Unigrams

6-class, discriminative features

Top100 features

Naïve Bayes

Time taken to build model: 0.31 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 5005 26.0392 %

Incorrectly Classified Instances 14216 73.9608 %

Kappa statistic 0.1105

Mean absolute error 0.2601

Root mean squared error 0.3765

Relative absolute error 93.7294 %

Root relative squared error 101.0765 %

Total Number of Instances 19221

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.297 0.151 0.283 0.297 0.290 0.144 0.641 0.273 Joey

0.210 0.115 0.279 0.210 0.239 0.106 0.604 0.257 Ross

0.264 0.172 0.235 0.264 0.248 0.088 0.598 0.223 Chandler

0.184 0.101 0.238 0.184 0.208 0.093 0.616 0.211 Phoebe

0.246 0.153 0.234 0.246 0.240 0.091 0.610 0.228 Monica

0.345 0.198 0.284 0.345 0.311 0.137 0.622 0.287 Rachel

Weighted Avg. 0.260 0.150 0.260 0.260 0.258 0.111 0.615 0.248

=== Confusion Matrix ===

a b c d e f <-- classified as

953 445 507 310 462 531 | a = Joey

633 707 613 315 514 592 | b = Ross

500 431 844 318 490 619 | c = Chandler

398 256 491 519 468 683 | d = Phoebe

428 320 571 321 756 673 | e = Monica

451 372 570 394 540 1226 | f = Rachel

J-48

=== Summary ===

Correctly Classified Instances 4866 25.3161 %

Incorrectly Classified Instances 14355 74.6839 %

Kappa statistic 0.1023

Mean absolute error 0.2608

Root mean squared error 0.3964

Relative absolute error 93.9934 %

Root relative squared error 106.4246 %

Total Number of Instances 19221

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.268 0.154 0.259 0.268 0.264 0.113 0.592 0.232 Joey

0.271 0.156 0.271 0.271 0.271 0.115 0.586 0.236 Ross

0.316 0.222 0.222 0.316 0.261 0.083 0.574 0.200 Chandler

0.187 0.105 0.234 0.187 0.207 0.090 0.574 0.191 Phoebe

0.193 0.114 0.243 0.193 0.215 0.086 0.574 0.203 Monica

0.270 0.147 0.294 0.270 0.282 0.127 0.591 0.251 Rachel

Weighted Avg. 0.253 0.151 0.255 0.253 0.252 0.103 0.582 0.220

=== Confusion Matrix ===

a b c d e f <-- classified as

861 574 710 321 312 430 | a = Joey

594 916 733 346 359 426 | b = Ross

533 502 1012 304 385 466 | c = Chandler

420 430 591 525 381 468 | d = Phoebe

405 472 740 341 591 520 | e = Monica

508 492 775 409 408 961 | f = Rachel

Top100to200

**Naïve Bayes**

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 3875 20.1602 %

Incorrectly Classified Instances 15346 79.8398 %

Kappa statistic 0.0252

Mean absolute error 0.277

Root mean squared error 0.372

Relative absolute error 99.809 %

Root relative squared error 99.8733 %

Total Number of Instances 19221

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.012 0.013 0.133 0.012 0.022 -0.004 0.539 0.162 Phoebe

0.031 0.021 0.226 0.031 0.054 0.024 0.546 0.189 Chandler

0.162 0.124 0.218 0.162 0.186 0.043 0.538 0.203 Ross

0.770 0.729 0.193 0.770 0.309 0.036 0.547 0.228 Rachel

0.137 0.085 0.244 0.137 0.175 0.067 0.566 0.205 Joey

0.007 0.003 0.315 0.007 0.015 0.026 0.538 0.181 Monica

Weighted Avg. 0.202 0.177 0.222 0.202 0.134 0.033 0.546 0.196

=== Confusion Matrix ===

a b c d e f <-- classified as

34 53 338 2125 259 6 | a = Phoebe

33 98 382 2384 298 7 | b = Chandler

40 74 547 2407 289 17 | c = Ross

47 68 428 2735 263 12 | d = Rachel

66 70 425 2201 438 8 | e = Joey

35 70 387 2309 245 23 | f = Monica

**J48**

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 4017 20.899 %

Incorrectly Classified Instances 15204 79.101 %

Kappa statistic 0.0386

Mean absolute error 0.2731

Root mean squared error 0.3756

Relative absolute error 98.4358 %

Root relative squared error 100.8348 %

Total Number of Instances 19221

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.040 0.029 0.193 0.040 0.067 0.024 0.533 0.160 Phoebe

