IE425 Data Mining

Homework 2

10.05.2023

Tevfik Buğra Türker 2019402120

Ömercan Mısırlıoğlu 2020402261

1st Question

a) Partition the dataset using the caTools package into training and test sets where 80% of the observations go into the training set and 20% goes into the test set.

Code:

```
set.seed(425)
split=sample.split(bank$y,SplitRatio=0.8)
banktr=subset(bank,split==TRUE)
bankte=subset(bank,split==FALSE)
```

b) Determine the best random forest (based on the random forest package) by using 10-fold cross validation five times with the caret package on the training set by playing with the mtry and ntree parameters. What are the best values of these two parameters?

Different values for mtry and ntree have been tried. For mtry values {3,4,5,6}, for ntree values {5,10,25,50,100,250,500} have been used. According to the results, best values for parameter mtry and ntree are "6" and "250" respectively.

Code:

```
set.seed(425)
models <- list() # An empty list for different models
ntrees = c(5,10,25,50,100,250,500) # A vector for ntree values
for (i in 1:length(ntrees)){
  ntree <- ntrees[i]
  ctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 1)
  rf_model <- train(y ~ ., data = banktr, metric = "Accuracy", method = "rf
", trControl = ctrl, ntree=ntree,tuneGrid = expand.grid(.mtry = (3:6)), imp
ortance = TRUE)
  models[[i]] = rf_model
}
rf = models[[1]]
rf$finalModel
```

Code Output:

```
[[1]]-> ntree=5
Random Forest
```

```
36169 samples
   16 predictor
    2 classes: 'no', 'yes'

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 32552, 32553, 32552, 32553, 32552, 32552, ...
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
  3     0.8889379  0.1468150
  4     0.8922556  0.2095459
  5     0.8947717  0.2683435
  6     0.8931957  0.3026539

Accuracy was used to select the optimal model using the largest value.
The final value used for the model was mtry = 5.


[[2]] -> ntree=10
Random Forest

36169 samples
   16 predictor
    2 classes: 'no', 'yes'

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 32552, 32552, 32552, 32552, 32553, 32552, ...
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
  3     0.8874727  0.09662731
  4     0.8923665  0.18471461
  5     0.8944956  0.24028799
  6     0.8972324  0.30401092

Accuracy was used to select the optimal model using the largest value.
The final value used for the model was mtry = 6.


[[3]] -> ntree=25
Random Forest

36169 samples
   16 predictor
    2 classes: 'no', 'yes'

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 32551, 32553, 32552, 32552, 32552, 32552, ...
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
  3     0.8891592  0.1126007
  4     0.8936382  0.1937233
  5     0.8961264  0.2437352
  6     0.8977024  0.2807499

Accuracy was used to select the optimal model using the largest value.
```

```
The final value used for the model was mtry = 6.


[[4]] -> ntree=50
Random Forest

36169 samples
   16 predictor
    2 classes: 'no', 'yes'

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 1 times)
Summary of sample sizes: 32552, 32552, 32552, 32552, 32551, 32552, ...
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
  3     0.8876110  0.08392626
  4     0.8944122  0.20220294
  5     0.8960711  0.23205737
  6     0.8981171  0.27879888

Accuracy was used to select the optimal model using the largest value.
The final value used for the model was mtry = 6.


[[5]] -> ntree=100
Random Forest

36169 samples
   16 predictor
    2 classes: 'no', 'yes'

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 1 times)
Summary of sample sizes: 32552, 32552, 32552, 32552, 32552, 32552, ...
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
  3     0.8881915  0.09344364
  4     0.8947993  0.20183108
  5     0.8964859  0.23938732
  6     0.8983659  0.27847490

Accuracy was used to select the optimal model using the largest value.
The final value used for the model was mtry = 6.


[[6]] -> ntree=250
Random Forest

36169 samples
   16 predictor
    2 classes: 'no', 'yes'

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 1 times)
Summary of sample sizes: 32552, 32552, 32552, 32552, 32552, 32552, ...
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
  3     0.8982829  0.2685273
```

```
  4     0.9037850  0.3620405
  5     0.9066881  0.4211082
  6     0.9076834  0.4489147

Accuracy was used to select the optimal model using the largest value.
The final value used for the model was mtry = 6.

[[7]] -> ntree=500
Random Forest

36169 samples
   16 predictor
    2 classes: 'no', 'yes'

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 1 times)
Summary of sample sizes: 32552, 32552, 32552, 32552, 32552, 32552, ...
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
  3     0.8980341  0.2664122
  4     0.9038956  0.3635779
  5     0.9061075  0.4163617
  6     0.9068816  0.4438828

Accuracy was used to select the optimal model using the largest value.
The final value used for the model was mtry = 6.
```

c) What is the out-of-bag accuracy? Comment on which input attributes are important in making predictions.

According to the random forest with **mtry** = **6** and **ntree** = **250**, following results have been obtained. Out of Bag Accuracy = 1 – Out of Bag Error, which means **Out Of Bag Accuracy** is **90.64**%.

```
Call:
 randomForest(x = x, y = y, ntree = ..1, mtry = param$mtry, importance = TR
UE)
               Type of random forest: classification
                     Number of trees: 250
No. of variables tried at each split: 6

        OOB estimate of  error rate: 9.36%
Confusion matrix:
       no  yes class.error
no  31155  783  0.02451625
yes  2604 1627  0.61545734
```

For the importance following code have been used. The attribute that has caused the biggest decrease in Gini is accepted as the most important, followed by the second, third, and fourth biggest decrease in Gini Index. According to the Gini Index, **duration(1818.7)** is the **most** important attribute followed by **age(569.1), balance(567.9),** and **day(503.4).** Values inside the paranthesis show the decrease amount in the Gini.

Code :

```
models[[6]]$finalModel
```

```
importance(models[[6]]$finalModel)
```

Code Output :

```
                          no          yes MeanDecreaseAccuracy MeanDecreaseGini
age                26.91053573  10.19323625           29.0691488       569.086310
jobblue-collar      5.42400628   5.83027359            8.0263153        56.989686
jobentrepreneur    -1.17813142   0.77992023           -0.5317839        24.397775
jobhousemaid        3.30497250  -2.19896409            1.7026405        21.969176
jobmanagement       4.93698099   1.03076760            5.3076629        65.453342
jobretired          3.71862393   2.11239581            5.5901936        32.912582
jobself-employed    2.07331670   0.33128013            1.8203112        29.860074
jobservices         3.38749781  -1.04880787            2.5114252        41.466291
jobstudent         10.76763911   2.01615113           11.5280334        35.218217
jobtechnician       4.70214785  -0.59286626            3.3699118        69.053452
jobunemployed       2.68229413  -0.96892394            1.8356864        31.574437
jobunknown          5.10675272   0.00936845            4.5338691         9.169252
maritalmarried      8.57746735   1.21780110            8.2209262        72.968999
maritalsingle       9.72263444   4.97985281           11.9076617        62.233017
educationsecondary  0.08750818   4.37087778            2.4082434        75.549160
educationtertiary   7.16928587   7.11294528            9.9919392        72.632002
educationunknown    4.32695593   0.46459875            4.0912244        34.771027
defaultyes          0.82509793   1.72569529            1.5695779        12.113001
balance             5.48855305   6.61330833            8.4780548       567.935208
housingyes         24.45564542  18.34530398           30.9062873       149.598637
loanyes             0.09647509  11.93697450            8.0792396        60.187203
contacttelephone    7.61444355   1.50595801            7.7756938        45.873346
contactunknown     32.91728939  10.97005095           34.5301659        94.963828
day                38.88359042   0.49191898           38.9672645       503.439992
monthaug           26.13490784  -0.82901840           26.4189488        71.282483
monthdec           15.74776037  13.11136540           18.8665220        32.778811
monthfeb           21.10607119  -1.89204569           21.0406210        54.312988
monthjan           21.86330436  -5.84681317           21.1884757        40.592916
monthjul           25.90925218  -0.74160197           25.9612577        64.129568
monthjun           23.74579934  -3.90925581           23.7087260        75.104667
monthmar           24.81843146  40.62810980           36.8449070       100.998324
monthmay           18.03667679   4.60066429           18.3948794        64.868899
monthnov           22.74149537  -3.37903630           22.2942820        54.412840
monthoct           25.22877653  19.65430647           29.9015705        78.276366
monthsep           19.03802089  14.56473168           22.3176010        58.455533
duration           85.56249536 138.97330074          130.1016993      1818.755796
campaign           16.46190932   7.86641188           18.7439069       218.769522
pdays              17.09330640  14.74724783           19.9481746       298.343135
previous           11.14335783   9.80088740           11.7045646       159.456261
poutcomeother       2.87343216   4.75165012            4.3771408        26.322315
poutcomesuccess     7.56299514  61.62519403           26.1932879       416.705234
poutcomeunknown    10.95187364   7.00302272           11.2387968        54.476950
```

d) Provide the Confusion Matrix along with sensitivity, specificity, precision, recall, and the F measure on the test set obtained by the best random forest. Does the out-of-bag accuracy provide a good estimate for the accuracy on the test set?

