



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

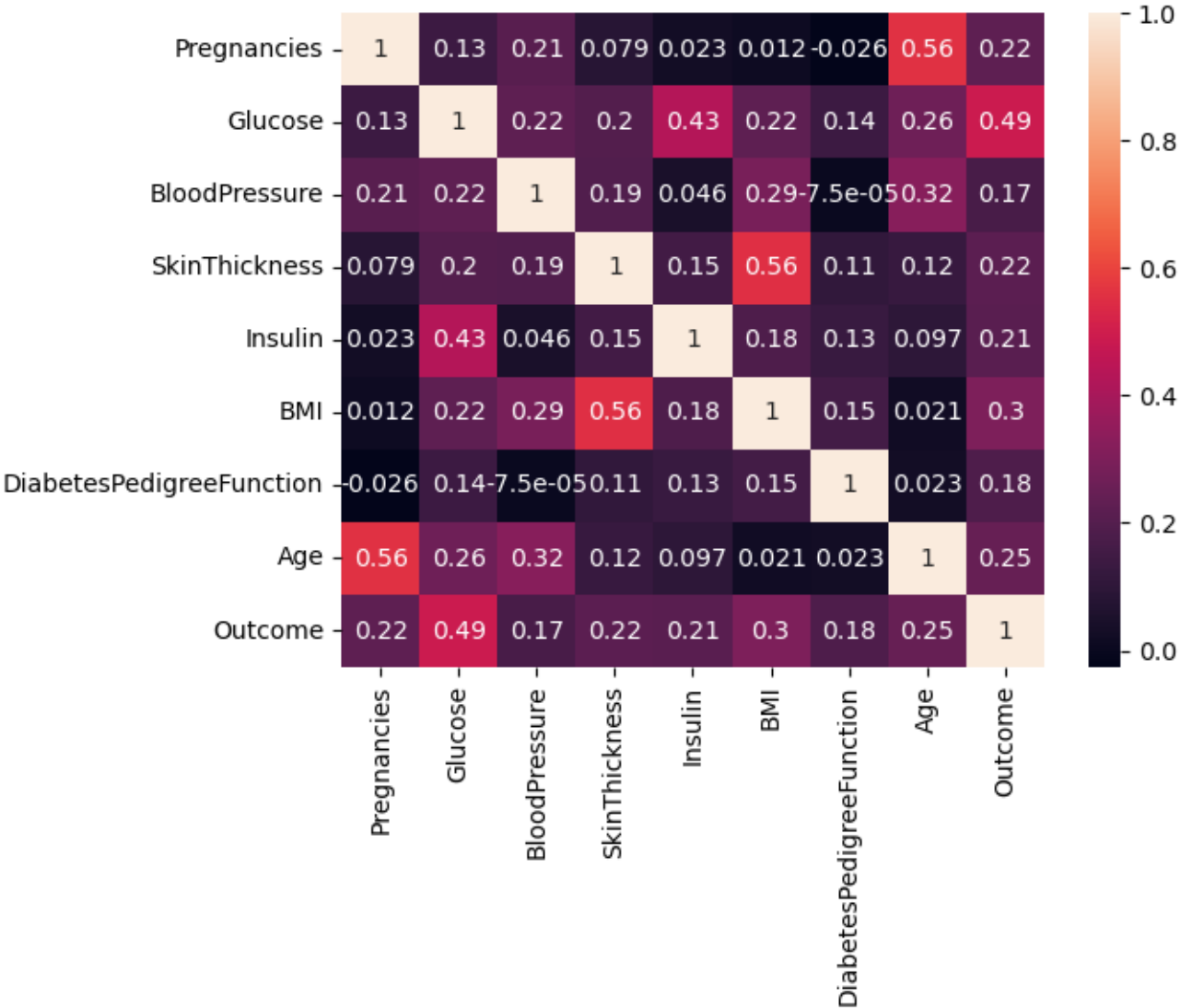
- For this project, 4 different machine learning models were used to predict diabetes. The models used were Random Forest, Decision Tree, SVM and Logistic Regression.
- After trying different parameters with the models, it was observed a similar accuracy among them. Accuracy calculates the number of 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)

Accuracy calculates the correct

For future analysis more features can be added to the models to try to improve accuracy



Confusion Matrix for diabetes prediction based on Insulin

