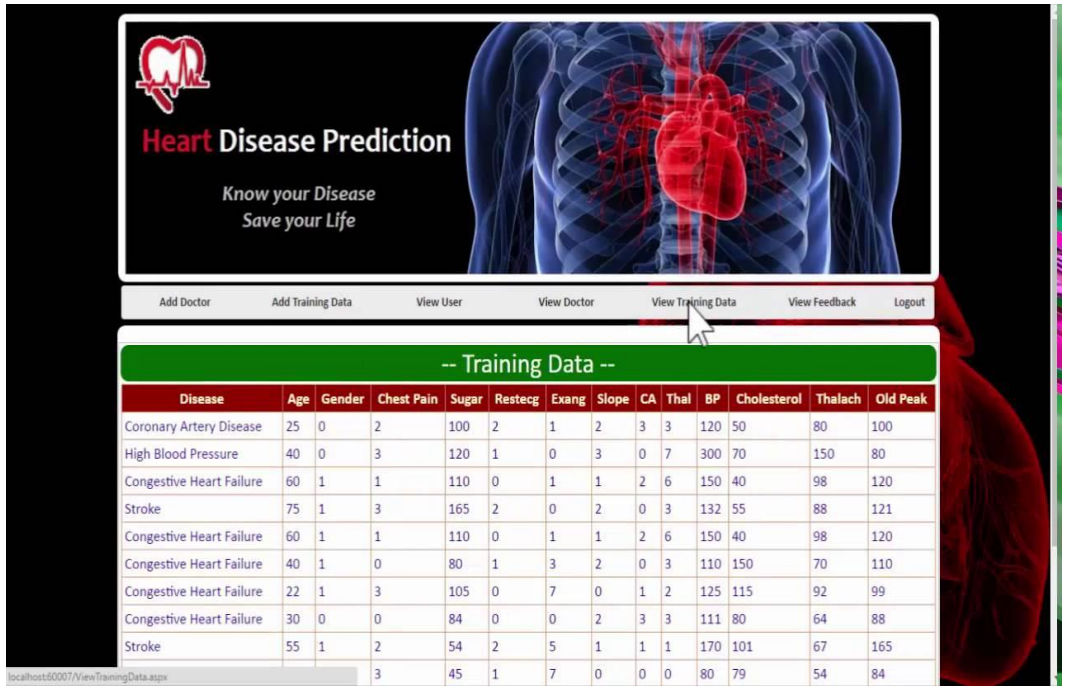
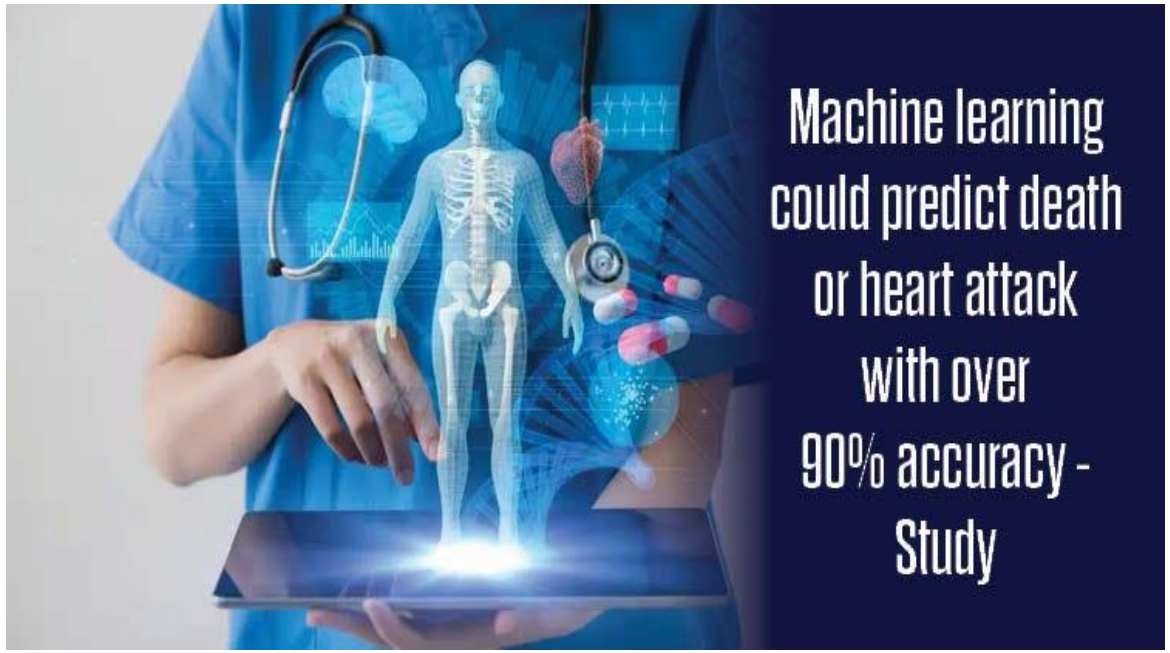
presented by :