According to the code, the values for the following performance measures are as follows:

Sensitivity (Recall) = 0.9765

Specifity = 0.3563

Precision (Pos. Pred. Value) = 0.9197

F-Measure = 0.9472

Since Out-of-Bag Accuracy is **90.64**% and the accuracy on the test set is **90.39%**, one can say it is a **good estimate** for the accuracy on the test set

Code:

```
predictions = predict(models[[6]], newdata = bankte)
confusionMatrix(predictions, bankte$y)
F_Value = (2*0.9197*0.9765)/(0.9197+0.9765)
```

Code Output:

```
Confusion Matrix and Statistics

          Reference
Prediction   no  yes
       no  7796  681
       yes  188  377

               Accuracy : 0.9039
                 95% CI : (0.8976, 0.9099)
    No Information Rate : 0.883
    P-Value [Acc > NIR] : 1.183e-10

                  Kappa : 0.4171

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.9765
            Specificity : 0.3563
         Pos Pred Value : 0.9197
         Neg Pred Value : 0.6673
             Prevalence : 0.8830
         Detection Rate : 0.8622
   Detection Prevalence : 0.9375
      Balanced Accuracy : 0.6664

       'Positive' Class : no

F_Value = 0.9472493
```

e) Repeat part b with the gradient boosting machine using the caret and gbm packages by playing with the interaction.depth, n.trees, shrinkage, and n.minobsinnode parameters. What are the best values of these four parameters?

Code:

```
set.seed(425)
ctrl1 <- trainControl(method = "cv", number = 10)
gbmGrid=expand.grid(interaction.depth = c(1, 3),
                    n.trees = (1:3)*10,
                    shrinkage = (1:2)*0.1,
                    n.minobsinnode = c(10))
gbm_bank = train(y~., data=banktr,
               method="gbm", metric="Accuracy",verbose = FALSE,
               trControl = ctrl1,tuneGrid = gbmGrid)


max_index = which.max(gbm_bank$results$Accuracy)
max_index # get the index of the case where the accuracy is the largest
```

Code Output:

According to the output **best values** for interaction.depth, n.trees, shrinkage, and n.minobsinnode parameters are as below:

n.trees = 60
interaction.depth = 5
shrinkage = 0.3
n.minobsinnode = 20

```
Stochastic Gradient Boosting

36169 samples
   16 predictor
    2 classes: 'no', 'yes'

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 32552, 32552, 32552, 32552, 32551, 32552, ...
Resampling results across tuning parameters:

  shrinkage  interaction.depth  n.minobsinnode  n.trees  Accuracy   Kappa
  0.1        1                  10               10      0.8852056  0.04172198
  0.1        1                  10               20      0.8873898  0.07924189
  0.1        1                  10               30      0.8903757  0.13387978
  0.1        1                  10               40      0.8920622  0.17193193
  0.1        1                  10               50      0.8950204  0.24393429
  0.1        1                  10               60      0.8967622  0.28074220
  0.1        1                  10               70      0.8976746  0.30115301
  0.1        1                  10               80      0.8984212  0.31721383
  0.1        1                  10               90      0.8997483  0.34348239
  0.1        1                  10              100      0.9005777  0.35975331
  0.1        1                  20               10      0.8852332  0.04245668
  0.1        1                  20               20      0.8873068  0.07870713
  0.1        1                  20               30      0.8894633  0.12221737
  0.1        1                  20               40      0.8921452  0.17344132
  0.1        1                  20               50      0.8951309  0.24798118
  0.1        1                  20               60      0.8964305  0.27571069
  0.1        1                  20               70      0.8975365  0.29722590
  0.1        1                  20               80      0.8994441  0.33111257
  0.1        1                  20               90      0.8995824  0.33969626
  0.1        1                  20              100      0.9004118  0.35491902
  0.1        1                  30               10      0.8852885  0.04360068
  0.1        1                  30               20      0.8869474  0.07117022
  0.1        1                  30               30      0.8897951  0.12973313
  0.1        1                  30               40      0.8918133  0.17042183
  0.1        1                  30               50      0.8951586  0.24891766
  0.1        1                  30               60      0.8965411  0.27731811
  0.1        1                  30               70      0.8974812  0.30056385
  0.1        1                  30               80      0.8991953  0.33117291
  0.1        1                  30               90      0.8995824  0.33968516
  0.1        1                  30              100      0.9003565  0.35383772
  0.1        3                  10               10      0.8860902  0.05701771
  0.1        3                  10               20      0.8977851  0.29071962
  0.1        3                  10               30      0.9009646  0.36840904
  0.1        3                  10               40      0.9019322  0.39050197
  0.1        3                  10               50      0.9018771  0.40171446
  0.1        3                  10               60      0.9029276  0.41540425
  0.1        3                  10               70      0.9033701  0.42531603
  0.1        3                  10               80      0.9037018  0.42951079
  0.1        3                  10               90      0.9043654  0.43659244
  0.1        3                  10              100      0.9048077  0.44181677
  0.1        3                  20               10      0.8862008  0.06497483
  0.1        3                  20               20      0.8971769  0.28380839
  0.1        3                  20               30      0.9009648  0.36459748
  0.1        3                  20               40      0.9015730  0.38842838
  0.1        3                  20               50      0.9022088  0.40244464
  0.1        3                  20               60      0.9027618  0.41141257
  0.1        3                  20               70      0.9030660  0.41900115
  0.1        3                  20               80      0.9035636  0.42653143
  0.1        3                  20               90      0.9038125  0.43214405
  0.1        3                  20              100      0.9043377  0.43838870
  0.1        3                  30               10      0.8844590  0.02707406
  0.1        3                  30               20      0.8971493  0.27974034
  0.1        3                  30               30      0.9014623  0.36905996
```

0.1	3	30	40	0.9017941	0.38987326
0.1	3	30	50	0.9023472	0.40312051
0.1	3	30	60	0.9031766	0.41654072
0.1	3	30	70	0.9035359	0.42486447
0.1	3	30	80	0.9039783	0.43083737
0.1	3	30	90	0.9042548	0.43763633
0.1	3	30	100	0.9041995	0.43907709
0.1	5	10	10	0.8886891	0.10693972
0.1	5	10	20	0.8995822	0.34248218
0.1	5	10	30	0.9037018	0.40805195
0.1	5	10	40	0.9047249	0.43340473
0.1	5	10	50	0.9048078	0.44087056
0.1	5	10	60	0.9056372	0.45152031
0.1	5	10	70	0.9061349	0.45743244
0.1	5	10	80	0.9061901	0.46089097
0.1	5	10	90	0.9064666	0.46517572
0.1	5	10	100	0.9065495	0.46674789
0.1	5	20	10	0.8900990	0.13182306
0.1	5	20	20	0.8999416	0.35101246
0.1	5	20	30	0.9028447	0.40366424
0.1	5	20	40	0.9041166	0.42818106
0.1	5	20	50	0.9048906	0.44052967
0.1	5	20	60	0.9056371	0.45215157
0.1	5	20	70	0.9056924	0.45638374
0.1	5	20	80	0.9063560	0.46273603
0.1	5	20	90	0.9064390	0.46375689
0.1	5	20	100	0.9065497	0.46601652
0.1	5	30	10	0.8876939	0.08903813
0.1	5	30	20	0.9006883	0.34936248
0.1	5	30	30	0.9031765	0.40734982
0.1	5	30	40	0.9037571	0.42438999
0.1	5	30	50	0.9050565	0.44291266
0.1	5	30	60	0.9053883	0.45115723
0.1	5	30	70	0.9056923	0.45610213
0.1	5	30	80	0.9061071	0.46290742
0.1	5	30	90	0.9065218	0.46646164
0.1	5	30	100	0.9069365	0.47050253
0.2	1	10	10	0.8877214	0.08650622
0.2	1	10	20	0.8934446	0.20773456
0.2	1	10	30	0.8968176	0.29071509
0.2	1	10	40	0.8991124	0.33400036
0.2	1	10	50	0.9004671	0.36136822
0.2	1	10	60	0.9020153	0.38500104
0.2	1	10	70	0.9022088	0.39422320
0.2	1	10	80	0.9025959	0.40384771
0.2	1	10	90	0.9028171	0.41025448
0.2	1	10	100	0.9029830	0.41289875
0.2	1	20	10	0.8881916	0.09499372
0.2	1	20	20	0.8933892	0.19511984
0.2	1	20	30	0.8971770	0.28755934
0.2	1	20	40	0.8998587	0.34413608
0.2	1	20	50	0.9009371	0.36359073
0.2	1	20	60	0.9021259	0.38520496
0.2	1	20	70	0.9018771	0.39123815
0.2	1	20	80	0.9020983	0.40035827
0.2	1	20	90	0.9024577	0.40742208
0.2	1	20	100	0.9028724	0.41334928
0.2	1	30	10	0.8881362	0.09197842
0.2	1	30	20	0.8928915	0.18846297
0.2	1	30	30	0.8971218	0.29062475
0.2	1	30	40	0.9000800	0.34324482
0.2	1	30	50	0.9016559	0.36904611
0.2	1	30	60	0.9020153	0.38361778
0.2	1	30	70	0.9021259	0.39367814
0.2	1	30	80	0.9022089	0.40097010
0.2	1	30	90	0.9031765	0.41281635
0.2	1	30	100	0.9032595	0.41587186
0.2	3	10	10	0.8985318	0.31212196
0.2	3	10	20	0.9019601	0.39675900
0.2	3	10	30	0.9030661	0.41549337
0.2	3	10	40	0.9041167	0.43396621
0.2	3	10	50	0.9046973	0.44242229
0.2	3	10	60	0.9046972	0.44543680
0.2	3	10	70	0.9051673	0.44967041
0.2	3	10	80	0.9060520	0.45705050
0.2	3	10	90	0.9059138	0.45960319
0.2	3	10	100	0.9063285	0.46273993

