Analiza finalnih odgovora

Merlin Andrija, Lisica Ivan

2024-01-24

Radimo analizu finalnih odgovora na 30 pitanja. Iz 29 konfiguracija, svaki od 30 odgovora sadrzi sljedece podatke koji su bitni za analizu: - time - blue score - rouge score: precision, recall, f1 - human evaluation: expression and logic, accuracy and relevance, overall quality and engagement

Učitamo podatke

```
#ucitvamo sve podatke u listu podatkovnih okvira
file_list <- list.files(path = "../answers", pattern = "*.json", full.names = TRUE)
config_list <- list()
#file_list
for (i in 1:length(file_list)) {
  file <- fromJSON(file_list[i])$answers
  file <- data.frame(file$question_number, file$time, file$evaluation$bleu_score, file$evaluation$rouge_
  config_list[[i]] <- file
}
```

Ispisujemo srednju vrijednost svih kategorija:

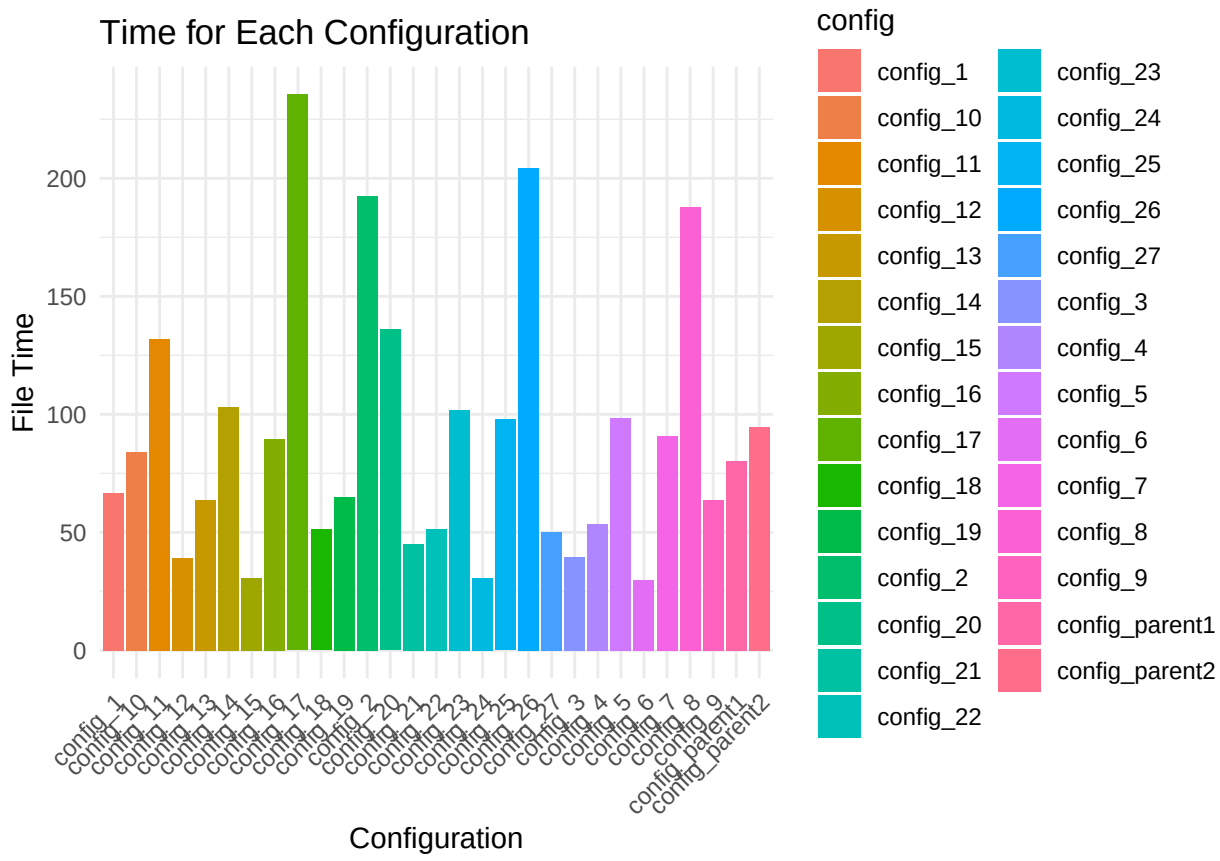
```
#ispisujemo srednju vrijednost svih kategorija
ocekivanje <- lapply(config_list, function(x) {
  sapply(x[,-1], function(y) {
    mean(as.numeric(y))
  })
})
ocekivanje <- as.data.frame(ocekivanje)
colnames(ocekivanje) <- c("config_1","config_10","config_11","config_12","config_13",
                          "config_14","config_15","config_16","config_17","config_18","config_19",
                          "config_2","config_20","config_21","config_22","config_23","config_24",
                          "config_25","config_26","config_27","config_3",
                          "config_4","config_5","config_6","config_7","config_8","config_9", "config_pa
                          "config_parent2")

ocekivanje
```

```
##                                config_1  config_10   config_11   config_12
## file.time                     66.5450000 84.1126667 132.0510000 39.13400000
## file.evaluation.bleu_score     0.1405987  0.1536526   0.1432737  0.09702275
## precision                      0.2719785  0.2771072   0.2826628  0.28091803
## recall                         0.3070516  0.3166548   0.3029405  0.25172807
## f1                             0.2607190  0.2551571   0.2586946  0.21283358
## file.evaluation.diversity      0.9770378  0.9451463   0.9450705  0.91675025
## expression_and_logic           4.7666667  4.6333333   4.5333333  4.50000000
```

