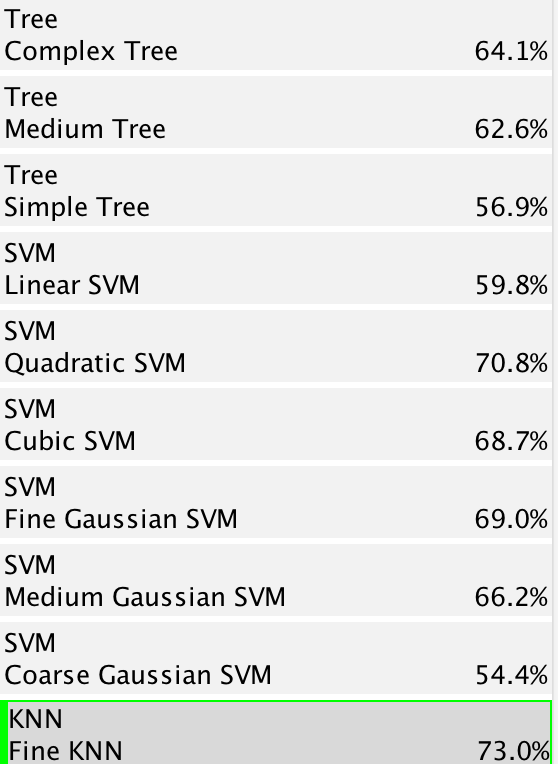
The entire table (containing multiple measurements for a patient) is considered. All the measurements are taken into consideration. Based on the maximum value for the TNT/TNT Hs, the binary outcome is assigned for a patient. All measurements for an individual patient are associated with the same binary outcome (based on the maximum value for TNT/ TNT Hs). The logistic regression script that I wrote in MATLAB used a regularization parameter=1000 (this is used penalize the coefficients in order to avoid overfitting).

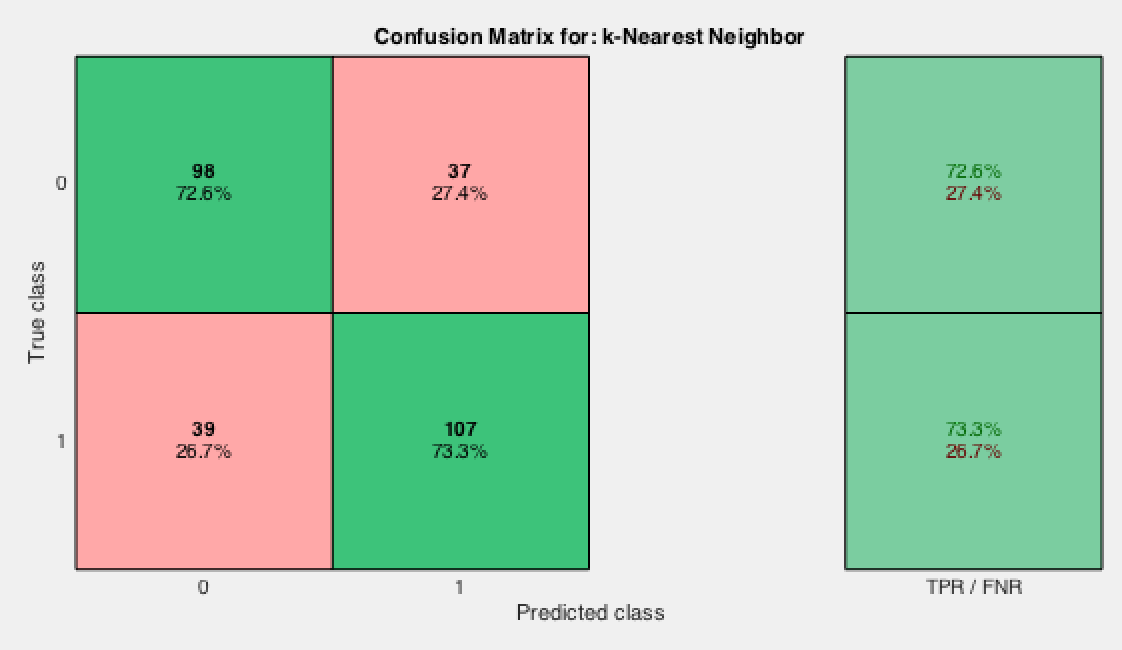
Remark: I have implemented logistic regression (glmfit) using 5-fold crossvalidation, repeated for 5 times. Trying to find out how to average the metrics across repeats.

