As Is 85.08

Balanced 58.55

Drop Column

Naïve Bayes : 59.93 : 63.35

SVM

Decision Trees

* CHAID : 79.13 : 79.29
* CART : 73.98 : 73.98
* C50 : 84.42 : 85.49

Random Forests : 85.19 : 88.71

|  |  |  |
| --- | --- | --- |
| Method | Without Feature Selection | With Feature Selection |
| Naïve Bayes | 59.93 % | 63.35 % |
| SVM |  |  |
| Decision Trees |  |  |
| * CHAID * CART * C50 | * 79.13 % * 73.98 % * 84.42 % | * 79.29 % * 73.98 % * 85.49 % |
| Random Forests | 85.19 % | 88.49 % |

|  |  |  |
| --- | --- | --- |
| Method | Without Feature Selection | With Feature Selection |
| Naïve Bayes |  | 57.89 % |
| SVM | 63.42 % |  |
| Decision Trees |  |  |
| * CHAID * CART * C50 |  | * 74.09 % * 68.61 % * 81.24 % |
| Random Forests |  | 85.37 % |

Balanced 58.55% : 65% :

Aggregated 62.89

Balanced 62.04

train\_file\_action = "Training Data 1801 - Drop Column.csv";

test\_file\_action = "Test Data 1801 - Drop Column.csv";

train\_file\_word = "Training Data 1801 - Aggregated.csv";

test\_file\_word = "Test Data 1801 - Aggregated.csv";

target\_column = "manual\_assessment";

> modelRandomForest = hitungRandomForest(train\_dataset1\_norm, test\_dataset1\_norm);

+

Loading required package: lattice

Loading required package: ggplot2

randomForest 4.6-12

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

Attaching package: `randomForest'

The following object is masked from `package:ggplot2':

margin

The following object is masked from `package:dplyr':

combine

pred 1 2 3 4 5

1 138 16 15 6 20

2 0 28 2 1 1

3 46 17 504 21 44

4 1 0 1 130 2

5 186 102 255 207 4626

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 138 16 15 6 20

2 0 28 2 1 1

3 46 17 504 21 44

4 1 0 1 130 2

5 186 102 255 207 4626

Overall Statistics

Accuracy : 0.8519

95% CI : (0.843, 0.8606)

No Information Rate : 0.7369

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

Kappa : 0.5919

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.37197 0.171779 0.64865 0.35616 0.9857

Specificity 0.99050 0.999355 0.97711 0.99933 0.5525

Pos Pred Value 0.70769 0.875000 0.79747 0.97015 0.8605

Neg Pred Value 0.96226 0.978697 0.95241 0.96231 0.9325

Prevalence 0.05825 0.025593 0.12200 0.05731 0.7369

Detection Rate 0.02167 0.004396 0.07913 0.02041 0.7263

Detection Prevalence 0.03062 0.005024 0.09923 0.02104 0.8441

Balanced Accuracy 0.68123 0.585567 0.81288 0.67775 0.7691

> hitungNaiveBayes(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungDecisionTree(train\_dataset1\_norm, test\_dataset1\_norm);

+

pred 1 2 3 4 5

1 53 12 42 17 71

2 178 92 305 161 850

3 23 11 103 18 133

4 8 7 6 5 75

5 109 41 321 164 3564

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 53 12 42 17 71

2 178 92 305 161 850

3 23 11 103 18 133

4 8 7 6 5 75

5 109 41 321 164 3564

Overall Statistics

Accuracy : 0.5993

95% CI : (0.5872, 0.6114)

No Information Rate : 0.7369

P-Value [Acc > NIR] : 1

Kappa : 0.198

Mcnemar's Test P-Value : <2e-16

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.142857 0.56442 0.13256 0.0136986 0.7594

Specificity 0.976325 0.75927 0.96692 0.9840107 0.6211

Pos Pred Value 0.271795 0.05801 0.35764 0.0495050 0.8488

Neg Pred Value 0.948494 0.98516 0.88916 0.9425654 0.4797

Prevalence 0.058251 0.02559 0.12200 0.0573088 0.7369

Detection Rate 0.008322 0.01444 0.01617 0.0007851 0.5596

Detection Prevalence 0.030617 0.24902 0.04522 0.0158581 0.6593

Balanced Accuracy 0.559591 0.66184 0.54974 0.4988546 0.6903

[1] "CHAID"

Loading required package: grid

Loading required package: mvtnorm

Loading required package: modeltools

Loading required package: stats4

Loading required package: strucchange

Loading required package: zoo

Attaching package: `zoo'

The following object is masked from `package:imputeTS':

na.locf

The following objects are masked from `package:base':

as.Date, as.Date.numeric

Loading required package: sandwich

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 84 10 18 14 27

2 3 13 6 2 13

3 25 11 300 10 66

4 3 2 4 81 25

5 256 127 449 258 4562

Overall Statistics

Accuracy : 0.7913

95% CI : (0.7811, 0.8013)

No Information Rate : 0.7369

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

Kappa : 0.3783

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.22642 0.079755 0.38610 0.22192 0.9721

