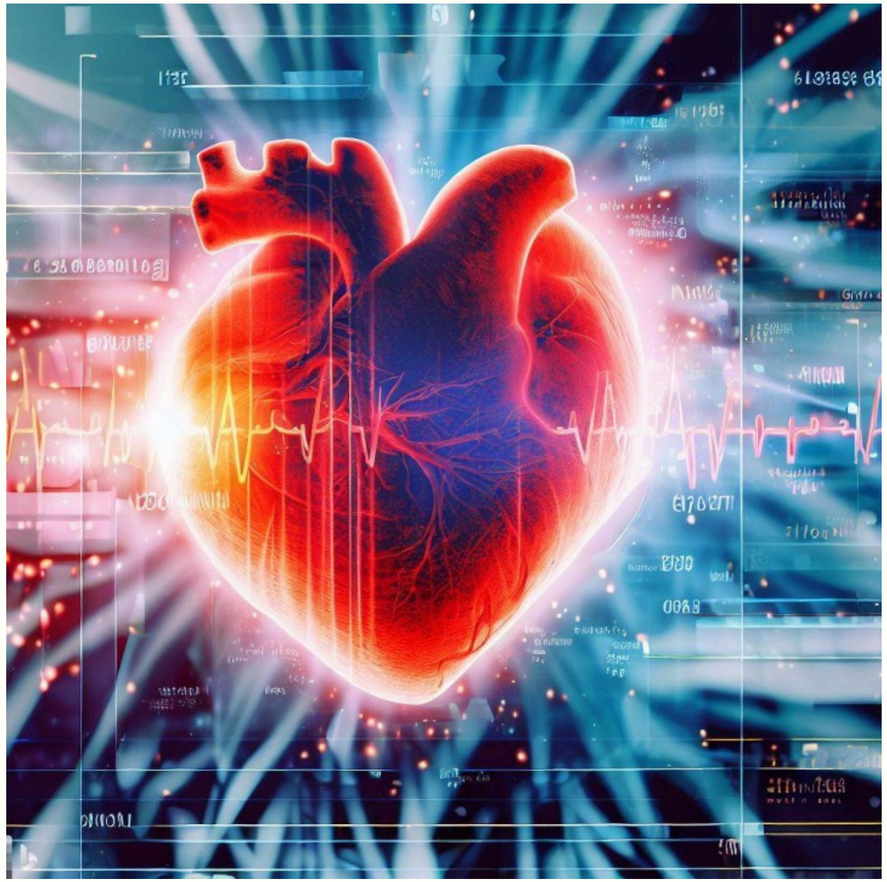
- The health care industries collect huge amounts of data that contain some hidden information, which is useful for making effective decisions. For providing appropriate results and making effective decisions on data, some advanced data mining techniques are used.
- In this study, a Heart Disease Prediction System (HDPS) is developed using logistic regression and correlation for predicting the risk level of heart disease. The system uses 13 medical parameters such as age, sex, blood pressure, cholesterol, and obesity for prediction.
- The HDPS predicts the likelihood of patients getting heart disease. It enables significant knowledge. E.g. Relationships between medical factors related to heart disease and patterns, to be established. We have employed the multilayer perceptron neural network with back propagation as the training algorithm.
- The obtained results have illustrated that the designed diagnostic system can effectively predict the risk level of heart diseases.



PROBLEM STATEMENT

Heart disease is one of the most common and costly diseases, often becoming dangerous and costly to treat in later stages. The goal is to develop a Heart Disease Prediction System using Machine Learning and Data Mining techniques. This system will predict the likelihood of patients developing heart problems based on their profiles, including parameters like blood pressure, age, sex, cholesterol, and blood sugar levels. The system aims to reduce healthcare costs and improve the quality of treatment by identifying heart disease early. It will utilize advanced data mining techniques to extract hidden information from healthcare data sets, allowing for more effective decision-making and improved patient outcomes.