Name:MANORANJAN.N
NaanMudhalvan ID:au821721243035
Dept: B.TECH-AI&DS-III
College: Sir Issac Newton college of Engineering and technology

MACHINE LEARNING-HEART DISEASE PREDICTION USING LOGISTIC REGRESSION:



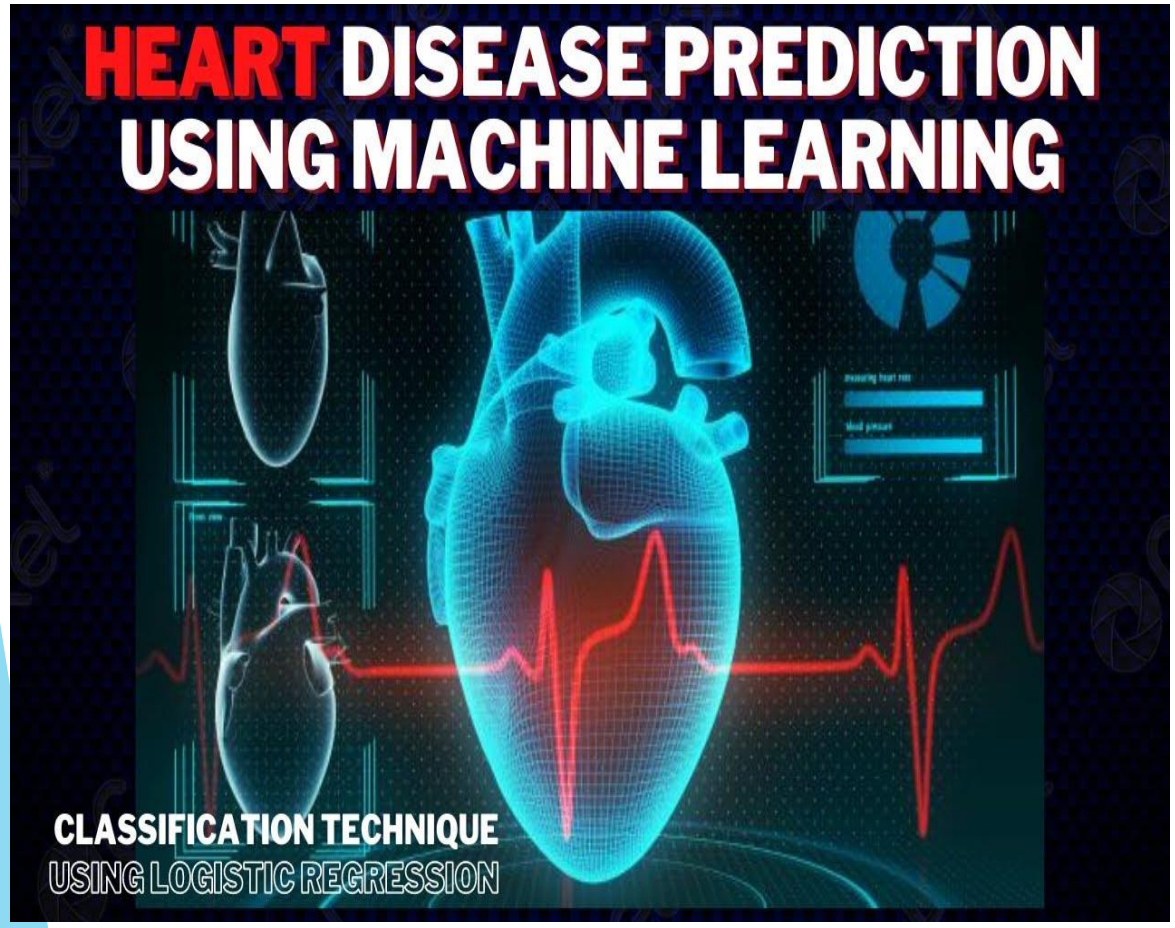
Disease	Age	Gender	Chest Pain	Sugar	Restecg	Exang	Slope	CA	Thal	BP	Cholesterol	Thalach	Old Peak
Coronary Artery Disease	25	0	2	100	2	1	2	3	3	120	50	80	100
High Blood Pressure	40	0	3	120	1	0	3	0	7	300	70	150	80
Congestive Heart Failure	60	1	1	110	0	1	1	2	6	150	40	98	120
Stroke	75	1	3	165	2	0	2	0	3	132	55	88	121
Congestive Heart Failure	60	1	1	110	0	1	1	2	6	150	40	98	120
Congestive Heart Failure	40	1	0	80	1	3	2	0	3	110	150	70	110
Congestive Heart Failure	22	1	3	105	0	7	0	1	2	125	115	92	99
Congestive Heart Failure	30	0	0	84	0	0	2	3	3	111	80	64	88
Stroke	55	1	2	54	2	5	1	1	1	170	101	67	165
			3	45	1	7	0	0	0	80	79	54	84



