Random Forest

Nabin Kumar Karki

October 10, 2018

```
#loading required packages.
require(randomForest)
```

```
## Loading required package: randomForest
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
require(rpart.plot)
```

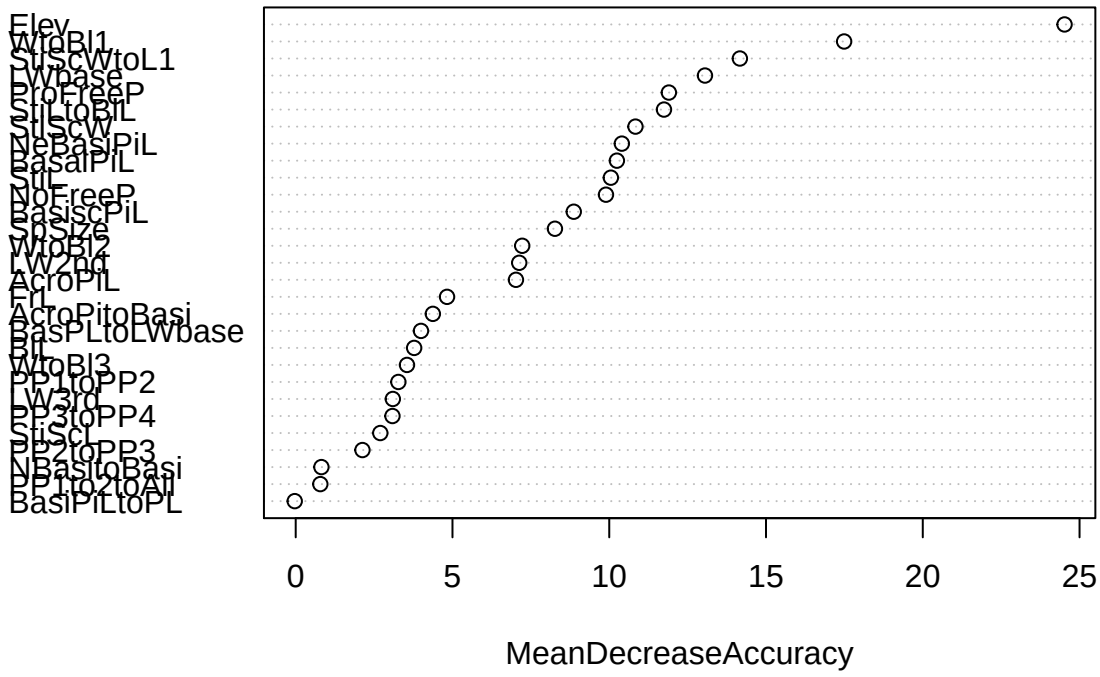
```
## Loading required package: rpart.plot
```

```
## Loading required package: rpart
```

