Naïve Bayes Classification Experiment Report

Distribution Analysis

I used histograms and conducted Shapiro-Wilk test to assess normality and distribution shapes.

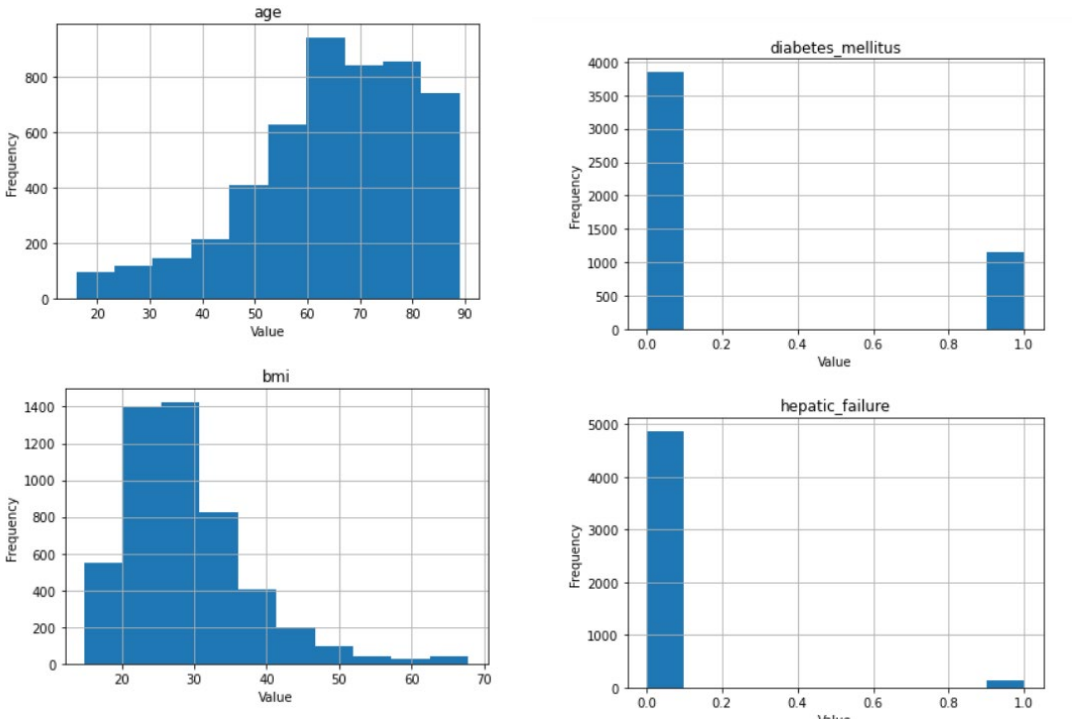

Fig 1. Distributions

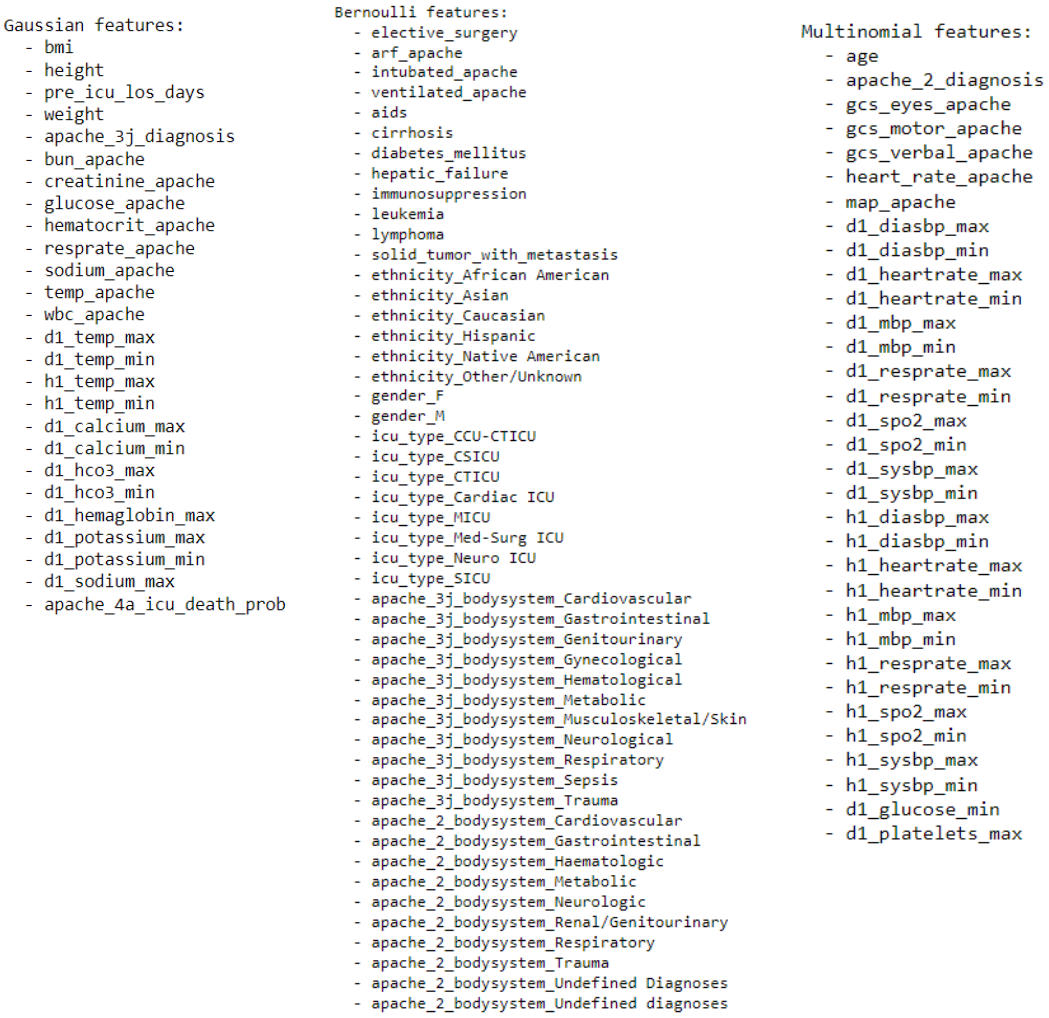

Fig 2. Analysis

Experimentation

Training Error: Calculated by predicting on the training data.

Cross-Validation Error: Computed using 5-fold cross-validation to assess the model's generalization capabilities.

Results

```
GaussianNB Cross-Validation Accuracy: 0.7101999999999999
BernoulliNB Cross-Validation Accuracy: 0.7070000000000001
MultinomialNB Cross-Validation Accuracy: 0.6652
Final Prediction Accuracy: 0.740200546946217
Test Error: 0.25979945305378305
```

Fig 3. Final results

Minimal-risk Bayesian decision criterion table

Parameter Choices

Loss Function

I considered that in medical field the consequences of a false negative can be far more severe than a false positive. As such, choosing a loss function that heavily penalizes false negatives. The ratio of 5:1 for false negatives to false positives was decided based on this rationale.

Loss function: a false negative is 5 times more costly than a false positive

L(1,0) (False positive) : Loss = 1

L(0,1) (False negative) : Loss = 5

Correct decision: Loss = 0

Predicted \ Actual	+ve (1)	-ve (0)
+ve (1)	0	1
-ve (0)	6	0

Result:
Minimal Risk Test Error: 0.25979945305378305

Discussion of Observations

Gaussian and Bernoulli classifiers performed rather similar and are better than the Multinomial classifier. This could be due to the nature of the ICU data where features are more predictive of the outcomes.

When applying the minimal-risk decision criterion, the test error remained more or less the same, which suggests that a cost-sensitive approach did not affect the model's performance very much. Probably it is because the increased cost of false negatives was balanced by the classifier's ability to predict true positives accurately.