Specificity 0.98850 0.996133 0.97997 0.99434 0.3496

Pos Pred Value 0.54902 0.351351 0.72816 0.70435 0.8071

Neg Pred Value 0.95383 0.976311 0.91993 0.95459 0.8173

Prevalence 0.05825 0.025593 0.12200 0.05731 0.7369

Detection Rate 0.01319 0.002041 0.04710 0.01272 0.7163

Detection Prevalence 0.02402 0.005809 0.06469 0.01806 0.8874

Balanced Accuracy 0.60746 0.537944 0.68304 0.60813 0.6609

[1] "CART"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 0 0 0 0 0

2 0 0 0 0 0

3 93 29 134 31 115

4 0 0 0 0 0

5 278 134 643 334 4578

Overall Statistics

Accuracy : 0.7398

95% CI : (0.7289, 0.7506)

No Information Rate : 0.7369

P-Value [Acc > NIR] : 0.2999

Kappa : 0.1384

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.00000 0.00000 0.17246 0.00000 0.9755

Specificity 1.00000 1.00000 0.95207 1.00000 0.1712

Pos Pred Value NaN NaN 0.33333 NaN 0.7672

Neg Pred Value 0.94175 0.97441 0.89224 0.94269 0.7139

Prevalence 0.05825 0.02559 0.12200 0.05731 0.7369

Detection Rate 0.00000 0.00000 0.02104 0.00000 0.7188

Detection Prevalence 0.00000 0.00000 0.06312 0.00000 0.9369

Balanced Accuracy 0.50000 0.50000 0.56227 0.50000 0.5734

[1] "C50"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 168 21 39 21 75

2 17 42 19 5 18

3 39 21 510 29 93

4 17 9 24 175 25

5 130 70 185 135 4482

Overall Statistics

Accuracy : 0.8442

95% CI : (0.8351, 0.8531)

No Information Rate : 0.7369

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

Kappa : 0.613

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.45283 0.257669 0.65637 0.47945 0.9550

Specificity 0.97399 0.990493 0.96745 0.98751 0.6897

Pos Pred Value 0.51852 0.415842 0.73699 0.70000 0.8960

Neg Pred Value 0.96642 0.980696 0.95297 0.96895 0.8456

Prevalence 0.05825 0.025593 0.12200 0.05731 0.7369

Detection Rate 0.02638 0.006594 0.08008 0.02748 0.7037

Detection Prevalence 0.05087 0.015858 0.10865 0.03925 0.7854

Balanced Accuracy 0.71341 0.624081 0.81191 0.73348 0.8224

train\_file\_action = "Training Data 1801 - Drop Column - Balanced.csv";

test\_file\_action = "Test Data 1801 - Drop Column - Balanced.csv";

train\_file\_word = "Training Data 1801 - Aggregated.csv";

test\_file\_word = "Test Data 1801 - Aggregated.csv";

target\_column = "manual\_assessment";

+ modelRandomForest = hitungRandomForest(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungNaiveBayes(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungDecisionTree(train\_dataset1\_norm, test\_dataset1\_norm);

+

pred 1 2 3 4 5

1 57 19 23 8 26

2 1 6 2 1 2

3 30 25 158 11 42

4 3 2 2 7 1

5 166 92 217 132 909

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 57 19 23 8 26

2 1 6 2 1 2

3 30 25 158 11 42

4 3 2 2 7 1

5 166 92 217 132 909

Overall Statistics

Accuracy : 0.5855

95% CI : (0.5632, 0.6075)

No Information Rate : 0.5046

P-Value [Acc > NIR] : 5.331e-13

Kappa : 0.2696

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.22179 0.041667 0.39303 0.044025 0.9276

Specificity 0.95490 0.996663 0.92987 0.995513 0.3690

Pos Pred Value 0.42857 0.500000 0.59398 0.466667 0.5996

Neg Pred Value 0.88944 0.928497 0.85442 0.921121 0.8333

Prevalence 0.13234 0.074150 0.20700 0.081874 0.5046

Detection Rate 0.02935 0.003090 0.08136 0.003605 0.4681

Detection Prevalence 0.06849 0.006179 0.13697 0.007724 0.7806

Balanced Accuracy 0.58834 0.519165 0.66145 0.519769 0.6483

pred 1 2 3 4 5

1 51 14 33 15 58

2 92 66 118 46 230

3 31 10 72 18 63

4 7 7 14 9 43

5 76 47 165 71 586

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 51 14 33 15 58

2 92 66 118 46 230

3 31 10 72 18 63

4 7 7 14 9 43

5 76 47 165 71 586

Overall Statistics

Accuracy : 0.4037

95% CI : (0.3818, 0.4259)

No Information Rate : 0.5046

P-Value [Acc > NIR] : 1

Kappa : 0.1453

Mcnemar's Test P-Value : <2e-16

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.19844 0.45833 0.17910 0.056604 0.5980

Specificity 0.92878 0.72970 0.92078 0.960179 0.6268

Pos Pred Value 0.29825 0.11957 0.37113 0.112500 0.6201

Neg Pred Value 0.88368 0.94388 0.81121 0.919441 0.6048

Prevalence 0.13234 0.07415 0.20700 0.081874 0.5046

Detection Rate 0.02626 0.03399 0.03708 0.004634 0.3018

Detection Prevalence 0.08805 0.28424 0.09990 0.041195 0.4866

Balanced Accuracy 0.56361 0.59402 0.54994 0.508392 0.6124

[1] "CHAID"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 60 29 44 17 56

2 0 0 0 0 0

3 13 12 84 7 45

4 0 0 0 2 0

5 184 103 274 133 879

Overall Statistics

Accuracy : 0.5278

95% CI : (0.5053, 0.5502)

