BaggingForests

Rukh Agha

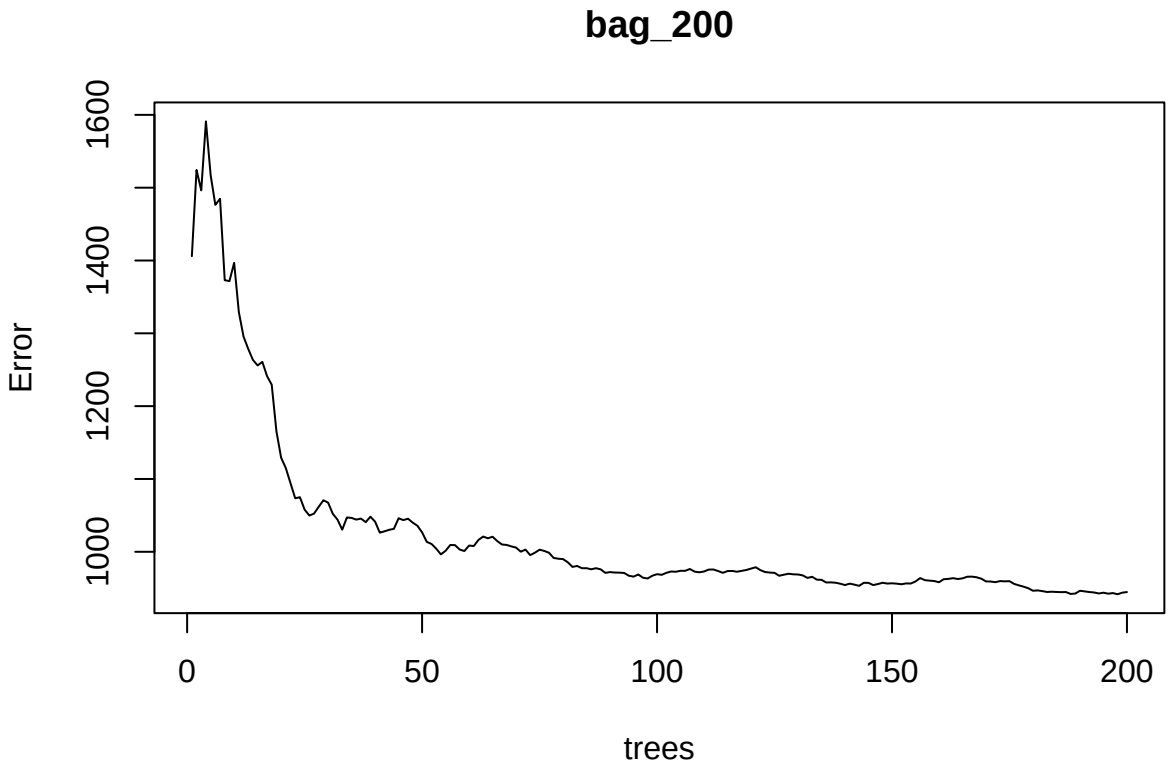
2022-07-31

#Bagging and Random Forests

Bagging - Create random forest with all variables (m=p=19)

