A classification task to predict breast cancer diagnosis using logistic regression.

**Documentation**

**Approach:**

1. **Data Loading and Exploration:**
   * Loaded the dataset using pandas and displayed the first few rows to understand its structure.
   * Used **df.info()** and **df.describe()** to get an overview of the data types, non-null counts, and summary statistics.
2. **Data Cleaning:**
   * Checked for and handled missing values by dropping rows with any missing data.
   * Dropped the 'id' column as it is not relevant for the prediction task.
   * Converted the 'diagnosis' column to numerical values ('M' -> 1, 'B' -> 0) for easier processing.
3. **Feature Selection:**
   * Computed the correlation matrix to identify features highly correlated with the target variable ('diagnosis').
   * Selected features with a correlation score greater than a specified threshold (0.5).
4. **Data Preprocessing:**
   * Split the data into training and testing sets with an 80-20 split.
   * Standardized the features using **StandardScaler** to improve model performance.
5. **Model Training and Evaluation:**
   * Trained a logistic regression model on the training data.
   * Evaluated the model using accuracy score, confusion matrix, and classification report.

**Insights:**

* The logistic regression model achieved an accuracy score of X.XX (replace with actual value).
* The confusion matrix and classification report provide detailed performance metrics, indicating how well the model distinguishes between malignant and benign cases.

Feel free to adjust the explanations or add more details based on your insights and findings. If you have any specific questions or need further assistance, let me know!

Top of Form