**Case-1**

The list of predictors used [WBC, RBCi, HGB, MCV, RDW, MCH, MCHC, HCT, PLT, Gender]

Outcome is based on TNT (0, 1)





Confusion matrix for Fine KNN

Using stepwise fit, the pvalues are

pval =

0.0263

0.2040

0.3826

0.0604

0.0127

0.0526

0.0058

0.3442

0.5228

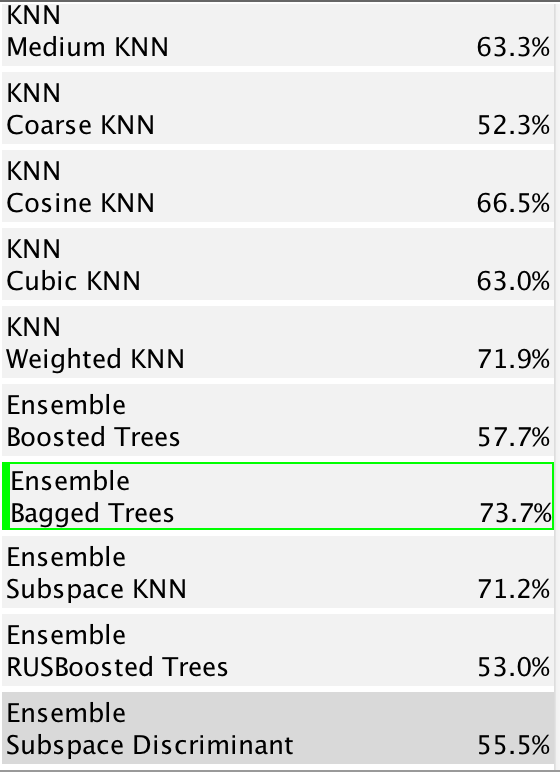
0.2945

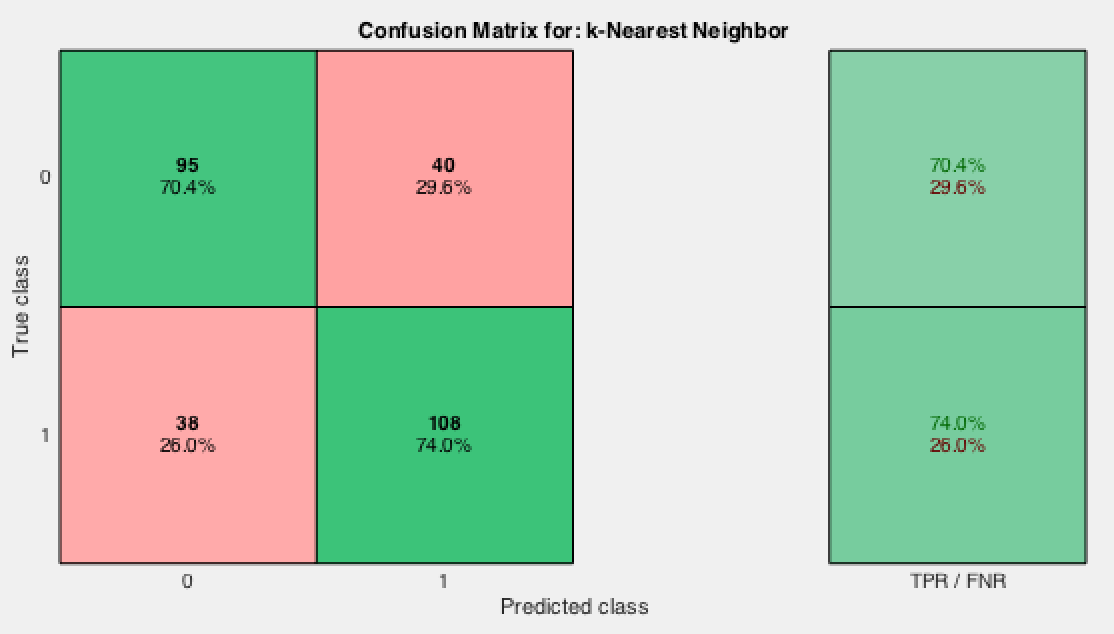
stats.rmse = 0.4887

Using the logistic regression model I wrote in MATLAB, there were 29 mismatches out of 81 while making predictions (used 200 to train and 81 to predict).