No Information Rate : 0.5046

P-Value [Acc > NIR] : 0.02168

Kappa : 0.1567

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.2335 0.00000 0.20896 0.01258 0.8969

Specificity 0.9134 1.00000 0.95000 1.00000 0.2786

Pos Pred Value 0.2913 NaN 0.52174 1.00000 0.5588

Neg Pred Value 0.8865 0.92585 0.82145 0.91907 0.7263

Prevalence 0.1323 0.07415 0.20700 0.08187 0.5046

Detection Rate 0.0309 0.00000 0.04325 0.00103 0.4526

Detection Prevalence 0.1061 0.00000 0.08290 0.00103 0.8100

Balanced Accuracy 0.5734 0.50000 0.57948 0.50629 0.5878

[1] "CART"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 31 19 23 4 25

2 0 0 0 0 0

3 17 10 39 6 22

4 0 0 0 0 0

5 209 115 340 149 933

Overall Statistics

Accuracy : 0.5165

95% CI : (0.494, 0.5389)

No Information Rate : 0.5046

P-Value [Acc > NIR] : 0.1536

Kappa : 0.0865

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.12062 0.00000 0.09701 0.00000 0.9520

Specificity 0.95786 1.00000 0.96429 1.00000 0.1549

Pos Pred Value 0.30392 NaN 0.41489 NaN 0.5344

Neg Pred Value 0.87717 0.92585 0.80357 0.91813 0.7602

Prevalence 0.13234 0.07415 0.20700 0.08187 0.5046

Detection Rate 0.01596 0.00000 0.02008 0.00000 0.4804

Detection Prevalence 0.05252 0.00000 0.04840 0.00000 0.8991

Balanced Accuracy 0.53924 0.50000 0.53065 0.50000 0.5535

[1] "C50"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 72 26 33 16 72

2 11 22 11 4 16

3 47 29 195 21 87

4 4 2 9 15 25

5 123 65 154 103 780

Overall Statistics

Accuracy : 0.5582

95% CI : (0.5358, 0.5804)

No Information Rate : 0.5046

P-Value [Acc > NIR] : 1.28e-06

Kappa : 0.2892

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.28016 0.15278 0.4851 0.094340 0.7959

Specificity 0.91276 0.97664 0.8805 0.977566 0.5374

Pos Pred Value 0.32877 0.34375 0.5145 0.272727 0.6367

Neg Pred Value 0.89263 0.93504 0.8676 0.923688 0.7211

Prevalence 0.13234 0.07415 0.2070 0.081874 0.5046

Detection Rate 0.03708 0.01133 0.1004 0.007724 0.4016

Detection Prevalence 0.11277 0.03296 0.1952 0.028321 0.6308

Balanced Accuracy 0.59646 0.56471 0.6828 0.535953 0.6667

train\_file\_action = "Training Data 1801.csv";

test\_file\_action = "Test Data 1801.csv";

train\_file\_word = "Training Data 1801 - Aggregated.csv";

test\_file\_word = "Test Data 1801 - Aggregated.csv";

target\_column = "manual\_assessment";

+ modelRandomForest = hitungRandomForest(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungNaiveBayes(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungDecisionTree(train\_dataset1\_norm, test\_dataset1\_norm);

+

pred 1 2 3 4 5

1 134 19 16 5 24

2 2 29 1 1 1

3 39 14 503 23 44

4 1 0 1 129 0

5 195 101 256 207 4624

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 134 19 16 5 24

2 2 29 1 1 1

3 39 14 503 23 44

4 1 0 1 129 0

5 195 101 256 207 4624

Overall Statistics

Accuracy : 0.8508

95% CI : (0.8419, 0.8595)

No Information Rate : 0.7369

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

Kappa : 0.5881

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.36119 0.177914 0.64736 0.35342 0.9853

Specificity 0.98933 0.999194 0.97854 0.99967 0.5471

Pos Pred Value 0.67677 0.852941 0.80738 0.98473 0.8590

Neg Pred Value 0.96159 0.978848 0.95231 0.96217 0.9300

Prevalence 0.05825 0.025593 0.12200 0.05731 0.7369

Detection Rate 0.02104 0.004553 0.07898 0.02025 0.7260

Detection Prevalence 0.03109 0.005338 0.09782 0.02057 0.8452

Balanced Accuracy 0.67526 0.588554 0.81295 0.67655 0.7662

pred 1 2 3 4 5

1 53 12 42 17 71

2 178 92 305 161 850

3 23 11 103 18 133

4 8 7 6 5 75

5 109 41 321 164 3564

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 53 12 42 17 71

2 178 92 305 161 850

3 23 11 103 18 133

4 8 7 6 5 75

5 109 41 321 164 3564

Overall Statistics

Accuracy : 0.5993

95% CI : (0.5872, 0.6114)

No Information Rate : 0.7369

P-Value [Acc > NIR] : 1

Kappa : 0.198

Mcnemar's Test P-Value : <2e-16

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.142857 0.56442 0.13256 0.0136986 0.7594

