1. **Decision Tree for Diabetes Dataset**

**DataPoints used:**

Number.of.times.pregnant

+Plasma.glucose.concentration

+Diastolic.blood.pressure

+Triceps.skin.fold.thickness

+X2.Hour.serum.insulin

+Body.mass.index

+Diabetes.pedigree.function

+Age..years

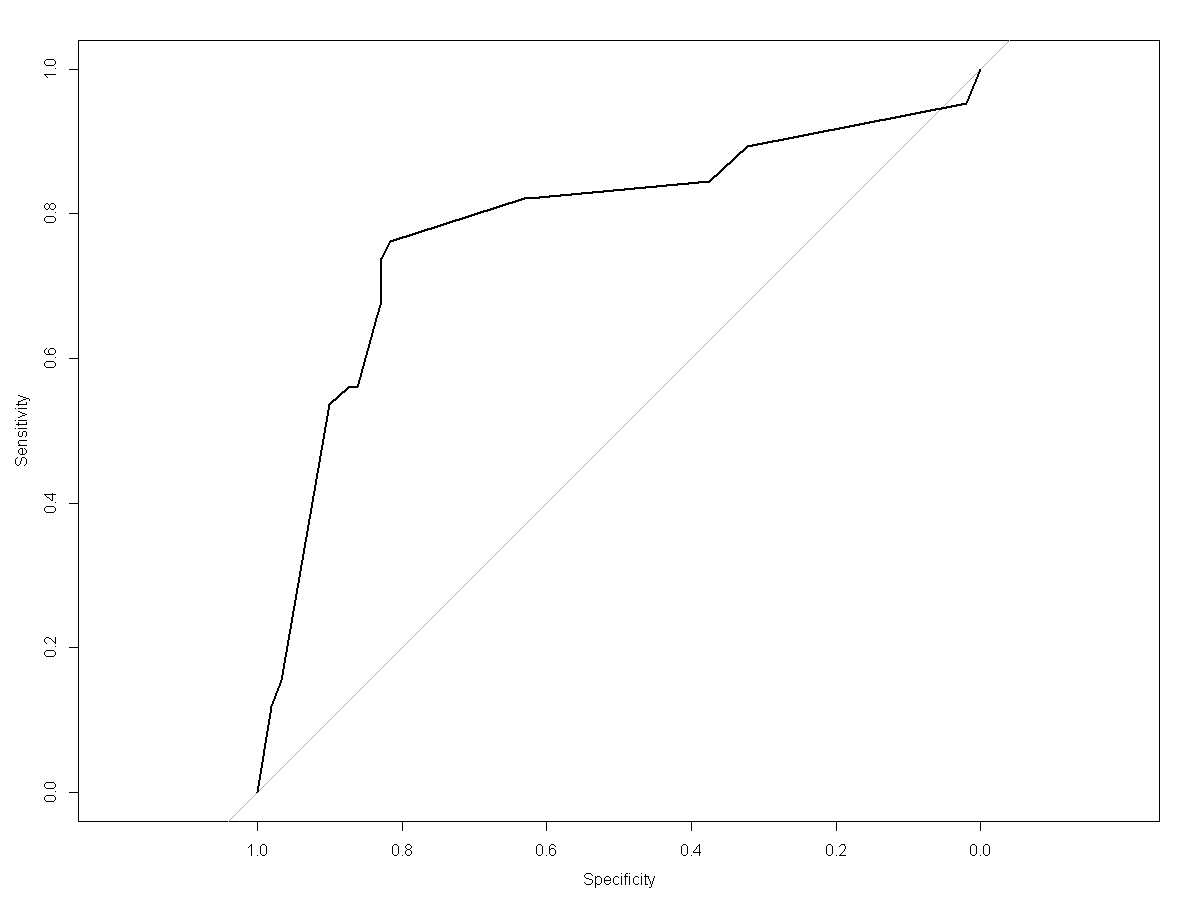
**Confusion Matrix for Decision Tree**

# NO YES

#NO 127 25

#YES 29 55

**#Accuracy of the above datapoints using Decision Tree - 77%**