OBJECTIVE



The primary objective of this project is to build a logistic regression model that accurately predicts the likelihood of heart disease in individuals using their medical data. By identifying key risk factors and understanding their impact on heart disease, we aim to create a tool that can assist healthcare professionals in early diagnosis and personalized treatment planning. The heart disease prediction project using logistic regression aims to develop a statistical model that can accurately predict a person's probability of heart disease based on various health-related characteristics.

OBJECTIVES

I

Identification of risk factors

Determine which health variables (eg, age, cholesterol, blood pressure) are important predictors of heart disease.

D

Developing models

Create a logistic regression model to classify people as having or not having heart disease based on their health data.

M

Model Evaluation

Evaluate model performance using metrics such as precision, accuracy, recall, and area under the ROC curve (AUC-ROC).

C

Clinical Application

Give healthcare professionals a tool to aid in early diagnosis and personalized treatments for patients at risk of heart disease.

METHODOLOGY

P

1. Problem Definition:

• Define the goal: To predict the probability of heart disease using health data.

• Identify the target variable (heart disease presence) and predictor variables (e.g., age, cholesterol, blood pressure).

D

2. Data Collection:

• Use publicly available datasets such as the Cleveland Heart Disease dataset from the UCI Machine Learning Repository.

• Ensure data is representative and includes relevant health indicators.

D

3. Data Preprocessing:

➢ Handle missing values and outliers.

➢ Normalize/standardize numerical features.

➢ Encode categorical variables.

➢ Split the dataset into training and test sets (e.g., 80-20 split).

METHODOLOGY

E

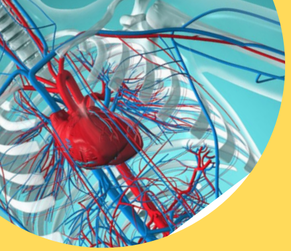
4. Exploratory Data Analysis (EDA):

- Perform descriptive statistics and visualizations to understand data distribution and relationships.
- Identify significant predictors through correlation analysis.

M

5. Model Building:

- Develop and train a logistic regression model using the training data.
- Perform feature selection to identify important predictors.



TECHNOLOGY USED

PROGRAMMING LANGUAGES

- **Python:** The primary language used for its extensive libraries and frameworks for data science and machine learning.

Data Manipulation and Analysis:

- **Pandas**: A powerful library for data manipulation and analysis, allowing for easy handling of data frames.
- **NumPy**: Provides support for large, multi-dimensional arrays and matrices, along with a collection of mathematical functions.

Data Visualization:

- **Matplotlib:** A plotting library for creating static, animated, and interactive visualizations in Python.
- **Seaborn:** Built on top of Matplotlib, it provides a high-level interface for drawing attractive and informative statistical graphics.

TECHNOLOGY USED

Machine Learning and Statistical Analysis:

- **Scikit-learn:** A robust library that offers simple and efficient tools for data mining and data analysis, including implementation of logistic regression and model evaluation metrics.
- **Stats models**: A library that complements Scikit-learn for statistical modeling and hypothesis testing.

Data Preprocessing:

- **Scikit-learn Preprocessing Modules:** For scaling, encoding, and normalizing data.

