Call:

randomForest(formula = neighbourhood\_group ~ ., data = train, )

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 2

OOB estimate of error rate: 0.12%

Confusion matrix:

Bronx Brooklyn Manhattan Queens Staten Island class.error

Bronx 654 0 6 0 0 0.0090909091

Brooklyn 0 12034 1 12 0 0.0010791068

Manhattan 2 1 12984 0 0 0.0002310002

Queens 0 13 1 3418 0 0.0040792541

Staten Island 0 0 0 0 211 0.0000000000

Random Forest

29337 samples

4 predictor

5 classes: 'Bronx', 'Brooklyn', 'Manhattan', 'Queens', 'Staten Island'

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 26403, 26403, 26405, 26402, 26402, 26404, ...

Resampling results across tuning parameters:

mtry Accuracy Kappa

2 0.9983298 0.9973106

3 0.9987048 0.9979145

5 0.9986707 0.9978597

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

The final value used for the model was mtry = 3.

Random Forest

29337 samples

4 predictor

5 classes: 'Bronx', 'Brooklyn', 'Manhattan', 'Queens', 'Staten Island'

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 26403, 26404, 26402, 26403, 26404, 26403, ...

Resampling results across tuning parameters:

mtry Accuracy Kappa

1 0.9200327 0.8677874

2 0.9976481 0.9962118

3 0.9986707 0.9978598

4 0.9987047 0.9979146

5 0.9983980 0.9974206

6 0.9983639 0.9973658

7 0.9983298 0.9973108

8 0.9983639 0.9973657

9 0.9985002 0.9975851

10 0.9983980 0.9974206

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

The final value used for the model was mtry = 4.

Call:

summary.resamples(object = results\_mtry)

Models: 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

Number of resamples: 10

Accuracy

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

5 0.8656666 0.8715479 0.8765976 0.8750376 0.8801125 0.8813906 0

6 0.8663484 0.8721446 0.8774499 0.8761284 0.8810802 0.8830948 0

7 0.8666894 0.8724003 0.8772794 0.8759579 0.8811350 0.8827539 0

8 0.8895329 0.8954995 0.8999485 0.8986256 0.9035446 0.9052488 0

9 0.9246505 0.9288271 0.9319925 0.9316558 0.9342252 0.9389911 0

10 0.9243354 0.9291681 0.9328447 0.9319967 0.9346512 0.9389911 0

11 0.9246505 0.9293386 0.9335263 0.9325419 0.9366214 0.9389911 0

12 0.9376278 0.9402489 0.9429007 0.9447785 0.9464894 0.9618399 0

13 0.9556919 0.9608762 0.9616499 0.9614479 0.9621645 0.9652352 0

14 0.9577369 0.9623349 0.9636891 0.9639358 0.9644683 0.9741056 0

15 0.9676210 0.9700895 0.9718766 0.9717761 0.9729039 0.9764906 0

Kappa

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

5 0.7802606 0.7908466 0.7982034 0.7960216 0.8043544 0.8063111 0

6 0.7813602 0.7918119 0.7997104 0.7978040 0.8060856 0.8090974 0

7 0.7819219 0.7922254 0.7994336 0.7975420 0.8057028 0.8085374 0

8 0.8201100 0.8306541 0.8372179 0.8353279 0.8431493 0.8459381 0

9 0.8770940 0.8841116 0.8893748 0.8887770 0.8930896 0.9006201 0

10 0.8769022 0.8846327 0.8907274 0.8893128 0.8937624 0.9006149 0

11 0.8770695 0.8849294 0.8918706 0.8902240 0.8970968 0.9005266 0

12 0.8985794 0.9027588 0.9071456 0.9101798 0.9129655 0.9381335 0

13 0.9280203 0.9365554 0.9375866 0.9373426 0.9383817 0.9434528 0

14 0.9312982 0.9387597 0.9408976 0.9413020 0.9420383 0.9580619 0

15 0.9471849 0.9513132 0.9541719 0.9540525 0.9559104 0.9618177 0

Call:

summary.resamples(object = results\_mtry)

