Machine Learning for Diabetes

Result Summary

My Model

	precision	recall	f1-score	support
0	0.84	0.81	0.82	99
1	0.68	0.73	0.70	55
accuracy			0.78	154
macro avg	0.76	0.77	0.76	154
weighted avg	0.78	0.78	0.78	154

Original Model

	precision	recall	f1-score	support
0	0.85	0.78	0.81	151
1	0.64	0.74	0.69	80
accuracy			0.77	231
macro avg	0.75	0.76	0.75	231
weighted avg	0.78	0.77	0.77	231

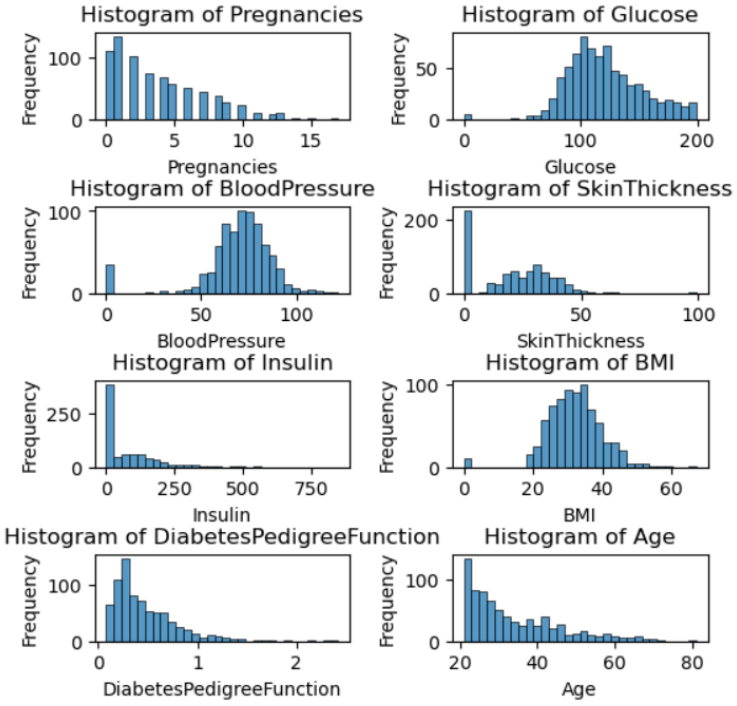
- Through specific hyperparameter tuning my Random Forest was able to improve on original model slightly.
- However, severe limitation was the small size of the data set. This was mentioned in the study. To get more precise results we would need a lot more data.

Preprocessing

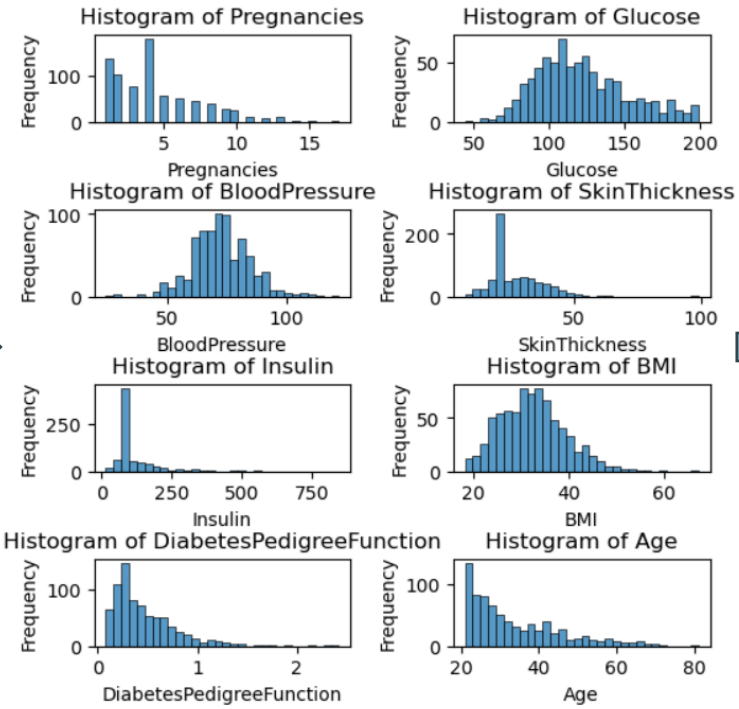
1. Replace 0 values with mean
2. Not altering outliers
3. Applying Power Transform to scale data
4. Examining Feature Correlation

