1.1 Please see python file

1.2 Please see python file

2.1 Please see python file

2.2 Accuracy: 0.7610536218250236

2.3 TP:809, TN:47, FP:175, FN:32

3.1 Please see python file

3.2 0.7177798682972719

3.3 TP:763, TN:164, FP:58, FN:78

4.1 Please see python file

4.2 Please refer to the attachment

4.3 0.7695202257761053

4.4 TP:818, TN:95, FP:127, FN:23

5

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | TP | TN | FP | FN | Accuracy | TPR | TNR |
| Naïve Bayesian | 809 | 47 | 175 | 32 | 0.76105 | 0.96195 | 0.21171 |
| Decision Tree | 763 | 164 | 58 | 78 | 0.71777 | 0.90725 | 0.73873 |
| Random Forest | 818 | 95 | 127 | 23 | 0.76952 | 0.97265 | 0.42792 |

Based on the table above, all three models can predict positives very well while predicting negatives are significantly lower. Out of all positive’s predictions, Random Forest performs the highest 97% accuracy while Decision Tree performs the highest 73% for all the negative predictions.

If I am deciding which model to be implemented based on the table above, I would certainly choose the Random Forest not only because it has the highest accuracy, but the lowest False Negative since we need to avoid telling patients negative results while the patients are actually being positive. Although TNR has only 42% vs Decision Tree’s 73%, it isn’t as relevant as TPR since patients with FP may still conduct further clinical testing to confirm.