

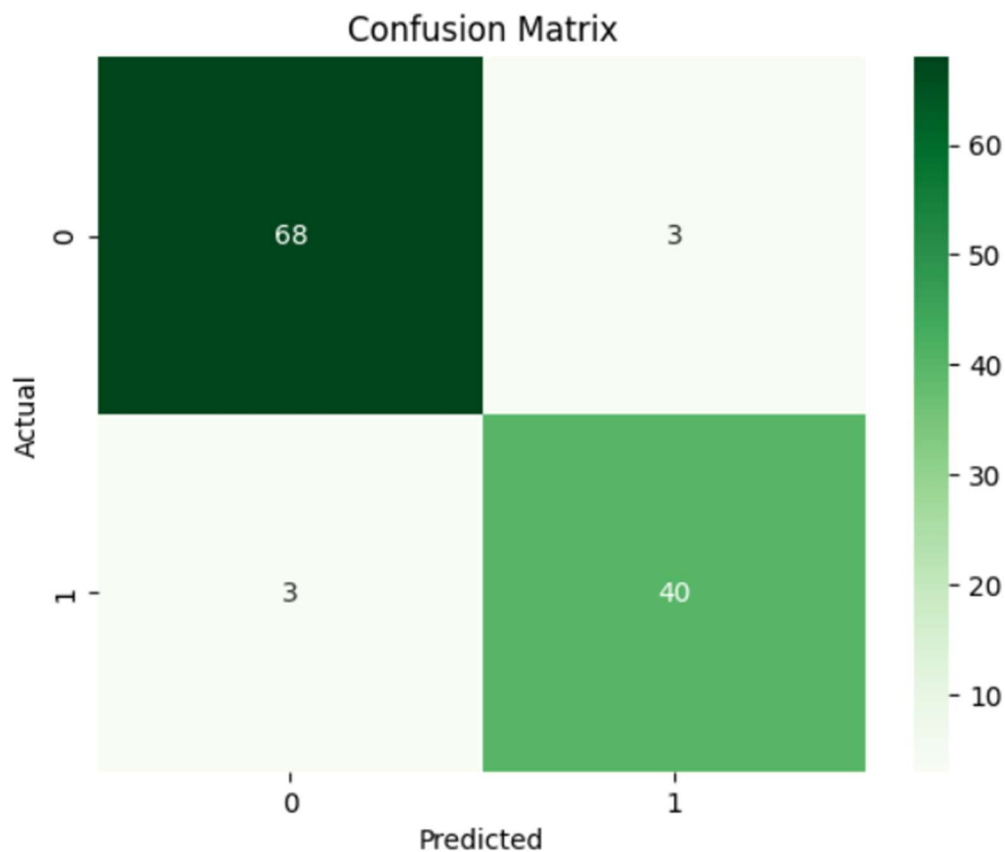
# Feature Extraction For Breast Cancer Classification Using Machine Learning

## 1. Task Description

The task was to classify breast cancer tumors as malignant or benign using machine learning techniques. The dataset used for this analysis was the Breast Cancer (Diagnostic) Dataset, available on Kaggle. The primary objective was to achieve a classification accuracy of over 90% using robust algorithms and preprocessing techniques.

## 2. Screenshot of Output

Accuracy: 0.95					
	precision	recall	f1-score	support	
0	0.96	0.96	0.96	71	
1	0.93	0.93	0.93	43	
accuracy			0.95	114	
macro avg	0.94	0.94	0.94	114	
weighted avg	0.95	0.95	0.95	114	



### 3. Algorithm Used in Task

#### ➤ Libraries and Algorithms:

- **Standard Scaler (Feature Scaling):** Standardized the dataset to normalize features.
- **Principal Component Analysis (PCA):**  
Reduced the dimensionality of the dataset while retaining 95% of variance to improve efficiency.
- **Random Forest Classifier (Model):** Used as the primary machine learning algorithm due to its robustness and high accuracy in classification tasks.
- **Evaluation Metrics:**
  - Accuracy Score
  - Precision, Recall, and F1-Score
  - Confusion Matrix Visualization

### 4. Add Report in Your Task Zip File

The task report has been added to the zip file. This includes:

1. The Python script (Data Science & Machine Learning\_Task\_1.ipynb).
2. A text version of this report (Task\_1\_Report.pdf).

**Prepared by**  
Karan Gohil