Results

Results obtained for accuracy are as follows-
Train set score after dimensionality reduction using KNN as classifier: 0.85
Test set score after dimensionality reduction using **KNN** as classifier: **0.84**
Train set score after dimensionality reduction using Decision Tree as classifier: 0.91
Test set score after dimensionality reduction using **Decision Tree** as classifier: **0.75**
Train set score after dimensionality reduction using Logistic Regression as classifier: 0.85
Test set score after dimensionality reduction using **Logistic Regression** as classifier: **0.83**

Metrics for K-Nearest-Neighbors model on testing data :
Accuracy: 0.8421052631579
Misclassification Rate: 0.1578947368421
Area Under Curve: 0.8369337979094
Root Average Squared Error: 0.3655036399936

Metrics for Decision Tree model on testing data:
Accuracy: 0.7500000000000
Misclassification Rate: 0.2500000000000
Area Under Curve: 0.7473867595819
Root Average Squared Error: 0.4330595232888

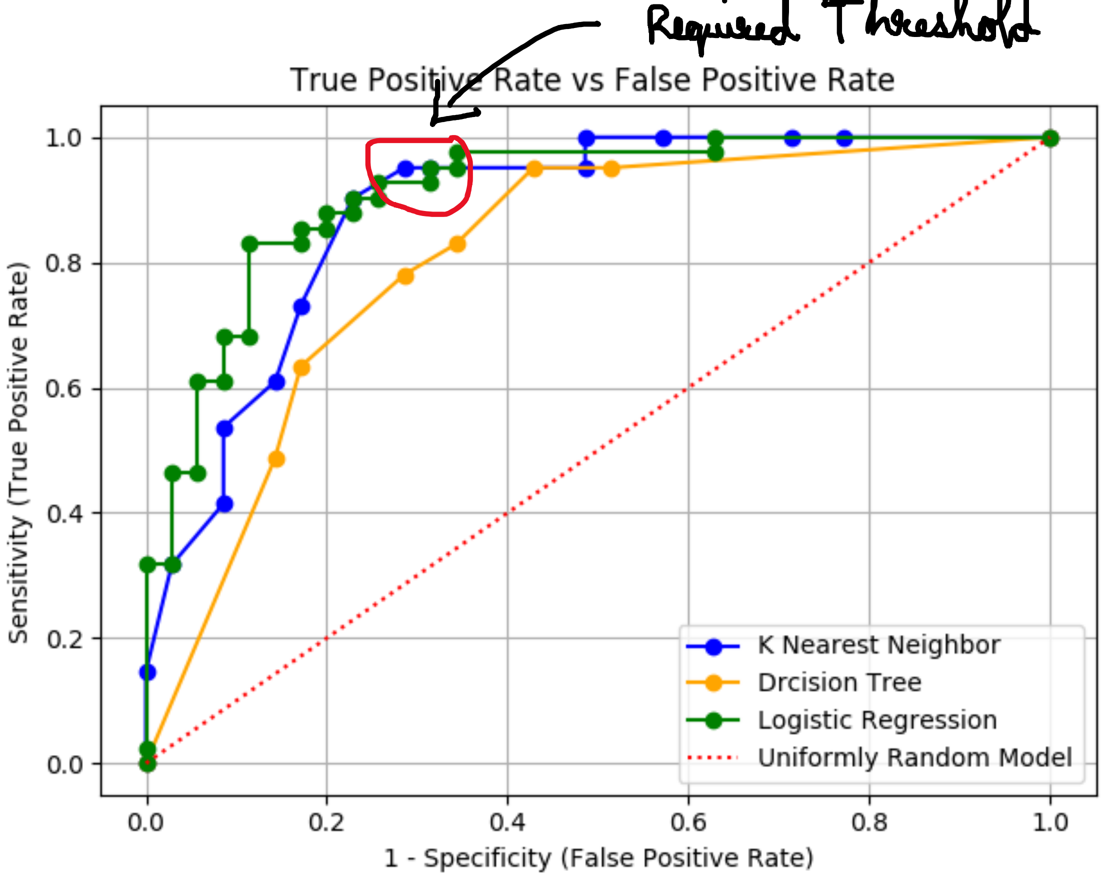
Metrics for Logistic Regression model on testing data
Accuracy: 0.8289473684211
Misclassification Rate: 0.1710526315789
Area Under Curve: 0.8205574912892
Root Average Squared Error: 0.3440307449803

Definitely, KNN and Logistic Regression are performing better than Decision tree. But we can notice from above results that Logistic Regression and KNN have comparable results. KNN has a little better Accuracy, Area Under Curve (AUC) and a bit lower misclassification rate than Logistic Regression model. Whereas, Logistic Regression has lower RASE value than KNN.

Now, we observe the ROC curves of three models together at common coordinates, which compares their True Positive Rate and False Positive Rate.

Results

In the below ROC curves obtained, we can notice the red circled point which will serve us the required threshold for True Positive Rate required by us from model. As if we choose the highest point (maximum True Positive Rate) of Green Curve (Logistic Regression) as our threshold, then we also have greater False Positive Rate(x-axis) along with it. Thus, we finally can decide that we do not need Logistic Regression and KNN is performing better with around 90%(more than) True Positive Rate and just around 30%(less than) False Positive Rate. As, finally physician should use the model which has greater True Positive Rate but along with it less False Positive Rate.


Required Threshold