**#Area under the curve: 0.7815**

**Datapoints Used: All**

**Confusion Matrix for all data points using Decision Tree**

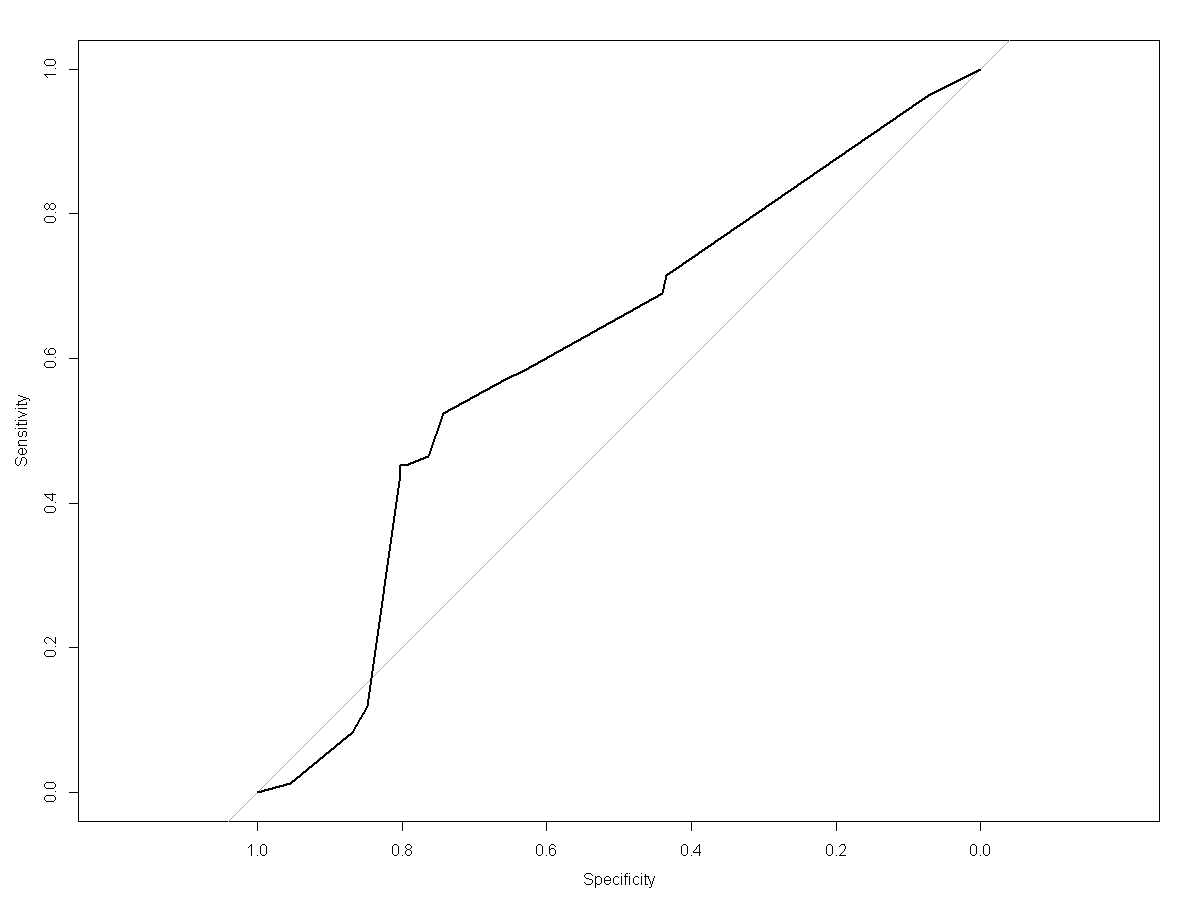
# NO YES

#NO 127 25

#YES 29 55

**#Accuracy of the above datapoints using Decision Tree - 65%**

**#Area under the curve: 0.61**



2. **Random Forest for Diabetes Dataset**

**DataPoints used:**

Number.of.times.pregnant

+Plasma.glucose.concentration

+Diastolic.blood.pressure