0.2	3	20	10	0.8979235	0.29935250
0.2	3	20	20	0.9022641	0.39510852
0.2	3	20	30	0.9034531	0.41958401
0.2	3	20	40	0.9037018	0.43035497
0.2	3	20	50	0.9043102	0.44154713
0.2	3	20	60	0.9049460	0.44881445
0.2	3	20	70	0.9059414	0.45807516
0.2	3	20	80	0.9059965	0.46190303
0.2	3	20	90	0.9065495	0.46677321
0.2	3	20	100	0.9066878	0.46797248
0.2	3	30	10	0.8980064	0.31726292
0.2	3	30	20	0.9007712	0.38729553
0.2	3	30	30	0.9027341	0.41727918
0.2	3	30	40	0.9037571	0.43080288
0.2	3	30	50	0.9043653	0.44171055
0.2	3	30	60	0.9048077	0.44830483
0.2	3	30	70	0.9051671	0.45199885
0.2	3	30	80	0.9058583	0.45868141
0.2	3	30	90	0.9055819	0.45902389
0.2	3	30	100	0.9061071	0.46273259
0.2	5	10	10	0.9012134	0.36491802
0.2	5	10	20	0.9041165	0.42647583
0.2	5	10	30	0.9045864	0.44588120
0.2	5	10	40	0.9047247	0.45339963
0.2	5	10	50	0.9059965	0.46405263
0.2	5	10	60	0.9064666	0.46970542
0.2	5	10	70	0.9067707	0.47347221
0.2	5	10	80	0.9074619	0.47797288
0.2	5	10	90	0.9074895	0.47951476
0.2	5	10	100	0.9072131	0.47677667
0.2	5	20	10	0.9000800	0.35630956
0.2	5	20	20	0.9039229	0.42809859
0.2	5	20	30	0.9051671	0.44745678
0.2	5	20	40	0.9058860	0.45960158
0.2	5	20	50	0.9070196	0.46920408
0.2	5	20	60	0.9067431	0.47096351
0.2	5	20	70	0.9072131	0.47570103
0.2	5	20	80	0.9071026	0.47806194
0.2	5	20	90	0.9068536	0.47724792
0.2	5	20	100	0.9069919	0.48010170
0.2	5	30	10	0.9011030	0.35972766
0.2	5	30	20	0.9035912	0.41981985
0.2	5	30	30	0.9061070	0.44970867
0.2	5	30	40	0.9067152	0.46328787
0.2	5	30	50	0.9068259	0.46717270
0.2	5	30	60	0.9068536	0.46877584
0.2	5	30	70	0.9072684	0.47277590
0.2	5	30	80	0.9076555	0.47694535
0.2	5	30	90	0.9075726	0.47688968
0.2	5	30	100	0.9076555	0.47927767
0.3	1	10	10	0.8908458	0.14631083
0.3	1	10	20	0.8973982	0.30320180
0.3	1	10	30	0.8998036	0.35488164
0.3	1	10	40	0.9021259	0.39003216
0.3	1	10	50	0.9019323	0.40080356
0.3	1	10	60	0.9022088	0.41053200
0.3	1	10	70	0.9027894	0.41837854
0.3	1	10	80	0.9027618	0.42003405
0.3	1	10	90	0.9029830	0.42427946
0.3	1	10	100	0.9025406	0.42239691
0.3	1	20	10	0.8924492	0.17704281
0.3	1	20	20	0.8973981	0.30371276
0.3	1	20	30	0.9007434	0.36938484
0.3	1	20	40	0.9011857	0.38543024
0.3	1	20	50	0.9017112	0.40257562
0.3	1	20	60	0.9018771	0.40780827
0.3	1	20	70	0.9027617	0.41671677
0.3	1	20	80	0.9031765	0.42293306
0.3	1	20	90	0.9034254	0.42650203
0.3	1	20	100	0.9027895	0.42225565
0.3	1	30	10	0.8907073	0.14785999
0.3	1	30	20	0.8985595	0.31774686
0.3	1	30	30	0.9011583	0.37536874
0.3	1	30	40	0.9017942	0.39351822
0.3	1	30	50	0.9025131	0.40696152
0.3	1	30	60	0.9022918	0.41001853
0.3	1	30	70	0.9028447	0.41676254

0.3	1	30	80	0.9030384	0.42244688
0.3	1	30	90	0.9032319	0.42632128
0.3	1	30	100	0.9034254	0.42734441
0.3	3	10	10	0.9006329	0.37577747
0.3	3	10	20	0.9024024	0.41423359
0.3	3	10	30	0.9046143	0.44153777
0.3	3	10	40	0.9054160	0.45336320
0.3	3	10	50	0.9058585	0.46091277
0.3	3	10	60	0.9050844	0.45660242
0.3	3	10	70	0.9060244	0.46481030
0.3	3	10	80	0.9061074	0.46655716
0.3	3	10	90	0.9065773	0.46887461
0.3	3	10	100	0.9066879	0.46909480
0.3	3	20	10	0.9006605	0.37724775
0.3	3	20	20	0.9026510	0.41396177
0.3	3	20	30	0.9038952	0.43479754
0.3	3	20	40	0.9044205	0.44372090
0.3	3	20	50	0.9054435	0.45623766
0.3	3	20	60	0.9051395	0.45671470
0.3	3	20	70	0.9058860	0.46263762
0.3	3	20	80	0.9061348	0.46488371
0.3	3	20	90	0.9058030	0.46129668
0.3	3	20	100	0.9059690	0.46290402
0.3	3	30	10	0.9014901	0.37961648
0.3	3	30	20	0.9022920	0.42037195
0.3	3	30	30	0.9042826	0.44253299
0.3	3	30	40	0.9050844	0.45266280
0.3	3	30	50	0.9058309	0.46264094
0.3	3	30	60	0.9059691	0.46475800
0.3	3	30	70	0.9064391	0.46806423
0.3	3	30	80	0.9055543	0.46497202
0.3	3	30	90	0.9053608	0.46429262
0.3	3	30	100	0.9061072	0.46767937
0.3	5	10	10	0.9017941	0.40161379
0.3	5	10	20	0.9036188	0.43847339
0.3	5	10	30	0.9050012	0.45553059
0.3	5	10	40	0.9054988	0.46349650
0.3	5	10	50	0.9059965	0.46843007
0.3	5	10	60	0.9059134	0.47110145
0.3	5	10	70	0.9064111	0.47371570
0.3	5	10	80	0.9059964	0.47377218
0.3	5	10	90	0.9054159	0.47167501
0.3	5	10	100	0.9049735	0.46882236
0.3	5	20	10	0.9029000	0.41725013
0.3	5	20	20	0.9052501	0.45297588
0.3	5	20	30	0.9064942	0.46821120
0.3	5	20	40	0.9071301	0.47554531
0.3	5	20	50	0.9075725	0.47976285
0.3	5	20	60	0.9079318	0.48348266
0.3	5	20	70	0.9075171	0.48275568
0.3	5	20	80	0.9077107	0.48498942
0.3	5	20	90	0.9076554	0.48564456
0.3	5	20	100	0.9074066	0.48430690
0.3	5	30	10	0.9022641	0.40432014
0.3	5	30	20	0.9046971	0.44520481
0.3	5	30	30	0.9059136	0.46368012
0.3	5	30	40	0.9067707	0.47075151
0.3	5	30	50	0.9061902	0.47085572
0.3	5	30	60	0.9070197	0.47828650
0.3	5	30	70	0.9064113	0.47458759
0.3	5	30	80	0.9064666	0.47683706
0.3	5	30	90	0.9072684	0.48383436
0.3	5	30	100	0.9072960	0.48422142
0.4	1	10	10	0.8954906	0.27050939
0.4	1	10	20	0.8996930	0.36512208
0.4	1	10	30	0.9015730	0.39700324
0.4	1	10	40	0.9019601	0.41118737
0.4	1	10	50	0.9023471	0.41940561
0.4	1	10	60	0.9024855	0.42176499
0.4	1	10	70	0.9023748	0.42238005
0.4	1	10	80	0.9023748	0.42427520