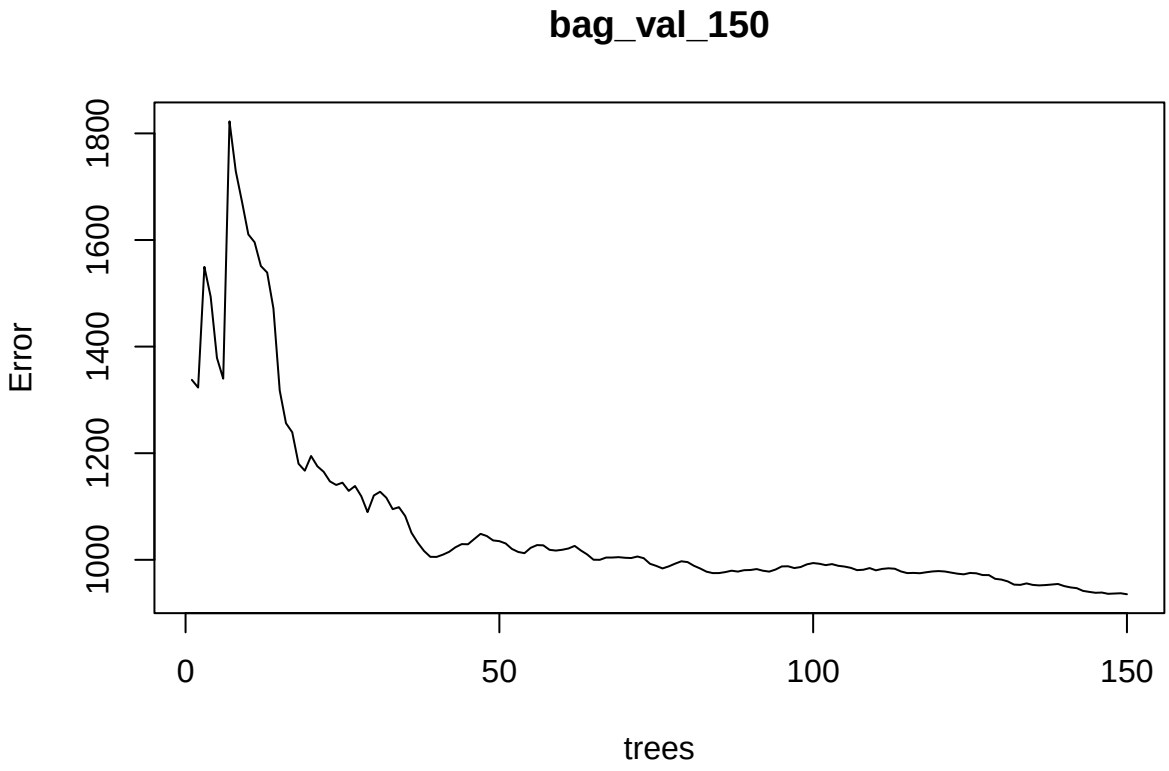
```
## [1] "Creating random forest of 200 trees.."
```

```
##
## Call:
##  randomForest(formula = TripCount ~ ., data = train, ntree = 200,      mtry = 19)
##                Type of random forest: regression
##                      Number of trees: 200
## No. of variables tried at each split: 19
##
##           Mean of squared residuals: 944.6872
##                     % Var explained: 83.48
```



