JAYPEE INSTITUTE OF INFORMATION TECHNOLOGY

OPEN SOURCE SOFTWARE LAB

Heart Failure Prediction Model

Submitted to : VIKASH SIR

SUBMITTED BY –

Astha Raghuwanshi (21103042)

Prerna (21103043)

Princi Agrawal (21103048)

Tanya Gupta (21103279)

INTRODUCTION

This report focuses on the analysis of a heart failure prediction model. The model is based on a logistic regression classifier and is designed to predict the likelihood of a heart failure event based on 12 patient attributes. We will cover various aspects of the model, including evaluation metrics, visualization of results, and how to use the model forpredictions.

Model Creation and Implementation

Importing DataSet

Firstly, we will mount the file containing the heart_failure.csv dataset from the documents and import it in our google colab using the pandas library by read_csv() function.

Data Preprocessing

Now we will distribute the dataset saved inform of data frames into the features variables and the prediction variable using the iloc function of the numpy library andthen divide the features variable into the train and test dataset.

Model Implementation and Training the data

Now we will implement our model that we are going to use in this project that is Logistic Regression from model_selection and then train our test_data to the Modelwith the fit() function.

So, after training and saving the logistic regression model, we have the ability to load the saved model and make predictions based on user input based on those 12 attributes.

Model Saving

The heart failure prediction model is saved using the pickle library. This allows for easy retrieval and reuse of the model in the future.

Model Loading

To use the saved model for making predictions, you need to load it from the saved file. The loaded model helps us to make predictions without needing to retrain the model.

Making Predictions for the model

With the loaded model and user's input that was previously stored as a list, is converted into a 2D NumPy array that can be used for prediction.

In our heart failure prediction code, the four tools (Confusion Matrix, ROC Curve, Precision-Recall Curve, and Correlation Heatmap) are usedto evaluate the performance of our logistic regression model and to gaininsights into the data.

This is how they are used:

1. Confusion Matrix:

- The confusion matrix is used to assess the model's performance interms of correctly and incorrectly classified instances. It is generated using ‘confusion_matrix’ from scikit-learn.

- The generated confusion matrix is visualized as a heatmap usingSeaborn and Matplotlib.

- This allows you to see how many true positives, true negatives, falsepositives, and false negatives our model produced.

2. ROC Curve:

- The ROC curve is used to assess the model's ability to distinguish between positive (heart failure) and negative (no heart failure) cases.

- The ROC curve is plotted using Matplotlib.

3. Precision-Recall Curve:

- The Precision-Recall curve and the area under the curve (AUC-PR)are used to assess the precision and recall trade-off of the model.

- The Precision-Recall curve is plotted using Matplotlib.

- A higher AUC-PR indicates a better precision-recall trade-off.

4. **Correlation Heatmap**:

- A correlation heatmap is generated to visualize the relationshipsbetween different variables in your dataset. The code calculates and displays the correlation matrix using Pandas and Seaborn:

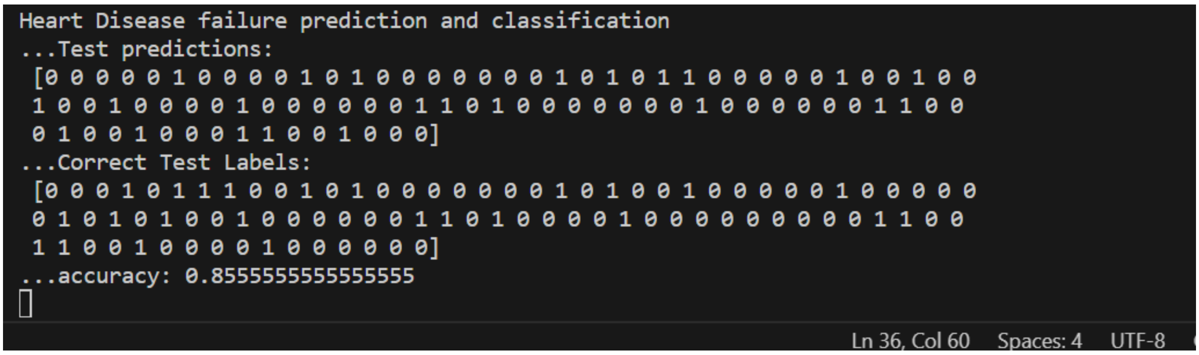
- This heatmap helps you understand the correlations between thefeatures in your dataset.

These tools and visualizations provide a comprehensive assessment of our logistic regression model's performance and offer insights into our data's characteristics. They help us to understand how well our model is making predictions and whether there are any dependencies between thefeatures in your dataset.