+Triceps.skin.fold.thickness

+X2.Hour.serum.insulin

+Body.mass.index

+Diabetes.pedigree.function

+Age..years

**Confusion Matrix for Random Forest**

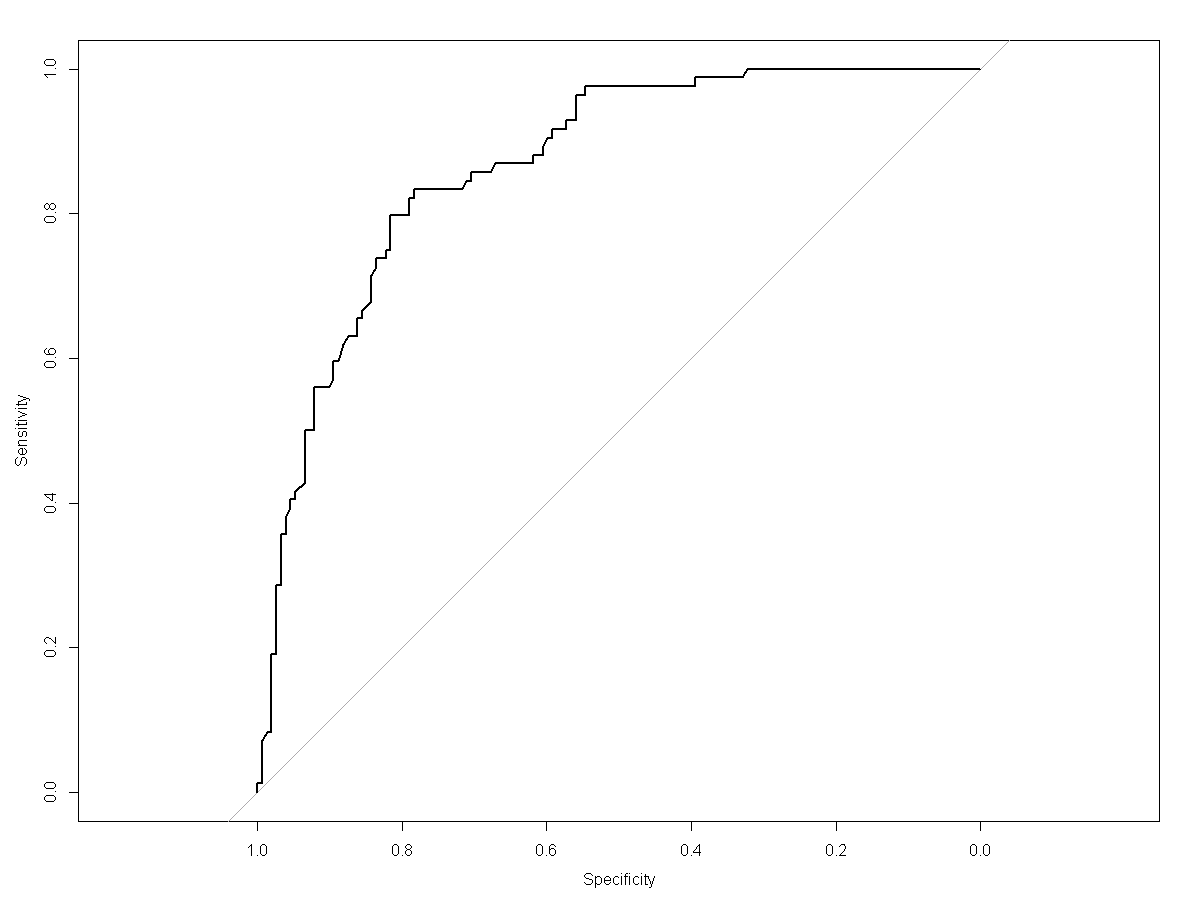
# NO YES

#NO 127 25

#YES 23 61

**#Accuracy of the above datapoints using Random Forest - 79.66%**

**#Area under the curve: 0.866**



**Datapoints Used: All**

**Confusion Matrix for all data points using Random Forest**

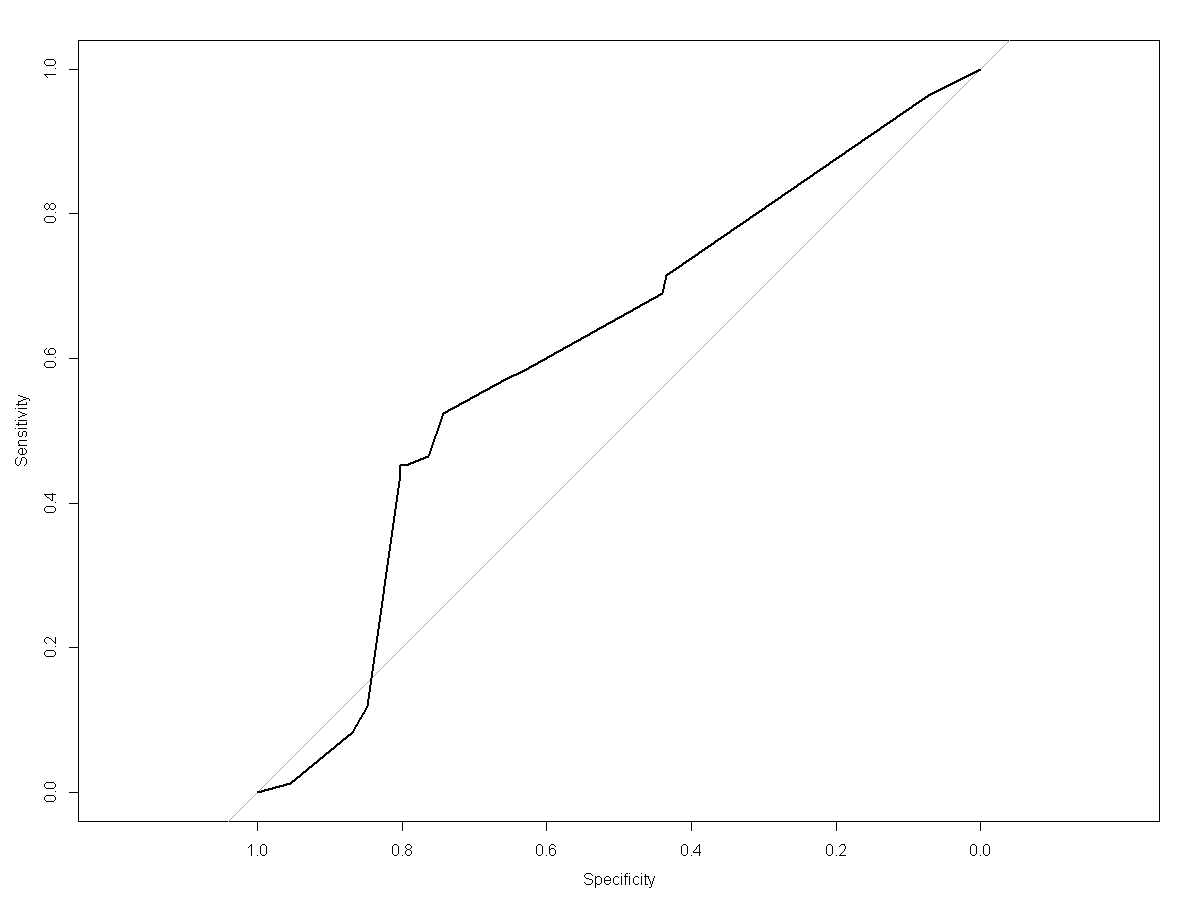