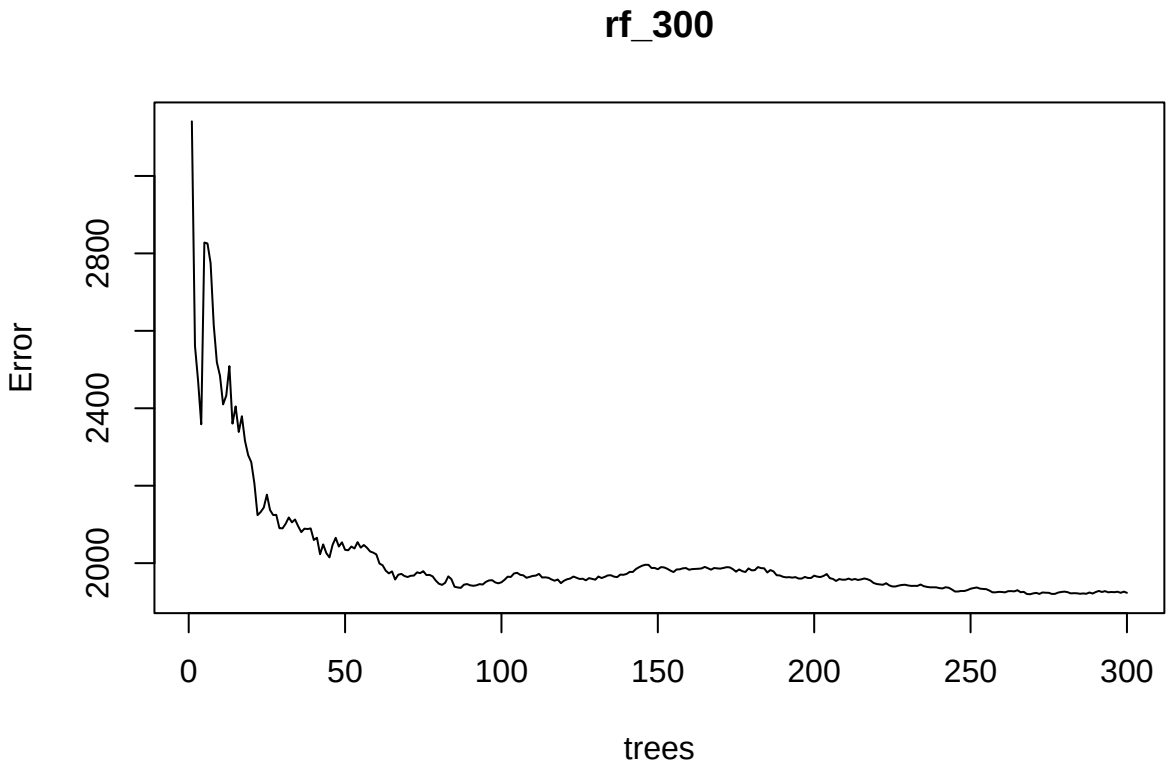
Bagging with validation set

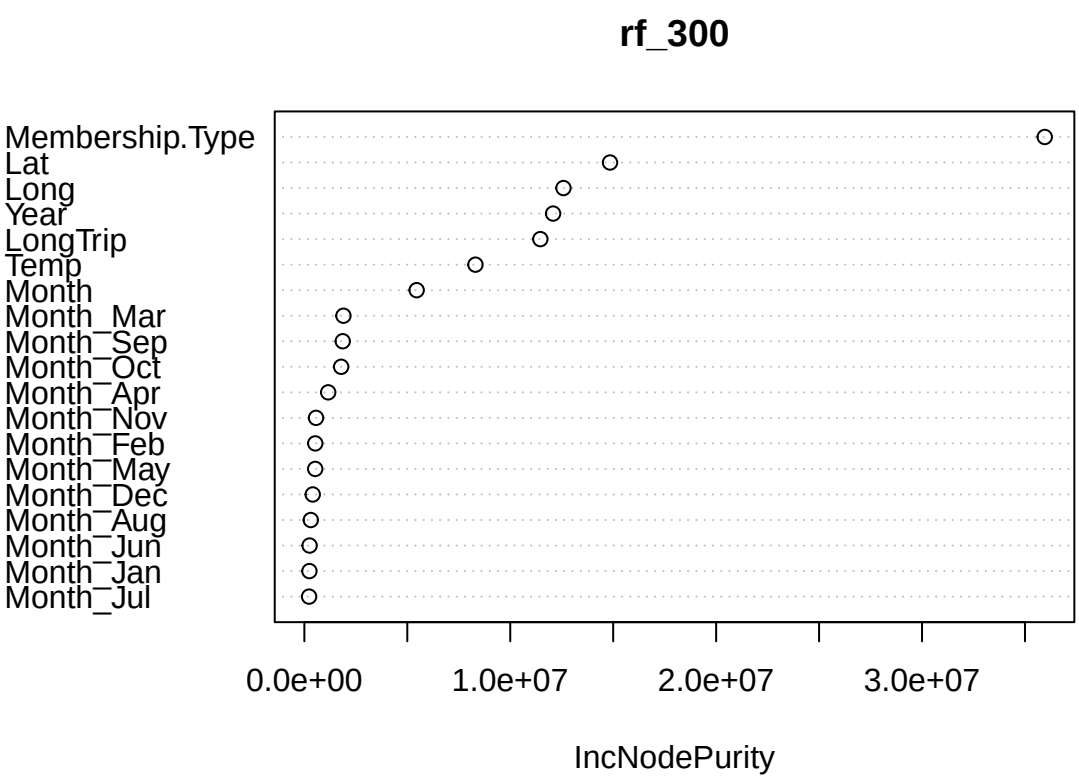
```
## [1] "Creating random forest of 150 trees.."

##
## Call:
##  randomForest(formula = TripCount ~ ., data = train, xtest = pred,      ytest = predY, ntree = 150, r
##                Type of random forest: regression
##                      Number of trees: 150
## No. of variables tried at each split: 19
##
##           Mean of squared residuals: 935.4027
##                     % Var explained: 83.64
##                        Test set MSE: 834.15
##                     % Var explained: 83.97
```



