TOOLS USED :-

R programming language, RStudio

METHODOLOGY :-

1) Bayes Classifier --> This classifier is based on Bayes theorem which is as follows :-

$P(C|Ai = ai) \sim P(Ai = ai|C)*P(C)$
where C represents a class and A represents an attribute and 'a' represents a value of attribute Ai

2) Decision Tree --> Split each data set (D) on some attribute (A) such that the information gain from splitting on attribute A is maximum among all other attributes.

$gain(D,A) = E(D) - EA(D)$
where E(D) is entropy of data set D before splitting and EA(D) is entropy of dataset after splitting on attribute A

3) SVM → Based on finding the maximum margin hyperplane which is of the form $W.X + b = 0$ and has margins of the form $W.X + b = 1$ and $W.X + b = -1$. A tuple Xi with $W.Xi + b > 0$ is classified as a positive example and a tuple Xi with $W.Xi + b < 0$ is classified as a negative example.

Parameters W and b can be computed using Legrange's Multiplier method along with KKT constraints to solve inequalities. Goal is to minimize $(|W|*|W|)/2$ under the constraint $yi(W.xi + b)>=1$.

$W = sum\_over\_all\_i(Lambda\_i*yi*xi)$
where Lambda_i is a Lagrange Multiplier

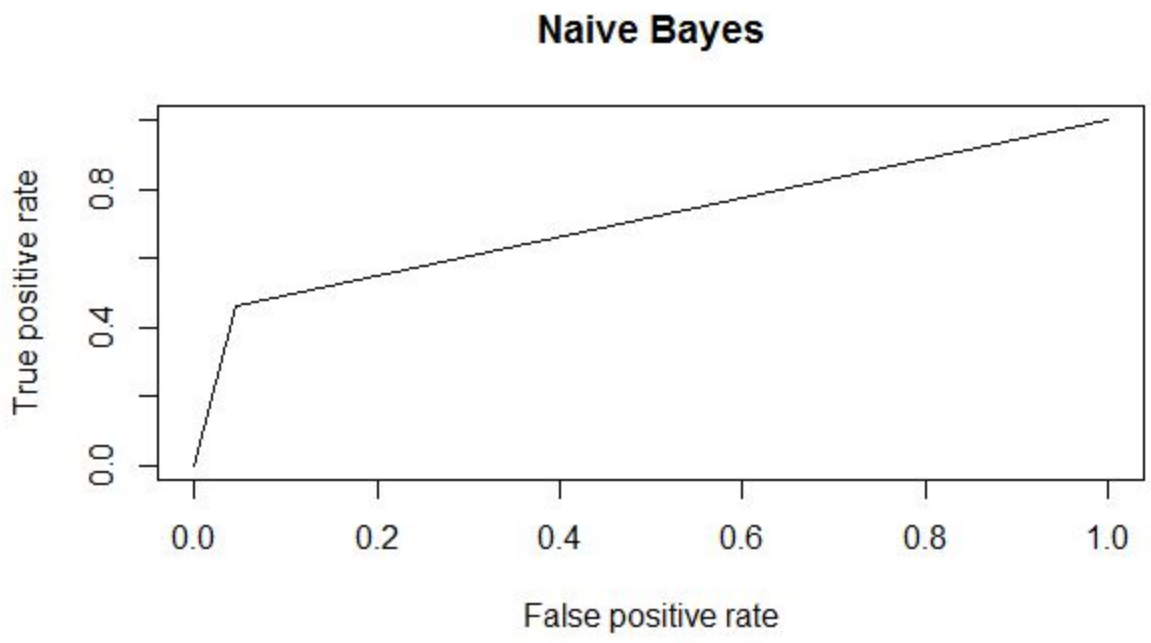
b can then be computed by using value of W, classifying training data using $W.X + b$ and taking average of all positive values of b found.