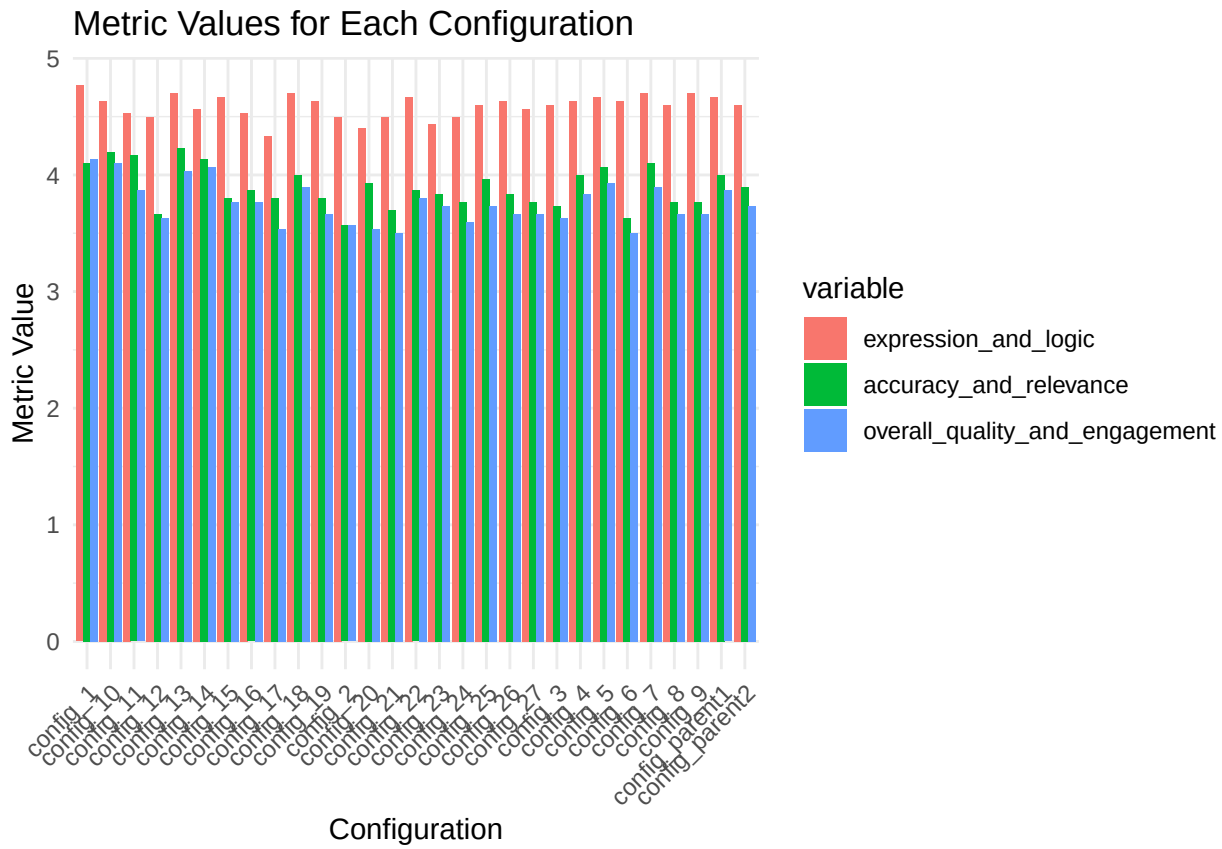
```
## accuracy_and_relevance          4.1000000  4.2000000   4.1666667  3.66666667
## overall_quality_and_engagement  4.1333333  4.1000000   3.8666667  3.63333333
##                                  config_13   config_14  config_15  config_16
## file.time                       63.8176667 103.2686667 30.7163333 89.4170000
## file.evaluation.bleu_score       0.1173203   0.1443221  0.1410021  0.1781632
## precision                        0.2732153   0.2913368  0.2760135  0.2890328
## recall                           0.2761418   0.3319495  0.2626451  0.3230017
## f1                               0.2360943   0.2787334  0.2294765  0.2698516
## file.evaluation.diversity        0.9165228   0.9710667  0.9217279  0.9445571
## expression_and_logic             4.7000000   4.5666667  4.6666667  4.5333333
## accuracy_and_relevance           4.2333333   4.1333333  3.8000000  3.8666667
## overall_quality_and_engagement   4.0333333   4.0666667  3.7666667  3.7666667
##                                  config_17  config_18  config_19    config_2
## file.time                       235.4840000 51.2470000 65.0566667 192.3570000
## file.evaluation.bleu_score       0.1684152  0.1149219  0.1241085   0.1333278
## precision                        0.3000582  0.2718421  0.2644181   0.2841638
## recall                           0.3167277  0.2599637  0.2754378   0.3032107
## f1                               0.2650807  0.2329508  0.2374296   0.2533930
## file.evaluation.diversity        0.9762983  0.9255431  0.9789061   0.9609918
## expression_and_logic             4.3333333  4.7000000  4.6333333   4.5000000
## accuracy_and_relevance           3.8000000  4.0000000  3.8000000   3.5666667
## overall_quality_and_engagement   3.5333333  3.9000000  3.6666667   3.5666667
##                                  config_20   config_21  config_22   config_23
## file.time                       135.9646667 44.96833333 51.6206667 101.8110000
## file.evaluation.bleu_score       0.1388235  0.08932692  0.1081324   0.1278615
## precision                        0.2876330  0.26179593  0.2599065   0.2929888
## recall                           0.3037571  0.25217184  0.2729479   0.3009453
## f1                               0.2552418  0.21463106  0.2266388   0.2637181
## file.evaluation.diversity        0.9453378  0.91773910  0.9185782   0.9861647
## expression_and_logic             4.4000000  4.50000000  4.6666667   4.4333333
## accuracy_and_relevance           3.9333333  3.70000000  3.8666667   3.8333333
## overall_quality_and_engagement   3.5333333  3.50000000  3.8000000   3.7333333
##                                  config_24  config_25   config_26   config_27
## file.time                       30.71866667 97.8340000 204.3756667 50.25766667
