Size of Data: (768, 9)

Statistical Summary: Pregnancies Glucose ... Age Outcome

count 768.000000 768.000000 ... 768.000000 768.000000

mean 3.845052 120.894531 ... 33.240885 0.348958

std 3.369578 31.972618 ... 11.760232 0.476951

min 0.000000 0.000000 ... 21.000000 0.000000

25% 1.000000 99.000000 ... 24.000000 0.000000

50% 3.000000 117.000000 ... 29.000000 0.000000

75% 6.000000 140.250000 ... 41.000000 1.000000

max 17.000000 199.000000 ... 81.000000 1.000000

[8 rows x 9 columns]

Count of class Labels: Outcome

0 500

1 268

dtype: int64

EDA : Count of zeros by column:

Pregnancies : 111

Glucose : 5

BloodPressure : 35

SkinThickness : 227

Insulin : 374

BMI : 11

DPedFunc : 0

Age : 0

Outcome : 500

Running Classifiers for All Features

kNN Classifier: best k: 19

Logistic Regression: best C: 1

Random Forest: Best N: 2 Best d: 6

kNN Classifier('All Features')

--------------------------------------------------

TN FP FN TP Accuracy Precision Recall f1\_score TNR

114 11 28 39 0.7969 0.78 0.5821 0.6667 0.912

Logistic Regression('All Features')

--------------------------------------------------

TN FP FN TP Accuracy Precision Recall f1\_score TNR

103 22 20 47 0.7812 0.6812 0.7015 0.6912 0.824

Feature Importances RF Classifier:

Glucose 0.416025

Age 0.206830

Insulin 0.098456

BMI 0.086641

DiabetesPedigreeFunction 0.062131

BloodPressure 0.060783

SkinThickness 0.034881

Pregnancies 0.034251

dtype: float64

Random Forest Classifier('All Features')

--------------------------------------------------

TN FP FN TP Accuracy Precision Recall f1\_score TNR

105 20 19 48 0.7969 0.7059 0.7164 0.7111 0.84

grid best params {'C': 100, 'gamma': 0.001}

grid best estimator SVC(C=100, class\_weight='balanced', gamma=0.001)

SVM with GridSearchCV('All Features')

--------------------------------------------------

TN FP FN TP Accuracy Precision Recall f1\_score TNR

103 22 21 46 0.776 0.6765 0.6866 0.6815 0.824

Running Classifier for Selected Features: " ['Glucose', 'BMI', 'Age', 'DiabetesPedigreeFunction']

kNN Classifier: best k: 17

Logistic Regression: best C: 0.001

Random Forest: Best N: 9 Best d: 3

kNN Classifier(Selected Features)

--------------------------------------------------

TN FP FN TP Accuracy Precision Recall f1\_score TNR

113 12 23 44 0.8177 0.7857 0.6567 0.7154 0.904

Logistic Regression(Selected Features)

--------------------------------------------------

TN FP FN TP Accuracy Precision Recall f1\_score TNR

103 22 19 48 0.7865 0.6857 0.7164 0.7007 0.824

Feature Importances RF Classifier:

Glucose 0.568863

BMI 0.226516

Age 0.157073

DiabetesPedigreeFunction 0.047548

dtype: float64

Random Forest Classifier(Selected Features)

--------------------------------------------------

TN FP FN TP Accuracy Precision Recall f1\_score TNR

98 27 14 53 0.7865 0.6625 0.791 0.7211 0.784

grid best params {'C': 10, 'gamma': 0.01}

grid best estimator SVC(C=10, class\_weight='balanced', gamma=0.01)

Best SVM with GridSearchCV(Selected Features)

--------------------------------------------------

TN FP FN TP Accuracy Precision Recall f1\_score TNR

101 24 19 48 0.776 0.6667 0.7164 0.6906 0.808