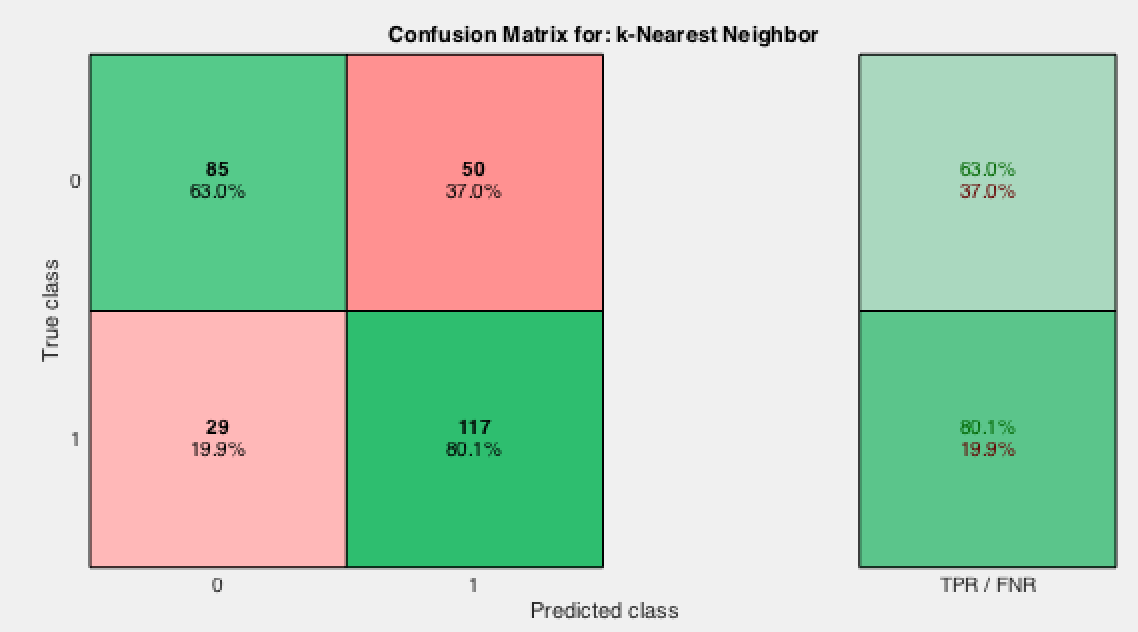
The list of predictors used [WBC, RBCi, HGB, MCV, RDW, MCHC, PLT, Gender]

Outcome is based on TNT (0, 1)