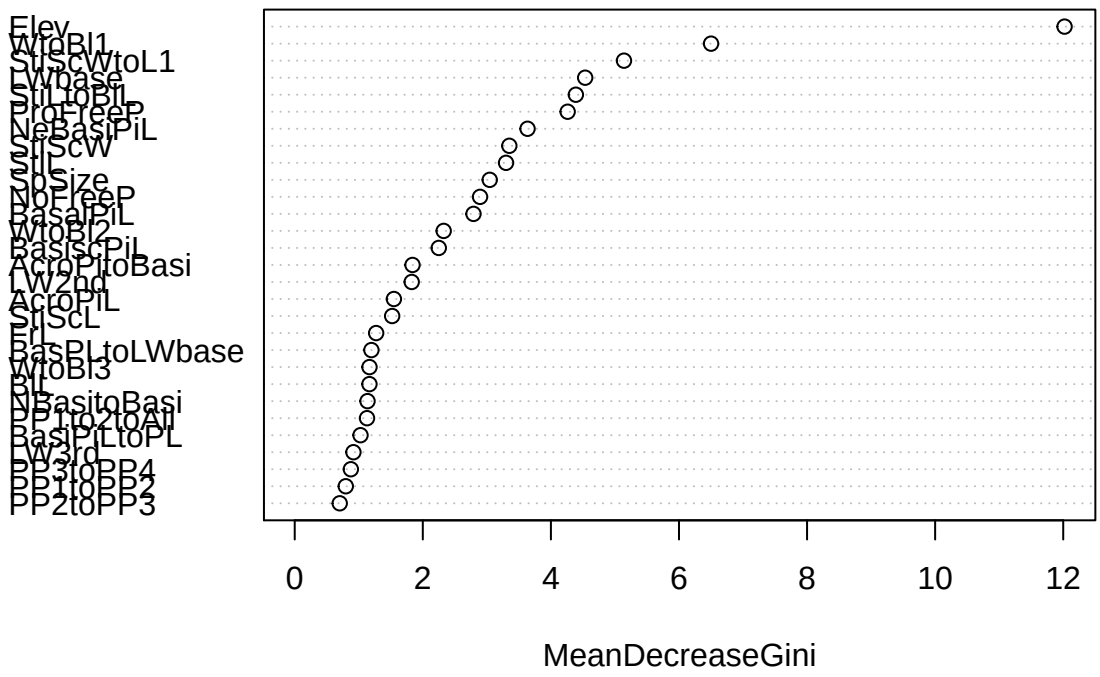
```
#importing the full data
four.Species <- read.csv("~/Capstone/four.csv")

##remove X1,spID
fourSpecies <- four.Species[-c(1,3)]

fourSpecies$SpCode<-as.factor(fourSpecies$SpCode)#change to factor

set.seed(12345)
four_all.rf<-randomForest(SpCode~.,fourSpecies, importance = TRUE) #with all variables
four_all.rf
```

```
##
## Call:
##  randomForest(formula = SpCode ~ ., data = fourSpecies, importance = TRUE)
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 5
##
##         OOB estimate of  error rate: 13.27%
## Confusion matrix:
##    1 2  3  4 class.error
## 1 37 0  0  4  0.09756098
## 2  3 5  4  0  0.58333333
## 3  2 0 18  0  0.10000000
## 4  2 0  0 38  0.05000000
```

```
varImpPlot(four_all.rf, type = 1) #Accuracy
```



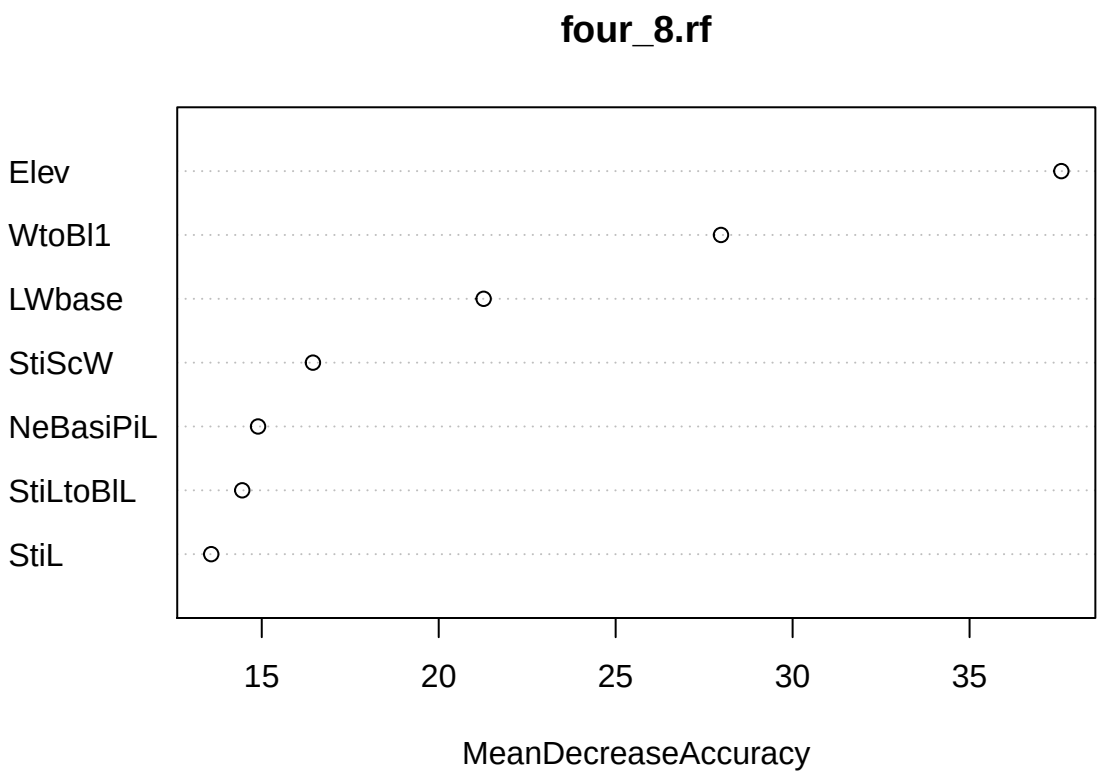

```
varImpPlot(four_all.rf, type = 2) #gini index
```