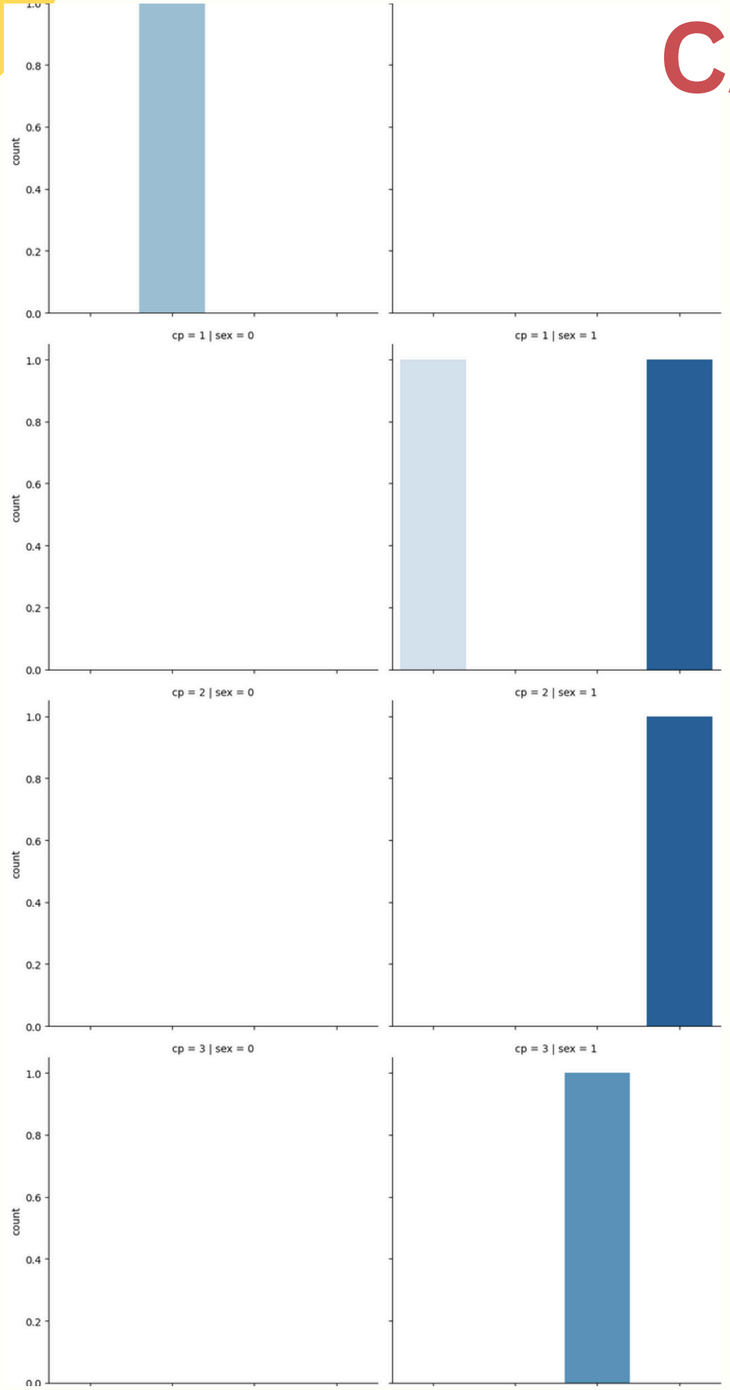
Model Evaluation

- **Scikit-learn Metrics**: For calculating accuracy, precision, recall, and AUC-ROC.