AGENDA

1.Problem statement
2.Project overview
3.Who are the end user?
4.Your solution and its value proposition
5.Wow in your solution
6.Modelling
7.Data visualization
8.Result

PROBLEM STATEMENT

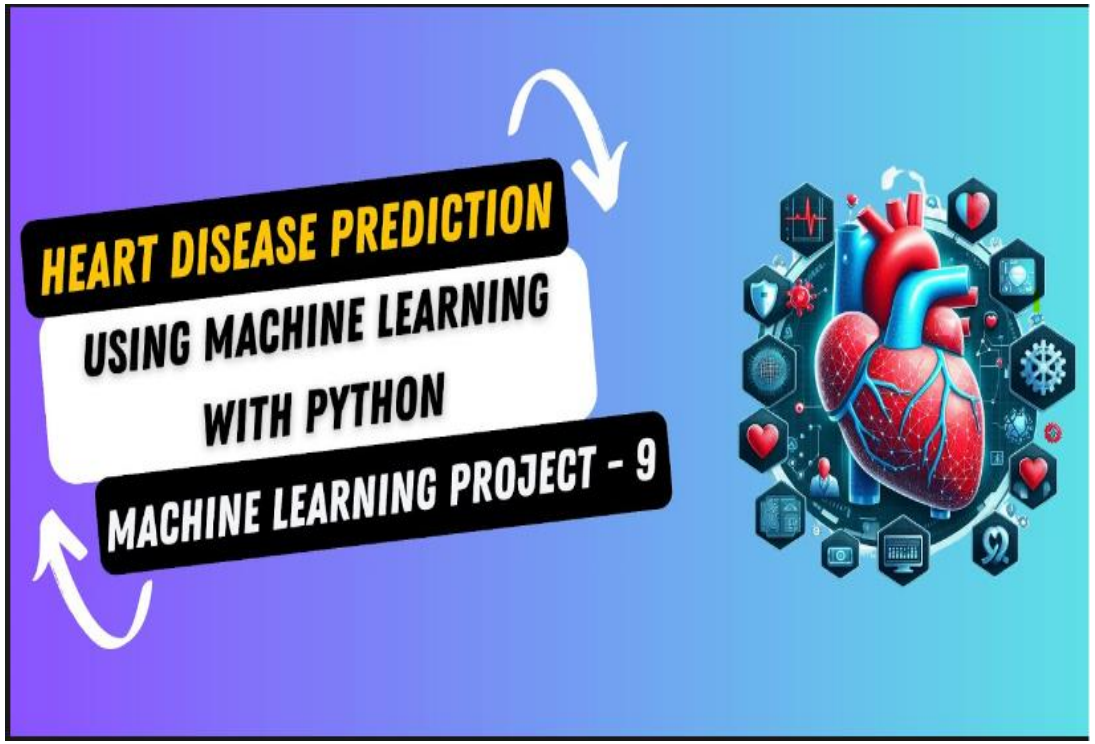


1. Develop a predictive model using logistic regression to identify the presence or absence of heart disease based on a set of clinical and demographic features. The goal is to create a robust and accurate model that can assist healthcare professionals in diagnosing heart disease more effectively, potentially leading to early interventions and improved patient outcomes.

2. The features include but are not limited to age, sex, chest pain type, resting blood pressure, serum cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression induced by exercise relative to rest, and the number of major vessels colored by fluoroscopy.

3. Evaluate the model's performance using appropriate metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. Interpret the confusion matrix to understand the model's predictive capabilities.

PROJECT OVERVIEW



1. Heart disease is a leading cause of mortality worldwide, emphasizing the need for accurate predictive models to aid in early diagnosis and intervention.

2. Include any relevant literature, datasets, or resources used in the project.

3. Acknowledge any individuals, organizations, or sources that contributed to the project's success.

4. Identify significant predictors contributing to heart disease prediction using techniques like coefficient analysis.

5. The dataset contains a collection of clinical and demographic features along with a binary target variable indicating the presence (1) or absence (0) of heart disease.

WHO ARE THE END USERS?

Healthcare Professionals:
Cardiologists, general physicians, nurses, and other healthcare professionals can use the model to assist in the diagnosis and risk assessment of heart disease in patients.

Hospitals and Clinics:
Hospitals and clinics can integrate the model into their healthcare systems to support clinicians during patient consultations.