Random Forests

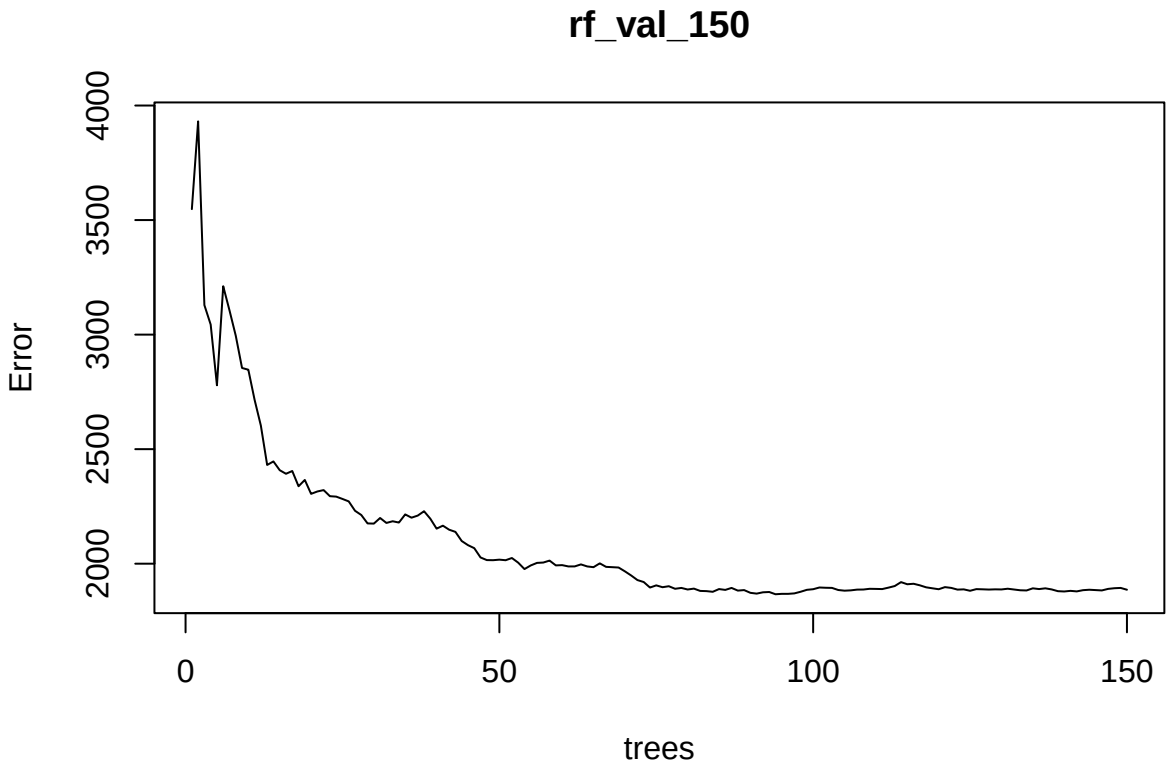
```
## [1] "Creating random forest of 300 trees.."

##
## Call:
##  randomForest(formula = TripCount ~ ., data = train, ntree = 300)
##                Type of random forest: regression
##                      Number of trees: 300
## No. of variables tried at each split: 6
##
##           Mean of squared residuals: 1923.4
##                     % Var explained: 66.36
```