.
.
.
.
.
.

```
.
.
.
.
.
{Observations between these values are omitted since it is too long and the optimal
values are not here}
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
  0.9        3                 10                50       0.9025129  0.45953533
  0.9        3                 10                60       0.9026789  0.46257075
  0.9        3                 10                70       0.9023471  0.46089036
  0.9        3                 10                80       0.9024854  0.46322027
  0.9        3                 10                90       0.9019322  0.46018984
  0.9        3                 10               100       0.9013792  0.46010117
  0.9        3                 20                10       0.9003011  0.43150087
  0.9        3                 20                20       0.9019321  0.44717477
  0.9        3                 20                30       0.9022366  0.45801165
  0.9        3                 20                40       0.9029277  0.45698582
  0.9        3                 20                50       0.9025407  0.45652027
  0.9        3                 20                60       0.9032595  0.46406615
  0.9        3                 20                70       0.9025959  0.46221742
  0.9        3                 20                80       0.9024301  0.46170572
  0.9        3                 20                90       0.9033147  0.46801044
  0.9        3                 20               100       0.9021537  0.46276819
  0.9        3                 30                10       0.9006608  0.44442477
  0.9        3                 30                20       0.9008267  0.44436417
  0.9        3                 30                30       0.9031491  0.46182655
  0.9        3                 30                40       0.9032596  0.46099772
  0.9        3                 30                50       0.9037572  0.46179921
  0.9        3                 30                60       0.9032319  0.46345974
  0.9        3                 30                70       0.9028171  0.46284907
  0.9        3                 30                80       0.9034805  0.47022420
  0.9        3                 30                90       0.9032593  0.46687523
  0.9        3                 30               100       0.9033425  0.46953195
  0.9        5                 10                10       0.9017387  0.45781559
  0.9        5                 10                20       0.9016557  0.45774364
  0.9        5                 10                30       0.9011305  0.45542615
  0.9        5                 10                40       0.9013517  0.46263129
  0.9        5                 10                50       0.8998863  0.45397523
  0.9        5                 10                60       0.8990569  0.45034468
  0.9        5                 10                70       0.8986423  0.45292715
  0.9        5                 10                80       0.8982552  0.45239001
  0.9        5                 10                90       0.8982827  0.45351164
  0.9        5                 10               100       0.8980063  0.45268844
  0.9        5                 20                10       0.9007987  0.45408702
  0.9        5                 20                20       0.9020428  0.46329126
  0.9        5                 20                30       0.9023470  0.46425489
  0.9        5                 20                40       0.9004668  0.45231336
  0.9        5                 20                50       0.9003010  0.45188791
  0.9        5                 20                60       0.9006880  0.45762105
  0.9        5                 20                70       0.9000521  0.45499946
  0.9        5                 20                80       0.9012410  0.46193877
  0.9        5                 20                90       0.9003840  0.46278517
  0.9        5                 20               100       0.8998587  0.45675002
  0.9        5                 30                10       0.9004393  0.44764105
  0.9        5                 30                20       0.9013794  0.45794358
  0.9        5                 30                30       0.9016559  0.45985053
  0.9        5                 30                40       0.9010752  0.46052684
  0.9        5                 30                50       0.9018770  0.46697851
  0.9        5                 30                60       0.9014346  0.46544205
  0.9        5                 30                70       0.9009647  0.46720035
  0.9        5                 30                80       0.9018218  0.46855815
```

```
  0.9        5                  30                90      0.9013240  0.46883776
  0.9        5                  30               100      0.9010199  0.46664046

Accuracy was used to select the optimal model using the largest value.
The final values used for the model were n.trees = 60, interaction.depth =
 5, shrinkage = 0.3 and n.minobsinnode = 20.
```

f) Provide the Confusion Matrix along with sensitivity, specificity, precision, recall, and the F measure on the test set obtained by the best boosting tree.

Code:

```
best_gbm_bank = gbm_bank
predict_gbm_bank = predict(best_gbm_bank, newdata = bankte)
confusionMatrix(predict_gbm_bank, bankte$y)
```

Code Output:

According to the code, for the best boosting tree the values for the following performance measures are as follows:

Sensitivity (Recall) = 0.9702

Specifity = 0.4045

Precision (Pos. Pred. Value) = 0.9248

F-Measure = 0.9469

```
Confusion Matrix and Statistics

          Reference
Prediction   no  yes
       no  7746  630
       yes  238  428

               Accuracy : 0.904
                 95% CI : (0.8977, 0.91)
    No Information Rate : 0.883
    P-Value [Acc > NIR] : 9.442e-11

                  Kappa : 0.4465

 Mcnemar's Test P-Value : < 2.2e-16

            Sensitivity : 0.9702
            Specificity : 0.4045
         Pos Pred Value : 0.9248
         Neg Pred Value : 0.6426
             Prevalence : 0.8830
         Detection Rate : 0.8567
   Detection Prevalence : 0.9263
      Balanced Accuracy : 0.6874

       'Positive' Class : no
```

2^{st} Question

a) Partition the dataset into training and test sets where 80% of goes into the training set and 20% goes into the test set.

Code:

```
set.seed(425)
split=sample.split(seoul$Rented.Bike.Count,SplitRatio=0.8)
seoultr=subset(seoul,split==TRUE)
seoulte=subset(seoul,split==FALSE)
```

b) Determine the best random forest (based on the random forest package) by using 10-fold cross validation five times with the caret package on the training set by playing with the mtry and ntree parameters. What are the best values of these two parameters?

Different values for **mtry** and **ntree** have been tried. For **mtry** values {3,4,5,6}, for **ntree** values {5,10,25,50,100,250,500} have been used. According to the results, best values for parameter mtry and ntree are “6” and “500” respectively.

Code:

```
set.seed(425)
models <- list() # An empty list for different models to be put into
ntrees = c(5,10,25,50,100,250,500) # A list for ntree values
for (i in 1:length(ntrees)){
  set.seed(425)
  ntree <- ntrees[i]
  ctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 5)
  rf_model <- train(Rented.Bike.Count ~ ., data = seoultr, metric = "RMSE", method
= "rf", trControl = ctrl, ntree=ntree,tuneGrid = expand.grid(.mtry = (3:6)), import
ance = TRUE)
  models[[i]] = rf_model
}
rf = models[[1]]
rf$finalModel
```

Code Output:

```
[[1]] -> ntree=5
Random Forest

7169 samples
  12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 6453, 6453, 6451, 6451, 6451, 6452, ...
Resampling results across tuning parameters:

  mtry  RMSE      Rsquared   MAE
  3     278.1983  0.8290577  183.5701
  4     267.8400  0.8406076  170.8468
  5     263.3504  0.8457519  166.1562
  6     259.1648  0.8507634  162.3520

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 6.
```

```
[[2]] -> ntree=10
Random Forest

7169 samples
  12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 6453, 6453, 6451, 6451, 6451, 6452, ...
Resampling results across tuning parameters:

  mtry  RMSE      Rsquared   MAE
  3     261.3647  0.8509066  173.1999
  4     250.8305  0.8609377  160.9677
  5     248.6693  0.8629406  156.7142
  6     246.5919  0.8651316  154.4505

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 6.

[[3]] -> ntree=25
Random Forest

7169 samples
  12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 6453, 6453, 6451, 6451, 6451, 6452, ...
Resampling results across tuning parameters:

  mtry  RMSE      Rsquared   MAE
  3     255.9711  0.8590109  169.3851
  4     243.2986  0.8698739  156.1816
  5     239.1002  0.8737389  151.1728
  6     239.6797  0.8727539  150.0777

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 5.

[[4]] -> ntree=50
Random Forest

7169 samples
  12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 6453, 6453, 6451, 6451, 6451, 6452, ...
Resampling results across tuning parameters:

  mtry  RMSE      Rsquared   MAE
  3     253.5652  0.8624430  168.0967
  4     240.5815  0.8729422  154.3109
  5     237.7346  0.8753142  150.4617
  6     236.9341  0.8757786  148.7324

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 6.

[[5]] -> ntree=100
Random Forest

7169 samples
```

```
  12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 6453, 6453, 6451, 6451, 6451, 6452, ...
Resampling results across tuning parameters:

  mtry  RMSE      Rsquared   MAE
  3     252.0887  0.8644190  166.8908
  4     239.2202  0.8745958  153.5617
  5     236.7383  0.8764939  149.7047
  6     235.4156  0.8774753  147.5584

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 6.

[[6]] -> ntree=250
Random Forest

7169 samples
  12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 6453, 6453, 6451, 6451, 6451, 6452, ...
Resampling results across tuning parameters:

  mtry  RMSE      Rsquared   MAE
  3     250.5667  0.8663862  166.1809
  4     238.8067  0.8752128  153.2732
  5     236.1089  0.8771753  149.1176
  6     235.0015  0.8779430  147.1396

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 6.

[[7]] -> ntree=500
Random Forest

7169 samples
  12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 1 times)
Summary of sample sizes: 6453, 6453, 6451, 6451, 6451, 6452, ...
Resampling results across tuning parameters:

  mtry  RMSE      Rsquared   MAE
  3     249.8261  0.8672291  165.7805
  4     239.0658  0.8749735  153.6216
  5     235.7659  0.8775535  148.9882
  6     234.7981  0.8781807  146.9703

RMSE was used to select the optimal model using the smallest value.
The final value used for the model was mtry = 6.

Call:
 randomForest(x = x, y = y, ntree = ..1, mtry = param$mtry, importance = TR
UE)
               Type of random forest: regression
                     Number of trees: 500
No. of variables tried at each split: 6

          Mean of squared residuals: 53953.54
                    % Var explained: 87.95
```