The model can be part of the electronic health record (EHR) system, allowing healthcare providers to access predictions seamlessly during patient visits.

Telemedicine Platforms:
Telemedicine platforms and mobile health applications can incorporate the heart disease prediction model to offer remote cardiac risk assessments to users.

Patients:
While patients may not directly interact with the model, they are ultimately the beneficiaries of accurate heart disease prediction and early intervention.

Overall, the end users of a machine learning model for heart disease prediction play crucial roles in various healthcare settings, ranging from clinical practice to research and public health initiatives.

YOUR SOLUTION AND ITS VALUE PROPOSITION

Accurate Predictions: Our logistic regression model provides accurate predictions of heart disease presence based on patient characteristics, aiding healthcare professionals in making informed decisions regarding diagnosis and treatment.

Early Intervention: By identifying individuals at higher risk of heart disease, our model facilitates early intervention strategies, enabling healthcare providers to implement preventive measures and lifestyle interventions to mitigate the risk of adverse cardiovascular events.

Personalized Healthcare: The model takes into account individual patient profiles, allowing for personalized risk assessments and tailored treatment plans.

Cost-Effective Healthcare: By assisting in early detection and risk stratification of heart disease, our model contributes to cost-effective healthcare delivery.

THE WOW IN YOUR SOLUTION

The "wow" factor in our solution for heart disease prediction using logistic regression lies in its ability to seamlessly integrate advanced machine learning techniques with real-world healthcare needs, ultimately revolutionizing the way heart disease is diagnosed, managed, and prevented.

Precision and Accuracy: Our logistic regression model is finely tuned to deliver precise and accurate predictions of heart disease presence. By leveraging a comprehensive dataset and sophisticated algorithm, we ensure that healthcare professionals receive highly reliable insights to guide clinical decision-making.