Conclusion

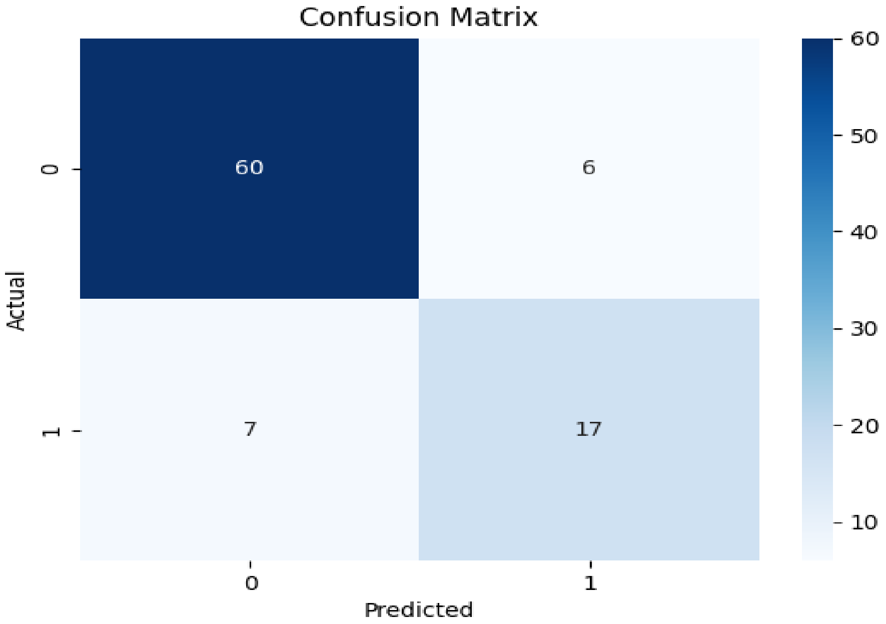
In this report, we have analyzed a heart failure prediction model, including its creation, evaluation, and how to use it for predictions. Themodel's performance was assessed using metrics like the confusion matrix, ROC curve, and Precision-Recall curve. Additionally, weexamined the relationships between patient attributes using a correlation heatmap. The model can be easily loaded from a saved file and used to make predictions based on user input. This analysis providesa comprehensive overview of the heart failure prediction model, enablingboth model assessment and real-world predictions.

OUTPUT

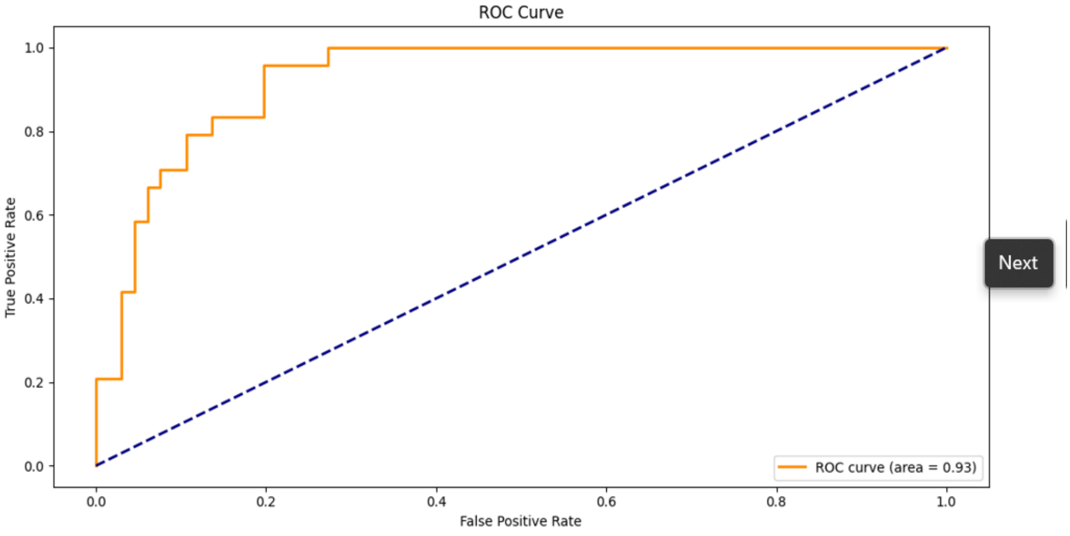


```
Heart Disease failure prediction and classification
...Test predictions:
 [0 0 0 0 0 1 0 0 0 0 1 0 1 0 0 0 0 0 0 0 1 0 1 0 1 1 0 0 0 0 0 1 0 0 1 0 0
 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 1 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 1 0 0
 0 1 0 0 1 0 0 0 1 1 0 0 1 0 0 0]
...Correct Test Labels:
 [0 0 0 1 0 1 1 1 0 0 1 0 1 0 0 0 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 1 0 0 0 0 0
 0 1 0 1 0 1 0 0 1 0 0 0 0 0 0 1 1 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 1 0 0
 1 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0]
...accuracy: 0.8555555555555555
▯
```