## file.evaluation.bleu_score       0.08178288 0.1368849   0.1435743  0.08550878
## precision                        0.25323358 0.2755484   0.2907648  0.25967801
## recall                           0.24609873 0.3048561   0.2989257  0.26200384
## f1                               0.21200125 0.2520362   0.2559777  0.21645963
## file.evaluation.diversity        0.92705780 0.9436299   0.9790664  0.92034195
## expression_and_logic             4.50000000 4.6000000   4.6333333  4.56666667
## accuracy_and_relevance           3.76666667 3.9666667   3.8333333  3.76666667
## overall_quality_and_engagement   3.60000000 3.7333333   3.6666667  3.66666667
##                                   config_3   config_4   config_5    config_6
## file.time                       39.7300000 53.5213333 98.3006667 29.95766667
## file.evaluation.bleu_score       0.0927931  0.1082026  0.1376816  0.07427984
## precision                        0.2777923  0.2661008  0.2797470  0.27373041
## recall                           0.2670281  0.2744313  0.2883045  0.24752817
## f1                               0.2155899  0.2337492  0.2479850  0.20761083
## file.evaluation.diversity        0.9132980  0.9177890  0.9545233  0.95684723
## expression_and_logic             4.6000000  4.6333333  4.6666667  4.63333333
## accuracy_and_relevance           3.7333333  4.0000000  4.0666667  3.63333333
## overall_quality_and_engagement   3.6333333  3.8333333  3.9333333  3.50000000
##                                   config_7    config_8    config_9
## file.time                       90.7496667 187.7790000 63.89900000
```

```
## file.evaluation.bleu_score      0.1534384   0.1527259  0.09703248
## precision                       0.2681331   0.2984042  0.27817471
## recall                          0.2902554   0.3157777  0.25693212
## f1                              0.2453083   0.2684531  0.21623200
## file.evaluation.diversity       0.9466677   0.9772701  0.91683640
## expression_and_logic            4.7000000   4.6000000  4.70000000
## accuracy_and_relevance          4.1000000   3.7666667  3.76666667
## overall_quality_and_engagement  3.9000000   3.6666667  3.66666667
##                                config_parent1 config_parent2
## file.time                          80.2300000     94.7403333
## file.evaluation.bleu_score          0.1169219      0.1285883
## precision                           0.3228032      0.2749796
## recall                              0.2976971      0.2858366
## f1                                  0.2661522      0.2410554
## file.evaluation.diversity           0.9419697      0.9062623
## expression_and_logic                4.6666667      4.6000000
## accuracy_and_relevance              4.0000000      3.9000000
## overall_quality_and_engagement      3.8666667      3.7333333
```