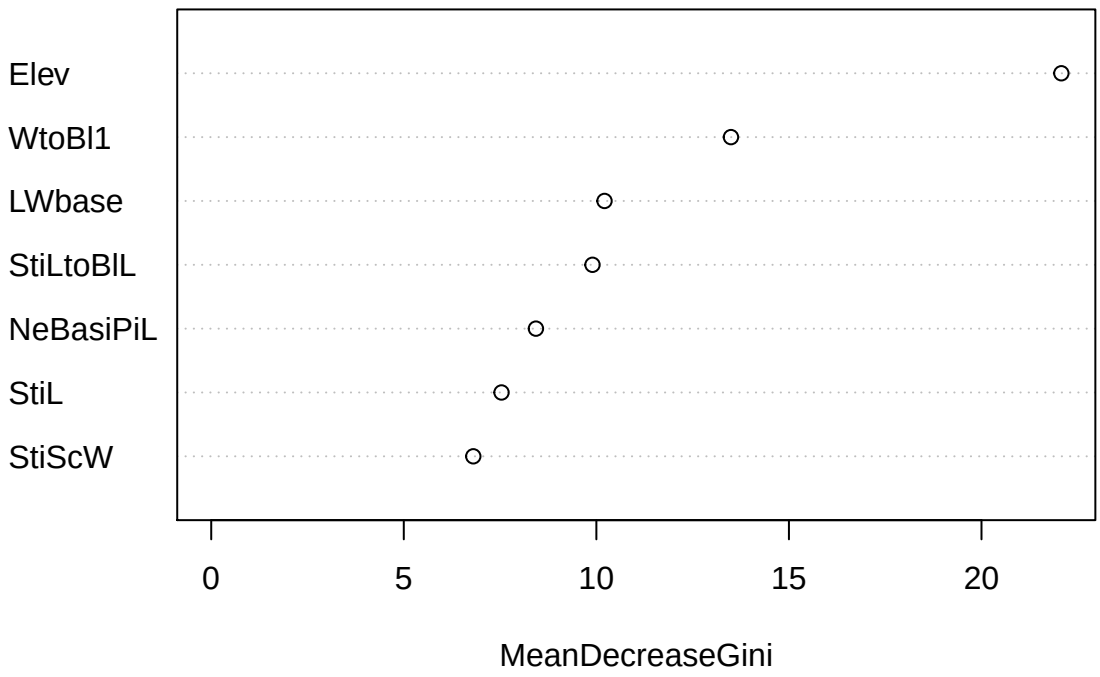

```
#top 8 variables are: StiLtoBlL+WtoBl1+Elev+StiScW+ LWbase +NeBasiPiL+StiL
four_8.rf<-randomForest(SpCode~StiLtoBlL+WtoBl1+Elev+StiScW+ LWbase +NeBasiPiL+StiL,fourSpecies, importa
four_8.rf
```

```
##
## Call:
##  randomForest(formula = SpCode ~ StiLtoBlL + WtoBl1 + Elev + StiScW +      LWbase + NeBasiPiL + StiL
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 2
##
##         OOB estimate of  error rate: 11.5%
## Confusion matrix:
##    1  2  3  4 class.error
## 1 35  1  1  4   0.1463415
## 2  1 10  0  1   0.1666667
## 3  1  0 19  0   0.0500000
## 4  4  0  0 36   0.1000000
```

```
varImpPlot(four_8.rf, type = 1) #Accuracy
```



