|  |  |  |  |
| --- | --- | --- | --- |
| N - Total Observations |  | **Prediction** | |
|  |  | Heart Disease | Healthy |
| **Actual** | Heart Disease | **True Positive (TP)** | **False Negative (FN)** |
| Healthy | **False Positive (FP)** | **True Negative (TN)** |
|  |  |  |  |
| **Accuracy** | How often were samples classified correctly? ***(TP + TN) / N*** | | |
| **Precision** | Of all the samples classified as Heart Disease, how many are Heart Disease? ***TP / (TP + FP)*** | | |
| **Recall** | Of all the actual Heart Disease samples, how many were classified as Heart Disease?  ***TP / (TP + FN)*** | | |
| **F1 Score** | Harmonic average of precision and recall where 1 is perfect and 0 is worst. ***2 \* ((Precision \* Recall)/(Precision + Recall))*** | | |

NOTES:

* Running more iterations increases the likelihood of identifying differences in models.
* NOTE:
  + Calculated Accuracy is Highest for NN, Majority, KNN
  + SciKit-Learn Scores are Highest for RF, NN
* Assuming we want to minimize False Negatives to error on the side of always finding heart disease then we would bias **Recall** over **Precision**
  + NN, KNN, Conservative Voting
* Looking at the Confusion Matrix Factors
  + Accuracy – NN
  + Precision – RF
  + Recall – Conservative
  + F1 Score - NN