

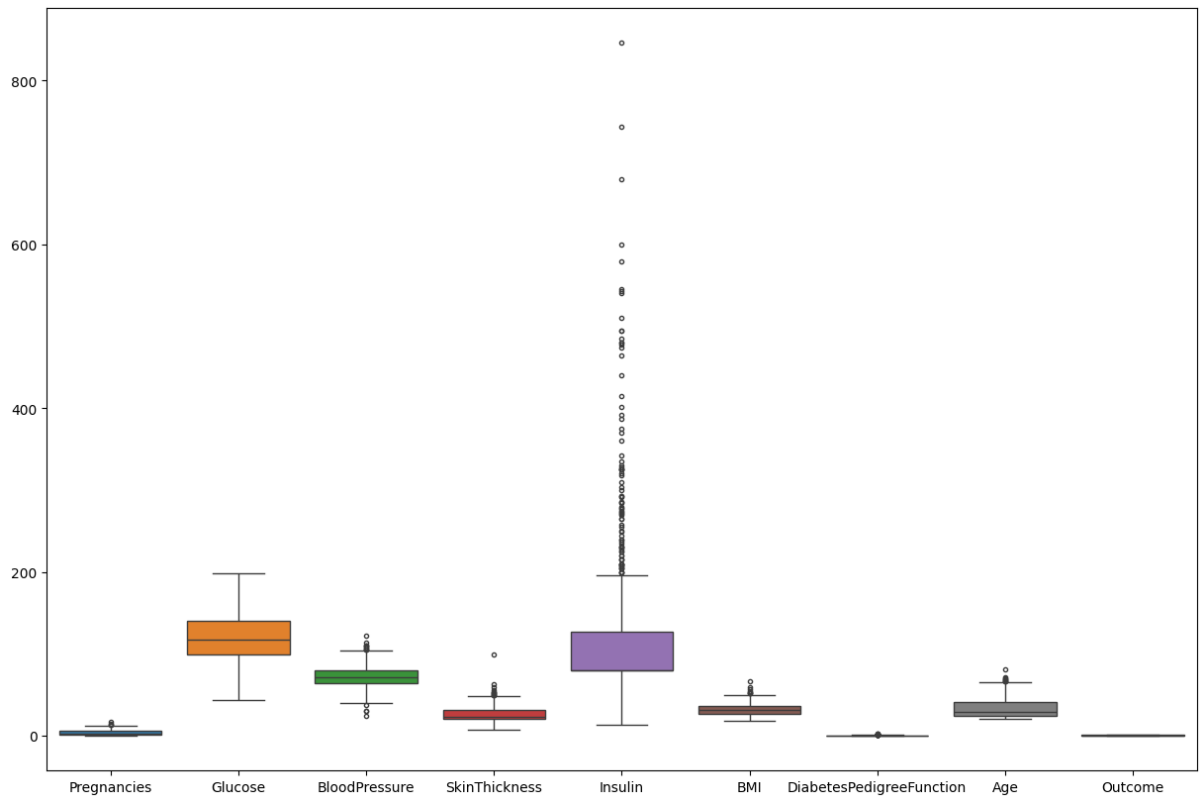
Summary Report for Logistic Regression for Diabetes Prediction

Prepared By: Mayankkumar Patel (C0883127)

I have built a logistic regression model to predict diabetes using the provided Database. Here's a summary of my key findings and model performance metrics:

1. Data Preparation:

- Replaced 0 values in 'Glucose', 'BloodPressure', 'BMI', 'SkinThickness', and 'Insulin' with their respective means.
- Did outliers Checking also using boxplot.



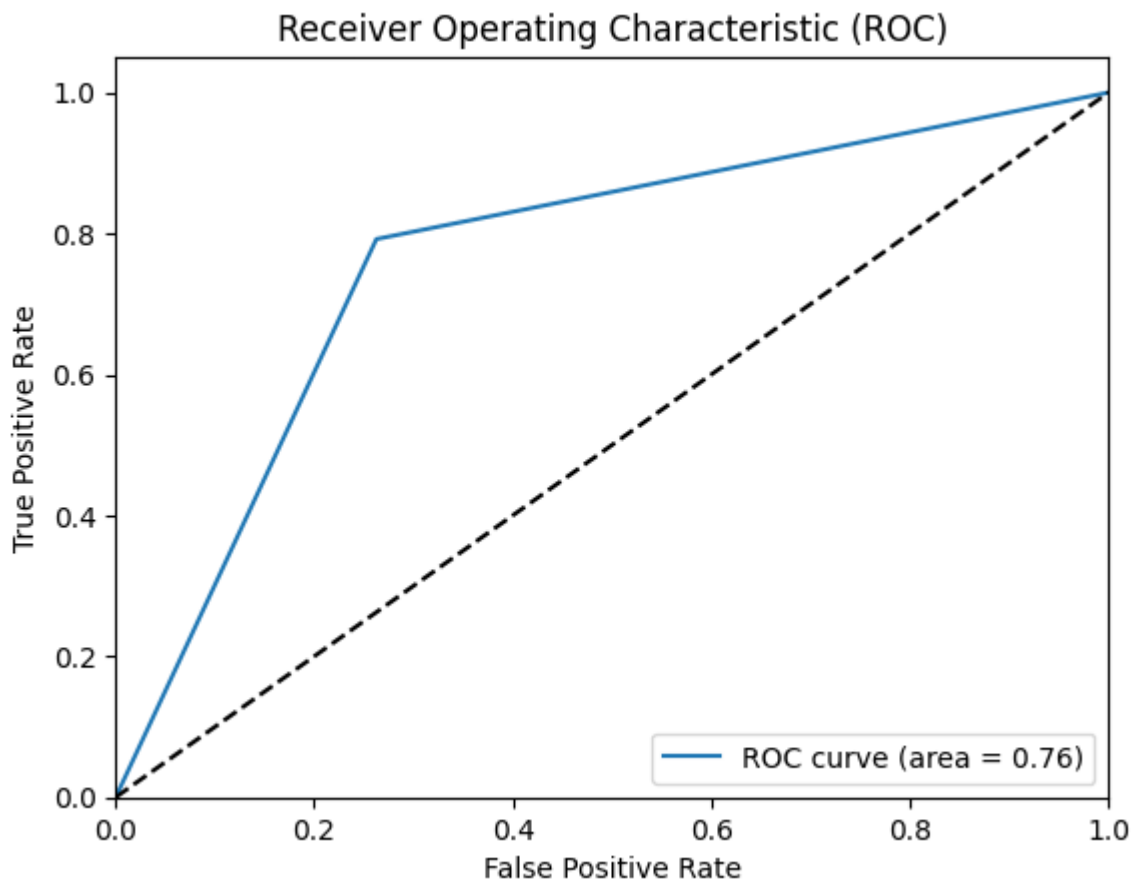
- Balanced the target classes using SMOTE oversampling.

2. Model Training:

- Used logistic regression with hyperparameter tuning using GridSearchCV.
- Selected hyperparameters: $C=1$, $\text{penalty}='l2'$.

3. Model Evaluation:

- Achieved an accuracy of 76.5% on the test set.
- Precision, Recall, and F1-score for class 1 were 0.75, 0.79, and 0.77, respectively.
- ROC AUC score was 0.765, indicating good model performance.



4. Interpretation of Coefficients:

- a. Pregnancies, Glucose, BMI, and DiabetesPedigreeFunction have a positive impact on the likelihood of diabetes.
- b. BloodPressure and Insulin have a negative impact.
- c. SkinThickness did not have a significant impact.

Overall, the logistic regression model performed reasonably well in predicting diabetes based on the provided features.