

Name: Pathan Firdos Mehraj

Roll no: 281073

Batch: A3

Assignment 4

Statement:

Q. Apply an appropriate ML algorithm on a dataset consisting of heart health details to predict the risk of heart disease. Create a confusion matrix based on the above data and find:

- a) Accuracy
- b) Precision
- c) Recall
- d) F1-score

Objective:

1. This assignment aims to apply a machine learning model for predicting the risk of heart disease.
2. Learn how to evaluate classification models using metrics such as accuracy, precision, recall, and F1-score.
3. Gain hands-on experience in data preprocessing, model training, and performance evaluation.

Resources used:

1. Software used: Google Colab
2. Libraries used: Pandas, Scikit-learn, Matplotlib, Seaborn

Introduction to Machine Learning for Heart Disease Prediction:

1. Machine learning is widely used in healthcare to analyze patient data and predict health risks like heart disease.
2. A classification model can help identify individuals at high risk of heart disease, enabling timely medical intervention.
3. The dataset contains various heart health attributes, such as age, sex, blood pressure, cholesterol, heart rate, and other factors that contribute to heart disease risk.

Methodology:

1. **Data Collection and Preprocessing:**
 - Load the dataset containing maternal health details.
 - Handle missing values and encode categorical variables.
 - Split the data into training and testing sets.
2. **Model Selection and Training:**
 - Choose a suitable classification algorithm (e.g., Logistic Regression, Decision Tree, Random Forest, or SVM).
 - Train the model on the training dataset.
3. **Model Evaluation:**
 - Generate predictions on the test dataset.
 - Compute the confusion matrix to analyze model performance.
 - Calculate accuracy, precision, recall, and F1-score.

Confusion Matrix and Performance Metrics:

1. **Accuracy:** Measures the overall correctness of the model.
 - Formula: $(TP + TN) / (TP + TN + FP + FN)$
2. **Precision:** Indicates how many of the predicted positive cases were actually positive.
 - Formula: $TP / (TP + FP)$
3. **Recall (Sensitivity):** Measures how well the model identifies actual positive cases.
 - Formula: $TP / (TP + FN)$
4. **F1-score:** Harmonic mean of precision and recall, balancing both metrics.
 - Formula: $2 * (Precision * Recall) / (Precision + Recall)$

Advantages ML in Heart Disease Prediction:

1. Helps in early detection of heart disease, which can save lives.
2. Enables data-driven decision-making for healthcare professionals.
3. Improves patient care and health outcomes by providing timely interventions.

Disadvantages:

1. Requires a large amount of quality data for better accuracy.
2. Model performance can be affected by imbalanced datasets.

Results:

```
Confusion Matrix:
[[ 91 104]
 [ 98 100]]
Accuracy: 0.4860050890585242
Precision: 0.49019607843137253
Recall: 0.5050505050505051
F1 Score: 0.4975124378109453
```

Conclusion:

In this assignment, we applied a machine learning classification model to predict the risk of heart disease. We used performance metrics such as accuracy, precision, recall, and F1-score to evaluate the model. By implementing this approach, healthcare professionals can enhance heart disease risk monitoring and provide timely interventions. These techniques are essential for improving healthcare outcomes and reducing the risks associated with heart disease.