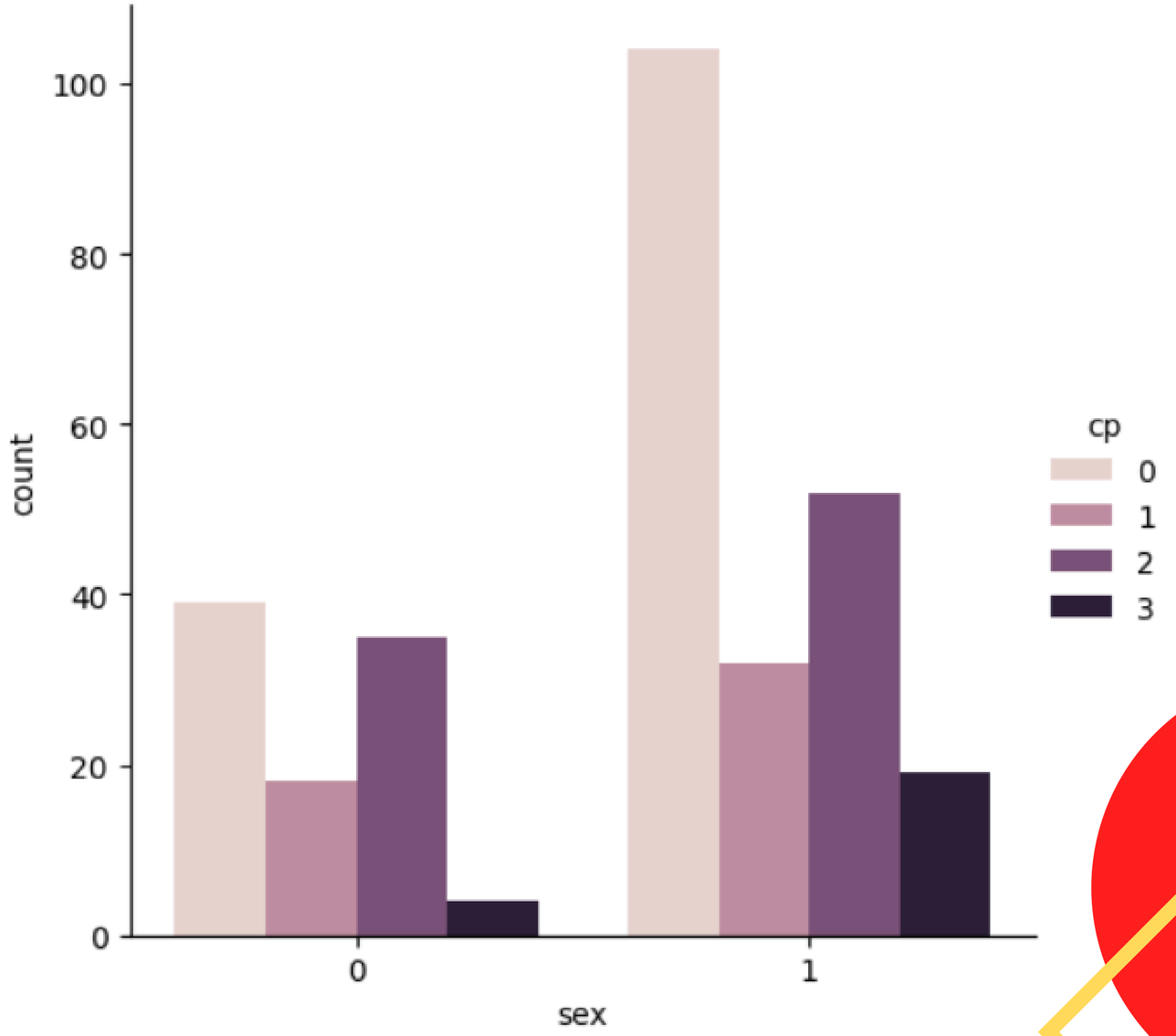
Integrated Development Environment (IDE):

- **Jupyter Notebook:** An open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text.
- **VS Code / PyCharm:** Alternative IDEs that provide advanced features for coding, debugging, and project management.