c) Comment on which input attributes are important in making predictions.

Code:

```
 imp_rf_seoul = models[[7]]$finalModel
importance(imp_rf_seoul)
```

Code Output:

According to IncNodePurity, **Hour (920352172)** is the most important attribute among all attributes followed by **Temperature (733403312)** and **Huminidity (294873148).**

```
                           %IncMSE      IncNodePurity
Hour                      185.480657     920352172
Temperature.C.             59.843056     733403312
Humidity...                63.100861     294873148
Wind.speed..m.s.           24.517482      67663688
Visibility..10m.           33.775400      84437355
Dew.point.temperature.C.   28.568475     196014523
Solar.Radiation..MJ.m2.    47.696008     228906317
Rainfall.mm.               52.903182     117637313
Snowfall..cm.              18.156318       2406530
SeasonsSpring              46.979365      27764550
SeasonsSummer               8.665176      24294378
SeasonsWinter              35.745767     212830884
HolidayNo Holiday          29.159877       9163431
Functioning.DayYes        150.513164     242708465
```

d) Make predictions in the test set and report the root mean square error rate and mean absolute error.

Code:

```
predictions_rf = predict(models[[7]], newdata = seoulte)

rmse_seoul = rmse(actual = seoulte$Rented.Bike.Count,predicted = prediction
s_rf) # RMSE
mae_seoul = mae(actual = seoulte$Rented.Bike.Count, predicted = predictions
_rf) #MAE
paste("RMSE:", rmse_seoul)
paste("MAE:", mae_seoul)
```

Code Output:

According to the code, these are the RMSE and MAE.

```
[1] "RMSE: 194.481670530118"
[1] "MAE: 121.739224009736"
```

e) Repeat part b with the gradient boosting using the caret and gbm packages by playing with the interaction.depth, n.trees, shrinkage, and n.minobsinnode parameters. What are the best values of these four parameters?

Code:

```
set.seed(425)
ctrl1 <- trainControl(method = "repeatedcv", number = 10, repeats = 5)
gbmGrid=expand.grid(interaction.depth = c(1, 3, 5),
                    n.trees = (1:10)*10,
                    shrinkage = (1:9)*0.1,
                    n.minobsinnode = c(10, 20, 30))
```

```
gbm_seoul = train(Rented.Bike.Count~., data=seoultr,
                  method="gbm", metric="RMSE",verbose = FALSE,
                  trControl = ctrl1,tuneGrid = gbmGrid)


gmbseoul
```

Code Output:

According to the output **best values** for interaction.depth, n.trees, shrinkage, and n.minobsinnode parameters are as below:

n.trees = 100
interaction.depth = 5
shrinkage = **0.3**
n.minobsinnode = 30

```
Stochastic Gradient Boosting

7169 samples
  12 predictor

No pre-processing
Resampling: Cross-Validated (10 fold, repeated 5 times)
Summary of sample sizes: 6453, 6453, 6451, 6451, 6451, 6452, ...
Resampling results across tuning parameters:

  shrinkage  interaction.depth  n.minobsinnode  n.trees  RMSE      Rsquared   MAE
  0.1        1                  10               10      568.5092  0.4419728  443.3075
  0.1        1                  10               20      518.0946  0.5142223  392.5379
  0.1        1                  10               30      483.4236  0.5774162  360.6806
  0.1        1                  10               40      457.6408  0.6071974  336.7740
  0.1        1                  10               50      438.8790  0.6257394  321.6159
  0.1        1                  10               60      425.5342  0.6354363  311.3819
  0.1        1                  10               70      415.8759  0.6425543  305.4463
  0.1        1                  10               80      408.2165  0.6515725  300.9271
  0.1        1                  10               90      401.9217  0.6601649  297.3670
  0.1        1                  10              100      396.5499  0.6667141  294.4890
  0.1        1                  20               10      568.8224  0.4431743  443.3770
  0.1        1                  20               20      518.6277  0.5151975  393.2294
  0.1        1                  20               30      483.9491  0.5769096  361.0228
  0.1        1                  20               40      457.7748  0.6082702  337.5381
  0.1        1                  20               50      438.8859  0.6257424  321.7782
  0.1        1                  20               60      425.5997  0.6352118  311.8298
  0.1        1                  20               70      415.8154  0.6422158  305.7380
  0.1        1                  20               80      408.0669  0.6520054  300.9789
  0.1        1                  20               90      401.7819  0.6597044  297.4534
  0.1        1                  20              100      396.2124  0.6673294  294.3016
  0.1        1                  30               10      568.7679  0.4433394  443.7428
  0.1        1                  30               20      518.7349  0.5138442  392.6842
  0.1        1                  30               30      484.0319  0.5776617  360.4722
  0.1        1                  30               40      458.2974  0.6067892  337.8746
  0.1        1                  30               50      439.5570  0.6245830  322.2754
  0.1        1                  30               60      425.8236  0.6344376  311.4094
  0.1        1                  30               70      416.3464  0.6417934  305.3904
  0.1        1                  30               80      408.4666  0.6510482  300.9863
  0.1        1                  30               90      402.0187  0.6593855  297.3132
  0.1        1                  30              100      396.6309  0.6657841  294.3491
  0.1        3                  10               10      468.6294  0.6741138  361.1043
  0.1        3                  10               20      384.4661  0.7384273  286.6010
  0.1        3                  10               30      343.1153  0.7690779  249.1971
  0.1        3                  10               40      320.2387  0.7895300  227.4261
  0.1        3                  10               50      305.5685  0.8036726  213.3068
  0.1        3                  10               60      294.9616  0.8147415  203.9494
  0.1        3                  10               70      288.0510  0.8217223  197.7109
  0.1        3                  10               80      282.1675  0.8279210  192.9101
  0.1        3                  10               90      277.1619  0.8331281  189.2341
  0.1        3                  10              100      273.0905  0.8376325  186.0261
  0.1        3                  20               10      467.9894  0.6780297  360.7274
  0.1        3                  20               20      383.4390  0.7385459  285.8354
  0.1        3                  20               30      341.9750  0.7707192  248.1718
  0.1        3                  20               40      319.7010  0.7897091  227.3921
  0.1        3                  20               50      305.3638  0.8037593  213.6705
  0.1        3                  20               60      295.3406  0.8140277  204.3807
  0.1        3                  20               70      288.1187  0.8215780  198.3068
  0.1        3                  20               80      281.7992  0.8284382  193.2546
  0.1        3                  20               90      276.8454  0.8338354  189.2405
  0.1        3                  20              100      272.7321  0.8380098  186.1082
  0.1        3                  30               10      468.7495  0.6716021  360.8330
```