Specificity 0.976325 0.75927 0.96692 0.9840107 0.6211

Pos Pred Value 0.271795 0.05801 0.35764 0.0495050 0.8488

Neg Pred Value 0.948494 0.98516 0.88916 0.9425654 0.4797

Prevalence 0.058251 0.02559 0.12200 0.0573088 0.7369

Detection Rate 0.008322 0.01444 0.01617 0.0007851 0.5596

Detection Prevalence 0.030617 0.24902 0.04522 0.0158581 0.6593

Balanced Accuracy 0.559591 0.66184 0.54974 0.4988546 0.6903

[1] "CHAID"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 84 10 18 14 27

2 3 13 6 2 13

3 25 11 300 10 66

4 3 2 4 81 25

5 256 127 449 258 4562

Overall Statistics

Accuracy : 0.7913

95% CI : (0.7811, 0.8013)

No Information Rate : 0.7369

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

Kappa : 0.3783

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.22642 0.079755 0.38610 0.22192 0.9721

Specificity 0.98850 0.996133 0.97997 0.99434 0.3496

Pos Pred Value 0.54902 0.351351 0.72816 0.70435 0.8071

Neg Pred Value 0.95383 0.976311 0.91993 0.95459 0.8173

Prevalence 0.05825 0.025593 0.12200 0.05731 0.7369

Detection Rate 0.01319 0.002041 0.04710 0.01272 0.7163

Detection Prevalence 0.02402 0.005809 0.06469 0.01806 0.8874

Balanced Accuracy 0.60746 0.537944 0.68304 0.60813 0.6609

[1] "CART"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 0 0 0 0 0

2 0 0 0 0 0

3 93 29 134 31 115

4 0 0 0 0 0

5 278 134 643 334 4578

Overall Statistics

Accuracy : 0.7398

95% CI : (0.7289, 0.7506)

No Information Rate : 0.7369

P-Value [Acc > NIR] : 0.2999

Kappa : 0.1384

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.00000 0.00000 0.17246 0.00000 0.9755

Specificity 1.00000 1.00000 0.95207 1.00000 0.1712

Pos Pred Value NaN NaN 0.33333 NaN 0.7672

Neg Pred Value 0.94175 0.97441 0.89224 0.94269 0.7139

Prevalence 0.05825 0.02559 0.12200 0.05731 0.7369

Detection Rate 0.00000 0.00000 0.02104 0.00000 0.7188

Detection Prevalence 0.00000 0.00000 0.06312 0.00000 0.9369

Balanced Accuracy 0.50000 0.50000 0.56227 0.50000 0.5734

[1] "C50"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 168 21 39 21 75

2 17 42 19 5 18

3 39 21 510 29 93

4 17 9 24 175 25

5 130 70 185 135 4482

Overall Statistics

Accuracy : 0.8442

95% CI : (0.8351, 0.8531)

No Information Rate : 0.7369

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

Kappa : 0.613

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.45283 0.257669 0.65637 0.47945 0.9550

Specificity 0.97399 0.990493 0.96745 0.98751 0.6897

Pos Pred Value 0.51852 0.415842 0.73699 0.70000 0.8960

Neg Pred Value 0.96642 0.980696 0.95297 0.96895 0.8456

Prevalence 0.05825 0.025593 0.12200 0.05731 0.7369

Detection Rate 0.02638 0.006594 0.08008 0.02748 0.7037

Detection Prevalence 0.05087 0.015858 0.10865 0.03925 0.7854

Balanced Accuracy 0.71341 0.624081 0.81191 0.73348 0.8224

> train\_file\_action = "Training Data 1801 - Balanced.csv";

+ test\_file\_action = "Test Data 1801 - Balanced.csv";

+ modelRandomForest = hitungRandomForest(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungNaiveBayes(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungDecisionTree(train\_dataset1\_norm, test\_dataset1\_norm);

+

pred 1 2 3 4 5

1 57 19 23 8 26

2 1 6 2 1 2

3 30 25 158 11 42

4 3 2 2 7 1

5 166 92 217 132 909

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 57 19 23 8 26

2 1 6 2 1 2

3 30 25 158 11 42

4 3 2 2 7 1

5 166 92 217 132 909

Overall Statistics

Accuracy : 0.5855

95% CI : (0.5632, 0.6075)

No Information Rate : 0.5046

P-Value [Acc > NIR] : 5.331e-13

Kappa : 0.2696

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.22179 0.041667 0.39303 0.044025 0.9276

Specificity 0.95490 0.996663 0.92987 0.995513 0.3690

Pos Pred Value 0.42857 0.500000 0.59398 0.466667 0.5996

Neg Pred Value 0.88944 0.928497 0.85442 0.921121 0.8333

Prevalence 0.13234 0.074150 0.20700 0.081874 0.5046

Detection Rate 0.02935 0.003090 0.08136 0.003605 0.4681

Detection Prevalence 0.06849 0.006179 0.13697 0.007724 0.7806

Balanced Accuracy 0.58834 0.519165 0.66145 0.519769 0.6483

pred 1 2 3 4 5

1 51 14 33 15 58

2 92 66 118 46 230

3 31 10 72 18 63

4 7 7 14 9 43

5 76 47 165 71 586

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 51 14 33 15 58

2 92 66 118 46 230

3 31 10 72 18 63

4 7 7 14 9 43

5 76 47 165 71 586

Overall Statistics

Accuracy : 0.4037

95% CI : (0.3818, 0.4259)

No Information Rate : 0.5046

P-Value [Acc > NIR] : 1

Kappa : 0.1453

Mcnemar's Test P-Value : <2e-16

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.19844 0.45833 0.17910 0.056604 0.5980