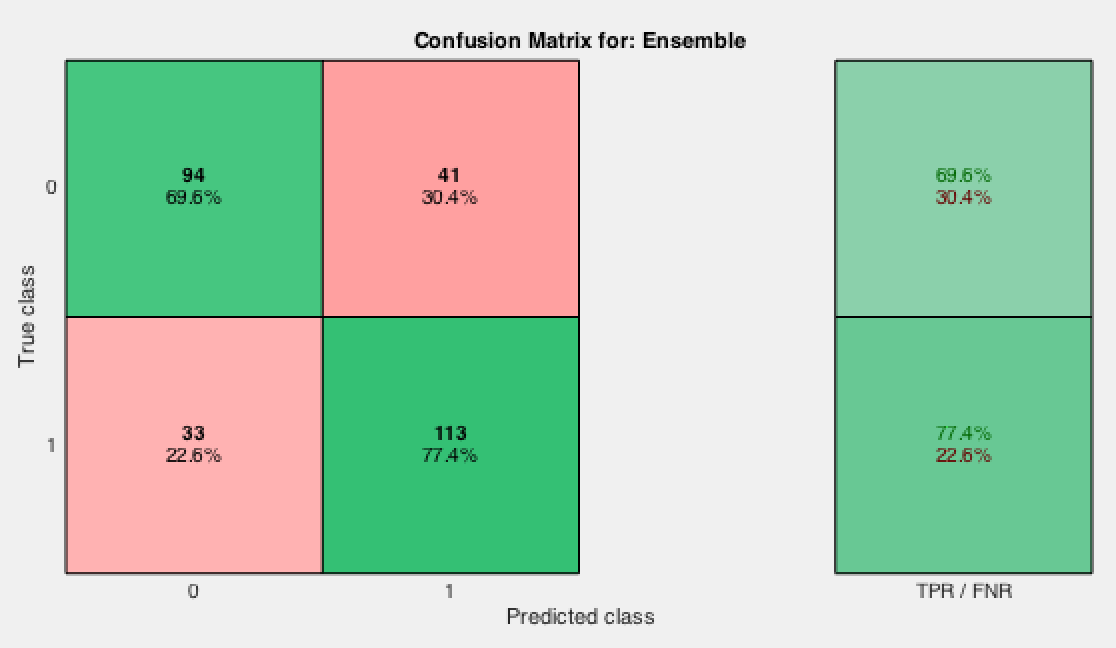




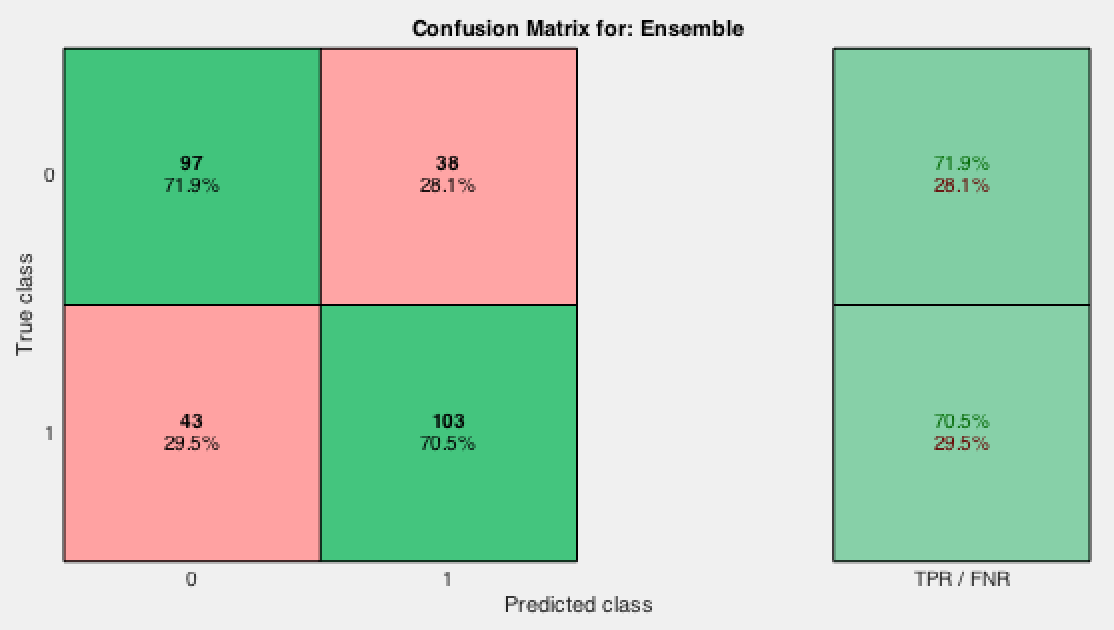
Confusion matrix for Fine KNN



Confusion matrix for Weighted KNN



Confusion matrix for Bagged Trees



Confusion matrix for Ensemble subspace KNN

