We list every calculation in every step below.

Step 2:

To consider LEAVE to be the Negative Class and STAY to be the Positive class, we could label the table that I output in R code in such a following form:

```
            stay.pred
stay.actual LEAVE STAY
      LEAVE   TN    FP
      STAY    FN    TP
```

$$FNR = \frac{FN}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

$$Accuracy = \frac{TN + TP}{TP + FP + TN + FN}$$

To calculate FNR, FPR and Accuracy, we write several functions in R:

R script codes:

```
> #Treating LEAVE as Negative and STAY as positive
> # now make a function for computing the accuracy
> accuracy = function(cm){  # input confusion matrix
+   return(sum(diag(cm))/sum(cm))  # accuracy
+ }
> # now make a function for computing FPR
> FPR = function(cm){ # input confusion matrix
+   return((cm[1,2])/(cm[1,2]+cm[1,1])) #FPR
+ }
> # now make a function for computing FNR
> FNR = function(cm){# input confusion matrix
+   return((cm[2,1])/(cm[2,1]+cm[2,2])) # FNR
+ }
```

We also use these functions to calculate FNR, FPR and Accuracy in step 3 and step 4.

Step 3: we use the following R code to find the CP which provides the lowest error:

```
lowestcp=fit.large$cptable[which.min(fit.large$cptable[,"xerror"]),"CP"]
```

Step 4: we use ROCR to find the best threshold at cut= 0.4359949, given cost.fn=490000*400, cost.fp=510000*600*0.5.

Step 5: Here is our idea to calculate expected value: the goal of our prediction model is to predict whether the customers leave or stay next year. So we only consider the influence of TN and FN.

For those customers who must stay in next year, we could get

$$\text{profit}_1 = \text{ the number of customers who must stay next year} \times \text{NPV} = 490000 \times 1000$$

For those customers who plan to leave in next year, we construct a prediction model to predict whether a customer will leave or stay and use the prediction model to offer the discount only to the customers who are predicted to leave. Offered the discount, a customer who was likely to leave will now stay with a 0.50 probability.

profit_2
$=$ the profit of cutomers who plan to leave next year but change their mind to stay
$-$ the loss of customers who actually stay but the prediction model incorrectly predicts them to leave
$-$ the loss of customers who plan to leave at first and get the discout but still leave
$=$ the number of cutomers who plan to leave next year but change their mind to stay
$\times$ (NPV $-$ discount)
$-$ the numbe of of customers who plan to leave at first and get the discout but still leave
$\times$ discount
$-$ the number of customers who actually stay but the prediction model incorrectly predicts them to leave
$\times$ discount $= \text{TN} \times 0.5 \times (1000 - 400) - \text{TN} \times 0.5 \times 400 - \text{FN} \times 0.5 \times 400$

Add profit1 and profit2 together, we could get

$$\begin{aligned}\text{Expexted value} &= \text{profit}_1 + \text{profit}_2 \\ &= 490000 \times 1000 + \text{TN} \times 0.5 \times (1000 - 400) - \text{TN} \times 0.5 \times 400 - \text{FN} \times 0.5 \times 400 = 490000 \times 1000 + \text{TN} \times 100 - \text{FN} \times 400\end{aligned}$$

Since $\text{TN} = (1 - \alpha) \times (\text{FP} + \text{TN})$ and $\text{FN} = \beta \times (\text{FN} + \text{TP})$, we could get Expected value in the required form: Expected value $= \text{v} - \text{a}\alpha - \text{b}\beta$

$$\begin{aligned}\text{Expexted value} &= 490000 \times 1000 + (1 - \alpha) \times (\text{TP} + \text{FN}) \times 100 - \beta \times (\text{FN} + \text{TP}) \times 400 \\ &= 490000 \times 1000 + (1 - \alpha) \times 510000 \times 100 - \beta \times 490000 \times 400 \\ &= 490000 \times 1000 + 100 \times 510000 - 510000 \times 100 \times \alpha - 490000 \times 400 \times \beta = 5.41 \times 10^8 - 5.1 \times 10^7 \alpha - 1.93 \times 10^8 \beta\end{aligned}$$

where v = 5.41×10^8, a = 5.1×10^7 and b=1.93×10^8.

The result table:

	FPR	FNR	Accuracy	Expected Value
BigTree	0.4052	0.3561	0.6194	450533759
Pruned Tree	0.2771	0.3456	0.6886	459122172
Best Threshold Pruned tree	0.9459	0.0123	0.5213	490348201

Also, for strategy a: the expected value = 490000*1000 = 490.0 million

For strategy b: the expected value = (490000+510000*0.5)*1000-400*1000000 = 345.0 million