Specificity 0.92878 0.72970 0.92078 0.960179 0.6268

Pos Pred Value 0.29825 0.11957 0.37113 0.112500 0.6201

Neg Pred Value 0.88368 0.94388 0.81121 0.919441 0.6048

Prevalence 0.13234 0.07415 0.20700 0.081874 0.5046

Detection Rate 0.02626 0.03399 0.03708 0.004634 0.3018

Detection Prevalence 0.08805 0.28424 0.09990 0.041195 0.4866

Balanced Accuracy 0.56361 0.59402 0.54994 0.508392 0.6124

[1] "CHAID"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 60 29 44 17 56

2 0 0 0 0 0

3 13 12 84 7 45

4 0 0 0 2 0

5 184 103 274 133 879

Overall Statistics

Accuracy : 0.5278

95% CI : (0.5053, 0.5502)

No Information Rate : 0.5046

P-Value [Acc > NIR] : 0.02168

Kappa : 0.1567

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.2335 0.00000 0.20896 0.01258 0.8969

Specificity 0.9134 1.00000 0.95000 1.00000 0.2786

Pos Pred Value 0.2913 NaN 0.52174 1.00000 0.5588

Neg Pred Value 0.8865 0.92585 0.82145 0.91907 0.7263

Prevalence 0.1323 0.07415 0.20700 0.08187 0.5046

Detection Rate 0.0309 0.00000 0.04325 0.00103 0.4526

Detection Prevalence 0.1061 0.00000 0.08290 0.00103 0.8100

Balanced Accuracy 0.5734 0.50000 0.57948 0.50629 0.5878

[1] "CART"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 31 19 23 4 25

2 0 0 0 0 0

3 17 10 39 6 22

4 0 0 0 0 0

5 209 115 340 149 933

Overall Statistics

Accuracy : 0.5165

95% CI : (0.494, 0.5389)

No Information Rate : 0.5046

P-Value [Acc > NIR] : 0.1536

Kappa : 0.0865

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.12062 0.00000 0.09701 0.00000 0.9520

Specificity 0.95786 1.00000 0.96429 1.00000 0.1549

Pos Pred Value 0.30392 NaN 0.41489 NaN 0.5344

Neg Pred Value 0.87717 0.92585 0.80357 0.91813 0.7602

Prevalence 0.13234 0.07415 0.20700 0.08187 0.5046

Detection Rate 0.01596 0.00000 0.02008 0.00000 0.4804

Detection Prevalence 0.05252 0.00000 0.04840 0.00000 0.8991

Balanced Accuracy 0.53924 0.50000 0.53065 0.50000 0.5535

[1] "C50"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 72 26 33 16 72

2 11 22 11 4 16

3 47 29 195 21 87

4 4 2 9 15 25

5 123 65 154 103 780

Overall Statistics

Accuracy : 0.5582

95% CI : (0.5358, 0.5804)

No Information Rate : 0.5046

P-Value [Acc > NIR] : 1.28e-06

Kappa : 0.2892

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.28016 0.15278 0.4851 0.094340 0.7959

Specificity 0.91276 0.97664 0.8805 0.977566 0.5374

Pos Pred Value 0.32877 0.34375 0.5145 0.272727 0.6367

Neg Pred Value 0.89263 0.93504 0.8676 0.923688 0.7211

Prevalence 0.13234 0.07415 0.2070 0.081874 0.5046

Detection Rate 0.03708 0.01133 0.1004 0.007724 0.4016

Detection Prevalence 0.11277 0.03296 0.1952 0.028321 0.6308

Balanced Accuracy 0.59646 0.56471 0.6828 0.535953 0.6667

> train\_file\_action = "Training Data 1801 - Aggregated.csv";

+ test\_file\_action = "Test Data 1801 - Aggregated.csv";

+

+ modelRandomForest = hitungRandomForest(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungNaiveBayes(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungDecisionTree(train\_dataset1\_norm, test\_dataset1\_norm);

+

pred 1 2 3 4 5

1 80 20 21 9 31

2 10 16 6 2 10

3 28 20 187 14 43

4 3 2 11 21 9

5 142 63 124 103 833

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 80 20 21 9 31

2 10 16 6 2 10

3 28 20 187 14 43

4 3 2 11 21 9

5 142 63 124 103 833

Overall Statistics

Accuracy : 0.6289

95% CI : (0.6061, 0.6512)

No Information Rate : 0.5122

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

Kappa : 0.375

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.30418 0.13223 0.5358 0.14094 0.8996

Specificity 0.94757 0.98340 0.9280 0.98493 0.5102

Pos Pred Value 0.49689 0.36364 0.6404 0.45652 0.6585

Neg Pred Value 0.88889 0.94048 0.8931 0.92736 0.8287

Prevalence 0.14546 0.06692 0.1930 0.08241 0.5122

Detection Rate 0.04425 0.00885 0.1034 0.01162 0.4607

Detection Prevalence 0.08905 0.02434 0.1615 0.02544 0.6997

Balanced Accuracy 0.62588 0.55782 0.7319 0.56294 0.7049

pred 1 2 3 4 5

1 66 21 34 14 51

2 27 21 25 17 80

3 45 26 102 17 72

4 4 6 13 9 34

5 121 47 175 92 689

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 66 21 34 14 51

2 27 21 25 17 80

3 45 26 102 17 72

4 4 6 13 9 34

5 121 47 175 92 689

Overall Statistics

Accuracy : 0.4906

95% CI : (0.4673, 0.5139)

