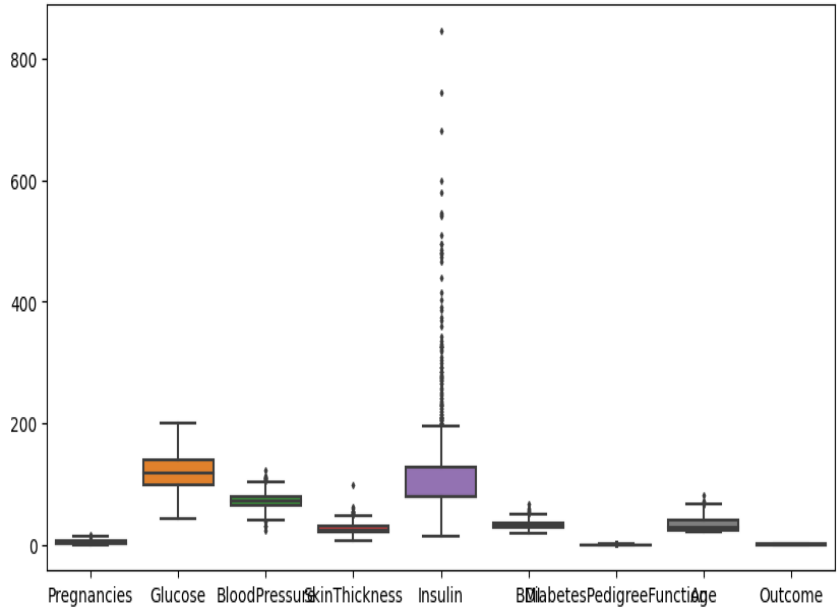
Summary Report of Logistic Regression for Diabetes Prediction

Prepared by: Komal Astawala (C0883392)

1. Data Preparation:

- Added their respective means to the 0 values for "Glucose," "BloodPressure," "BMI," “SkinThickness," and "Insulin."
- Used Boxplot to check outliers as well.



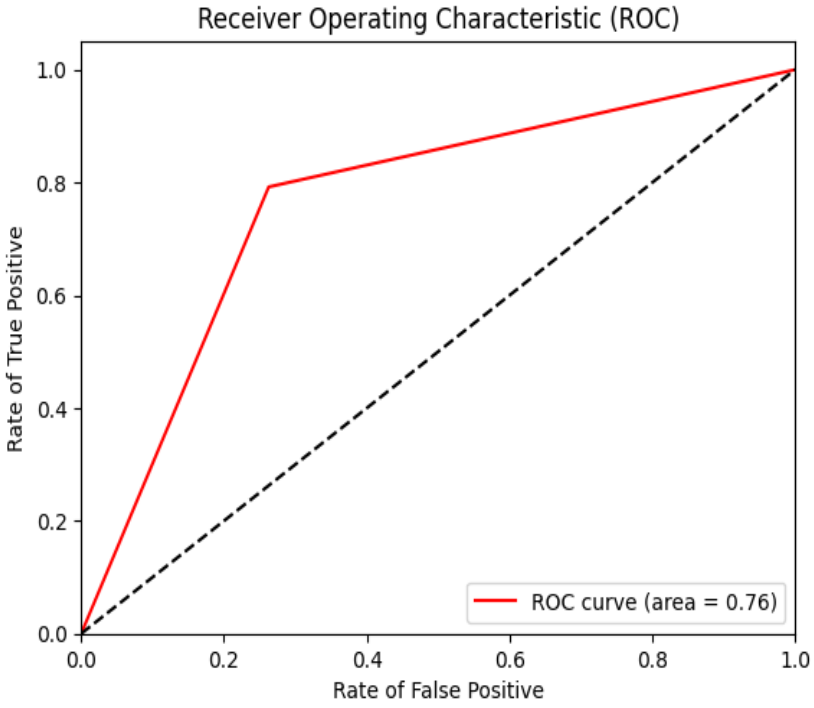
- Employed SMOTE oversampling to achieve target class balance.

2. Model Training:

- Made use of GridSearchCV for hyperparameter adjustment in logistic regression.
- Specified hyperparameters: penalty='l2', C = 1.

3. Model Evaluation:

- Accomplished 76.5% accuracy on the test set
- For class 1, the corresponding precision, recall, and F1-score were 0.75, 0.79, and 0.77.
- The model performed well, as indicated by the ROC AUC value of 0.765.



4. Interpretation of Coefficients:

- Diabetes, BMI, glucose, and pregnanciesDiabetes risk is positively impacted by pedigree function.
- Insulin and blood pressure have an adverse effect.
- Skin Thickness had no appreciable effect.

Based on the available variables, the logistic regression model predicted diabetes quite well overall.