0.1	3	30	20	385.0350	0.7349757	286.4197
0.1	3	30	30	344.0764	0.7662878	249.5702
0.1	3	30	40	321.2220	0.7868162	228.3025
0.1	3	30	50	306.9295	0.8012363	214.5067
0.1	3	30	60	297.0317	0.8116984	205.3270
0.1	3	30	70	289.2026	0.8199779	198.6126
0.1	3	30	80	283.0356	0.8264724	193.8779
0.1	3	30	90	278.1023	0.8319308	189.9012
0.1	3	30	100	274.4198	0.8358321	186.6797
0.1	5	10	10	436.1472	0.7111602	335.4677
0.1	5	10	20	346.9986	0.7782795	255.6780
0.1	5	10	30	306.5858	0.8086685	218.4492
0.1	5	10	40	285.2793	0.8283196	197.7754
0.1	5	10	50	272.5526	0.8398614	186.0453
0.1	5	10	60	264.3920	0.8475406	178.7135
0.1	5	10	70	258.9555	0.8529541	173.9881
0.1	5	10	80	255.2534	0.8565375	170.7682
0.1	5	10	90	252.4676	0.8592915	168.5586
0.1	5	10	100	250.4188	0.8613361	166.6521
0.1	5	20	10	435.5112	0.7120544	335.3414
0.1	5	20	20	346.4977	0.7776413	255.4861
0.1	5	20	30	307.0450	0.8081078	219.1912
0.1	5	20	40	286.6380	0.8261471	200.2462
0.1	5	20	50	273.5727	0.8386523	187.4970
0.1	5	20	60	265.9781	0.8456984	180.0202
0.1	5	20	70	260.5863	0.8510797	175.1704
0.1	5	20	80	256.7419	0.8550034	171.6123
0.1	5	20	90	253.7252	0.8579645	169.0178
0.1	5	20	100	251.8460	0.8598704	167.3498
0.1	5	30	10	437.3284	0.7077260	335.6563
0.1	5	30	20	346.5320	0.7777652	254.8414
0.1	5	30	30	306.9877	0.8074910	218.6000
0.1	5	30	40	287.0730	0.8254164	199.2016
0.1	5	30	50	275.0180	0.8369858	188.2009
0.1	5	30	60	266.8436	0.8447406	180.5380
0.1	5	30	70	261.5779	0.8498437	176.0466
0.1	5	30	80	257.7681	0.8536514	172.9045
0.1	5	30	90	254.7539	0.8566039	170.5188
0.1	5	30	100	252.5458	0.8589072	168.9027
0.2	1	10	10	516.3446	0.5036968	389.1800
0.2	1	10	20	455.0636	0.6021966	335.0793
0.2	1	10	30	423.4586	0.6324176	310.8772
0.2	1	10	40	407.2039	0.6506468	300.5378
0.2	1	10	50	395.5370	0.6661053	293.9449
0.2	1	10	60	387.5704	0.6762697	288.9852
0.2	1	10	70	381.4456	0.6843957	285.0048
0.2	1	10	80	376.2898	0.6926783	281.3111
0.2	1	10	90	371.5579	0.6990273	278.2055
0.2	1	10	100	367.2429	0.7054269	275.2195
0.2	1	20	10	515.5348	0.5097047	388.7593
0.2	1	20	20	454.8239	0.6084019	334.6025
0.2	1	20	30	423.4298	0.6351471	309.6342
0.2	1	20	40	406.6105	0.6522744	299.9096
0.2	1	20	50	395.1761	0.6667781	293.7156
0.2	1	20	60	387.1074	0.6775634	288.5329
0.2	1	20	70	380.5758	0.6865814	284.3542
0.2	1	20	80	375.3586	0.6945314	280.3496
0.2	1	20	90	370.9873	0.7005004	277.3985
0.2	1	20	100	367.0226	0.7060063	274.7366
0.2	1	30	10	515.3544	0.5026172	387.3057
0.2	1	30	20	455.2810	0.6056248	334.7540
0.2	1	30	30	423.7090	0.6315992	310.6740
0.2	1	30	40	406.9826	0.6517284	300.4575
0.2	1	30	50	395.8209	0.6650747	294.5606
0.2	1	30	60	387.7578	0.6763323	289.4433
0.2	1	30	70	381.3716	0.6850861	285.1845
0.2	1	30	80	376.0029	0.6925886	281.1098
0.2	1	30	90	371.3517	0.6997619	277.6332
0.2	1	30	100	367.4418	0.7057337	275.0240
0.2	3	10	10	380.9720	0.7361517	282.6448
0.2	3	10	20	320.5067	0.7851906	227.3864
0.2	3	10	30	296.4773	0.8109546	205.3927
0.2	3	10	40	283.5922	0.8250510	194.3742
0.2	3	10	50	275.5562	0.8339071	188.2115
0.2	3	10	60	268.8614	0.8410198	183.5747
0.2	3	10	70	264.0749	0.8460543	179.9391
0.2	3	10	80	260.8953	0.8495530	177.6283
0.2	3	10	90	258.5535	0.8520898	175.7824
0.2	3	10	100	256.9831	0.8537428	174.3418
0.2	3	20	10	382.9008	0.7311878	283.2739
0.2	3	20	20	320.9164	0.7841497	226.8410
0.2	3	20	30	297.4806	0.8097388	205.6665
0.2	3	20	40	283.5851	0.8249642	194.3061
0.2	3	20	50	275.4905	0.8335545	187.8374
0.2	3	20	60	270.4895	0.8389105	183.9013
0.2	3	20	70	266.4813	0.8432270	180.5146
0.2	3	20	80	263.0767	0.8468240	178.3992

0.2	3	20	90	260.6972	0.8495950	176.7553
0.2	3	20	100	259.2002	0.8512215	175.4915
0.2	3	30	10	381.1998	0.7355104	283.4526
0.2	3	30	20	320.8000	0.7852905	228.2870
0.2	3	30	30	296.1814	0.8116769	205.4124
0.2	3	30	40	283.6856	0.8251521	194.7412
0.2	3	30	50	274.4530	0.8352894	187.6274
0.2	3	30	60	268.4927	0.8415593	182.9135
0.2	3	30	70	264.5767	0.8456121	179.7233
0.2	3	30	80	260.8807	0.8494874	176.9786
0.2	3	30	90	258.3036	0.8524132	175.1478
0.2	3	30	100	256.5448	0.8542388	173.7392
0.2	5	10	10	345.3865	0.7720324	253.1349
0.2	5	10	20	288.4775	0.8220038	200.4792
0.2	5	10	30	268.4584	0.8420394	182.2838
0.2	5	10	40	260.3410	0.8505387	175.6048
0.2	5	10	50	255.7235	0.8551464	171.7163
0.2	5	10	60	252.0607	0.8590767	168.9617
0.2	5	10	70	249.4786	0.8618745	166.8972
0.2	5	10	80	248.5147	0.8629152	166.0000
0.2	5	10	90	247.4436	0.8641567	164.9306
0.2	5	10	100	246.3159	0.8653459	164.0561
0.2	5	20	10	345.1541	0.7709682	252.1155
0.2	5	20	20	287.9534	0.8225892	199.6525
0.2	5	20	30	268.2839	0.8420813	182.0608
0.2	5	20	40	260.2860	0.8504126	174.8765
0.2	5	20	50	254.9742	0.8561658	171.0028
0.2	5	20	60	252.1723	0.8589703	168.8576
0.2	5	20	70	250.0232	0.8613281	167.1032
0.2	5	20	80	248.2814	0.8632419	165.9319
0.2	5	20	90	247.3830	0.8641683	165.3730
0.2	5	20	100	246.2548	0.8654935	164.5187
0.2	5	30	10	344.1870	0.7744014	252.0921
0.2	5	30	20	287.6978	0.8227371	199.8241
0.2	5	30	30	268.8606	0.8413942	182.8192
0.2	5	30	40	259.8363	0.8507559	174.7714
0.2	5	30	50	255.3314	0.8555882	171.1168
0.2	5	30	60	252.1785	0.8589390	168.8593
0.2	5	30	70	250.0367	0.8612403	167.0336
0.2	5	30	80	248.4250	0.8629805	166.1621
0.2	5	30	90	246.6806	0.8648787	165.1029
0.2	5	30	100	245.8454	0.8657244	164.5782
0.3	1	10	10	476.2824	0.5733762	352.4708
0.3	1	10	20	421.2036	0.6330582	308.5621
0.3	1	10	30	399.2291	0.6590769	296.2783
0.3	1	10	40	386.0061	0.6763915	288.4453
0.3	1	10	50	377.4627	0.6896493	281.7990
0.3	1	10	60	370.3263	0.7006162	276.8112
0.3	1	10	70	364.2629	0.7099349	272.5144
0.3	1	10	80	359.6779	0.7163377	269.1597
0.3	1	10	90	355.4006	0.7226430	265.9832
0.3	1	10	100	351.7865	0.7270685	263.7571
0.3	1	20	10	477.6938	0.5734814	355.8297
0.3	1	20	20	421.8929	0.6310770	310.1320
0.3	1	20	30	398.9319	0.6602035	296.7591
0.3	1	20	40	386.7083	0.6759093	289.6551
0.3	1	20	50	377.9545	0.6886170	283.1851
0.3	1	20	60	371.0915	0.6986042	278.5910
0.3	1	20	70	365.3702	0.7073249	274.2391
0.3	1	20	80	360.4243	0.7144319	271.0496
0.3	1	20	90	356.2768	0.7212398	267.4506
0.3	1	20	100	352.8377	0.7258157	264.4846
0.3	1	30	10	477.0164	0.5790398	355.4579
0.3	1	30	20	421.7064	0.6305885	309.6514
0.3	1	30	30	399.7880	0.6586603	296.5110
0.3	1	30	40	386.7707	0.6754341	289.1116
0.3	1	30	50	378.1112	0.6883789	282.3992
0.3	1	30	60	371.3434	0.6989377	277.5082
0.3	1	30	70	365.4693	0.7079717	273.0489
0.3	1	30	80	360.4760	0.7144870	270.2726
0.3	1	30	90	356.1181	0.7209399	266.9961
0.3	1	30	100	352.5712	0.7259847	264.4571
0.3	3	10	10	342.5067	0.7569085	246.5126
0.3	3	10	20	299.8956	0.8047347	207.7056
0.3	3	10	30	281.2079	0.8262937	192.7573
0.3	3	10	40	271.9338	0.8367760	185.1214
0.3	3	10	50	265.9393	0.8434322	180.8924
0.3	3	10	60	262.5353	0.8471670	178.1266
0.3	3	10	70	259.9484	0.8500339	176.1370
0.3	3	10	80	258.7988	0.8512482	174.9085
0.3	3	10	90	257.1369	0.8531713	173.9602
0.3	3	10	100	255.7512	0.8546734	172.7162
0.3	3	20	10	340.3687	0.7602248	244.4894
0.3	3	20	20	298.5070	0.8067686	206.2120
0.3	3	20	30	282.0718	0.8252268	192.2777
0.3	3	20	40	270.1723	0.8386774	184.4561
0.3	3	20	50	264.5141	0.8450744	180.2655