No Information Rate : 0.5122

P-Value [Acc > NIR] : 0.9684

Kappa : 0.1906

Mcnemar's Test P-Value : <2e-16

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.2510 0.17355 0.29226 0.060403 0.7441

Specificity 0.9223 0.91168 0.89034 0.965642 0.5068

Pos Pred Value 0.3548 0.12353 0.38931 0.136364 0.6130

Neg Pred Value 0.8785 0.93895 0.84023 0.919633 0.6535

Prevalence 0.1455 0.06692 0.19303 0.082412 0.5122

Detection Rate 0.0365 0.01162 0.05642 0.004978 0.3811

Detection Prevalence 0.1029 0.09403 0.14491 0.036504 0.6217

Balanced Accuracy 0.5866 0.54262 0.59130 0.513022 0.6254

[1] "CHAID"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 38 13 24 11 35

2 0 0 0 0 3

3 40 14 77 12 39

4 4 6 5 7 10

5 181 88 243 119 839

Overall Statistics

Accuracy : 0.5315

95% CI : (0.5082, 0.5547)

No Information Rate : 0.5122

P-Value [Acc > NIR] : 0.05222

Kappa : 0.1526

Mcnemar's Test P-Value : < 2e-16

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.14449 0.000000 0.22063 0.046980 0.9060

Specificity 0.94628 0.998222 0.92803 0.984931 0.2846

Pos Pred Value 0.31405 0.000000 0.42308 0.218750 0.5707

Neg Pred Value 0.86663 0.932964 0.83272 0.920045 0.7426

Prevalence 0.14546 0.066925 0.19303 0.082412 0.5122

Detection Rate 0.02102 0.000000 0.04259 0.003872 0.4640

Detection Prevalence 0.06692 0.001659 0.10066 0.017699 0.8131

Balanced Accuracy 0.54538 0.499111 0.57433 0.515955 0.5953

[1] "CART"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 42 8 24 8 20

2 0 0 0 0 0

3 17 10 62 9 32

4 0 0 0 0 0

5 204 103 263 132 874

Overall Statistics

Accuracy : 0.5409

95% CI : (0.5176, 0.5641)

No Information Rate : 0.5122

P-Value [Acc > NIR] : 0.007664

Kappa : 0.1362

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.15970 0.00000 0.17765 0.00000 0.9438

Specificity 0.96117 1.00000 0.95339 1.00000 0.2041

Pos Pred Value 0.41176 NaN 0.47692 NaN 0.5546

Neg Pred Value 0.87046 0.93308 0.82896 0.91759 0.7759

Prevalence 0.14546 0.06692 0.19303 0.08241 0.5122

Detection Rate 0.02323 0.00000 0.03429 0.00000 0.4834

Detection Prevalence 0.05642 0.00000 0.07190 0.00000 0.8717

Balanced Accuracy 0.56043 0.50000 0.56552 0.50000 0.5740

[1] "C50"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 57 22 19 13 54

2 22 22 16 5 18

3 46 17 164 26 82

4 5 2 9 16 21

5 133 58 141 89 751

Overall Statistics

Accuracy : 0.5586

95% CI : (0.5354, 0.5817)

No Information Rate : 0.5122

P-Value [Acc > NIR] : 4.176e-05

Kappa : 0.2805

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.21673 0.18182 0.46991 0.10738 0.8110

Specificity 0.93010 0.96384 0.88280 0.97770 0.5227

Pos Pred Value 0.34545 0.26506 0.48955 0.30189 0.6408

Neg Pred Value 0.87462 0.94261 0.87441 0.92422 0.7248

Prevalence 0.14546 0.06692 0.19303 0.08241 0.5122

Detection Rate 0.03153 0.01217 0.09071 0.00885 0.4154

Detection Prevalence 0.09126 0.04591 0.18529 0.02931 0.6482

Balanced Accuracy 0.57341 0.57283 0.67636 0.54254 0.6668

> train\_file\_action = "Training Data 1801 – Aggregated Balanced.csv";

+ test\_file\_action = "Test Data 1801 – Aggregated Balanced.csv";

+ modelRandomForest = hitungRandomForest(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungNaiveBayes(train\_dataset1\_norm, test\_dataset1\_norm);

+ hitungDecisionTree(train\_dataset1\_norm, test\_dataset1\_norm);

+

pred 1 2 3 4 5

1 94 17 20 11 35

2 13 14 8 3 9

3 26 19 193 17 36

4 2 2 11 18 14

5 138 73 138 97 807

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 94 17 20 11 35

2 13 14 8 3 9

3 26 19 193 17 36

4 2 2 11 18 14

5 138 73 138 97 807

Overall Statistics

Accuracy : 0.6204

95% CI : (0.5976, 0.6428)

No Information Rate : 0.4964

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

Kappa : 0.3737

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.34432 0.112000 0.5216 0.123288 0.8957

Specificity 0.94617 0.980473 0.9322 0.982624 0.5120

Pos Pred Value 0.53107 0.297872 0.6632 0.382979 0.6441

Neg Pred Value 0.89072 0.937217 0.8839 0.927602 0.8327

Prevalence 0.15041 0.068871 0.2039 0.080441 0.4964

Detection Rate 0.05179 0.007713 0.1063 0.009917 0.4446

Detection Prevalence 0.09752 0.025895 0.1603 0.025895 0.6904

Balanced Accuracy 0.64525 0.546237 0.7269 0.552956 0.7039

pred 1 2 3 4 5

1 64 16 37 16 59

2 45 22 31 14 77

3 45 22 91 17 65

4 8 12 15 8 48

5 111 53 196 91 652

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 64 16 37 16 59

2 45 22 31 14 77

3 45 22 91 17 65

4 8 12 15 8 48

5 111 53 196 91 652

Overall Statistics

Accuracy : 0.4612

95% CI : (0.438, 0.4844)

