Logistic Regression Model Report on Breast Cancer Dataset

Data Overview

- Total Samples: **569**
- Target Column: **diagnosis**
    - M (Malignant) → 1
    - B (Benign) → 0
- Features: **30 numerical features** related to tumor characteristics (e.g., radius, texture, perimeter).
- Dropped columns:
    - id: Identifier, not useful for prediction.
    - Unnamed: 32: Empty column (all NaN).

Data Preprocessing

- Encoded the diagnosis column using:

```
df['diagnosis'] = df['diagnosis'].map({'B': 0, 'M': 1})
```

Split dataset using:

```
X_train, X_test, y_train, y_test = train_test_split(..., test_size=0.2)
```

Model Training

- Trained a Logistic Regression model:

```
model = LogisticRegression(max_iter=10000)
model.fit(X_train, y_train)
```

Evaluation Results

Accuracy Score: Accuracy: ~96%

Confusion Matrix (Seaborn Heatmap):

This shows correct and incorrect predictions.

ROC Curve:

- **AUC Score**: Shows how well the model distinguishes classes.
- AUC value close to **1.0** indicates strong performance.

📈 Visualizations

- Confusion Matrix: Clear separation between correctly predicted malignant and benign cases.
- ROC Curve: High TPR and low FPR, indicating good classification capability.

✅ Conclusion

The logistic regression model performs very well with high accuracy and AUC. It effectively distinguishes between benign and malignant tumours. This model can be a useful baseline for medical diagnostic systems.