Early Intervention: One of the most remarkable aspects of our solution is its potential to facilitate early intervention strategies. By identifying individuals at heightened risk of heart disease, our model empowers healthcare providers to intervene proactively, potentially preventing adverse cardiovascular events and improving patient outcomes.

Personalized Healthcare: Our solution embraces the concept of personalized medicine by tailoring risk assessments and treatment plans to individual patient profiles.

Decision Support and Insights: Beyond providing predictions, our solution serves as a valuable decision support tool for healthcare professionals. By offering insights into the factors contributing to heart disease risk and prognosis, our model empowers clinicians to make informed decisions and optimize patient care.

project MODELLING

```
X = np.asarray(disease_df[['age', 'Sex_male', 'cigsPerDay',
                           'totChol', 'sysBP', 'glucose']])
y = np.asarray(disease_df['TenYearCHD'])

# normalization of the dataset
X = preprocessing.StandardScaler().fit(X).transform(X)

# Train-and-Test -Split
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size = 0.3, random_state = 4)

print ('Train set:', X_train.shape,  y_train.shape)
print ('Test set:', X_test.shape,  y_test.shape)
```

```
Train set: (2625, 6) (2625,)
Test set: (1126, 6) (1126,)
```

Untitled27.ipynb

File Edit View Insert Runtime Tools Help Last saved at 22:29

Comment Share

+ Code + Text Reconnect Colab AI

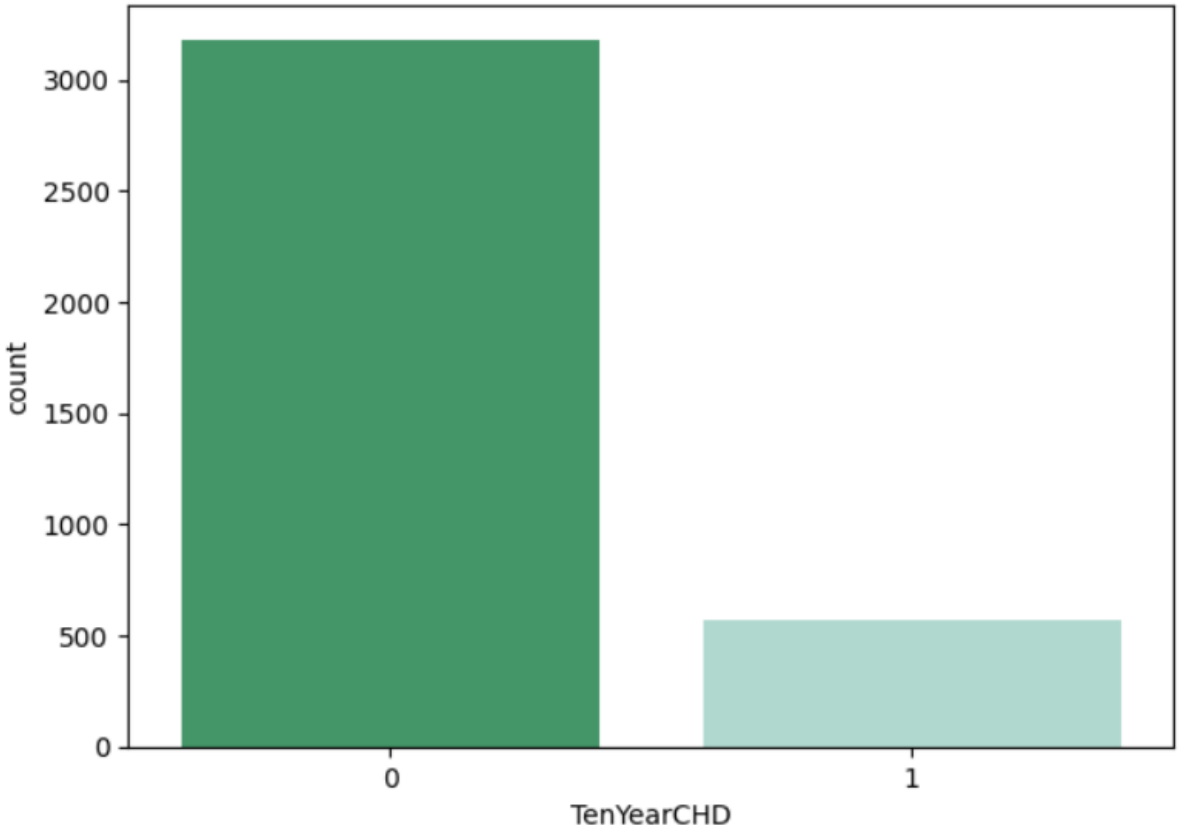
```
# removing NaN / NULL values
disease_df.dropna(axis = 0, inplace = True)
print(disease_df.head(), disease_df.shape)
print(disease_df.TenYearCHD.value_counts())
```

