project of Data Analysis 2

Name	ID
Mawadh Saleem Aloufi	444005971
Reham Hameed Alqurashi	444002142

Naive Bayes Report

Objective:

The Naive Bayes classifier was applied to predict whether a patient has diabetes using various medical parameters. The dataset was split into 80% training data and 20% testing data.

1. Dataset Overview:

The dataset contains data about various health parameters related to diabetes prediction. The goal is to predict whether a patient has diabetes based on several medical attributes. The dataset includes 9 columns, which are as follows:

- **Pregnancies**: Number of times the patient has been pregnant.

- **Glucose**: Plasma glucose concentration (measured after a 2-hour oral glucose tolerance test).

- **BloodPressure**: Diastolic blood pressure (mm Hg).

- **SkinThickness**: Triceps skinfold thickness (mm).

- **Insulin**: 2-hour serum insulin (mu U/ml).

- **BMI**: Body mass index (weight in kg/(height in m)^2).

- **DiabetesPedigreeFunction**: A function that scores the likelihood of diabetes based on family history.

- **Age**: Age of the patient (years).

- **Outcome**: Class variable (0 or 1), where 1 indicates the presence of diabetes and 0 indicates no diabetes.

Results:

- Accuracy: The model achieved an accuracy of 76.62%, meaning it correctly predicted whether a patient had diabetes in 76.62% of the cases.

- **Classification Metrics:**
1. **Precision:** Measures the proportion of positive identifications that were actually correct.

For class 0 (No Diabetes): 83%
For class 1 (Diabetes): 66%

2. **Recall:** Measures the proportion of actual positives that were correctly identified.

For class 0 (No Diabetes): 80%
For class 1 (Diabetes): 71%

3. **F1-score:** Harmonic mean of precision and recall, providing a balanced measure.

For class 0 (No Diabetes): 81%
For class 1 (Diabetes): 68%

4. **Support:** Number of actual instances for each class in the dataset.

Class 0: 99 instances
Class 1: 55 instances

Conclusion:
The model performs well for predicting class 0 (No Diabetes), with higher precision and recall. However, class 1 (Diabetes) predictions show lower precision and recall, indicating room for improvement in identifying diabetic patients.