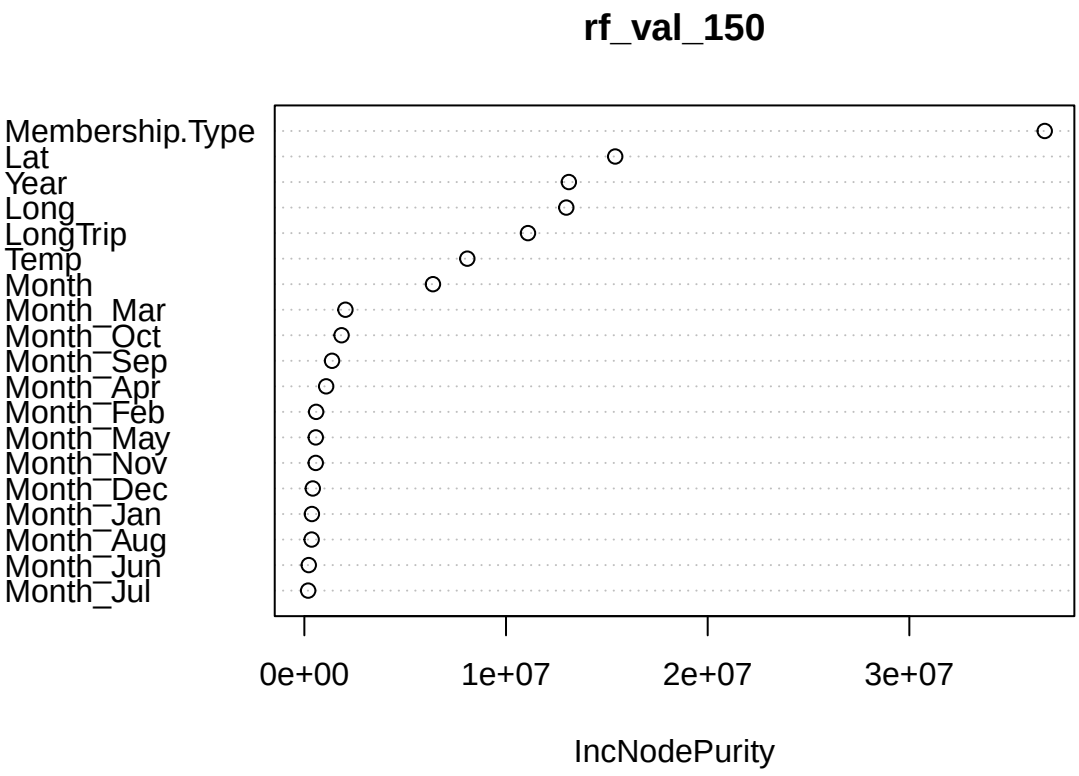

rf_300
Error
2000
2400
2800
0
50
100
150
200
250
300
trees



Random Forest Validation

```
##
## Call:
##  randomForest(formula = TripCount ~ ., data = train, xtest = pred,      ytest = predY, ntree = 150)
##                Type of random forest: regression
##                      Number of trees: 150
## No. of variables tried at each split: 6
##
##           Mean of squared residuals: 1886.264
##                     % Var explained: 67.01
##                        Test set MSE: 1598.5
##                     % Var explained: 69.28
```