No Information Rate : 0.4964

P-Value [Acc > NIR] : 0.9988

Kappa : 0.1636

Mcnemar's Test P-Value : <2e-16

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.23443 0.17600 0.24595 0.054795 0.7236

Specificity 0.91699 0.90118 0.89689 0.950270 0.5066

Pos Pred Value 0.33333 0.11640 0.37917 0.087912 0.5911

Neg Pred Value 0.87123 0.93665 0.82286 0.919954 0.6503

Prevalence 0.15041 0.06887 0.20386 0.080441 0.4964

Detection Rate 0.03526 0.01212 0.05014 0.004408 0.3592

Detection Prevalence 0.10579 0.10413 0.13223 0.050138 0.6077

Balanced Accuracy 0.57571 0.53859 0.57142 0.502532 0.6151

[1] "CHAID"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 8 3 7 3 6

2 0 0 0 0 0

3 61 16 83 9 50

4 0 0 0 0 0

5 204 106 280 134 845

Overall Statistics

Accuracy : 0.5157

95% CI : (0.4924, 0.5389)

No Information Rate : 0.4964

P-Value [Acc > NIR] : 0.05265

Kappa : 0.1098

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.029304 0.00000 0.22432 0.00000 0.9378

Specificity 0.987678 1.00000 0.90588 1.00000 0.2079

Pos Pred Value 0.296296 NaN 0.37900 NaN 0.5386

Neg Pred Value 0.851790 0.93113 0.82018 0.91956 0.7724

Prevalence 0.150413 0.06887 0.20386 0.08044 0.4964

Detection Rate 0.004408 0.00000 0.04573 0.00000 0.4656

Detection Prevalence 0.014876 0.00000 0.12066 0.00000 0.8645

Balanced Accuracy 0.508491 0.50000 0.56510 0.50000 0.5729

[1] "CART"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 39 5 17 6 20

2 0 0 0 0 0

3 16 8 60 8 26

4 0 0 0 0 0

5 218 112 293 132 855

Overall Statistics

Accuracy : 0.5256

95% CI : (0.5023, 0.5488)

No Information Rate : 0.4964

P-Value [Acc > NIR] : 0.006851

Kappa : 0.1202

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.14286 0.00000 0.16216 0.00000 0.9489

Specificity 0.96887 1.00000 0.95986 1.00000 0.1740

Pos Pred Value 0.44828 NaN 0.50847 NaN 0.5311

Neg Pred Value 0.86458 0.93113 0.81732 0.91956 0.7756

Prevalence 0.15041 0.06887 0.20386 0.08044 0.4964

Detection Rate 0.02149 0.00000 0.03306 0.00000 0.4711

Detection Prevalence 0.04793 0.00000 0.06501 0.00000 0.8871

Balanced Accuracy 0.55586 0.50000 0.56101 0.50000 0.5615

[1] "C50"

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 87 23 33 12 74

2 16 21 21 10 28

3 40 27 161 17 66

4 13 4 9 12 34

5 117 50 146 95 699

Overall Statistics

Accuracy : 0.5399

95% CI : (0.5167, 0.5631)

No Information Rate : 0.4964

P-Value [Acc > NIR] : 0.0001136

Kappa : 0.2772

Mcnemar's Test P-Value : 1.501e-13

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.31868 0.16800 0.43514 0.082192 0.7758

Specificity 0.90791 0.95562 0.89619 0.964050 0.5536

Pos Pred Value 0.37991 0.21875 0.51768 0.166667 0.6314

Neg Pred Value 0.88272 0.93950 0.86104 0.923121 0.7147

Prevalence 0.15041 0.06887 0.20386 0.080441 0.4964

Detection Rate 0.04793 0.01157 0.08871 0.006612 0.3851

Detection Prevalence 0.12617 0.05289 0.17135 0.039669 0.6099

Balanced Accuracy 0.61330 0.56181 0.66566 0.523121 0.6647

EFFECT OF FEATURE SELECTION

+ importances$threshMDA[importances$MeanDecreaseAccuracy >= meanMDA] = 1;

+ importances$threshMDG[importances$MeanDecreaseGini >= meanMDA] = 1;

> featuresMDA %>% arrange(-MeanDecreaseAccuracy) %>% select(kolom,MeanDecreaseAccuracy) %>% as.data.frame();

kolom MeanDecreaseAccuracy

1 origin\_mean\_duration 39.42194

2 origin\_total\_duration 38.00041

3 origin\_median\_duration 37.30728

4 origin\_mean\_vote\_down 36.35652

5 origin\_total\_vote\_up 35.23317

6 translation\_n\_char 34.17344

7 origin\_mean\_n\_gram\_translation 34.14435

8 origin\_mean\_n\_char\_translation 33.73457

9 origin\_n\_char 32.80776

10 origin\_word\_entropy 30.46255

11 origin\_most\_common\_n\_char 29.43154

12 translation\_lv\_distance 29.13930

13 origin\_n\_diff\_translation 24.78504

14 origin\_num\_most\_common\_translation 23.15738

15 origin\_n\_show 23.05295

16 origin\_n\_user 22.06135

17 origin\_total\_vote\_down 21.78589

18 translation\_total\_duration 21.21700

19 translation\_n\_gram 19.36492

20 translation\_mean\_duration 18.95396

21 action\_duration 18.36732

22 origin\_n\_gram 18.12109

23 translation\_mean\_vote\_down 17.68911

24 translation\_median\_duration 17.63961

25 translation\_mean\_vote\_up 17.28689

26 origin\_readibility 16.92213

27 translation\_jaro\_distance 16.76472

28 origin\_most\_common\_n\_gram 16.42106

+ hitungRandomForest(train\_dataset1\_normMDA, test\_dataset1\_normMDA);

