Goal:

* Naudojant klusterius, patikrinti atskyrimo kokybę Known/Unknown

Kaip:

* Dataset: known (11452) ir unknow(11452)
* Features: paskutininis 54 prekių klasifikatoriaus sluoksnis (t.y. tikimybės)
* K-means suformuoti klasterius (nuo 2 iki 54)
* Klusteriams priskirti “dažniausią klasę”
* Pamatuoti naudojant confusion matricą, accuracy

|  |  |
| --- | --- |
| # clusters | Accuracy  Confusion matrix |
| 2 | Accuracy: 0.5071603213412504  [[10482 970]  [10318 1134]] |
| 3 | Accuracy: 0.5508208173244848  [[5035 6417]  [3871 7581]] |
| 5 | Accuracy: 0.6193677960181627  [[6062 5390]  [3328 8124]] |
| 8 | Accuracy: 0.5614739783443939  [[10005 1447]  [ 8597 2855]] |
| 13 | Accuracy: 0.6131243450925602  [[9679 1773]  [7088 4364]] |
| 21 | Accuracy: 0.6037809989521481  [[6779 4673]  [4402 7050]] |
| 34 | Accuracy: 0.6548637792525324  [[8403 3049]  [4856 6596]] |
| 54 | Accuracy: 0.**6843782745371988**  [[7634 3818]  [3411 8041]] |

Klusterių nestabilumas: - naudojant 1000 iteracijų K-means:

Cluster 0: 3818 (**60.0%**) known, 2549 (40.0%) unknown

Cluster 1: 7634 (**46.2%**) known, 8903 (53.8%) unknown

Cluster 0: 7364 (**48.1%**) known, 7950 (51.9%) unknown

Cluster 1: 4088 (**53.9%**) known, 3502 (46.1%) unknown

--Debug info

Accuracy: 0.5071603213412504

[[10482 970]

[10318 1134]]

Cluster 0: 1336 (71.7%) known, 528 (28.3%) unknown. Assigned class 1

Cluster 1: 6245 (51.5%) known, 5889 (48.5%) unknown. Assigned class 1

Cluster 2: 3871 (43.5%) known, 5035 (56.5%) unknown. Assigned class 0

Accuracy: 0.5508208173244848

[[5035 6417]

[3871 7581]]

Cluster 0: 755 (52.4%) known, 687 (47.6%) unknown. Assigned class 1

Cluster 1: 3328 (35.4%) known, 6062 (64.6%) unknown. Assigned class 0

Cluster 2: 1834 (60.6%) known, 1194 (39.4%) unknown. Assigned class 1

Cluster 3: 178 (66.2%) known, 91 (33.8%) unknown. Assigned class 1

Cluster 4: 5357 (61.0%) known, 3418 (39.0%) unknown. Assigned class 1

Accuracy: 0.6193677960181627

[[6062 5390]

[3328 8124]]

C:\Program Files\Python36\lib\site-packages\scipy\cluster\vq.py:525: UserWarning: One of the clusters is empty. Re-run kmeans with a different initialization.

warnings.warn("One of the clusters is empty. "

Cluster 0: 702 (47.4%) known, 778 (52.6%) unknown. Assigned class 0

Cluster 1: 6839 (46.7%) known, 7797 (53.3%) unknown. Assigned class 0

Cluster 2: 1011 (80.4%) known, 246 (19.6%) unknown. Assigned class 1

Cluster 3: 213 (54.5%) known, 178 (45.5%) unknown. Assigned class 1

Cluster 4: 415 (61.5%) known, 260 (38.5%) unknown. Assigned class 1

Cluster 5: 1056 (42.5%) known, 1430 (57.5%) unknown. Assigned class 0

Cluster 6: 1216 (61.4%) known, 763 (38.6%) unknown. Assigned class 1

Cluster 7: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Accuracy: 0.5614739783443939

[[10005 1447]

[ 8597 2855]]

Cluster 0: 2137 (75.2%) known, 706 (24.8%) unknown. Assigned class 1

Cluster 1: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 2: 2045 (45.8%) known, 2419 (54.2%) unknown. Assigned class 0

Cluster 3: 2082 (43.9%) known, 2658 (56.1%) unknown. Assigned class 0

Cluster 4: 447 (81.1%) known, 104 (18.9%) unknown. Assigned class 1

Cluster 5: 373 (78.4%) known, 103 (21.6%) unknown. Assigned class 1

Cluster 6: 218 (46.7%) known, 249 (53.3%) unknown. Assigned class 0

Cluster 7: 206 (78.3%) known, 57 (21.7%) unknown. Assigned class 1

Cluster 8: 796 (63.4%) known, 460 (36.6%) unknown. Assigned class 1

Cluster 9: 1448 (35.3%) known, 2659 (64.7%) unknown. Assigned class 0

Cluster 10: 665 (38.7%) known, 1054 (61.3%) unknown. Assigned class 0

Cluster 11: 630 (49.6%) known, 640 (50.4%) unknown. Assigned class 0

Cluster 12: 405 (54.1%) known, 343 (45.9%) unknown. Assigned class 1

Accuracy: 0.6131243450925602

[[9679 1773]

