IMPLEMENTATION REPORT

Heart Disease Prediction using MACHINE LEARNING

1.Data Loading and Exploration:

* A pandas DataFrame called dataframe is used to hold the data that has been loaded from the CSV file "heart - UCI.csv".
* Dataframe.head(10) is used to display the DataFrame's top 10 rows.
* Dataframe.info () is used to display DataFrame information.
* The DataFrame's missing values are examined with dataframe.isna(). sum().
* Plotting and picture saving are done for the correlation matrix heatmap and feature histogram.

2.Preprocessing of the Data:

* The target variable and features are taken out of the DataFrame.
* Variable X is allocated features, and variable y is assigned the target variable.
* Train\_test\_split () from sklearn.model\_selection divides the data into training and testing sets.

3. Model Evaluation Function:

* To assess classification models, a function called evaluate\_model() is defined.
* Its inputs are test data (X\_test, y\_test) and the trained model.
* Accuracy, sensitivity, specificity, precision, and a confusion matrix are all calculated and shown.

4.Decision Tree Model:

* A Decision Tree classifier is created with the criteria "entropy" and max\_depth=5.
* X\_train and y\_train, the training data, are used to train the model.
* To assess the Decision Tree model using the test data, the evaluate\_model() function is utilized.

5.Logistic Regression Model:

* Using GridSearchCV, grid search is used to determine the optimal hyperparameters for logistic regression.
* Model and parameter grid definitions are made.
* Five-fold cross-validation is used in grid search.
* Grid search lr.best\_params\_ is used to obtain the best parameters.
* Using the evaluate\_model() function, the best-parameter Logistic Regression model is trained and assessed.

6. Random Forest Model:

* GridSearchCV is used to do a grid search to determine the optimal hyperparameters for Random Forest.
* Model and parameter grid definitions are made.
* Five-fold cross-validation is used in grid search.
* Grid search rf.best\_params\_ is used to obtain the best parameters.
* Using the evaluate\_model() function, the Random Forest model with the optimal parameters is trained and assessed.

7.Support Vector Machine (SVM) Model:

* First, a classifier for the Support Vector Machine is instantiated with C=12 and kernel='linear'.
* X\_train and y\_train, the training data, are used to train the model.
* For the SVM model, the accuracy, classification report, and confusion matrix are shown.

CONCLUSION:

For the data mentioned above, the Support Vector Machine classifier has an

accuracy of roughly 82%. Based on our findings, four machine learning

algorithms—Random Forest, Decision Trees, Support Vector Machines, and

Logistic Regression—were used. The end findings showed that Decision Trees

scored 92%, Random Forest scored 80%, Support Vector Machines scored 82%,

and Logistic Regression scored 80%. With an accuracy of almost 92%, we may

infer that the Decision Tree method is the optimal choice for our model.

GITHUB LINK:

https://github.com/Sowsonu/AI-PROJECT/upload