RESULTS:-

BAYESIAN CONFUSION MATRIX

	False	True
False	1220	116
True	67	96

ROC curve :-



DECISION TREE CONFUSION MATRIX

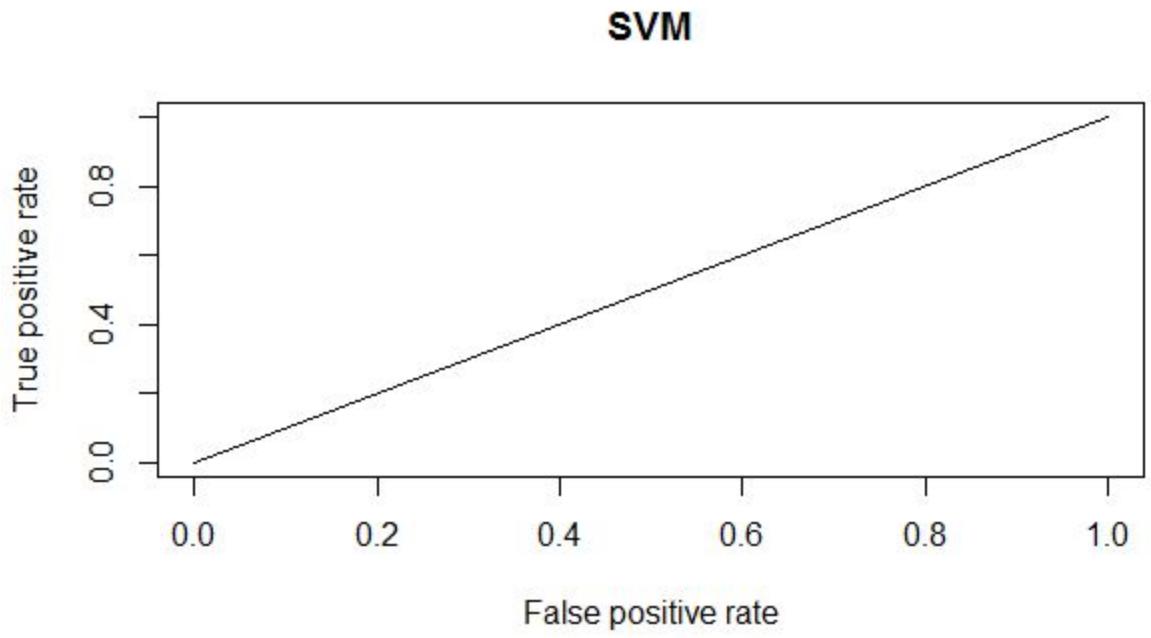
	False	True
False	1245	163
True	42	49

ROC curve :-

SVM CONFUSION MATRIX

	False	True
False	1287	212
True	0	0

ROC curve

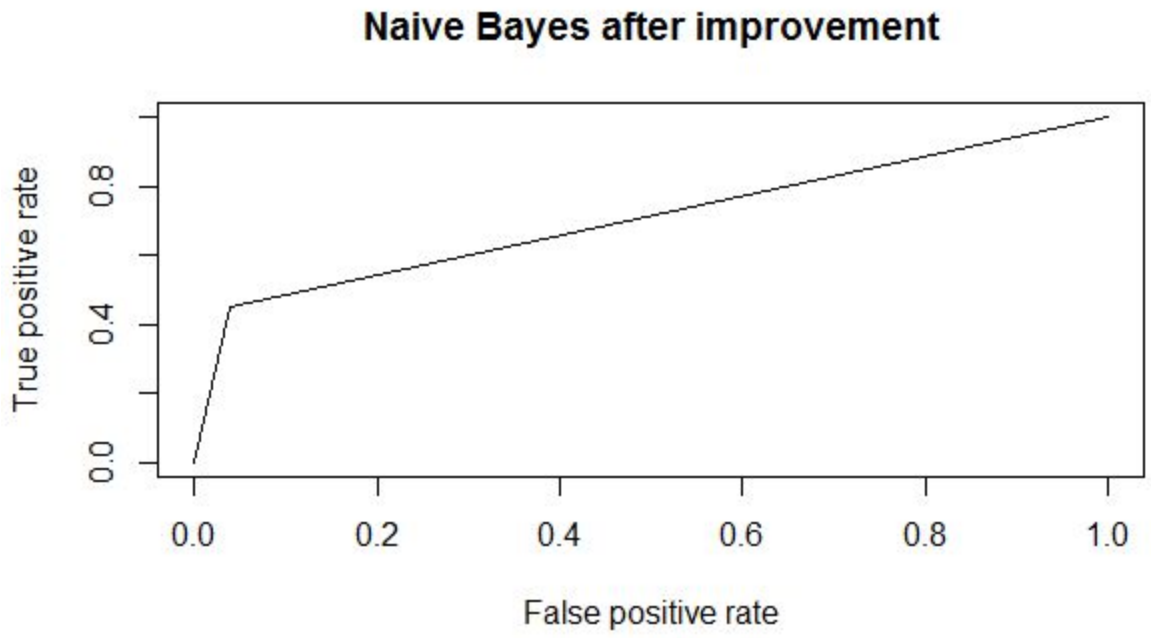


