|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Faculty of Engineering & Technology | | | | |
| Ramaiah University of Applied Sciences | | | | |
| Department | Computer Science and Engineering | | | |
| Programme | B. Tech. in Machine Learning and Intelligent Systems | | | |
| Batch | Full-Time  2021 | | Course Start Date | 5-06-2023 |
| Course Code | | 21AIC215A | | |
| Course Title | | Machine Learning-1 | | |
| Course Leader(s) | | Dr.Divya BS | | |

|  |  |  |  |
| --- | --- | --- | --- |
| Assignment Assessment | | | |
| Reg.No. | 21ETAI410033 | Name of the Student | Bhoomika K |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1. A Report on Assignment | | | | |
| Sections | Problem Statement | | Marks | |
| Max Marks | Examiner Marks |
|  | 2.1 | Design a machine learning algorithm (MLA) on the dataset of your interest | 10 |  |
| 2.2 | Visualise the dataset using Matplotlib and Seaborn | 5 |  |
| 2.3 | Discuss the performance of the MLA on the dataset | 5 |  |
| 2.4 | Demonstration and Conclusion | 5 |  |
|  | **Max Marks** | 25 |  |
|  |  |  | 25 |  |

**Instructions to students:**

1. The assignment has to be neatly word processed as per the prescribed format.
2. The maximum number of pages should be restricted to **35**.
3. Use only SI units
4. **Submission Date: 08/09/23**
5. **Submission after the due date is not permitted.**
6. Method of evaluation as per the submission and marking scheme
7. At the end, you are required to comment on -
   1. Benefits you have derived by solving this assignment
   2. Whether assignment was able to assess *module learning outcomes* or not?
8. IMPORTANT: It is essential that all the sources used in preparation of the assignment must be suitably referenced in the text.

#### 

#### Guidelines for Submission

*Citation* of the *references* in the text is necessary in all the parts.

* ***Please note: Marks will be awarded only to the sections and sub-sections clearly indicated as per the problem statement.***

🙚🙘

Introduction to Machine Learning

Machine learning is programming computers to optimize a performance criterion using example data or past experience. We have a model defined up to some parameters, and learning is the execution of a computer program to optimize the parameters of the model using the training data or past experience. The model may be predictive to make predictions in the future, or descriptive to gain knowledge from data.

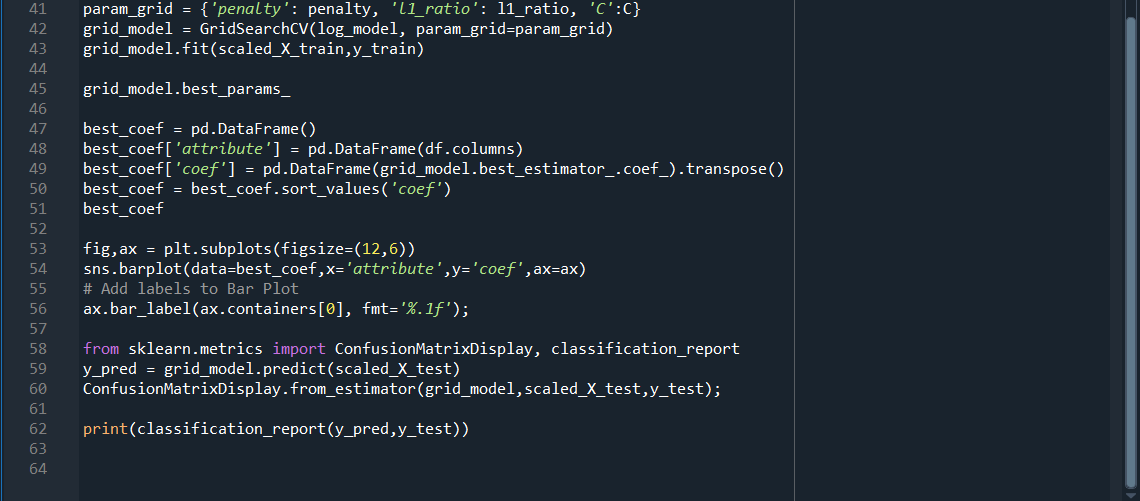
The field of study known as machine learning is concerned with the question of how to construct computer programs that automatically improve with experience.