Models: 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30

Number of resamples: 10

Accuracy

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

5 0.8656666 0.8715479 0.8765976 0.8750376 0.8801125 0.8813906 0

6 0.8663484 0.8721446 0.8774499 0.8761284 0.8810802 0.8830948 0

7 0.8666894 0.8724003 0.8772794 0.8759579 0.8811350 0.8827539 0

8 0.8895329 0.8954995 0.8999485 0.8986256 0.9035446 0.9052488 0

9 0.9246505 0.9288271 0.9319925 0.9316558 0.9342252 0.9389911 0

10 0.9243354 0.9291681 0.9328447 0.9319967 0.9346512 0.9389911 0

11 0.9246505 0.9293386 0.9335263 0.9325419 0.9366214 0.9389911 0

12 0.9376278 0.9402489 0.9429007 0.9447785 0.9464894 0.9618399 0

13 0.9556919 0.9608762 0.9616499 0.9614479 0.9621645 0.9652352 0

14 0.9577369 0.9623349 0.9636891 0.9639358 0.9644683 0.9741056 0

15 0.9676210 0.9700895 0.9718766 0.9717761 0.9729039 0.9764906 0

16 0.9669393 0.9699191 0.9723879 0.9720147 0.9737560 0.9757927 0

17 0.9679618 0.9703401 0.9720470 0.9721169 0.9736708 0.9761499 0

18 0.9689843 0.9734083 0.9749446 0.9745712 0.9761418 0.9792164 0

19 0.9720518 0.9746912 0.9773307 0.9770255 0.9796353 0.9812479 0

20 0.9723926 0.9759694 0.9785240 0.9778435 0.9801466 0.9819298 0

21 0.9747699 0.9775885 0.9798876 0.9794797 0.9821019 0.9833049 0

22 0.9792022 0.9804823 0.9821033 0.9820361 0.9834682 0.9846678 0

23 0.9798909 0.9837239 0.9844894 0.9846608 0.9869600 0.9877342 0

24 0.9829526 0.9857691 0.9868758 0.9870129 0.9882413 0.9904535 0

25 0.9832936 0.9869632 0.9887506 0.9880696 0.9893508 0.9907944 0

26 0.9836345 0.9877301 0.9889192 0.9883423 0.9893490 0.9914763 0

27 0.9832936 0.9878153 0.9890915 0.9885468 0.9894360 0.9914763 0

28 0.9836345 0.9882413 0.9890915 0.9889557 0.9900298 0.9921635 0

29 0.9884117 0.9888349 0.9904550 0.9907624 0.9926714 0.9942078 0

30 0.9887487 0.9928406 0.9942059 0.9935235 0.9952277 0.9959100 0

Kappa

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

5 0.7802606 0.7908466 0.7982034 0.7960216 0.8043544 0.8063111 0

6 0.7813602 0.7918119 0.7997104 0.7978040 0.8060856 0.8090974 0

7 0.7819219 0.7922254 0.7994336 0.7975420 0.8057028 0.8085374 0

8 0.8201100 0.8306541 0.8372179 0.8353279 0.8431493 0.8459381 0

9 0.8770940 0.8841116 0.8893748 0.8887770 0.8930896 0.9006201 0

10 0.8769022 0.8846327 0.8907274 0.8893128 0.8937624 0.9006149 0

11 0.8770695 0.8849294 0.8918706 0.8902240 0.8970968 0.9005266 0

12 0.8985794 0.9027588 0.9071456 0.9101798 0.9129655 0.9381335 0

13 0.9280203 0.9365554 0.9375866 0.9373426 0.9383817 0.9434528 0

14 0.9312982 0.9387597 0.9408976 0.9413020 0.9420383 0.9580619 0

15 0.9471849 0.9513132 0.9541719 0.9540525 0.9559104 0.9618177 0

16 0.9460847 0.9509720 0.9549663 0.9543924 0.9572343 0.9606380 0

17 0.9477418 0.9516688 0.9543976 0.9545595 0.9570974 0.9612346 0

18 0.9494291 0.9566474 0.9591933 0.9585922 0.9611581 0.9662681 0

19 0.9544810 0.9588131 0.9631404 0.9626357 0.9668982 0.9695821 0

20 0.9550385 0.9609128 0.9650824 0.9639740 0.9677329 0.9706983 0

21 0.9589161 0.9635676 0.9673327 0.9666633 0.9709516 0.9729664 0

22 0.9662215 0.9683124 0.9709669 0.9708625 0.9731938 0.9751807 0

23 0.9673861 0.9736159 0.9748722 0.9751518 0.9788992 0.9801809 0

24 0.9723834 0.9769503 0.9787634 0.9789840 0.9809769 0.9845705 0

25 0.9729353 0.9789103 0.9817960 0.9806995 0.9827877 0.9851240 0

26 0.9734871 0.9801559 0.9820741 0.9811422 0.9827731 0.9862306 0

27 0.9729353 0.9802937 0.9823499 0.9814733 0.9829256 0.9862304 0

28 0.9734976 0.9809785 0.9823536 0.9821395 0.9838786 0.9873548 0

29 0.9812530 0.9819633 0.9845639 0.9850726 0.9881660 0.9906606 0

30 0.9818284 0.9884391 0.9906586 0.9895489 0.9923063 0.9934078 0

Call:

summary.resamples(object = results\_tree)

