CS 39543: HW 2

Theadora Williams

February 24, 2022

1 Baseline Results

1.1 Full Data

- J48: Correctly classfied 73.83%
- Naive Bayes: Correctly classfied 76.30%
- Bayes Net: Correctly classfied 74.35%
- Logistic Regression: Correctly classfied 77.21%

1.2 Training Data (all 0's deleted)

- J48: Correctly classfied 79.34%
- Naive Bayes: Correctly classfied 77.81%
- Bayes Net: Correctly classfied 76.53%
- Logistic Regression: Correctly classfied 77.81%

2 Predicting missing values

2.1 Predicting missing values using Mean

- J48: Correctly classfied 73.30%
- Naive Bayes: Correctly classfied 74.61%
- Bayes Net: Correctly classfied 74.87%
- Logistic Regression: Correctly classfied 77.34%

2.2 Predicting missing values using Median

- J48: Correctly classfied 73.57%
- Naive Bayes: Correctly classfied 74.87%
- Bayes Net: Correctly classfied 74.48%
- Logistic Regression: Correctly classfied 77.47%

2.3 Predicting missing values using Linear Regression

- J48: Correctly classfied 73.57%
- Naive Bayes: Correctly classfied 76.30%
- Bayes Net: Correctly classfied 74.87%
- Logistic Regression: Correctly classfied 77.34%

3 Randomness of missing skin and insuling values

- The presence of the skin and insulin attributes are correlated at 66%

I binarized data in skin and insu based off it's presence, here are the results for predicting their presence:

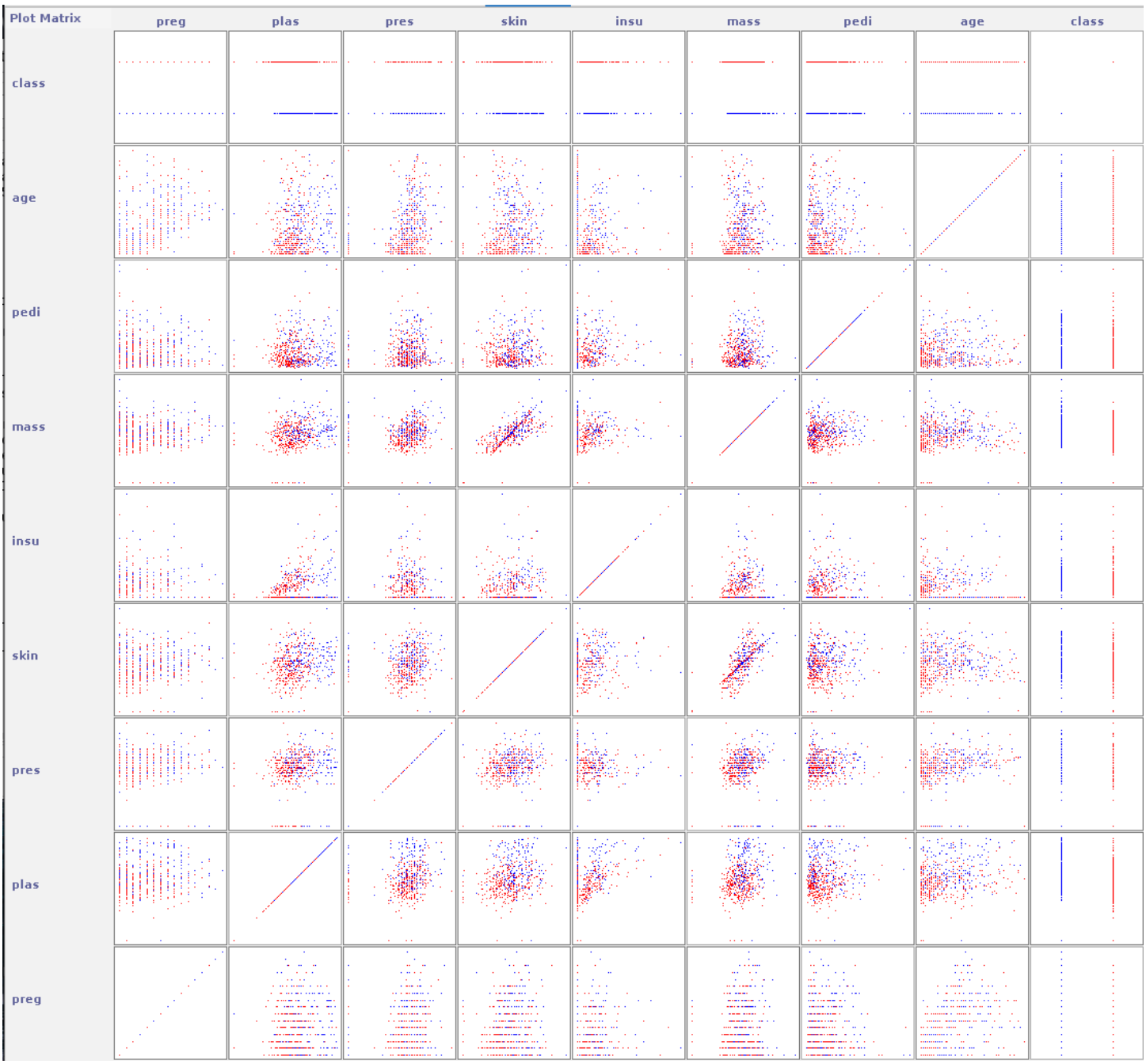
3.1 Predicting insu

- J48: Correctly classfied 79.30%
- Naive Bayes: Correctly classfied 78.78%
- Bayes Net: Correctly classfied 79.95%
- Logistic Regression: Correctly classfied 81.38%

3.2 Predicting skin

- J48: Correctly classfied 80.2083%
- Naive Bayes: Correctly classfied 82.16%
- Bayes Net: Correctly classfied 81.38%
- Logistic Regression: Correctly classfied 81.25%

4 Visualization with means



Full Code Available at `https://github.com/theadorabledev/data_mining_class`