Stock Price Prediction Project

Disha An

**Problem**

* **Select 1 stock and 10 different technical indicators.**

I select Walt Disney Co (DIS). The 10 technical indicators I chose are Momentum, DEMA, Volume, RSI, ROC, ZLEMA, WMA, DPO, CCI, and CMO.

The time horizon I used for analyzing is from Jan 1, 2015 to Jan 1, 2016.

The code is as follows,

library("quantmod")

#Allows us to import the data we need and calculate the technical indicators

library("rpart")

#Gives us access to the decision trees we will be using.

library("rpart.plot")

startDate = as.Date("2015-01-01")

#The beginning of the date range we want to look at

endDate = as.Date("2016-01-01")

#The end of the date range we want to look at

getSymbols("DIS", src = "yahoo", from = startDate, to = endDate, auto.assign = TRUE)

# Select 10 different technical indicators

# 1. Momentum

Momentum3 <- momentum(Op(DIS), n = 3)

# 2. Moving Average (DEMA)

DEMA3 <- DEMA(Op(DIS), n = 3)

# 3. Volume

Volume<- Vo(DIS)

# 4. Relative Strength Index

RSI3<-RSI(Op(DIS), n = 3)

# 5. Rate of Change

ROC3 <- ROC(Op(DIS), n = 3)

# 6. ZLEMA

ZLEMA3 <- ZLEMA(Op(DIS), n = 3)

# 7. WMA

WMA3 <- WMA(Op(DIS), n = 3)

# 8. DPR

DPO3 <- DPO(Op(DIS), n = 3)

# 9. CCI

CCI3 <- CCI(Op(DIS), n = 3)

# 10. CMF

CMO3 <- CMO(Op(DIS), n = 3)

* **Construct a decision tree and plot it.**

The code we used for constructing a decision tree is as follows,

# Construct a decision tree

#Calculate the difference between the close price and open price

PriceChange<- Cl(DIS) - Op(DIS)

# Construct a class for the price goes up and goes down

Class<-ifelse(PriceChange>0,"UP","DOWN")

#Create our data set

DataSet<-data.frame(Momentum3,DEMA3,Volume,RSI3,ROC3,ZLEMA3,WMA3,DPO3,CCI3,CMO3,Class)

#Name the columns

colnames(DataSet)<-c("Momentum","DEMA","Volume","RSI","ROC","ZLEMA","WMA","DPO","CCI","CMO","Class")

#Get rid of the data where the indicators are being calculated

DataSet<-DataSet[-c(1:4),]

DataSet <- DataSet[-c(247:248),]

nrow(DataSet)

#Use 2/3 of the data to build the tree

TrainingSet<-DataSet[1:164,]

#And leave out 1/3 data to test our strategy

TestSet<-DataSet[165:246,]

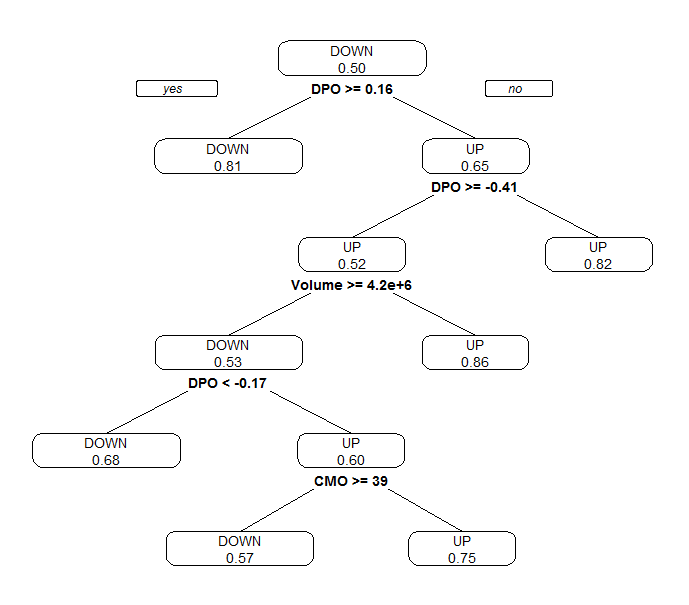
DecisionTree<-rpart(Class~Momentum+DEMA+Volume+RSI+ROC+ZLEMA+WMA+DPO+CCI+CMO,data = TrainingSet,cp=.001)

#Specifying the indicators to we want to use to predict the class and controlling the growth of the tree by setting the minimum amount of information gained (cp) needed to justify a split.

prp(DecisionTree,type=2,extra=8)

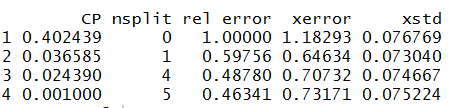
#Nice plotting tool with a couple parameters to make it look good.

The plot of this decision tree is as follows,



* **Prune the tree and calculate the best fit**

Calculate the minimal cp for each trees of each size. The result is as follows,



I am selecting the complexity parameter (cp) that has the lowest cross-validated error (xerror), which is 0.001000.