# NO YES

#NO 135 17

#YES 32 52

**#Accuracy of the above datapoints using Random Forest - 79%**

**#Area under the curve: 0.61**



3. **Naïve Bias for Diabetes Dataset**

**DataPoints used:**

Number.of.times.pregnant

+Plasma.glucose.concentration

+Diastolic.blood.pressure

+Triceps.skin.fold.thickness

+X2.Hour.serum.insulin

+Body.mass.index

+Diabetes.pedigree.function

+Age..years

**Confusion Matrix for Random Forest**

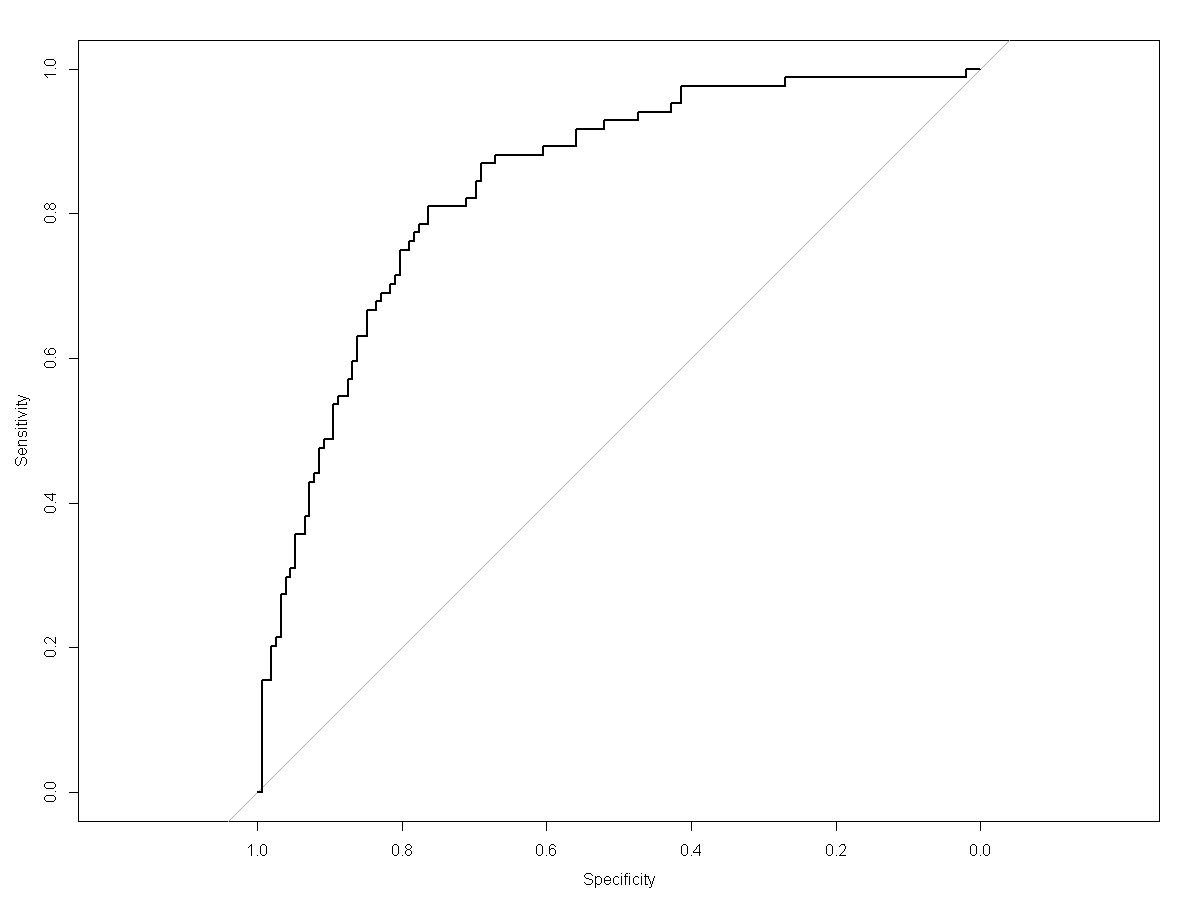
# NO YES

#NO 129 23

#YES 28 56

**#Accuracy of the above datapoints using Naïve Bias - 78.3%**

**#Area under the curve: 0.8398**



**Datapoints Used: All**

**Confusion Matrix for all data points using Naïve Bias**

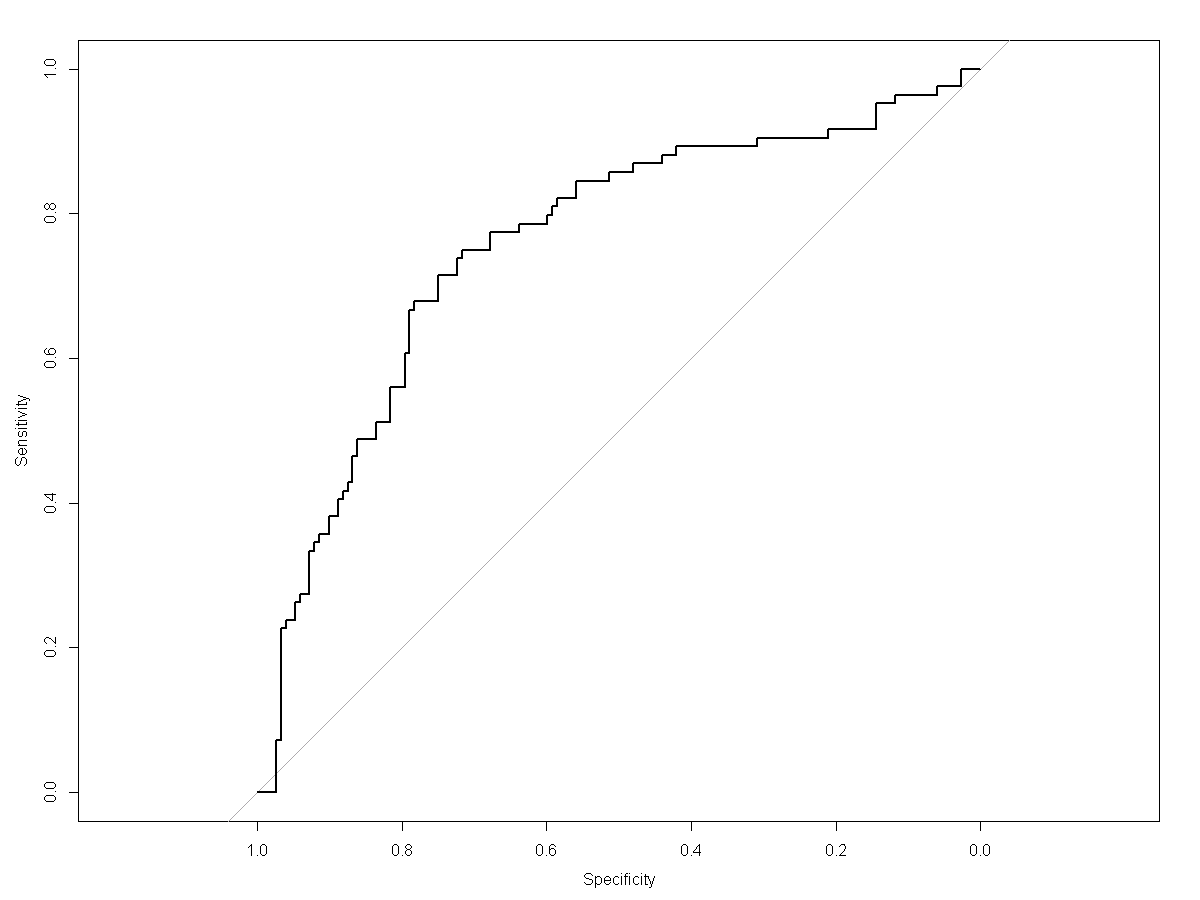
# NO YES

NO 120 32

YES 29 55

**#Accuracy of the above datapoints using Decision Tree - 74.3%**

**#Area under the curve: 0.7564**



4. **Logistic Regression for Diabetes Dataset**

**DataPoints used:**

Number.of.times.pregnant

+Plasma.glucose.concentration

+Diastolic.blood.pressure

+Triceps.skin.fold.thickness

+X2.Hour.serum.insulin

+Body.mass.index

+Diabetes.pedigree.function

+Age..years

**Confusion Matrix for Logistic Regression**

survival

0 1

0 135 17

1 27 57

**#Accuracy of the above datapoints using Logistic regression - 81%**

**#Area under the curve: 0.7834**

