ML
Supervised Learning

Project: Diabetes Prediction

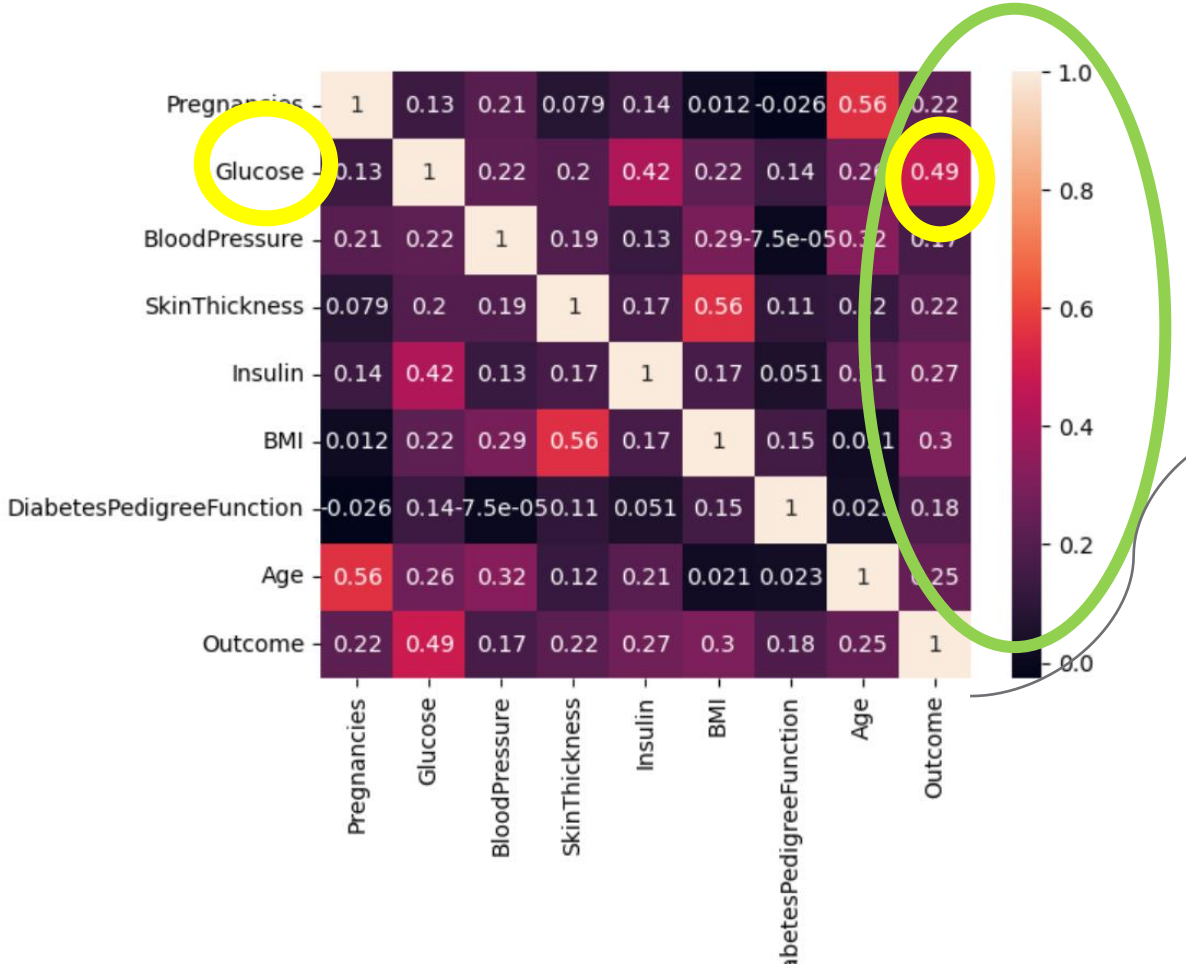
Conclusions

- For this project the target was to predict: “diabetes or not diabetes” using machine learning models
- The following were the variables available in the data set: 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.

Conclusions (Continuation)

- 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:
  - .77 SVM
  - .76 Logistic Regression
  - .73 Decision Tree
  - .72 Random Forest
  - Accuracy calculates the correct
- For future analysis more features can be added to the models to try to improved accuracy.

Correlation Matrix



The closer the number is to 1, The higher correlation

Outcome is the target