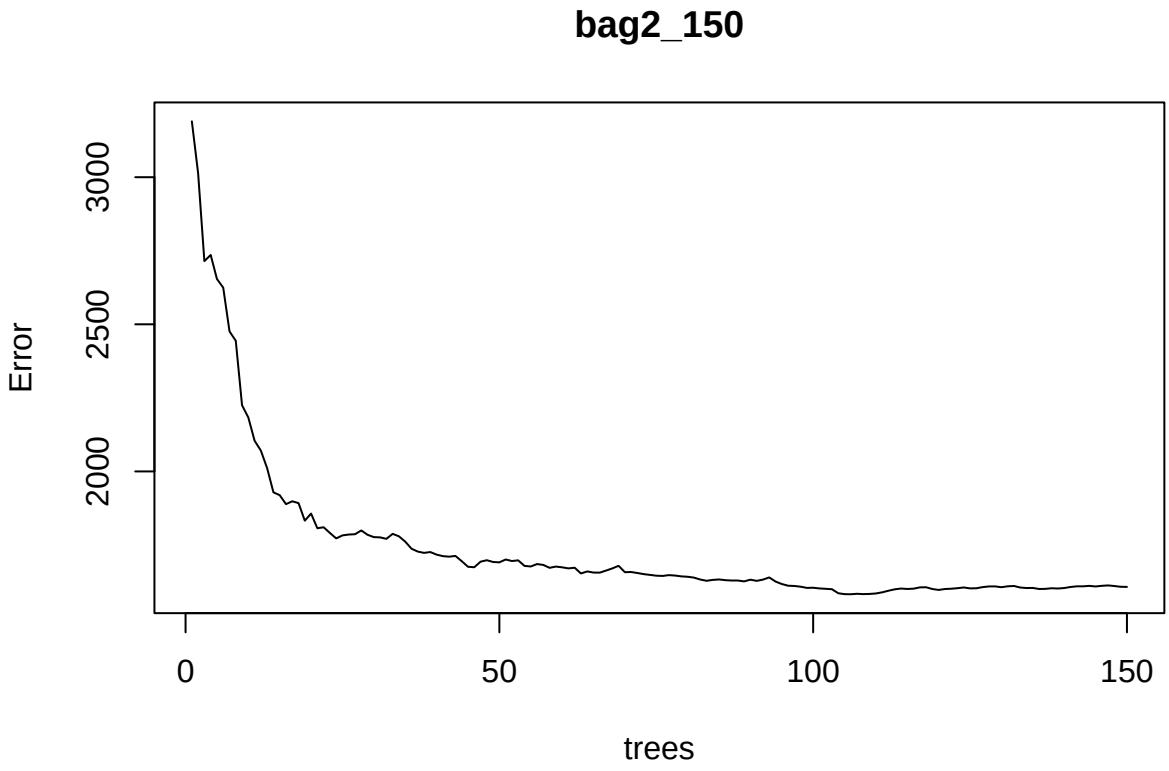

rf_val_150
Error
2000
2500
3000
3500
4000
0
50
100
150
trees

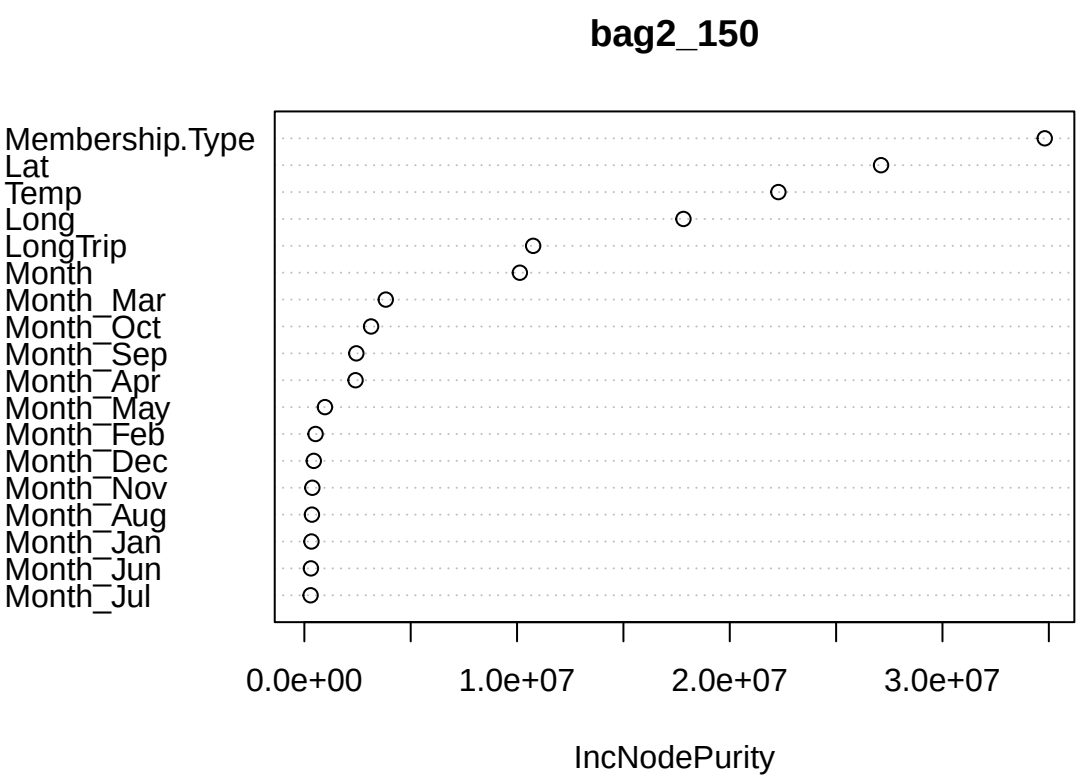


Appendix

Bagging without year

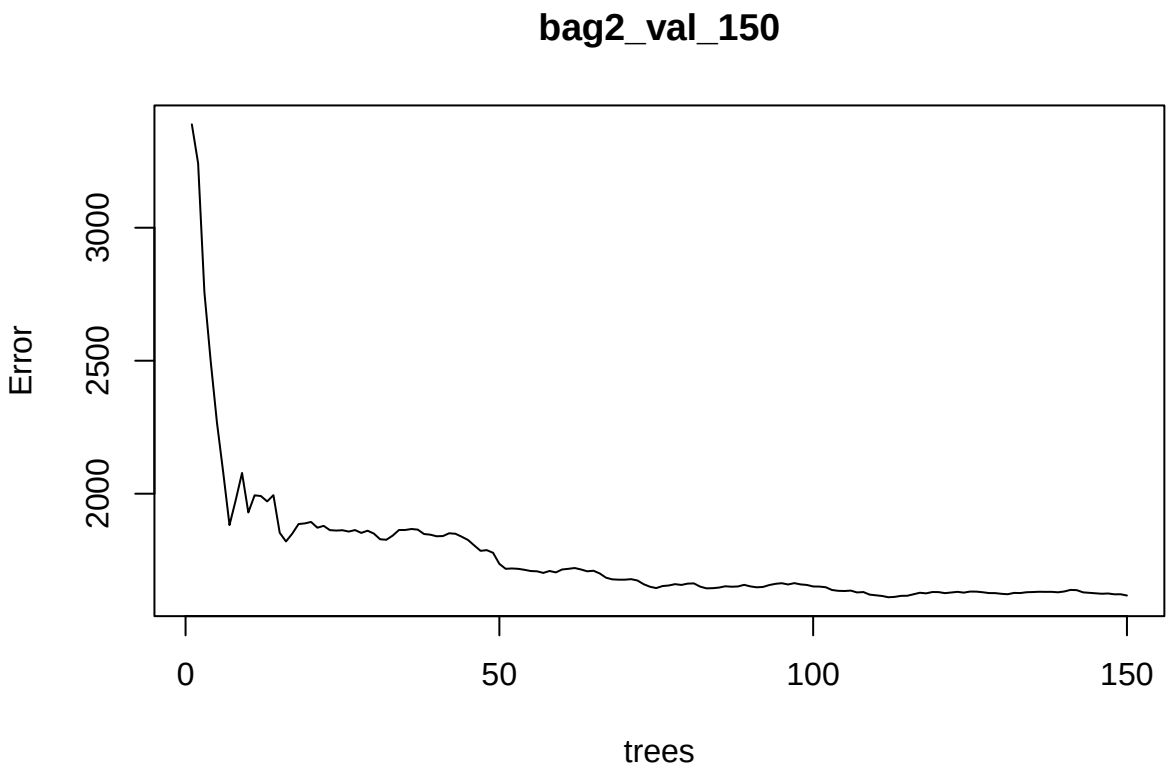
```
## [1] "Creating random forest of 150 trees.."

##
## Call:
##  randomForest(formula = TripCount ~ ., data = train, ntree = 150,      mtry = 18)
##                Type of random forest: regression
##                      Number of trees: 150
## No. of variables tried at each split: 18
##
##           Mean of squared residuals: 1607.479
##                     % Var explained: 71.89
```