So, based on this CP, I construct the pruned decision tree. The code is as follows,

printcp(DecisionTree)

#shows the minimal cp for each trees of each size.

plotcp(DecisionTree,upper="splits")

#plots the average geometric mean for trees of each size.

# Prue the tree and calculate the best fit

PrunedDecisionTree<-prune(DecisionTree,cp=0.00100)

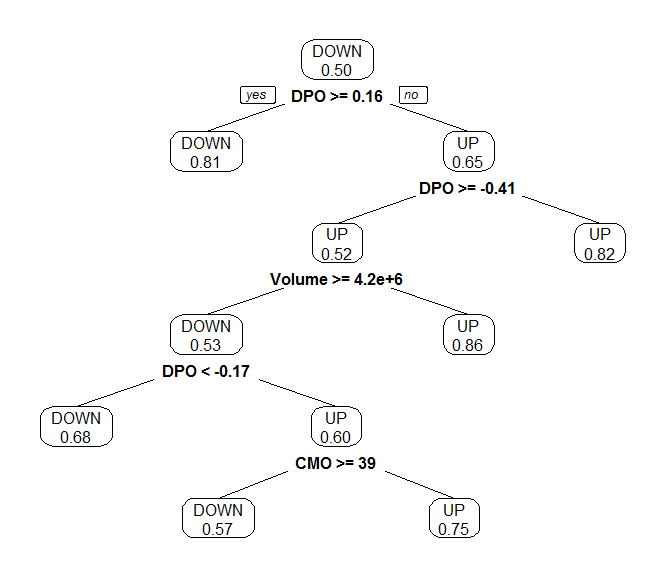
#I am selecting the complexity parameter (cp) that has the lowest cross-validated error (xerror)

prp(PrunedDecisionTree, type=2, extra=8)

# Calculate the accuracy of prediction and present the confusion matrix

printcp(PrunedDecisionTree)

The pruned decision tree plot is as follows. It’s almost as same as the previous one.



Calculate the accuracy of prediction and present the confusion matrix.

The confusion matrix is as follows,

|  |  |  |
| --- | --- | --- |
| Actual  Predicted | Down | Up |
| Down | 34 | 11 |
| Up | 5 | 32 |

The accuracy of prediction is calculated as follows,

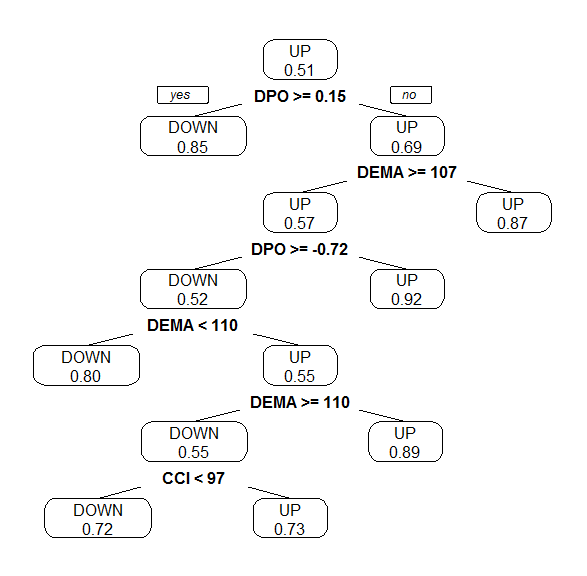
* **Propose alternative models for the selection of indicators such that the accuracy of predictions is increased.**

I follow the previous steps to get the alternative models.

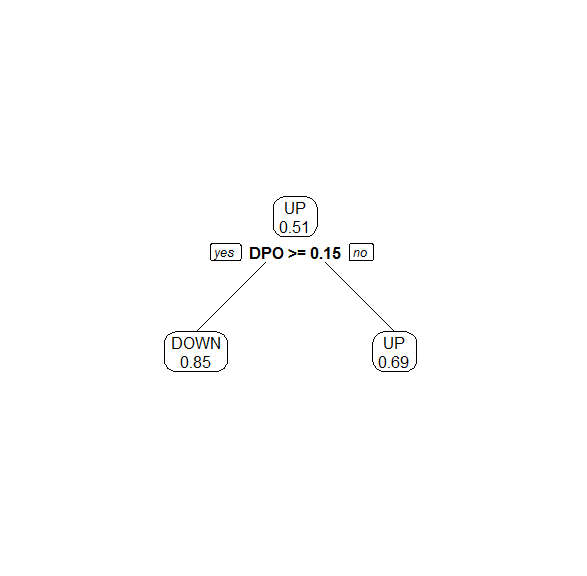
The technical indicators I chose are: DEMA, DPO, RSI, EMAcross, MACDsignal, SMI, ALMA, WMA, DPO, CCI, and CMO.

Repeat the code we used for the previous questions, we then get the following plot.

The alternative decision tree is:



The pruned alternative tree is:



The confusion matrix is as follows:

|  |  |  |
| --- | --- | --- |
| Actual  Predicted | Down | Up |
| Down | 31 | 4 |
| Up | 4 | 32 |

So the accuracy is:

Compared the one we calculate before, the accuracy is increased.