Naive Bayes Report

Objective

To discover whether a patient has heart disease based on several medical attributes.

1. Dataset Overview

The dataset contains data about health parameters related to Heart Disease.

The dataset consisted with the following key columns:

- Pregnancies: Refers to the number of times the patient has been pregnant.
- Glucose: Represents the plasma glucose concentration.
- Blood Pressure: Measures the diastolic blood pressure.
- Skin Thickness: Indicates the thickness of the triceps skinfold.
- Insulin: Refers to the serum insulin level after two hours
- BMI: body mass index.
- Diabetes Pedigree Function: A function that assesses the likelihood of diabetes based on family history.
- Age: Represents the age of the patient in years.
- Outcome: A variable indicating the condition (0 or 1), where 1 indicates the presence of diabetes and 0 indicates its absence.

2.Methodology

1-Data Preprocessing:

- drop rows with missing values.
- The dataset is divided into training and testing sets, with 80% allocated for training the model and 20% reserved for testing its performance.

2- Model Building and Training:

- create an instance of the Gaussian Naive Bayes classifier and train it using the training data This trained model can later be used for making predictions on new data.

3- Classification Metrics:

to evaluates the performance of the trained model using accuracy, F1 score, and a classification report. These metrics help understand how well the model performs on the test data

3.**Result**

Accuracy: 0.77

- This means that 77% of the predictions made by the model were correct.

Classification Metrics:

- The classification report provides detailed metrics for each class (0 and 1):

Class 0 (No Heart Disease) , Class 1 (Heart Disease):

Precision :

- Class 0 : it is correct (66%) of the time.
- Class 1 : it is correct(83%)of the time.

Recall:

- Class 0 : it is correctly identifies(80%) of the actual class.
- Class 1: it is correctly identifies(71%) of the actual class .

F1 Score:

- Class 0 :reflects a good balance of precision and recall (81%).
- Class 1: model is less effective for this class compared to class 0, it still performs reasonably well(68%).

Support:

- Class 0: (99%)
- Class 1 : (55%)

4.Conclusion:

By using Naive Bayes in the analysis of heart patients, accurate and rapid classifications can be achieved, helping to improve healthcare and inform treatment decisions. This algorithm is a valuable tool in the medical data analysis process.

Name:

Jori Bajahnoun (444006482)

Taghreed Alzahrani (443007366)