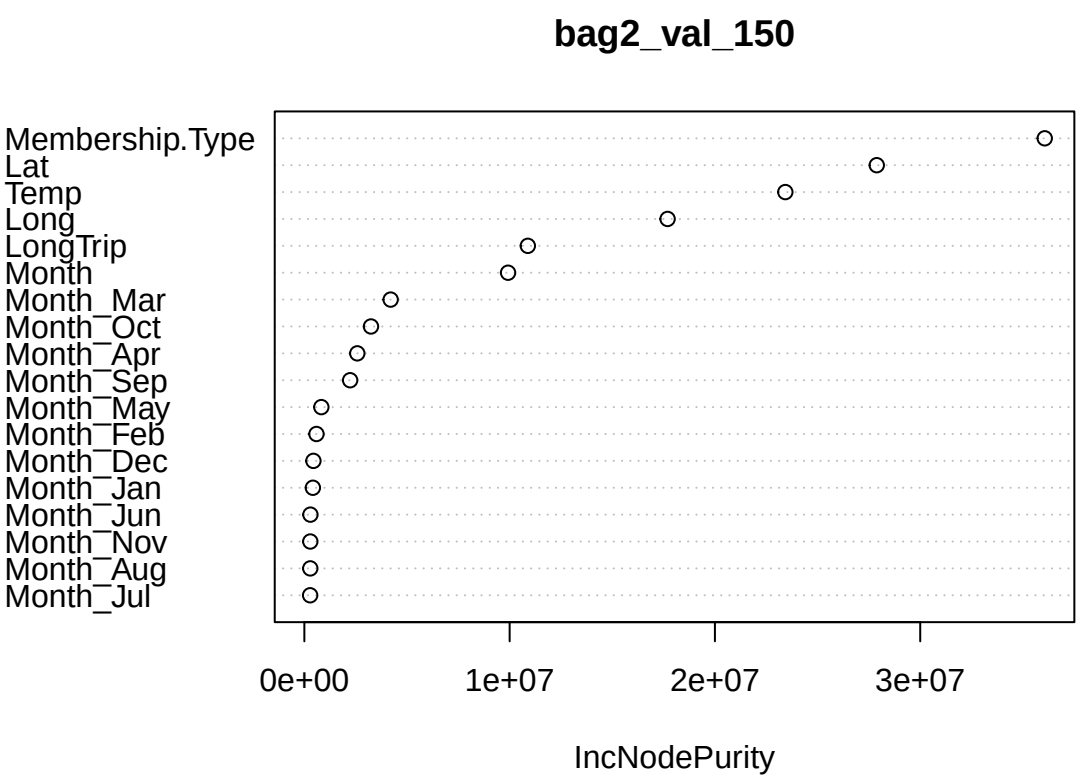

bag2_150
Error
2000
2500
3000
0
50
100
150
trees



Validation set

```
## [1] "Creating random forest of 150 trees.."

##
## Call:
##  randomForest(formula = TripCount ~ ., data = train, xtest = pred,      ytest = predY, ntree = 150, r
##                Type of random forest: regression
##                      Number of trees: 150
## No. of variables tried at each split: 18
##
##           Mean of squared residuals: 1617.19
##                     % Var explained: 71.72
##                        Test set MSE: 1639.21
##                     % Var explained: 68.5
```


bag2_val_150
Error
2000
2500
3000
0
50
100
150
trees


bag2_val_150
Membership.Type
Lat
Temp
Long
LongTrip
Month
Month_Mar
Month_Oct
Month_Apr
Month_Sep
Month_May
Month_Feb
Month_Dec
Month_Jan
Month_Jun
Month_Nov
Month_Aug
Month_Jul
0e+00
1e+07
2e+07
3e+07
IncNodePurity