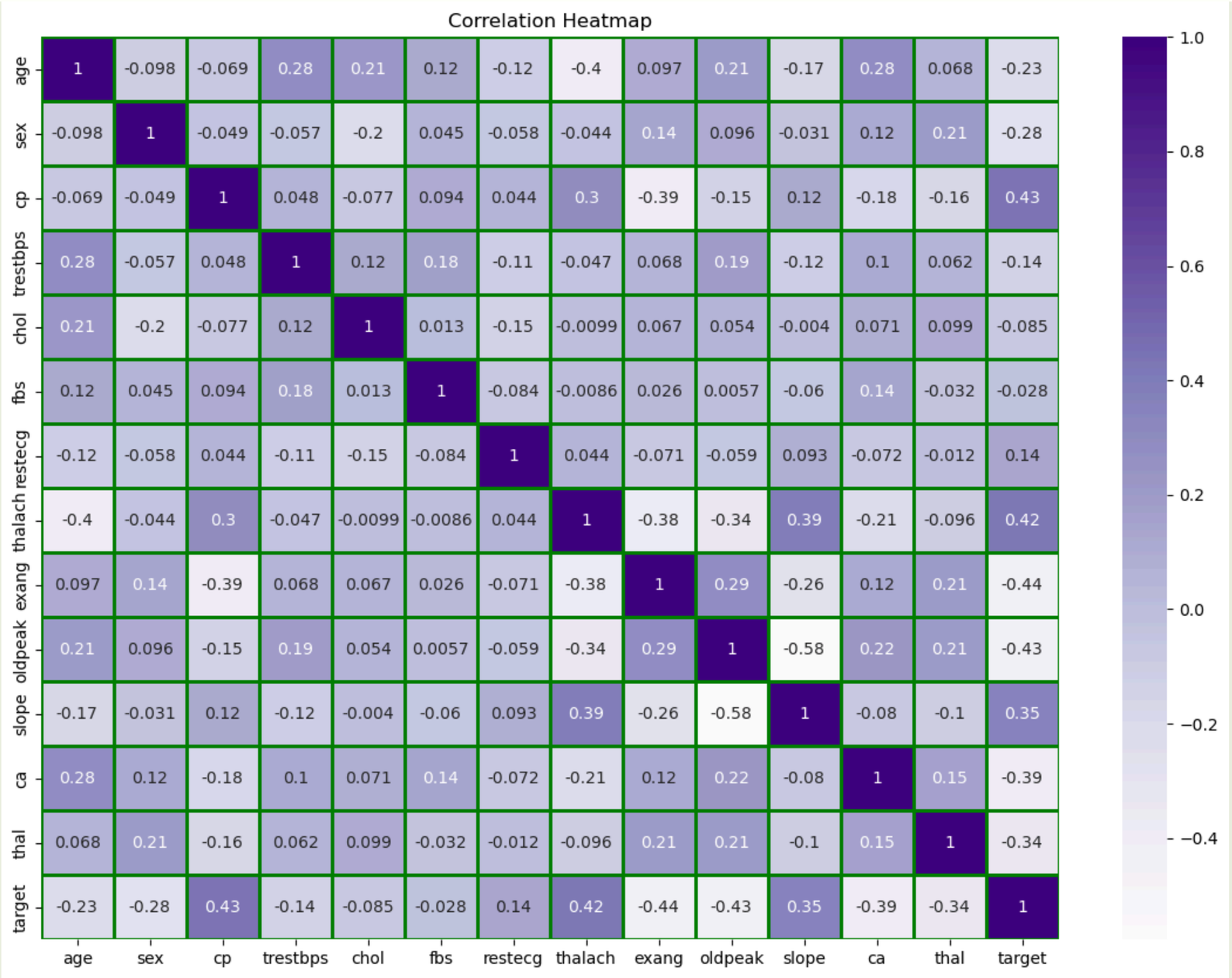
CAT PLOT



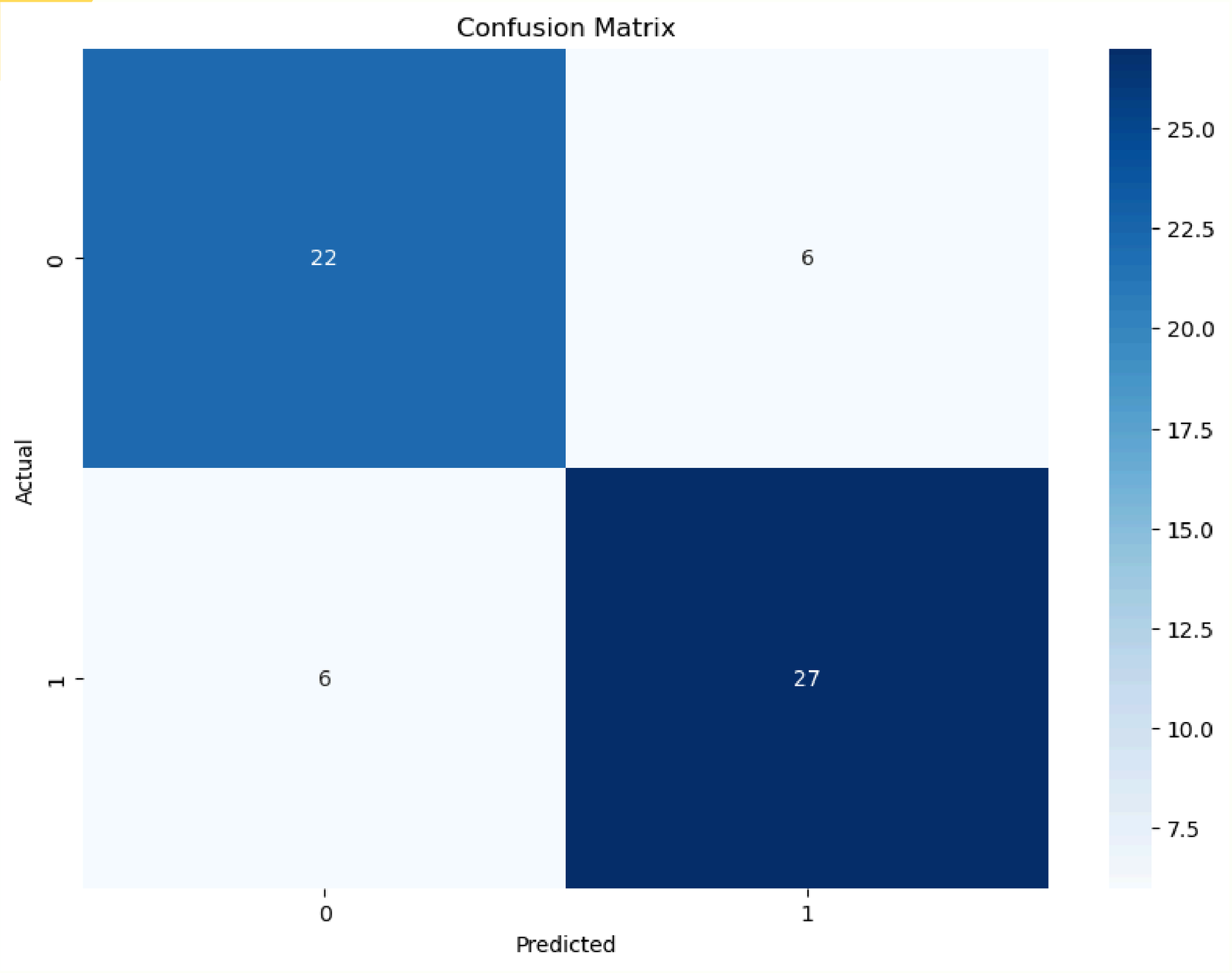

```
4]: sns.catplot(data=df, kind='count', x='sex',hue='cp')
    plt.show()
```



CORRELATION HEATMAP



PAIR PLOT GRAPH



CONFUSION MATRIX

CONCLUSION

The Heart Disease Prediction System (HDPS) uses machine learning and data mining techniques to predict the likelihood of heart disease, helping reduce treatment costs and improve patient outcomes. By analyzing 13 medical parameters such as age, sex, blood pressure, cholesterol, and blood sugar, the system can identify potential heart disease in its early stages. Logistic regression and correlation matrix are employed for prediction and explanation, ensuring accurate and insightful results. Although the current system does not include a web interface, it effectively aids in medical decision-making by revealing hidden patterns and relationships in the data. The performance of the system is validated using a confusion matrix, which helps calculate key metrics like accuracy, precision, and recall, demonstrating the system's high performance and reliability.

REFERENCES

- https://github.com/Ravjot03/Heart-Disease-Prediction
- https://www.researchgate.net/publication/368848738_Heart_Disease_Prediction_Using_Logistic_Regression.
- Mozaffarian, D., Benjamin, E., Go, A., Arnett, D., Blaha, M.Cushmans, M. et al. (2015). Heart Disease and Stroke Statistics—2015, Update. Circulation, 131(4). doi: 10.1161/cir.0000000000000152.
- Das, S., Dey, A., Pal, A., & Roy, N. (2015). Applications of Artificial Intelligence in Machine Learning: Review and Prospect. International Journal of Computer Applications, 115(9), 31-41.doi: 10.5120/20182-2402
- Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. (2019). Applications of Artificial Intelligence in Transport: An Overview. Sustainability, 11(1), 189.doi: 10.3390/su11010189

Thank you