Transforming Data Distribution Shape

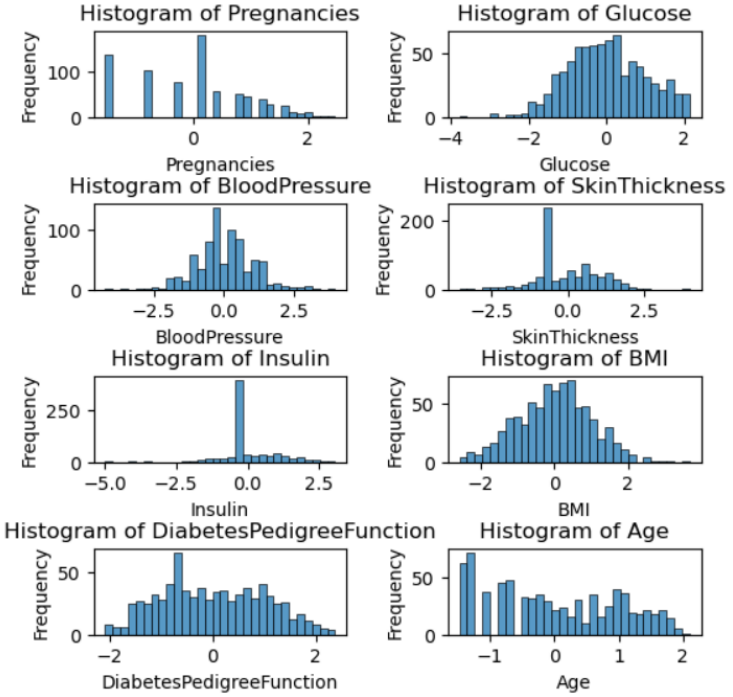
Original Data



Changing 0 to mean



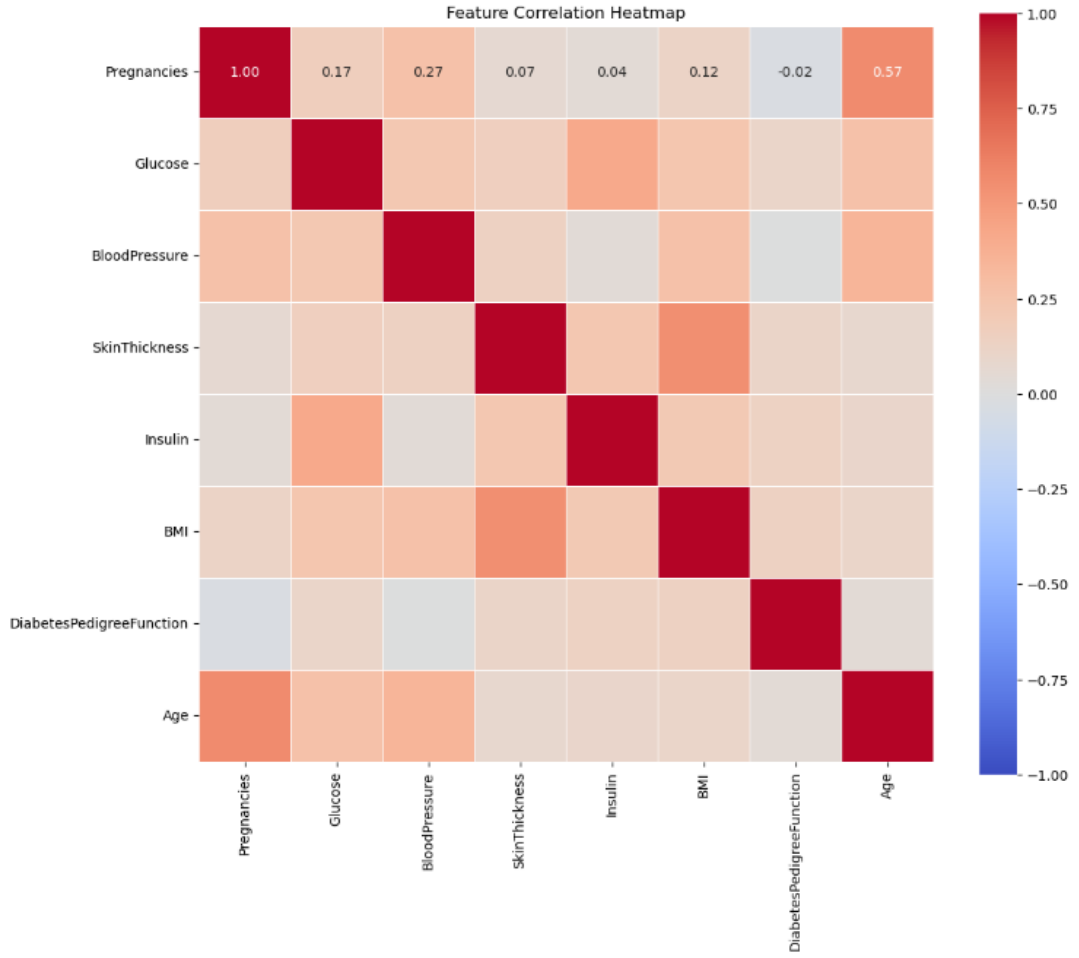
Power Transform



Transforming data we seen final data looks like normal distribution mostly and data is centralised, which is ideal for modelling.