Using stepwise the p-values are

pval =

0.0263

0.2040

0.3826

0.0604

0.0127

0.0058

0.5228

0.2945

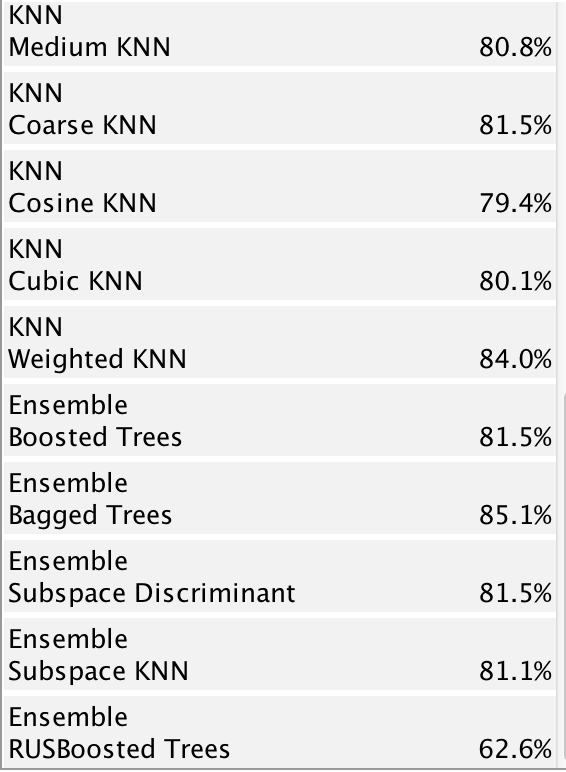
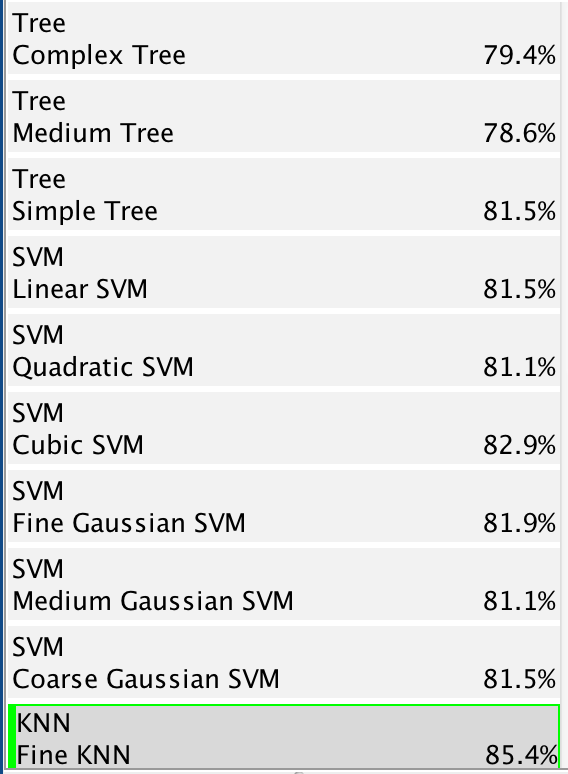
stats.rmse = 0.4887