[7088 4364]]

Cluster 0: 202 (51.1%) known, 193 (48.9%) unknown. Assigned class 1

Cluster 1: 211 (70.1%) known, 90 (29.9%) unknown. Assigned class 1

Cluster 2: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 3: 1184 (49.2%) known, 1224 (50.8%) unknown. Assigned class 0

Cluster 4: 1081 (71.9%) known, 422 (28.1%) unknown. Assigned class 1

Cluster 5: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 6: 233 (80.3%) known, 57 (19.7%) unknown. Assigned class 1

Cluster 7: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 8: 2267 (39.6%) known, 3459 (60.4%) unknown. Assigned class 0

Cluster 9: 450 (60.2%) known, 297 (39.8%) unknown. Assigned class 1

Cluster 10: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 11: 1671 (54.0%) known, 1422 (46.0%) unknown. Assigned class 1

Cluster 12: 391 (66.4%) known, 198 (33.6%) unknown. Assigned class 1

Cluster 13: 189 (64.9%) known, 102 (35.1%) unknown. Assigned class 1

Cluster 14: 608 (52.4%) known, 552 (47.6%) unknown. Assigned class 1

Cluster 15: 346 (62.0%) known, 212 (38.0%) unknown. Assigned class 1

Cluster 16: 414 (73.1%) known, 152 (26.9%) unknown. Assigned class 1

Cluster 17: 415 (59.2%) known, 286 (40.8%) unknown. Assigned class 1

Cluster 18: 634 (53.9%) known, 543 (46.1%) unknown. Assigned class 1

Cluster 19: 205 (58.2%) known, 147 (41.8%) unknown. Assigned class 1

Cluster 20: 951 (31.2%) known, 2096 (68.8%) unknown. Assigned class 0

Accuracy: 0.6037809989521481

[[6779 4673]

[4402 7050]]

Cluster 0: 446 (56.7%) known, 340 (43.3%) unknown. Assigned class 1

Cluster 1: 430 (62.7%) known, 256 (37.3%) unknown. Assigned class 1

Cluster 2: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 3: 211 (70.6%) known, 88 (29.4%) unknown. Assigned class 1

Cluster 4: 216 (86.7%) known, 33 (13.3%) unknown. Assigned class 1

Cluster 5: 195 (75.0%) known, 65 (25.0%) unknown. Assigned class 1

Cluster 6: 202 (51.5%) known, 190 (48.5%) unknown. Assigned class 1

Cluster 7: 192 (52.5%) known, 174 (47.5%) unknown. Assigned class 1

Cluster 8: 381 (86.0%) known, 62 (14.0%) unknown. Assigned class 1

Cluster 9: 400 (39.5%) known, 613 (60.5%) unknown. Assigned class 0

Cluster 10: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 11: 504 (30.9%) known, 1128 (69.1%) unknown. Assigned class 0

Cluster 12: 666 (66.3%) known, 338 (33.7%) unknown. Assigned class 1

Cluster 13: 384 (66.3%) known, 195 (33.7%) unknown. Assigned class 1

Cluster 14: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 15: 553 (48.1%) known, 597 (51.9%) unknown. Assigned class 0

Cluster 16: 201 (83.8%) known, 39 (16.2%) unknown. Assigned class 1

Cluster 17: 2036 (33.6%) known, 4015 (66.4%) unknown. Assigned class 0

Cluster 18: 163 (56.2%) known, 127 (43.8%) unknown. Assigned class 1

Cluster 19: 206 (80.2%) known, 51 (19.8%) unknown. Assigned class 1

Cluster 20: 218 (75.2%) known, 72 (24.8%) unknown. Assigned class 1

Cluster 21: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 22: 411 (82.4%) known, 88 (17.6%) unknown. Assigned class 1

Cluster 23: 460 (79.2%) known, 121 (20.8%) unknown. Assigned class 1

Cluster 24: 398 (58.0%) known, 288 (42.0%) unknown. Assigned class 1

Cluster 25: 214 (35.7%) known, 385 (64.3%) unknown. Assigned class 0

Cluster 26: 189 (63.9%) known, 107 (36.1%) unknown. Assigned class 1

Cluster 27: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 28: 1149 (40.8%) known, 1665 (59.2%) unknown. Assigned class 0

Cluster 29: 399 (68.0%) known, 188 (32.0%) unknown. Assigned class 1

Cluster 30: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 31: 219 (96.1%) known, 9 (3.9%) unknown. Assigned class 1