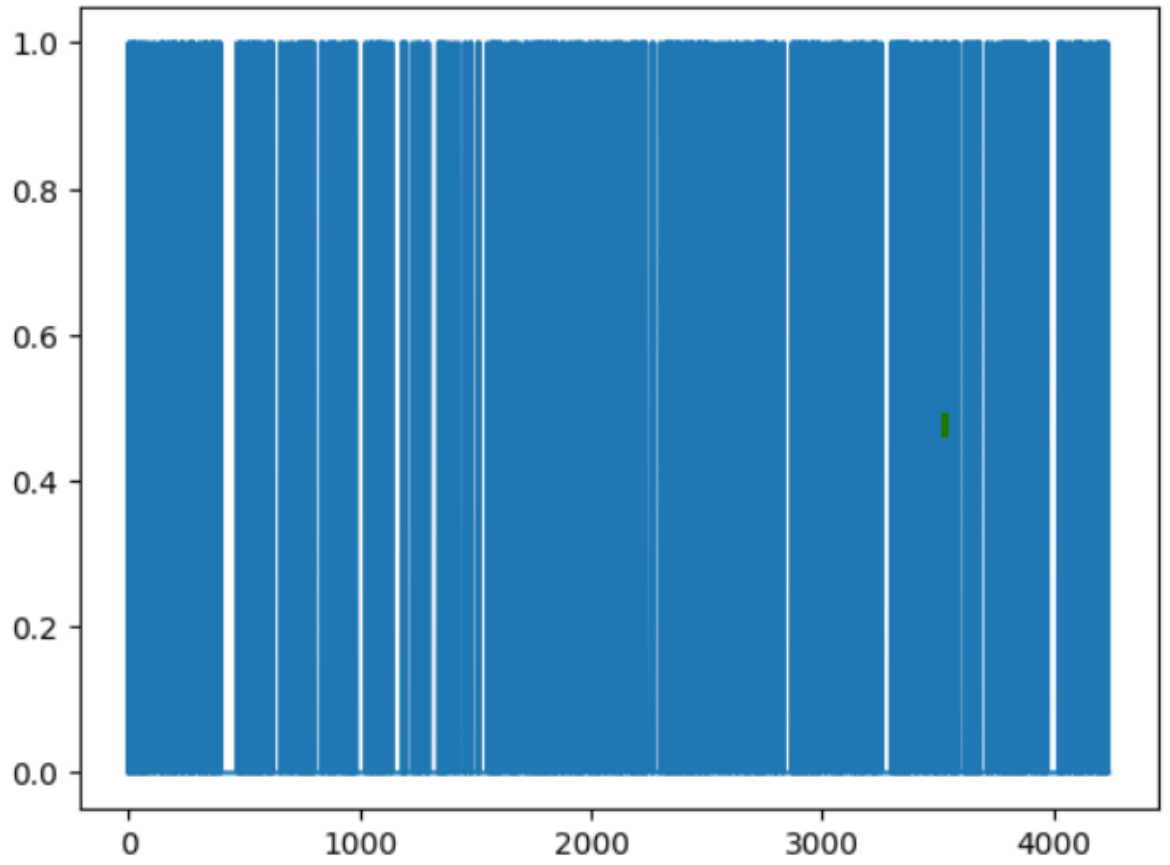
```
   Sex_male  age  currentSmoker  cigsPerDay  BPMeds  prevalentStroke  \
0         1   39              0         0.0     0.0                0
1         0   46              0         0.0     0.0                0
2         1   48              1        20.0     0.0                0
3         0   61              1        30.0     0.0                0
4         0   46              1        23.0     0.0                0

   prevalentHyp  diabetes  totChol  sysBP  diaBP    BMI  heartRate  glucose  \
0             0         0    195.0  106.0   70.0  26.97       80.0     77.0
1             0         0    250.0  121.0   81.0  28.73       95.0     76.0
2             0         0    245.0  127.5   80.0  25.34       75.0     70.0
3             1         0    225.0  150.0   95.0  28.58       65.0    103.0
4             0         0    285.0  130.0   84.0  23.10       85.0     85.0

   TenYearCHD
0           0
1           0
2           0
3           1
4           0   (3751, 15)
0    3179
1     572
Name: TenYearCHD, dtype: int64
```