Correlation Between Features

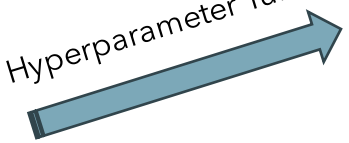
- None of the features were highly correlated so none had to be dropped.



Hyperparameter Tuning

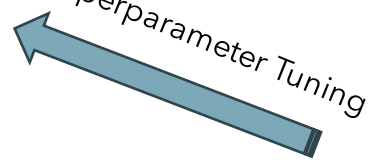
Grid Search Results

	precision	recall	f1-score	support
0	0.81	0.80	0.81	99
1	0.65	0.67	0.66	55
accuracy			0.75	154
macro avg	0.73	0.74	0.73	154
weighted avg	0.76	0.75	0.75	154



Best Model Results

	precision	recall	f1-score	support
0	0.84	0.81	0.82	99
1	0.68	0.73	0.70	55
accuracy			0.78	154
macro avg	0.76	0.77	0.76	154
weighted avg	0.78	0.78	0.78	154

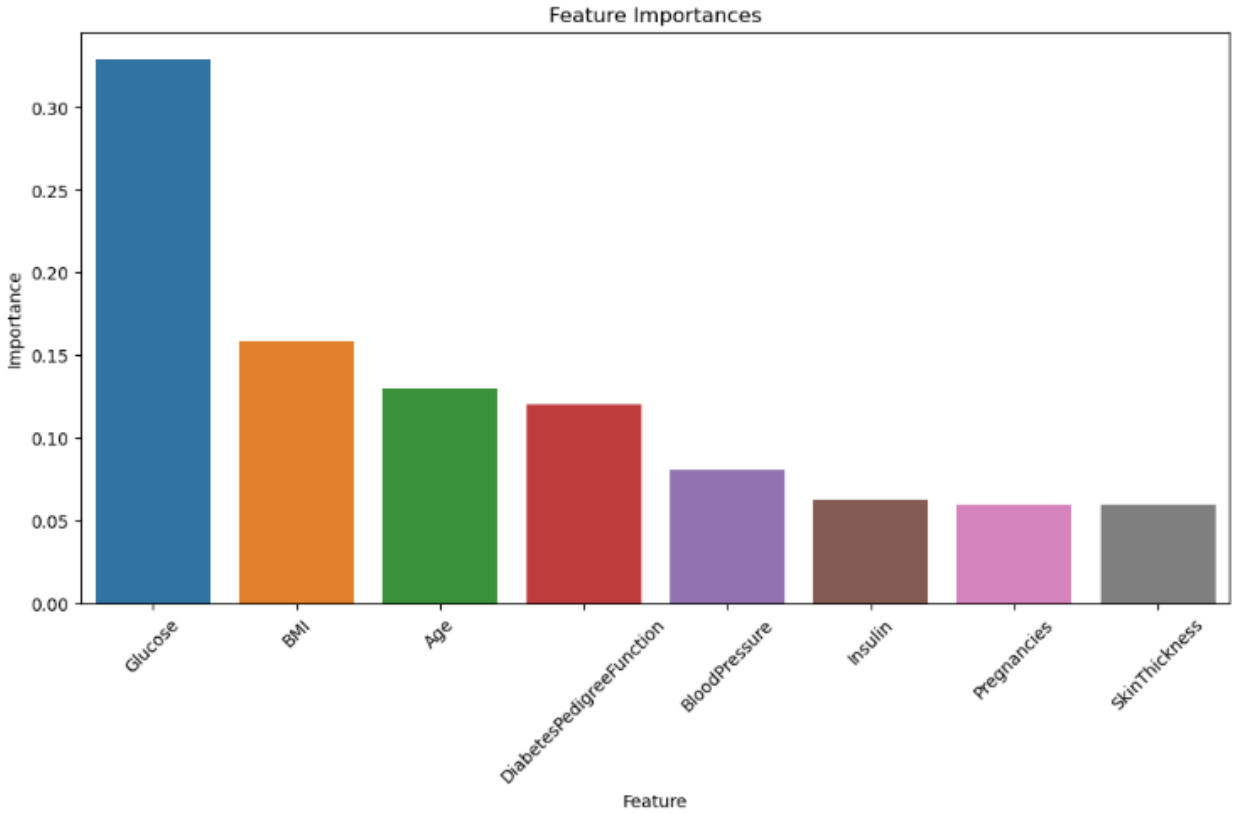


Original Results

	precision	recall	f1-score	support
0	0.85	0.78	0.81	151
1	0.64	0.74	0.69	80
accuracy			0.77	231
macro avg	0.75	0.76	0.75	231
weighted avg	0.78	0.77	0.77	231

- Combining Grid Search and extra hyperparameter tuning results in best model results.

Feature Importance



- Glucose levels is by far the most important feature in predicting diabetes, more than twice the importance of second most important feature, BMI.