0.3	3	20	60	261.0815	0.8488716	177.3511
0.3	3	20	70	258.3065	0.8520406	175.6857
0.3	3	20	80	256.4577	0.8540150	174.3368
0.3	3	20	90	254.8231	0.8558016	173.1226
0.3	3	20	100	254.2311	0.8565113	172.6010
0.3	3	30	10	342.1582	0.7569863	245.7073
0.3	3	30	20	298.9005	0.8055035	207.1547
0.3	3	30	30	281.7518	0.8253425	192.9408
0.3	3	30	40	270.2863	0.8382152	184.4144
0.3	3	30	50	265.0610	0.8442835	180.0134
0.3	3	30	60	261.2245	0.8484196	177.2545
0.3	3	30	70	258.7937	0.8512088	175.7713
0.3	3	30	80	256.6838	0.8535702	174.4423
0.3	3	30	90	254.7433	0.8556661	173.0011
0.3	3	30	100	254.0709	0.8564238	172.4847
0.3	5	10	10	305.8266	0.8021169	215.3512
0.3	5	10	20	269.8402	0.8393780	183.0904
0.3	5	10	30	259.4852	0.8506942	174.0228
0.3	5	10	40	254.2636	0.8563603	170.2388
0.3	5	10	50	251.3073	0.8596550	167.4294
0.3	5	10	60	250.1272	0.8609116	166.5624
0.3	5	10	70	248.7976	0.8623301	165.7568
0.3	5	10	80	247.4885	0.8639346	164.9220
0.3	5	10	90	247.1271	0.8643476	164.5712
0.3	5	10	100	245.8803	0.8655922	163.7537
0.3	5	20	10	309.1423	0.7967650	218.7248
0.3	5	20	20	271.6079	0.8374817	185.0836
0.3	5	20	30	260.6796	0.8492811	176.0001
0.3	5	20	40	255.7305	0.8546663	172.0668
0.3	5	20	50	252.6902	0.8582364	170.2940
0.3	5	20	60	250.5708	0.8606072	169.1864
0.3	5	20	70	248.1665	0.8631297	167.6109
0.3	5	20	80	246.8796	0.8646342	166.3848
0.3	5	20	90	247.5761	0.8639154	166.8564
0.3	5	20	100	247.0673	0.8645021	166.3558
0.3	5	30	10	306.1366	0.8026213	217.1229
0.3	5	30	20	269.2117	0.8405079	183.4128
0.3	5	30	30	257.1396	0.8535573	173.6056
0.3	5	30	40	252.6880	0.8583628	169.9336
0.3	5	30	50	249.2904	0.8620274	167.5356
0.3	5	30	60	247.5626	0.8639105	166.4541
0.3	5	30	70	246.5981	0.8649799	165.5766
0.3	5	30	80	246.8312	0.8648271	165.4265
0.3	5	30	90	246.3537	0.8652925	165.2820
0.3	5	30	100	245.6844	0.8660550	165.2470
0.4	1	10	10	448.9503	0.5937728	328.0142
0.4	1	10	20	405.6422	0.6465640	300.0308
0.4	1	10	30	387.0987	0.6738514	289.3946
0.4	1	10	40	376.2231	0.6897787	282.1012
0.4	1	10	50	368.0406	0.7022576	275.9034
0.4	1	10	60	361.0172	0.7127335	271.1833
0.4	1	10	70	355.8664	0.7201242	267.8563
0.4	1	10	80	351.5623	0.7266276	264.4109
0.4	1	10	90	347.8672	0.7319941	261.6085
0.4	1	10	100	345.3168	0.7355106	259.8974
0.4	1	20	10	450.8636	0.5903337	329.7706
0.4	1	20	20	406.7135	0.6455927	302.5901
0.4	1	20	30	388.7650	0.6708828	291.0906
0.4	1	20	40	377.8244	0.6871447	283.6298
0.4	1	20	50	369.4202	0.6999662	277.7901
0.4	1	20	60	362.2911	0.7113119	272.6602
0.4	1	20	70	356.5547	0.7194046	268.9175
0.4	1	20	80	351.7311	0.7267233	265.1571
0.4	1	20	90	347.7709	0.7326471	261.7485
0.4	1	20	100	344.8713	0.7365371	259.6415
0.4	1	30	10	448.9452	0.5966646	327.5768

.
.
.
.
.
.
.
.
.

Observations here are omitted in the report since the optimal values are not here

.
.
.
.
.
.
.

0.8	1	30	90	338.8175	0.7442568	255.0691
0.8	1	30	100	336.9466	0.7472221	254.1810
0.8	3	10	10	313.3991	0.7814719	222.5172
0.8	3	10	20	289.7158	0.8141022	203.2292

