Assignment 2

Step 1

Firstly, the data was separated into training, validation and test data.

Step 2

```
## Misclassification on deviance with test:  0.268
```

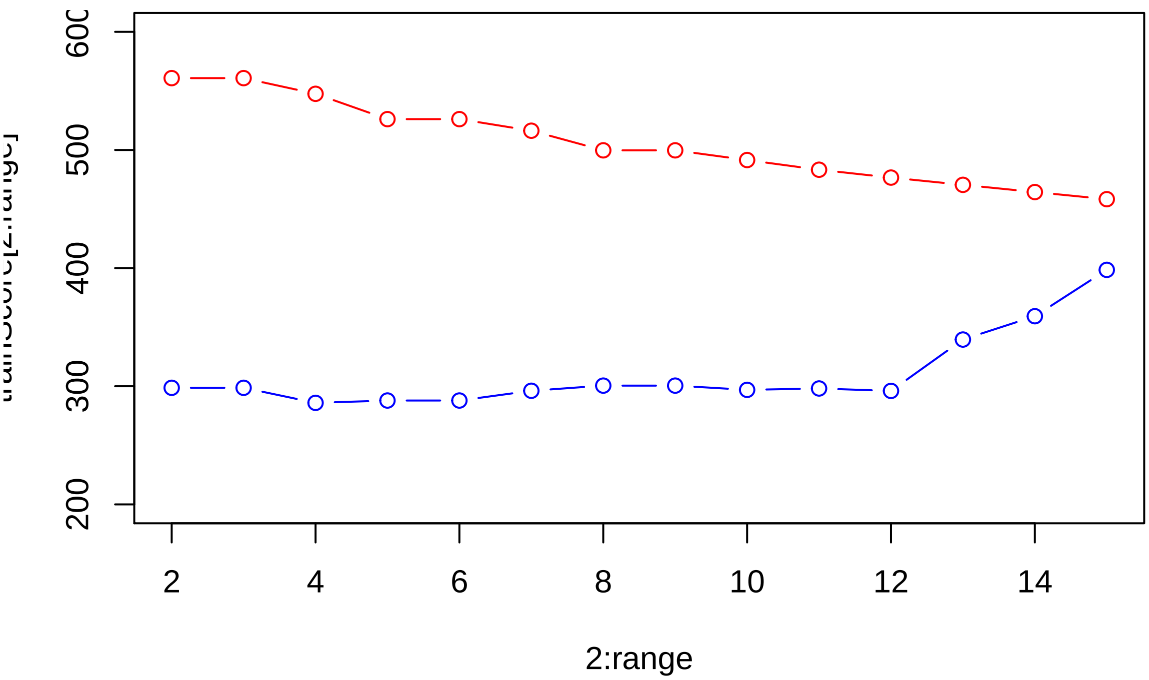
```
## Misclassification on deviance with train:  0.212
```

```
## Misclassification on gini with test:  0.368
```

```
## Misclassification on gini with train:  0.24
```

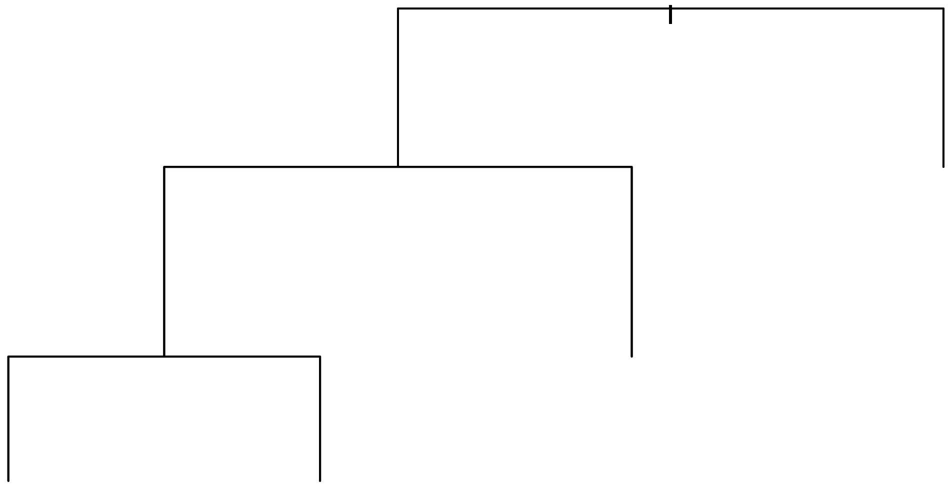
Here, misclassification rate is lower when using deviance as measure of impurity.