0.104 0.074 0.221 0.104 0.142 0.042 0.549 0.194 Chandler

0.132 0.081 0.259 0.132 0.175 0.069 0.529 0.206 Ross

0.635 0.596 0.195 0.635 0.298 0.031 0.543 0.218 Rachel

0.164 0.099 0.250 0.164 0.198 0.078 0.548 0.205 Joey

0.111 0.084 0.202 0.111 0.144 0.036 0.544 0.181 Monica

Weighted Avg. 0.209 0.171 0.220 0.209 0.175 0.047 0.541 0.195

=== Confusion Matrix ===

a b c d e f <-- classified as

114 201 207 1778 272 243 | a = Phoebe

80 334 254 1944 332 258 | b = Chandler

90 249 446 1938 347 304 | c = Ross

139 256 291 2255 333 279 | d = Rachel

87 251 278 1795 526 271 | e = Joey

80 222 248 1880 297 342 | f = Monica

**Top200-300**

**Naïve Bayes**

Time taken to build model: 0.15 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 3659 19.0365 %

Incorrectly Classified Instances 15562 80.9635 %

Kappa statistic 0.0089

Mean absolute error 0.2773

Root mean squared error 0.3723

Relative absolute error 99.9413 %

Root relative squared error 99.9478 %

Total Number of Instances 19221

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.050 0.035 0.219 0.050 0.081 0.028 0.545 0.194 Joey

0.869 0.858 0.187 0.869 0.307 0.012 0.525 0.198 Rachel

0.009 0.008 0.182 0.009 0.017 0.004 0.531 0.183 Chandler

0.107 0.085 0.211 0.107 0.142 0.030 0.522 0.190 Ross

0.007 0.004 0.242 0.007 0.014 0.015 0.535 0.179 Monica

0.000 0.000 0.250 0.000 0.001 0.004 0.530 0.158 Phoebe

Weighted Avg. 0.190 0.182 0.214 0.190 0.100 0.016 0.531 0.185

=== Confusion Matrix ===

a b c d e f <-- classified as

159 2733 31 268 17 0 | a = Joey

125 3087 26 293 21 1 | b = Rachel

126 2760 28 280 7 1 | c = Chandler

107 2867 27 362 11 0 | d = Ross

106 2629 24 287 22 1 | e = Monica

102 2459 18 222 13 1 | f = Phoebe

Binary classification

**Chandler v/s rest**

Time taken to build model: 0.19 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 15938 82.9197 %

Incorrectly Classified Instances 3283 17.0803 %

Kappa statistic 0.0163

Mean absolute error 0.2623

Root mean squared error 0.377

Relative absolute error 94.4622 %

Root relative squared error 101.1746 %

Total Number of Instances 19221

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.991 0.981 0.835 0.991 0.906 0.037 0.576 0.867 Other

0.019 0.009 0.300 0.019 0.036 0.037 0.576 0.208 Chandler

Weighted Avg. 0.829 0.819 0.746 0.829 0.761 0.037 0.576 0.758

=== Confusion Matrix ===

a b <-- classified as

15877 142 | a = Other

3141 61 | b = Chandler

**Joey v/s rest**

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 15367 79.949 %

Incorrectly Classified Instances 3854 20.051 %

Kappa statistic 0.1313

Mean absolute error 0.2865

Root mean squared error 0.3897

Relative absolute error 103.0261 %

Root relative squared error 104.5034 %

Total Number of Instances 19221

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.187 0.078 0.325 0.187 0.237 0.138 0.624 0.260 Joey

0.922 0.813 0.850 0.922 0.885 0.138 0.624 0.882 Other

Weighted Avg. 0.799 0.691 0.762 0.799 0.777 0.138 0.624 0.778

=== Confusion Matrix ===

a b <-- classified as

599 2609 | a = Joey

1245 14768 | b = Other

**Ross v/s rest**

Correctly Classified Instances 15497 80.6254 %

Incorrectly Classified Instances 3724 19.3746 %

Kappa statistic 0.1124

Mean absolute error 0.2909

Root mean squared error 0.3862

Relative absolute error 100.5037 %

Root relative squared error 101.5178 %

Total Number of Instances 19221

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.949 0.864 0.838 0.949 0.890 0.130 0.605 0.868 Other

0.136 0.051 0.362 0.136 0.197 0.130 0.605 0.258 Ross

Weighted Avg. 0.806 0.721 0.754 0.806 0.768 0.130 0.605 0.761

=== Confusion Matrix ===

a b <-- classified as

15039 808 | a = Other

2916 458 | b = Ross

Monica v/s rest

START ALL OVER AGAIN