0.8	3	10	30	280.2622	0.8260961	196.4278
0.8	3	10	40	276.7010	0.8307488	192.6261
0.8	3	10	50	274.5698	0.8330853	190.8585
0.8	3	10	60	274.2301	0.8335607	190.0633
0.8	3	10	70	272.7424	0.8355866	189.2421
0.8	3	10	80	273.9360	0.8343341	190.1812
0.8	3	10	90	274.0187	0.8342887	190.1070
0.8	3	10	100	274.8005	0.8336533	190.7652
0.8	3	20	10	308.4399	0.7883900	219.0137
0.8	3	20	20	286.1045	0.8186758	200.3174
0.8	3	20	30	277.2642	0.8293385	193.2618
0.8	3	20	40	272.6071	0.8353040	189.2445
0.8	3	20	50	272.0916	0.8362252	189.3189
0.8	3	20	60	269.0043	0.8399586	187.2786
0.8	3	20	70	268.8816	0.8399157	186.4559
0.8	3	20	80	268.7171	0.8402866	185.8517
0.8	3	20	90	267.4710	0.8416283	184.6320
0.8	3	20	100	266.8108	0.8427629	184.6071
0.8	3	30	10	310.0595	0.7863736	220.4242
0.8	3	30	20	286.6350	0.8178124	201.8285
0.8	3	30	30	276.7609	0.8304706	194.2356
0.8	3	30	40	273.6052	0.8344169	191.6044
0.8	3	30	50	270.5048	0.8379865	188.8110
0.8	3	30	60	268.5048	0.8405929	187.5475
0.8	3	30	70	267.7081	0.8419326	187.3892
0.8	3	30	80	266.7968	0.8429300	186.8339
0.8	3	30	90	265.0868	0.8452370	185.7087
0.8	3	30	100	264.4959	0.8456057	184.8629
0.8	5	10	10	287.6896	0.8170406	199.1870
0.8	5	10	20	277.1024	0.8305732	190.2491
0.8	5	10	30	276.4306	0.8322566	189.1176
0.8	5	10	40	273.7170	0.8357182	186.4819
0.8	5	10	50	277.7425	0.8314013	188.7027
0.8	5	10	60	278.5808	0.8302239	189.4871
0.8	5	10	70	280.8233	0.8281134	190.5736
0.8	5	10	80	281.8110	0.8266652	191.0360
0.8	5	10	90	283.3678	0.8255242	192.2027
0.8	5	10	100	285.1257	0.8233990	194.1169
0.8	5	20	10	288.7310	0.8150670	200.7708
0.8	5	20	20	278.0336	0.8290892	192.0483
0.8	5	20	30	273.6218	0.8345402	189.1972
0.8	5	20	40	272.6594	0.8364340	188.3820
0.8	5	20	50	273.0732	0.8363782	188.7574
0.8	5	20	60	274.4159	0.8349000	189.2368
0.8	5	20	70	275.4811	0.8337394	190.5148
0.8	5	20	80	277.0171	0.8321988	191.9016
0.8	5	20	90	278.7610	0.8307158	193.3844
0.8	5	20	100	280.9790	0.8274947	194.5412
0.8	5	30	10	290.7331	0.8126536	202.9289
0.8	5	30	20	278.3583	0.8284188	192.7321
0.8	5	30	30	275.6259	0.8324878	190.1391
0.8	5	30	40	274.3555	0.8341441	189.1691
0.8	5	30	50	273.3999	0.8352286	188.6528
0.8	5	30	60	273.5831	0.8350514	189.3824
0.8	5	30	70	275.6703	0.8331254	190.3269
0.8	5	30	80	275.9762	0.8330368	190.6929
0.8	5	30	90	278.3859	0.8306432	191.4052
0.8	5	30	100	278.6630	0.8305647	191.4195
0.9	1	10	10	414.8063	0.6170671	312.2511
0.9	1	10	20	386.7889	0.6673867	289.4774
0.9	1	10	30	371.0998	0.6933155	278.8785
0.9	1	10	40	358.6394	0.7134588	271.6798
0.9	1	10	50	352.8975	0.7226282	266.4527
0.9	1	10	60	346.8112	0.7322188	260.5744
0.9	1	10	70	343.2689	0.7376695	258.4233
0.9	1	10	80	340.6728	0.7415661	256.0447
0.9	1	10	90	338.5686	0.7448857	253.9967
0.9	1	10	100	336.7018	0.7474526	252.2800
0.9	1	20	10	411.3201	0.6233645	310.3184
0.9	1	20	20	384.1182	0.6714674	288.5440
0.9	1	20	30	367.6075	0.6990348	276.3208
0.9	1	20	40	356.4158	0.7171393	270.6802
0.9	1	20	50	348.7361	0.7292542	264.1105
0.9	1	20	60	344.6519	0.7354166	261.2004
0.9	1	20	70	341.2455	0.7405939	257.5663
0.9	1	20	80	339.2597	0.7437875	256.2598
0.9	1	20	90	337.4297	0.7465094	254.0291
0.9	1	20	100	335.7686	0.7488797	252.3995
0.9	1	30	10	413.1364	0.6201569	313.7252
0.9	1	30	20	385.6340	0.6683476	290.6420
0.9	1	30	30	369.0839	0.6966603	277.7134
0.9	1	30	40	356.7791	0.7160088	269.9758
0.9	1	30	50	350.2945	0.7268614	265.0344
0.9	1	30	60	345.2038	0.7346804	261.2523
0.9	1	30	70	341.5533	0.7400813	257.1691
0.9	1	30	80	339.4933	0.7432158	254.9206
0.9	1	30	90	338.0695	0.7455819	254.1001

```
  0.9        1                  30                100      335.4525  0.7493191  252.2678
  0.9        3                  10                 10      319.5001  0.7759612  226.6134
  0.9        3                  10                 20      293.0955  0.8103760  205.9286
  0.9        3                  10                 30      286.2675  0.8191614  198.9493
  0.9        3                  10                 40      283.6010  0.8224724  195.6560
  0.9        3                  10                 50      281.4391  0.8252741  194.2805
  0.9        3                  10                 60      280.9465  0.8261519  194.0485
  0.9        3                  10                 70      280.4002  0.8271139  194.3529
  0.9        3                  10                 80      280.5707  0.8268796  194.2447
  0.9        3                  10                 90      282.6415  0.8245501  195.2303
  0.9        3                  10                100      280.8485  0.8266050  193.8663
  0.9        3                  20                 10      312.2636  0.7843654  222.5697
  0.9        3                  20                 20      287.0123  0.8177614  201.5902
  0.9        3                  20                 30      277.2554  0.8301448  194.4609
  0.9        3                  20                 40      274.4417  0.8336108  191.2741
  0.9        3                  20                 50      271.4743  0.8371259  189.8565
  0.9        3                  20                 60      270.2224  0.8383144  189.0029
  0.9        3                  20                 70      270.5947  0.8379415  188.9530
  0.9        3                  20                 80      271.1089  0.8375700  189.4954
  0.9        3                  20                 90      271.1707  0.8380059  189.1691
  0.9        3                  20                100      269.8467  0.8391103  188.2031
  0.9        3                  30                 10      313.7603  0.7824111  225.0689
  0.9        3                  30                 20      292.3928  0.8114701  205.9293
  0.9        3                  30                 30      279.4600  0.8278250  196.1871
  0.9        3                  30                 40      276.5619  0.8312569  192.6903
  0.9        3                  30                 50      275.0239  0.8327729  191.7414
  0.9        3                  30                 60      275.0152  0.8333142  191.9388
  0.9        3                  30                 70      274.7634  0.8336889  192.4005
  0.9        3                  30                 80      274.2921  0.8344854  191.7144
  0.9        3                  30                 90      275.1057  0.8337253  192.4825
  0.9        3                  30                100      275.6726  0.8333818  192.4697
  0.9        5                  10                 10      302.9615  0.7987095  209.2456
  0.9        5                  10                 20      290.6225  0.8147022  198.2590
  0.9        5                  10                 30      288.2319  0.8176689  196.0519
  0.9        5                  10                 40      287.0664  0.8198248  195.7296
  0.9        5                  10                 50      289.9651  0.8168013  197.1681
  0.9        5                  10                 60      290.3648  0.8164664  198.6217
  0.9        5                  10                 70      294.2038  0.8126069  200.2874
  0.9        5                  10                 80      293.8755  0.8131283  200.8489
  0.9        5                  10                 90      296.1632  0.8101984  203.5367
  0.9        5                  10                100      297.5063  0.8087714  205.5411
  0.9        5                  20                 10      292.1216  0.8117084  204.1867
  0.9        5                  20                 20      280.0749  0.8274069  194.5892
  0.9        5                  20                 30      278.4701  0.8294453  194.4135
  0.9        5                  20                 40      277.9348  0.8305645  193.5764
  0.9        5                  20                 50      277.6396  0.8309728  193.0416
  0.9        5                  20                 60      279.0383  0.8298544  194.0891
  0.9        5                  20                 70      280.3732  0.8287552  194.9731
  0.9        5                  20                 80      283.1612  0.8252848  197.2889
  0.9        5                  20                 90      283.9032  0.8251842  198.8021
  0.9        5                  20                100      283.8740  0.8254374  198.6089
  0.9        5                  30                 10      299.1913  0.8028538  208.4769
  0.9        5                  30                 20      284.9982  0.8207952  198.4173
  0.9        5                  30                 30      279.4695  0.8284036  195.1479
  0.9        5                  30                 40      278.6529  0.8289859  194.0039
  0.9        5                  30                 50      280.6441  0.8263762  195.3877
  0.9        5                  30                 60      282.5047  0.8245211  196.9083
  0.9        5                  30                 70      282.5603  0.8254042  197.3762
  0.9        5                  30                 80      283.2600  0.8243877  198.9525
  0.9        5                  30                 90      284.8748  0.8228574  199.6747
  0.9        5                  30                100      285.6253  0.8225291  200.1894

RMSE was used to select the optimal model using the smallest value.
The final values used for the model were n.trees = 100, interaction.depth = 5, shrinkage = 0.3
and n.minobsinnode = 30.
```

f) Make predictions in the test set and report the root mean square error rate and mean absolute error.

Code:

```
best_gbm_seoul = gbm_seoul
predict_gbm_seoul = predict(best_gbm_seoul, newdata = seoulte)

rmse_seoul = rmse(actual = seoulte$Rented.Bike.Count,predicted = predict_gbm_seoul) # RMSE
mae_seoul = mae(actual = seoulte$Rented.Bike.Count, predicted = predict_gbm_seoul) #MAE
paste("RMSE:", rmse_seoul)
paste("MAE", mae_seoul)
```

Code Output:

Gradient boosting trees have the potential to outperform random forests in terms of accuracy. This is due to their ability to collaborate in correcting each other's errors and capture intricate patterns in the data. Nevertheless, when dealing with noisy data, gradient boosting trees can be prone to overfitting and mistakenly modeling the noise present in the data.

```
[1] "RMSE: 219.843722972971"
[1] "MAE 147.649248497018"
```