Step 3




```
## Misclassification on optimal tree:  0.256
```

```
## Tree depth is 3 as can be seen in the plot.
```



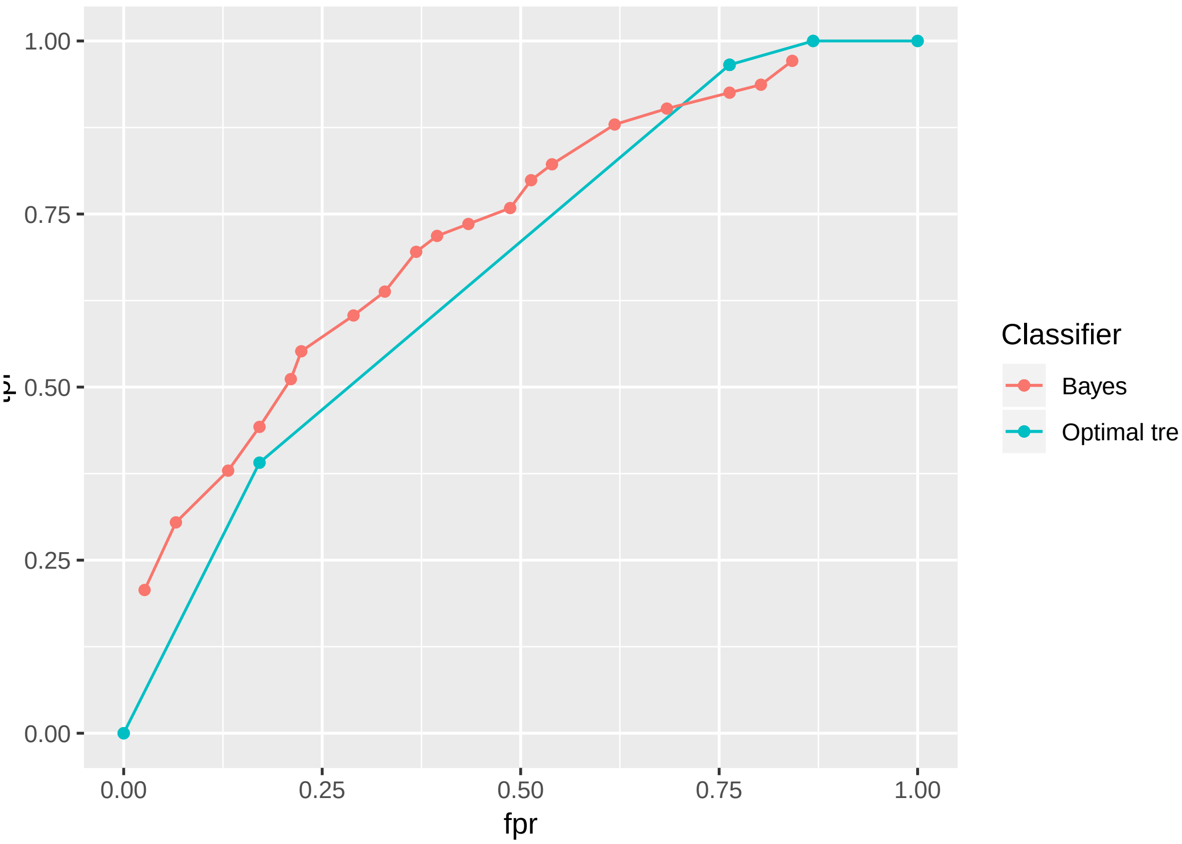
```
## Used variables in optimal tree:
## 'savings' 'duration' 'history'
```

Step 4

```
## Confusion table of naïve bayes (test):

##          Actual
## Predicted bad good
##      bad   46   49
##      good  30  125

## Misclassification with naïve bayes (test):  0.316

## Confusion table of naïve bayes (train):

##          Actual
## Predicted bad good
##      bad   95   98
##      good  52  255

## Misclassification with naïve bayes (train):  0.3
```

Naïve Bayes has much better result than in step 3.

Step 5



Naïve Bayes has better ratio between TPR and FPR. The only exception is around $\pi = 0.75$, as can be seen in the graph.

Step 6

Using loss matrix with naïve bayes.

```
## Confusion table of naïve bayes (using test data):
```

```
##          Actual
## Predicted bad good
##      bad   71  122
##      good   5   52
```

```
## Misclassification with naïve bayes (using test data):  0.508
```

```
## Confusion table of naïve bayes (using train data):
```

```
##          Actual
## Predicted bad good
##      bad  137  263
##      good  10   90
```

```
## Misclassification with naïve bayes (using train data):  0.546
```

The misclassification is much greater. But the confusion matrix is more favorable from an economic point of view for a company since less are predicted good that are actually bad.