Models: 250, 300, 350, 400, 450, 500, 550, 600, 800, 1000, 2000

Number of resamples: 10

Accuracy

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

250 0.9826176 0.9857666 0.9867054 0.9863312 0.9878122 0.9884078 0

300 0.9826117 0.9857666 0.9865349 0.9862971 0.9875659 0.9884078 0

350 0.9826176 0.9857666 0.9865349 0.9862971 0.9877363 0.9884078 0

400 0.9839755 0.9856814 0.9865349 0.9865016 0.9876470 0.9880668 0

450 0.9836345 0.9857666 0.9868758 0.9866379 0.9879826 0.9884196 0

500 0.9839755 0.9857666 0.9867054 0.9867061 0.9879826 0.9887487 0

550 0.9839755 0.9857666 0.9870462 0.9867402 0.9876511 0.9887487 0

600 0.9843164 0.9857666 0.9870462 0.9867744 0.9876470 0.9890897 0

800 0.9843164 0.9860259 0.9868823 0.9869789 0.9882373 0.9890897 0

1000 0.9843164 0.9861075 0.9868758 0.9868425 0.9878996 0.9887487 0

2000 0.9843164 0.9864519 0.9870527 0.9870811 0.9879816 0.9887526 0

Kappa

Min. 1st Qu. Median Mean 3rd Qu. Max. NA's

250 0.9717933 0.9769595 0.9784741 0.9778731 0.9802956 0.9812810 0

300 0.9718130 0.9769595 0.9781983 0.9778203 0.9798863 0.9812669 0

350 0.9717933 0.9769480 0.9782209 0.9778170 0.9801619 0.9812669 0

400 0.9740219 0.9768000 0.9782227 0.9781537 0.9800127 0.9807336 0

450 0.9734745 0.9769470 0.9787755 0.9783747 0.9805705 0.9812766 0

500 0.9740266 0.9769470 0.9785001 0.9784851 0.9805705 0.9818339 0

550 0.9740219 0.9769595 0.9790259 0.9785409 0.9800288 0.9818339 0

600 0.9745788 0.9769595 0.9790310 0.9785958 0.9800127 0.9823815 0

800 0.9745788 0.9773733 0.9787621 0.9789286 0.9809803 0.9823815 0

1000 0.9745871 0.9774997 0.9787755 0.9787082 0.9804313 0.9818339 0

2000 0.9745934 0.9780462 0.9790601 0.9790966 0.9805729 0.9818339 0

* mtry=4: 4 features is chosen for each iteration
* maxnodes = 30: Maximum 30 nodes in the terminal nodes (leaves)
* ntree =2000: 2000 trees will be trained

> fit\_rf

Random Forest

29337 samples

4 predictor

5 classes: 'Bronx', 'Brooklyn', 'Manhattan', 'Queens', 'Staten Island'

No pre-processing

Resampling: Cross-Validated (10 fold)

Summary of sample sizes: 26402, 26403, 26404, 26404, 26403, 26402, ...

Resampling results:

Accuracy Kappa

0.9943075 0.9908197

Tuning parameter 'mtry' was held constant at a value of 4

> conf

Confusion Matrix and Statistics

Reference

Prediction Bronx Brooklyn Manhattan Queens Staten Island

Bronx 423 0 6 0 0

Brooklyn 0 8055 4 50 0

Manhattan 8 0 8664 30 0

Queens 0 2 0 2154 0

Staten Island 0 0 0 0 162

Overall Statistics

Accuracy : 0.9949

95% CI : (0.9938, 0.9958)

No Information Rate : 0.4435

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.9917

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: Bronx Class: Brooklyn Class: Manhattan Class: Queens

Sensitivity 0.98144 0.9998 0.9988 0.9642

Specificity 0.99969 0.9953 0.9965 0.9999

Pos Pred Value 0.98601 0.9933 0.9956 0.9991

Neg Pred Value 0.99958 0.9998 0.9991 0.9954

Prevalence 0.02204 0.4120 0.4435 0.1142

Detection Rate 0.02163 0.4119 0.4430 0.1101

Detection Prevalence 0.02193 0.4146 0.4449 0.1102

Balanced Accuracy 0.99056 0.9975 0.9977 0.9820

Class: Staten Island

Sensitivity 1.000000

Specificity 1.000000

Pos Pred Value 1.000000

Neg Pred Value 1.000000

Prevalence 0.008283

Detection Rate 0.008283

Detection Prevalence 0.008283

Balanced Accuracy 1.000000

> conf\_default

Confusion Matrix and Statistics

Reference

Prediction Bronx Brooklyn Manhattan Queens Staten Island

Bronx 429 0 1 0 0

Brooklyn 0 8056 0 12 0

Manhattan 2 0 8673 1 0

Queens 0 1 0 2221 0

Staten Island 0 0 0 0 162

Overall Statistics

Accuracy : 0.9991

95% CI : (0.9986, 0.9995)

No Information Rate : 0.4435

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.9986

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: Bronx Class: Brooklyn Class: Manhattan Class: Queens

Sensitivity 0.99536 0.9999 0.9999 0.9942

Specificity 0.99995 0.9990 0.9997 0.9999

Pos Pred Value 0.99767 0.9985 0.9997 0.9995

Neg Pred Value 0.99990 0.9999 0.9999 0.9993

Prevalence 0.02204 0.4120 0.4435 0.1142

Detection Rate 0.02193 0.4119 0.4435 0.1136

Detection Prevalence 0.02199 0.4125 0.4436 0.1136

Balanced Accuracy 0.99765 0.9994 0.9998 0.9971

Class: Staten Island

Sensitivity 1.000000

Specificity 1.000000

Pos Pred Value 1.000000

Neg Pred Value 1.000000

Prevalence 0.008283

Detection Rate 0.008283

Detection Prevalence 0.008283

Balanced Accuracy 1.000000