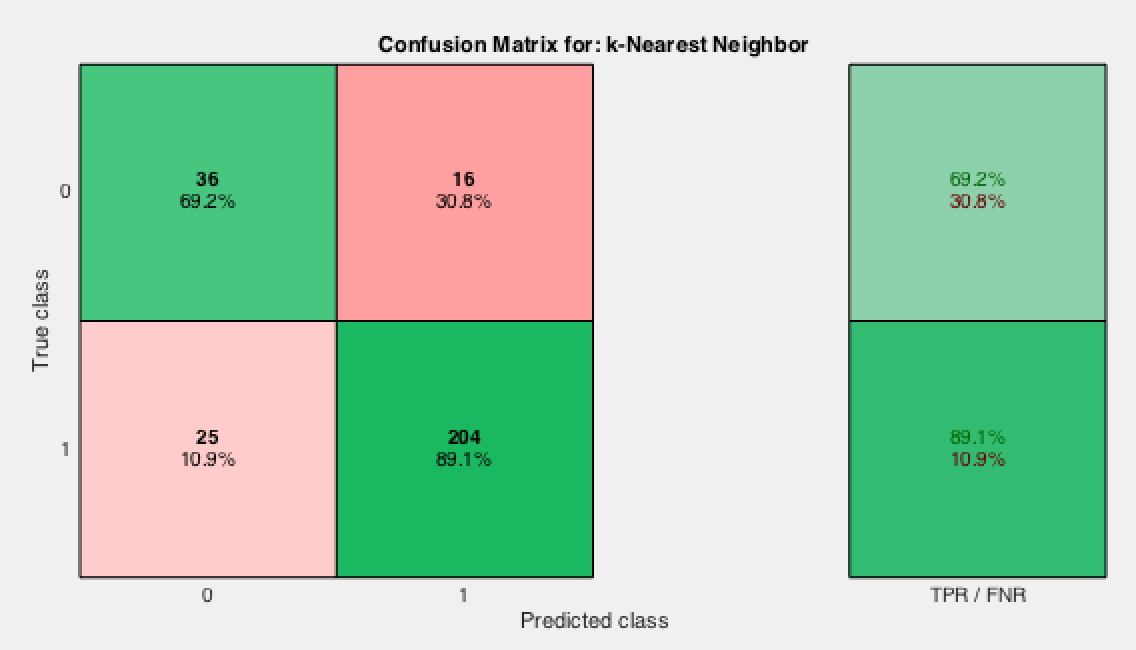
Using the logistic regression model I wrote in MATLAB, there were 31 mismatches out of 81 while making predictions (used 200 to train and 81 to predict). This value is varying between 27-42 for each run (due to the random selection of training and testing instances).

**Case-2**

The list of predictors used [WBC, RBCi, HGB, MCV, RDW, MCH, MCHC, HCT, PLT, Gender]

Outcome is based on HsTNT (0, 1)





Confusion matrix for Fine KNN

Using stepwisefit, the p-values are

pval =

0.0000

0.1364

0.2707

0.5058

0.3447

0.3886

0.3932

0.2734

0.0022

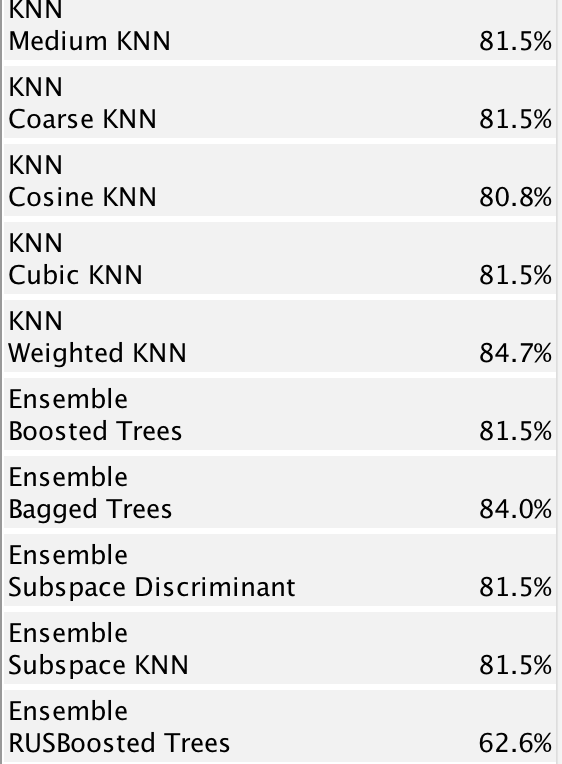
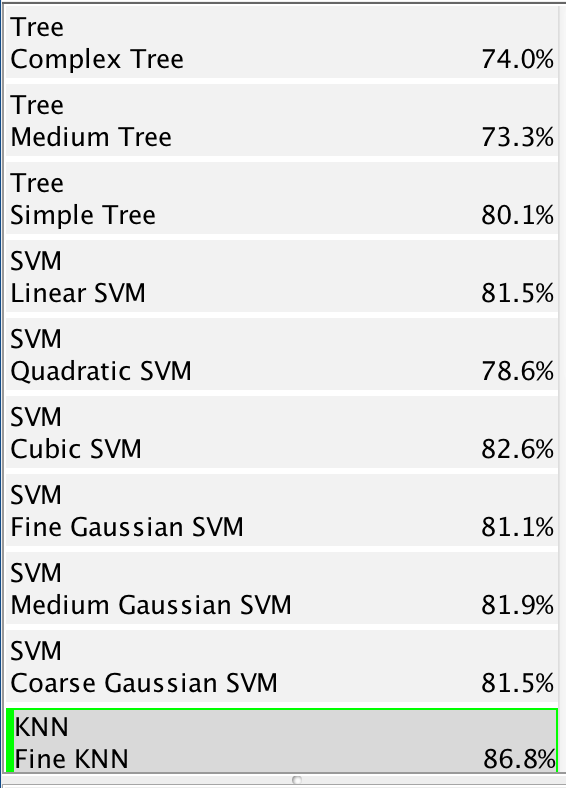
0.0904