+

pred 1 2 3 4 5

1 88 20 25 9 35

2 11 23 3 2 6

3 23 22 234 11 43

4 7 7 11 34 6

5 128 72 129 103 890

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 88 20 25 9 35

2 11 23 3 2 6

3 23 22 234 11 43

4 7 7 11 34 6

5 128 72 129 103 890

Overall Statistics

Accuracy : 0.6535

95% CI : (0.6318, 0.6746)

No Information Rate : 0.5046

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

Kappa : 0.4267

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

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.34241 0.15972 0.5821 0.21384 0.9082

Specificity 0.94718 0.98776 0.9357 0.98261 0.5509

Pos Pred Value 0.49718 0.51111 0.7027 0.52308 0.6732

Neg Pred Value 0.90425 0.93622 0.8956 0.93340 0.8548

Prevalence 0.13234 0.07415 0.2070 0.08187 0.5046

Detection Rate 0.04531 0.01184 0.1205 0.01751 0.4583

Detection Prevalence 0.09114 0.02317 0.1715 0.03347 0.6807

Balanced Accuracy 0.64480 0.57374 0.7589 0.59823 0.7295

Call:

randomForest(formula = respon ~ ., data = train\_data, importance = TRUE)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 5

OOB estimate of error rate: 33.17%

Confusion matrix:

1 2 3 4 5 class.error

1 396 40 127 12 465 0.6192308

2 77 80 65 11 272 0.8415842

3 83 22 882 29 494 0.4158940

4 26 8 63 110 430 0.8273155

5 134 31 132 42 3695 0.0840357

+ importances$threshMDA[importances$MeanDecreaseAccuracy >= q85MDA] = 1;

+ importances$threshMDG[importances$MeanDecreaseGini >= q85MDA] = 1;

+

> featuresMDA %>% arrange(-MeanDecreaseAccuracy) %>% select(kolom, MeanDecreaseAccuracy) %>% as.data.frame();

+

kolom MeanDecreaseAccuracy

1 origin\_mean\_duration 39.42194

2 origin\_total\_duration 38.00041

3 origin\_median\_duration 37.30728

4 origin\_mean\_vote\_down 36.35652

5 origin\_total\_vote\_up 35.23317

6 translation\_n\_char 34.17344

7 origin\_mean\_n\_gram\_translation 34.14435

8 origin\_mean\_n\_char\_translation 33.73457

9 origin\_n\_char 32.80776

10 origin\_word\_entropy 30.46255

11 origin\_most\_common\_n\_char 29.43154

12 translation\_lv\_distance 29.13930

13 origin\_n\_diff\_translation 24.78504

14 origin\_num\_most\_common\_translation 23.15738

+ hitungRandomForest(train\_dataset1\_normMDA, test\_dataset1\_normMDA);

+

pred 1 2 3 4 5

1 105 16 32 10 47

2 15 36 11 10 29

3 24 22 250 13 57

4 18 9 15 41 22

5 95 61 94 85 825

Confusion Matrix and Statistics

Reference

Prediction 1 2 3 4 5

1 105 16 32 10 47

2 15 36 11 10 29

3 24 22 250 13 57

4 18 9 15 41 22

5 95 61 94 85 825

Overall Statistics

Accuracy : 0.6473

95% CI : (0.6256, 0.6685)

No Information Rate : 0.5046

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

Kappa : 0.4462

Mcnemar's Test P-Value : 3.075e-13

Statistics by Class:

Class: 1 Class: 2 Class: 3 Class: 4 Class: 5

Sensitivity 0.40856 0.25000 0.6219 0.25786 0.8418

Specificity 0.93769 0.96385 0.9247 0.96411 0.6518

Pos Pred Value 0.50000 0.35644 0.6831 0.39048 0.7112

Neg Pred Value 0.91224 0.94134 0.9036 0.93576 0.8018

Prevalence 0.13234 0.07415 0.2070 0.08187 0.5046

Detection Rate 0.05407 0.01854 0.1287 0.02111 0.4248

Detection Prevalence 0.10814 0.05201 0.1885 0.05407 0.5973

Balanced Accuracy 0.67312 0.60692 0.7733 0.61098 0.7468

Call:

randomForest(formula = respon ~ ., data = train\_data, importance = TRUE)

Type of random forest: classification

Number of trees: 500

No. of variables tried at each split: 3

OOB estimate of error rate: 34.74%

Confusion matrix:

1 2 3 4 5 class.error

1 394 60 136 39 411 0.6211538

2 71 113 72 26 223 0.7762376

3 120 54 927 52 357 0.3860927

4 37 29 65 174 332 0.7268446

5 204 81 189 126 3434 0.1487357