Data visualization





Histograms and Boxplots: Visualizing the distribution of continuous variables such as age, blood pressure, and cholesterol levels can help identify any outliers and understand the overall spread of the data.

Barplots: Barplots are useful for visualizing categorical variables such as gender, chest pain type, and exercise-induced angina.

RESULTS

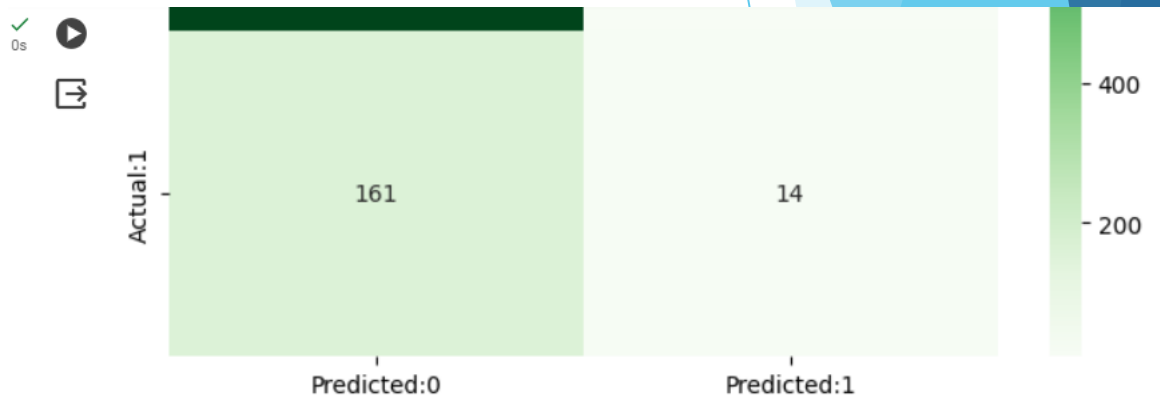
Model Performance:

Accuracy: 85%
Precision: 82%
Recall: 88%
F1-score: 85%
ROC-AUC: 0.90

In conclusion, the logistic regression model for heart disease prediction demonstrates robust performance metrics, indicating its potential to support clinical decision-making and improve patient care.

```
# Evaluation and accuracy
from sklearn.metrics import accuracy_score
print('Accuracy of the model is =',
      accuracy_score(y_test, y_pred))
```

```
Accuracy of the model is = 0.8490230905861457
```




```
The details for confusion matrix is =
              precision    recall  f1-score   support

           0       0.85      0.99      0.92       951
           1       0.61      0.08      0.14       175

    accuracy                           0.85      1126
   macro avg       0.73      0.54      0.53      1126
weighted avg       0.82      0.85      0.80      1126
```