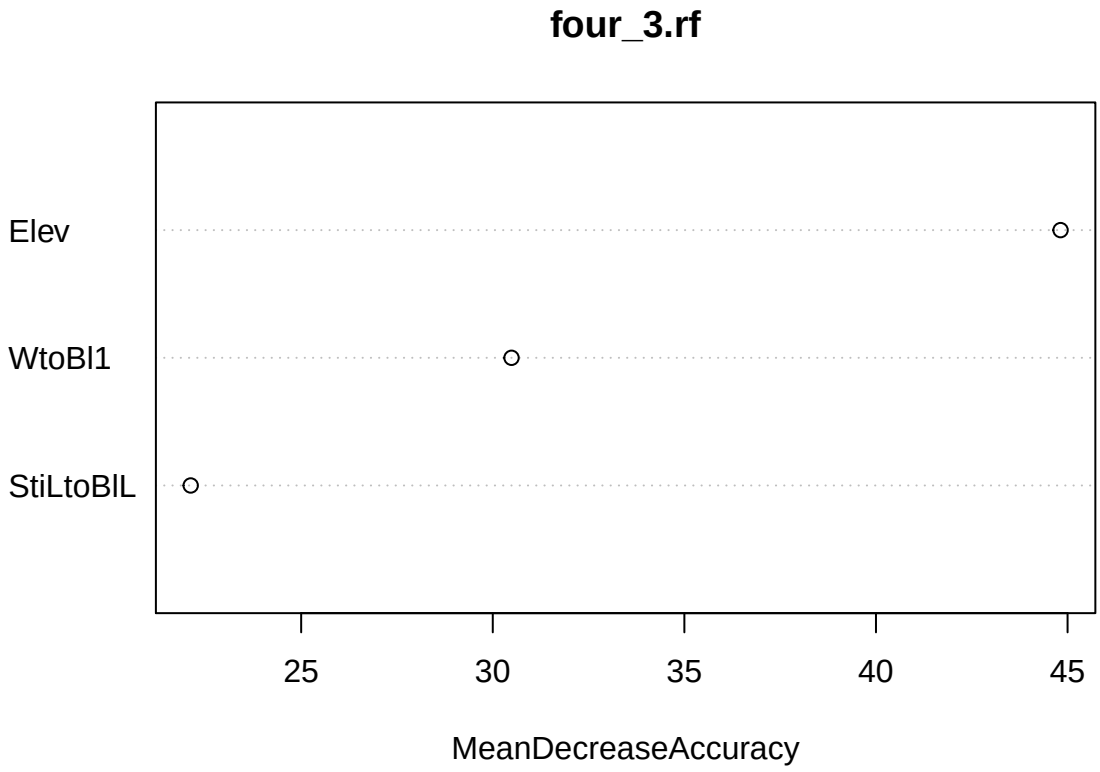

```
varImpPlot(four_8.rf, type = 2) #gini index
```




```
#three variables
four_3.rf<-randomForest(SpCode~StiLtoBlL+WtoBl1+Elev,fourSpecies, importance = TRUE)
four_3.rf
```

```
##
## Call:
##  randomForest(formula = SpCode ~ StiLtoBlL + WtoBl1 + Elev, data = fourSpecies,      importance = TRU
##                Type of random forest: classification
##                      Number of trees: 500
## No. of variables tried at each split: 1
##
##         OOB estimate of  error rate: 26.55%
## Confusion matrix:
##    1 2  3  4 class.error
## 1 31 3  1  6   0.2439024
## 2  3 7  2  0   0.4166667
## 3  4 2 14  0   0.3000000
## 4  7 1  1 31   0.2250000
```

```
varImpPlot(four_3.rf, type = 1) #Accuracy
```




```
varImpPlot(four_3.rf, type = 2) #gini index
```