Machine learning Supervised Learning Project Diabetes Prediction

Zeynab Akolade

Conclusion

- For this project the target was to predict: “diabetes or not diabetes” using machine learning models.

- The following were the variables available in the dataset: Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age and Outcome.

- “Outcome” was defined as the target to be predict.

- After analysis it was identified that Glucose had a higher correlation with the target.

- “Glucose” was the variable to be use as predictor.

Results

- For this project, 4 different machine learning models were used to predict diabetes. The models used were Random Forest, Decision Tree, SVM and Logistic Regression.

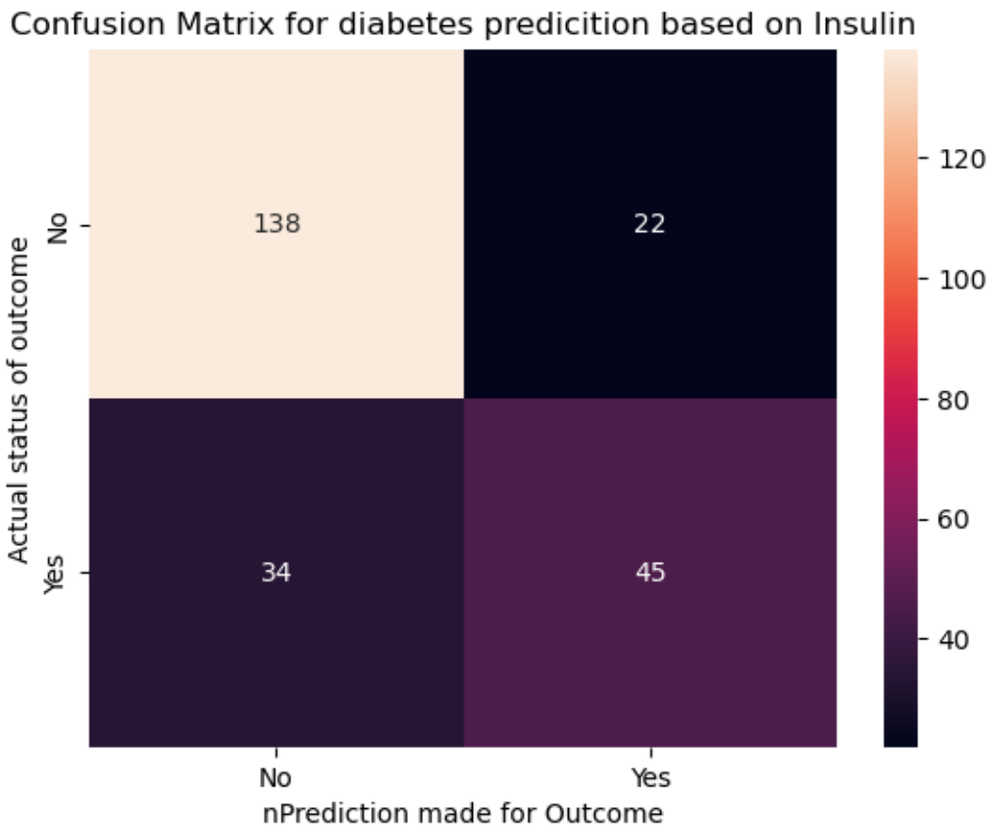
- After trying different parameters with the models, it was observed a similar accuracy among them. Accuracy calculates the number of correct predictions made by the model.

Accuracy results as follow:

- 0.72 = Random Forest
- 0.73 = Decision Tree
- 0.77 = Support Vector Model (SVM)
- 0.77 = Logistic Regression Model (LRM)

Regarding the dataset's key attributes, significant associations with the detection outcome were identified in Glucose levels, BMI, and Age, indicating their importance. However, further exploration is recommended for the Insulin feature for further investigation as it displayed a notable presence of outliers, which may require specialized handling with additional domain knowledge. For future analysis more features can be added to the models to try to improve accuracy.


Pregnancies 1 0.13 0.21 0.079 0.023 0.012 -0.026 0.56 0.22
Glucose 0.13 1 0.22 0.2 0.43 0.22 0.14 0.26 0.49
BloodPressure 0.21 0.22 1 0.19 0.046 0.29 -7.5e-05 0.32 0.17
SkinThickness 0.079 0.2 0.19 1 0.15 0.56 0.11 0.12 0.22
Insulin 0.023 0.43 0.046 0.15 1 0.18 0.13 0.097 0.21
BMI 0.012 0.22 0.29 0.56 0.18 1 0.15 0.021 0.3
DiabetesPedigreeFunction -0.026 0.14 -7.5e-05 0.11 0.13 0.15 1 0.023 0.18
Age 0.56 0.26 0.32 0.12 0.097 0.021 0.023 1 0.25
Outcome 0.22 0.49 0.17 0.22 0.21 0.3 0.18 0.25 1
Pregnancies
Glucose
BloodPressure
SkinThickness
Insulin
BMI
DiabetesPedigreeFunction
Age
Outcome
1.0
0.8
0.6
0.4
0.2
0.0


Confusion Matrix for diabetes predicition based on Insulin
Actual status of outcome
No
Yes
138
22
34
45
No
Yes
nPrediction made for Outcome
120
100
80
60
40