Cluster 32: 205 (58.2%) known, 147 (41.8%) unknown. Assigned class 1

Cluster 33: 204 (74.2%) known, 71 (25.8%) unknown. Assigned class 1

Accuracy: 0.6548637792525324

[[8403 3049]

[4856 6596]]

Cluster 0: 143 (6.5%) known, 2051 (93.5%) unknown. Assigned class 0

Cluster 1: 192 (95.5%) known, 9 (4.5%) unknown. Assigned class 1

Cluster 2: 495 (31.7%) known, 1068 (68.3%) unknown. Assigned class 0

Cluster 3: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 4: 216 (71.3%) known, 87 (28.7%) unknown. Assigned class 1

Cluster 5: 202 (51.5%) known, 190 (48.5%) unknown. Assigned class 1

Cluster 6: 38 (35.5%) known, 69 (64.5%) unknown. Assigned class 0

Cluster 7: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 8: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 9: 195 (85.5%) known, 33 (14.5%) unknown. Assigned class 1

Cluster 10: 68 (76.4%) known, 21 (23.6%) unknown. Assigned class 1

Cluster 11: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 12: 189 (79.1%) known, 50 (20.9%) unknown. Assigned class 1

Cluster 13: 162 (66.7%) known, 81 (33.3%) unknown. Assigned class 1

Cluster 14: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 15: 233 (81.2%) known, 54 (18.8%) unknown. Assigned class 1

Cluster 16: 201 (83.8%) known, 39 (16.2%) unknown. Assigned class 1

Cluster 17: 412 (83.4%) known, 82 (16.6%) unknown. Assigned class 1

Cluster 18: 220 (70.5%) known, 92 (29.5%) unknown. Assigned class 1

Cluster 19: 1006 (59.5%) known, 684 (40.5%) unknown. Assigned class 1

Cluster 20: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 21: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 22: 206 (80.2%) known, 51 (19.8%) unknown. Assigned class 1

Cluster 23: 191 (53.2%) known, 168 (46.8%) unknown. Assigned class 1

Cluster 24: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 25: 182 (95.8%) known, 8 (4.2%) unknown. Assigned class 1

Cluster 26: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 27: 211 (71.3%) known, 85 (28.7%) unknown. Assigned class 1

Cluster 28: 211 (39.4%) known, 325 (60.6%) unknown. Assigned class 0

Cluster 29: 213 (79.2%) known, 56 (20.8%) unknown. Assigned class 1

Cluster 30: 417 (52.0%) known, 385 (48.0%) unknown. Assigned class 1

Cluster 31: 433 (55.2%) known, 352 (44.8%) unknown. Assigned class 1

Cluster 32: 189 (67.7%) known, 90 (32.3%) unknown. Assigned class 1

Cluster 33: 160 (58.8%) known, 112 (41.2%) unknown. Assigned class 1

Cluster 34: 151 (51.7%) known, 141 (48.3%) unknown. Assigned class 1

Cluster 35: 217 (61.3%) known, 137 (38.7%) unknown. Assigned class 1

Cluster 36: 427 (42.1%) known, 588 (57.9%) unknown. Assigned class 0

Cluster 37: 186 (65.0%) known, 100 (35.0%) unknown. Assigned class 1

Cluster 38: 424 (41.1%) known, 607 (58.9%) unknown. Assigned class 0

Cluster 39: 201 (43.0%) known, 266 (57.0%) unknown. Assigned class 0

Cluster 40: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 41: 499 (48.6%) known, 528 (51.4%) unknown. Assigned class 0

Cluster 42: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 43: 181 (42.1%) known, 249 (57.9%) unknown. Assigned class 0

Cluster 44: 450 (39.9%) known, 677 (60.1%) unknown. Assigned class 0

Cluster 45: 265 (75.5%) known, 86 (24.5%) unknown. Assigned class 1

Cluster 46: 227 (82.5%) known, 48 (17.5%) unknown. Assigned class 1

Cluster 47: 0 (0.0%) known, 0 (0.0%) unknown. Assigned class 0

Cluster 48: 199 (48.7%) known, 210 (51.3%) unknown. Assigned class 0

Cluster 49: 438 (53.6%) known, 379 (46.4%) unknown. Assigned class 1

Cluster 50: 205 (91.9%) known, 18 (8.1%) unknown. Assigned class 1

Cluster 51: 431 (81.2%) known, 100 (18.8%) unknown. Assigned class 1

Cluster 52: 143 (12.6%) known, 996 (87.4%) unknown. Assigned class 0

Cluster 53: 423 (84.1%) known, 80 (15.9%) unknown. Assigned class 1

Accuracy: 0.6843782745371988

[[7634 3818]

[3411 8041]]