AFTER CHOOSING THE RIGHT SET OF PREDICTOR VARIABLES

BAYESIAN CONFUSION MATRIX

	False	True
False	1234	122
True	53	90

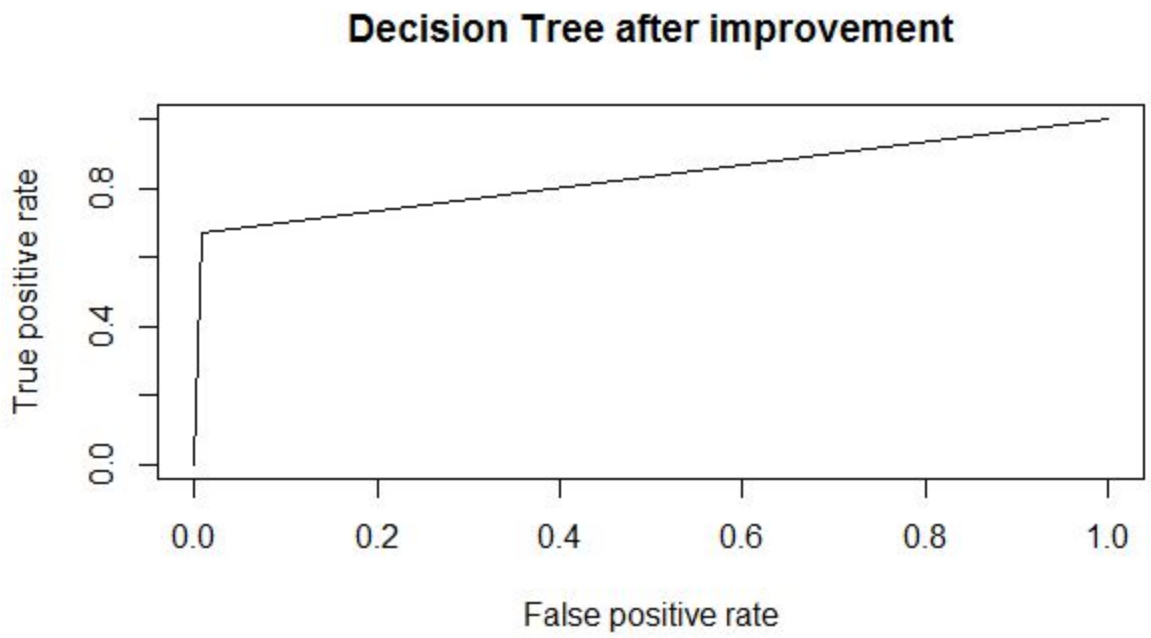
ROC curve



DECISION TREE CONFUSION MATRIX

	False	True
False	1270	81
True	17	131

ROC curve



SVM CONFUSION MATRIX

	False	True
False	1283	163
True	4	49

ROC curve