# Feature Engineering and Model Evaluation on Pima Indians Diabetes Dataset

## Introduction

The Pima Indians Diabetes dataset is widely used for demonstrating various machine learning techniques. This report covers feature selection, transformation, scaling, and model evaluation metrics, followed by practical exercises involving feature engineering and model evaluation.

## Feature Transformation and Scaling

1. **Standardization**: Features were standardized to have a mean of 0 and a standard deviation of

2. **Normalization**: Features were scaled to a range of [0, 1].

## Model Evaluation Metrics

1. **Accuracy**: The ratio of correctly predicted instances to the total instances.

2. **Precision**: The ratio of true positive predictions to the total predicted positives.

3. **Recall**: The ratio of true positive predictions to all actual positives.

4. **F1 Score**: The harmonic mean of precision and recall.

5. **ROC-AUC**: The area under the Receiver Operating Characteristic curve.

## Practical Exercises: Feature Engineering and Model Evaluation

### Model Training and Evaluation

**Logistic Regression**: A Logistic Regression model was trained and evaluated. The following metrics were obtained:

- Accuracy: 0.727

- Precision: 0.614

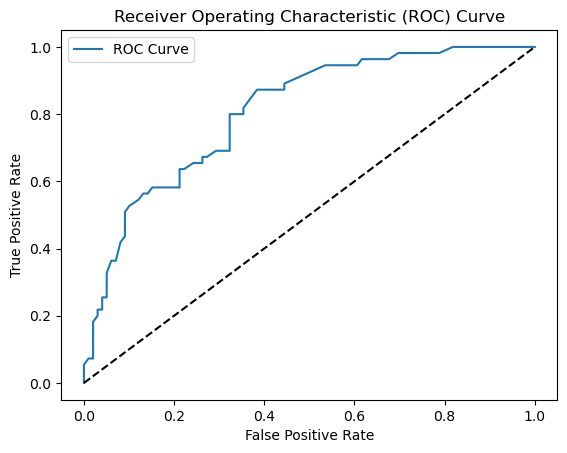
- Recall: 0.636

- F1 Score: 0.625

- ROC-AUC: 0.808

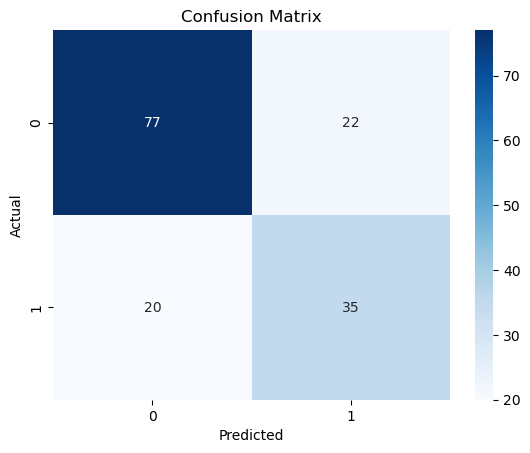
### Visualization

**ROC Curve**

****

The Receiver Operating Characteristic (ROC) curve was plotted to visualize the performance of the model.

**Confusion Matrix**



The Confusion Matrix was plotted to visualize the performance of the model in terms of true positives, false positives, true negatives, and false negatives.