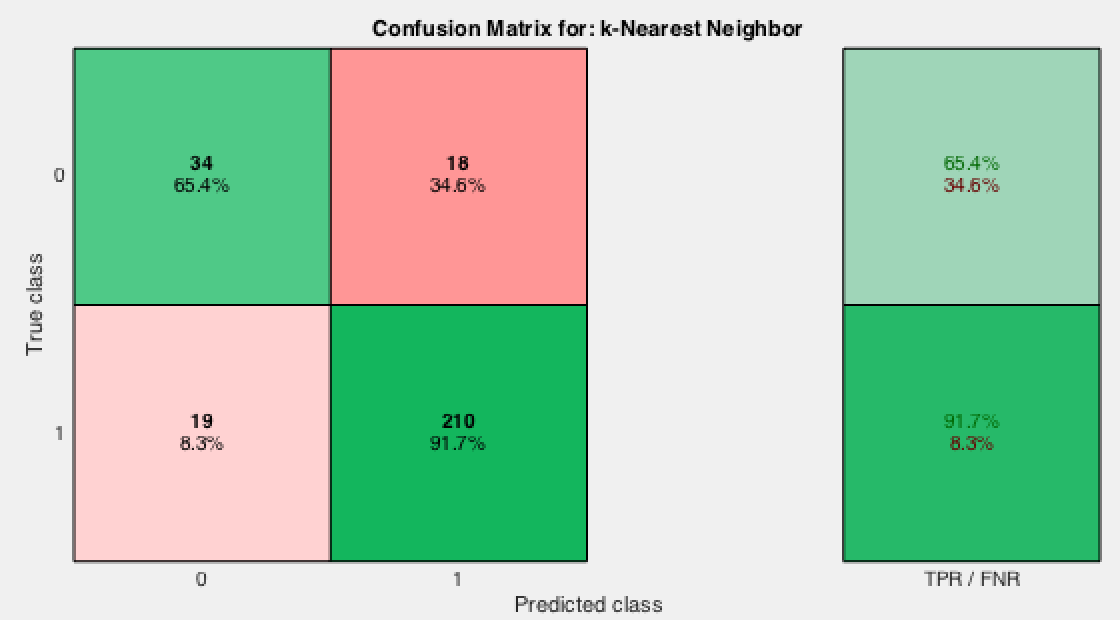
stats.rmse=0.3727

Using the logistic regression model I wrote in MATLAB, there were 13 mismatches out of 81 while making predictions (used 200 to train and 81 to predict). This value is varying between 11-16 for each run (due to the random selection of training and testing instances).

The list of predictors used [WBC, RBCi, HGB, MCV, RDW, MCHC, PLT, Gender]

Outcome is based on HsTNT (0, 1)





Confusion matrix for Fine KNN

Using stepwise the p-values are

pval =

0.0000

0.1364

0.2707

0.5058

0.3447

0.3932

0.0022

0.0904

stats.rmse = 0.3727

Using the logistic regression model I wrote in MATLAB, there were 14 mismatches out of 81 while making predictions (used 200 to train and 81 to predict). This value is varying between 10-20 for each run (due to the random selection of training and testing instances).