Chosen Dataset: Heart Diseases

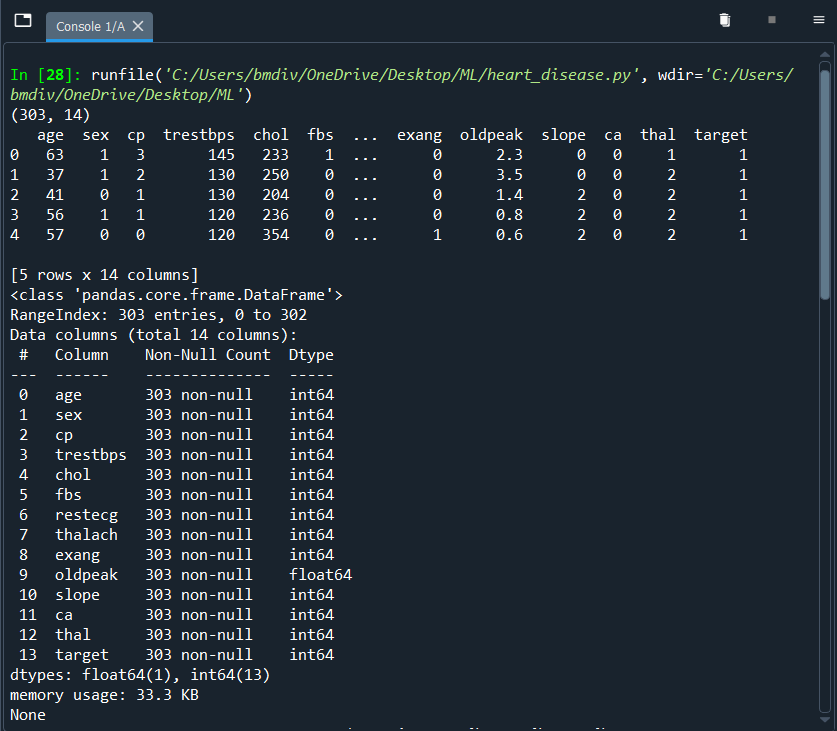
We have a data which classified if patients have heart disease or not according to features in it. We will try to use this data to create a model which tries predict if a patient has this disease or not. We will use logistic regression (classification) algorithm.

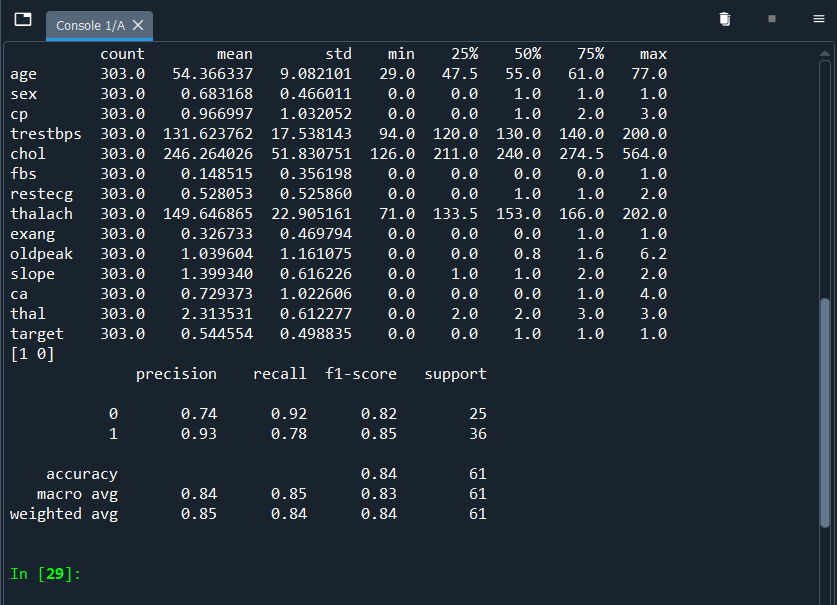
Used Code:

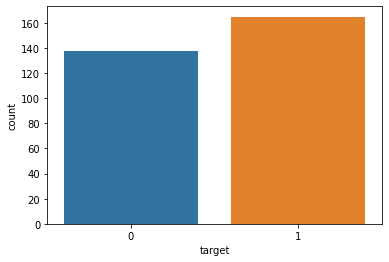


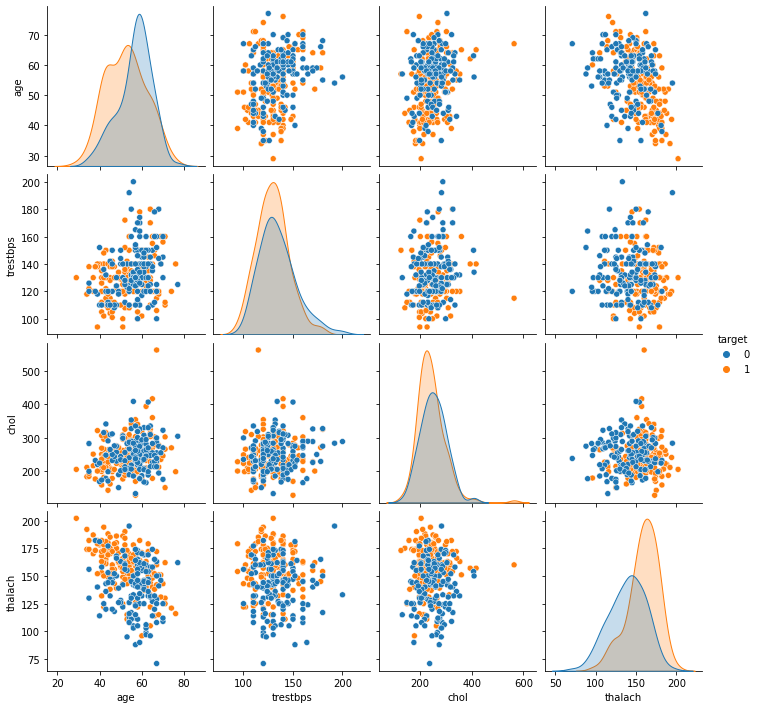


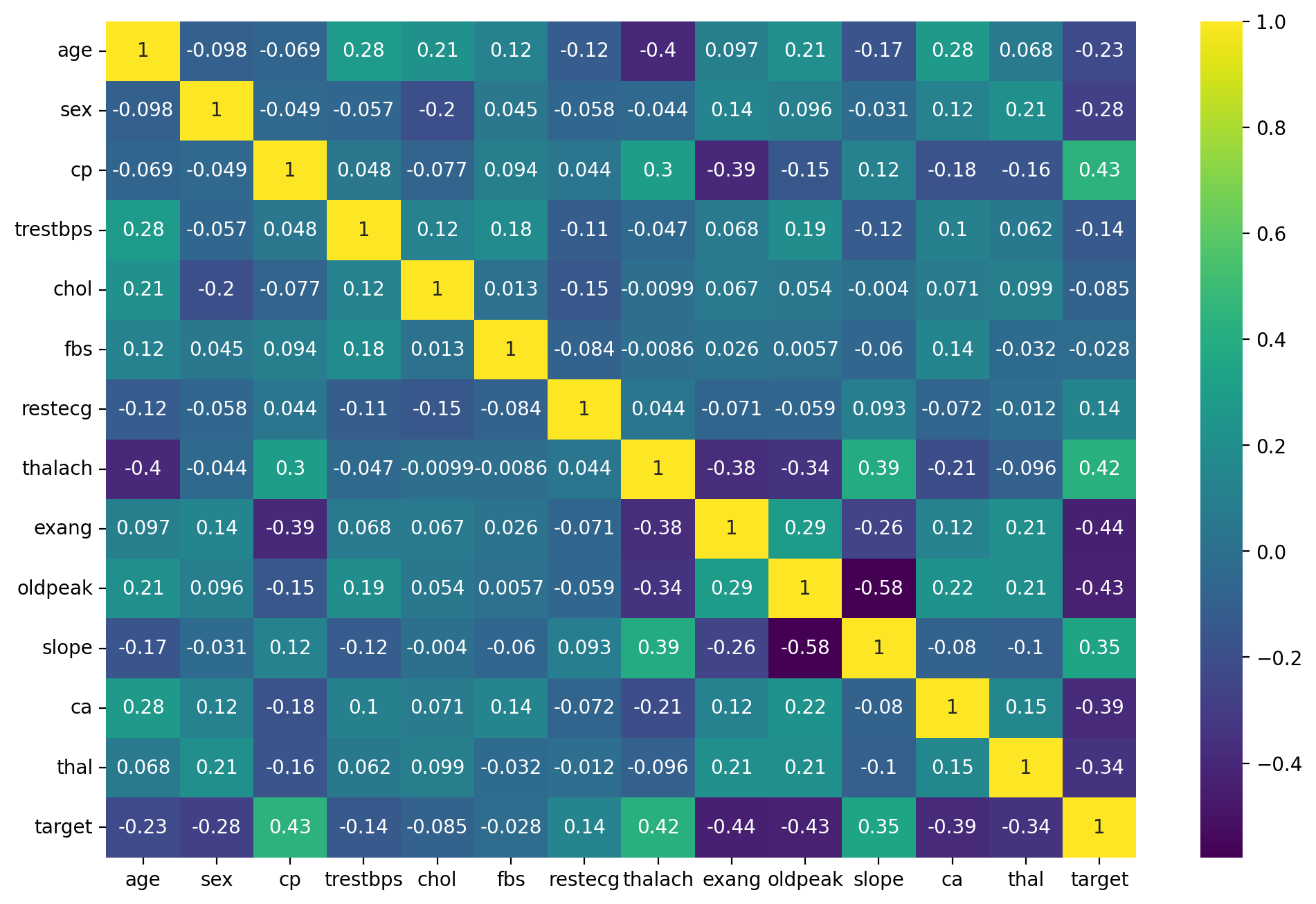
Results:

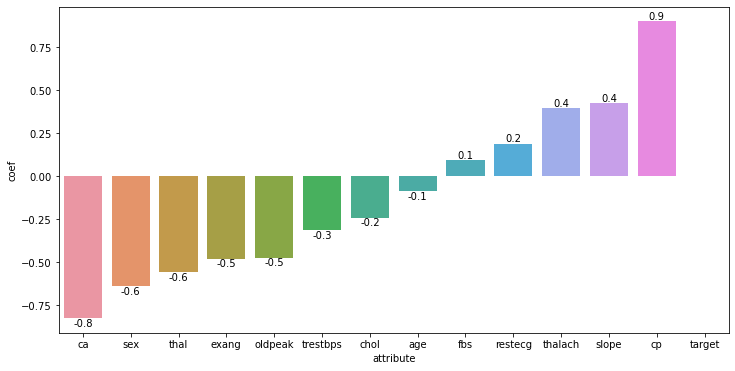


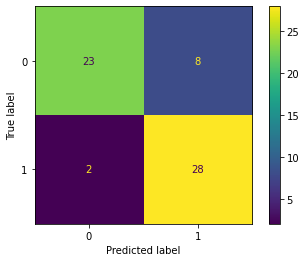












**Performance of the Machine Learning Algorithm (MLA):**

1. **Data Exploration:** The code began by exploring the dataset's shape and displaying the first few rows to get an initial understanding of the data. You also checked for missing data and provided a summary of the dataset's statistics. This is a good practice for understanding your data.
2. **Data Visualization:** We used Seaborn to create visualizations. The countplot showed the distribution of the target variable, which indicates the presence or absence of heart disease. Pairplots allowed us to visualize relationships between selected attributes with respect to the target variable. The heatmap displayed the correlation matrix of the features. These visualizations help in understanding the dataset's characteristics and potential relationships.
3. **Data Preprocessing:** We split the dataset into training and testing sets and applied standard scaling to the features, which is crucial for many machine learning algorithms to work effectively.
4. **Model Selection and Hyperparameter Tuning:** We chose logistic regression as the initial machine learning algorithm and used GridSearchCV to perform hyperparameter tuning. The grid search explored different penalty types, L1 ratio, and regularization strengths (C) to find the best combination of hyperparameters.
5. **Model Evaluation:** We visualized the coefficients of the best model attributes and provided a bar plot with labels. This helps in understanding which features contribute the most to the model's predictions.
6. **Confusion Matrix and Classification Report:** We displayed the confusion matrix and a classification report for the model's predictions on the test set, which includes metrics such as precision, recall, and F1-score for each class.

**Demonstration and Conclusion:**

The code demonstrates a comprehensive machine learning workflow, from data exploration and preprocessing to model training and evaluation. Here's a summary of the key findings and conclusions:

* The model's hyperparameters were tuned using GridSearchCV, which should lead to improved model performance compared to using default hyperparameters.
* We observed the feature coefficients, which indicated which attributes have the most influence on the model's predictions. This can be valuable for domain experts seeking insights into heart disease prediction.
* The confusion matrix and classification report provided an assessment of the model's performance. You can see metrics like accuracy, precision, recall, and F1-score for each class, allowing you to understand how well the model predicts both the presence and absence of heart disease.
* To provide a complete conclusion, you should summarize the key insights gained from the analysis, mention any limitations or areas for improvement, and consider potential next steps, such as trying different algorithms or feature engineering techniques.