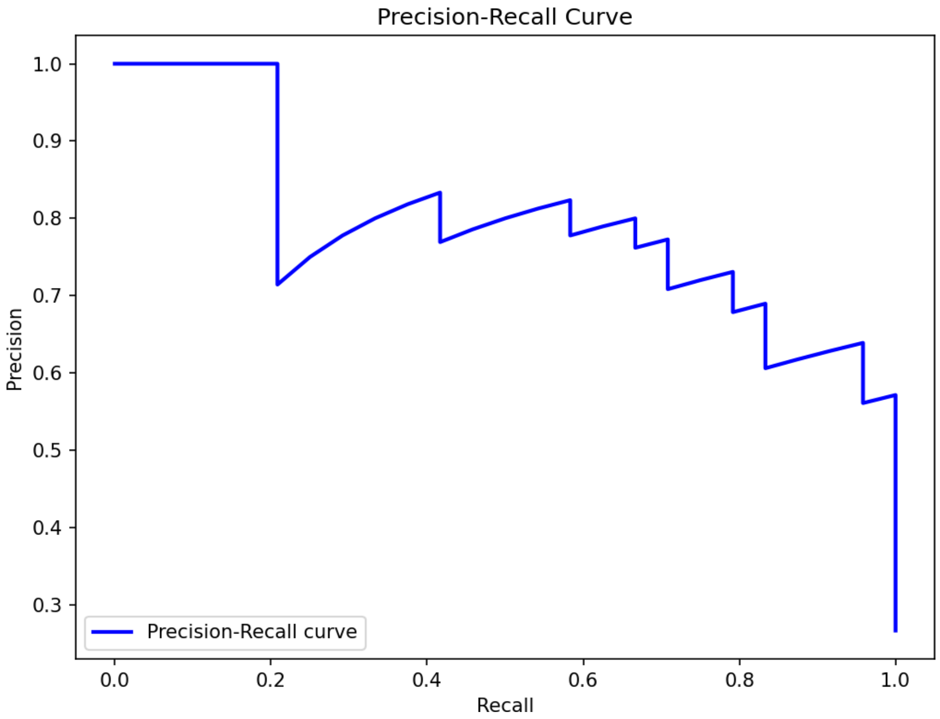
CONFUSION MATRIX



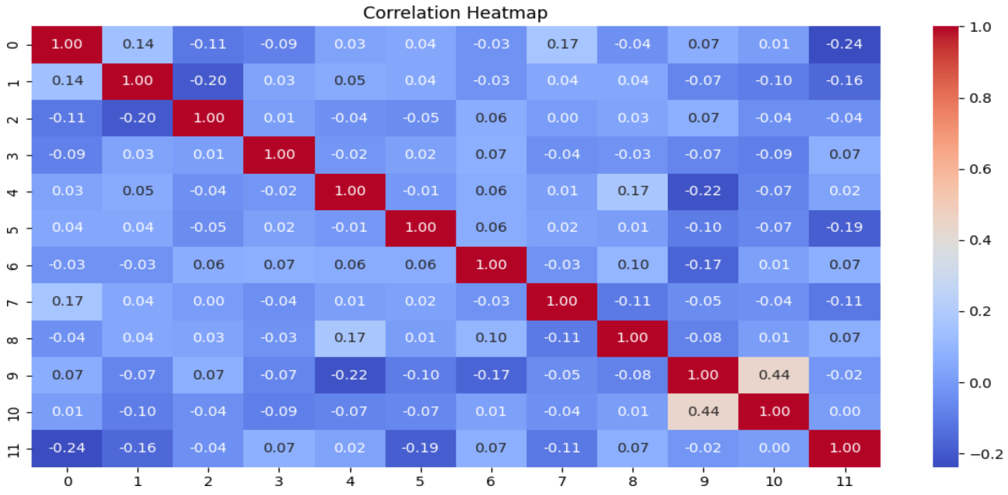
ROC CURVE:



PRECISION-RECALL CURVE:



CORRELATION HEATMAP:



	0	1	2	3	4	5	6	7	8	9	10	11
0	1.00	0.14	-0.11	-0.09	0.03	0.04	-0.03	0.17	-0.04	0.07	0.01	-0.24
1	0.14	1.00	-0.20	0.03	0.05	0.04	-0.03	0.04	0.04	-0.07	-0.10	-0.16
2	-0.11	-0.20	1.00	0.01	-0.04	-0.05	0.06	0.00	0.03	0.07	-0.04	-0.04
3	-0.09	0.03	0.01	1.00	-0.02	0.02	0.07	-0.04	-0.03	-0.07	-0.09	0.07
4	0.03	0.05	-0.04	-0.02	1.00	-0.01	0.06	0.01	0.17	-0.22	-0.07	0.02
5	0.04	0.04	-0.05	0.02	-0.01	1.00	0.06	0.02	0.01	-0.10	-0.07	-0.19
6	-0.03	-0.03	0.06	0.07	0.06	0.06	1.00	-0.03	0.10	-0.17	0.01	0.07
7	0.17	0.04	0.00	-0.04	0.01	0.02	-0.03	1.00	-0.11	-0.05	-0.04	-0.11
8	-0.04	0.04	0.03	-0.03	0.17	0.01	0.10	-0.11	1.00	-0.08	0.01	0.07
9	0.07	-0.07	0.07	-0.07	-0.22	-0.10	-0.17	-0.05	-0.08	1.00	0.44	-0.02
10	0.01	-0.10	-0.04	-0.09	-0.07	-0.07	0.01	-0.04	0.01	0.44	1.00	0.00
11	-0.24	-0.16	-0.04	0.07	0.02	-0.19	0.07	-0.11	0.07	-0.02	0.00	1.00

References:

- https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data