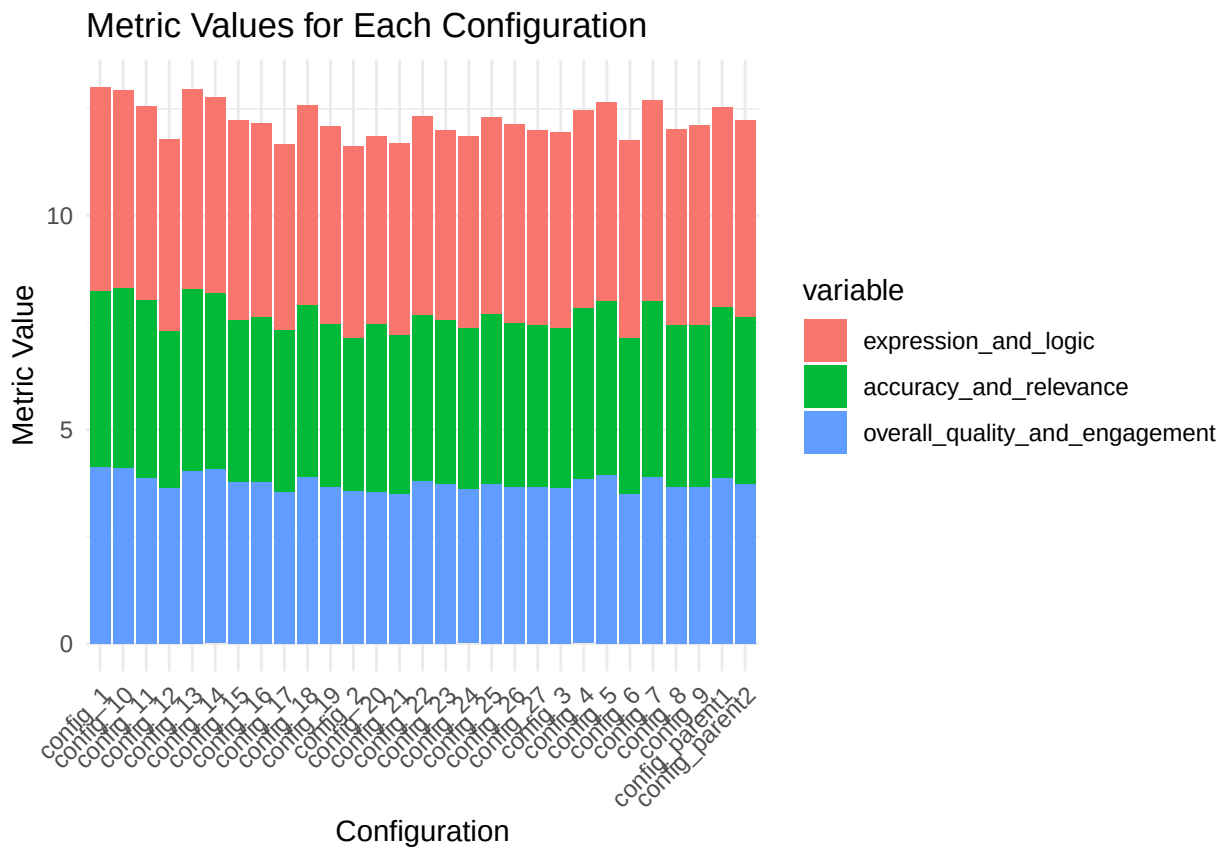
```r
library(ggplot2)

time_data <- data.frame(config = colnames(ocekivanje), value = as.numeric(ocekivanje["file.time", ]))

# Create a bar plot for time
ggplot(time_data, aes(x = config, y = value, fill = config)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Time for Each Configuration",
       x = "Configuration", y = "File Time") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```r
ocekivanje_last_three <- tail(ocekivanje, 3)


transposed_ocekivanje <- t(ocekivanje_last_three)
plot_data <- as.data.frame(transposed_ocekivanje)
plot_data$config <- rownames(plot_data)

melted_data <- reshape2::melt(plot_data, id.vars = "config")

# Create a bar plot
ggplot(melted_data, aes(x = config, y = value, fill = variable)) +
  geom_bar(stat = "identity",position = "dodge") +
  labs(title = "Metric Values for Each Configuration",
       x = "Configuration", y = "Metric Value") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```




```r
# Create a bar plot
ggplot(melted_data, aes(x = config, y = value, fill = variable)) +
  geom_bar(stat = "identity") +
  labs(title = "Metric Values for Each Configuration",
       x = "Configuration", y = "Metric Value") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Najmanje vrijeme: Nalazimo top 5 kandidata u bitnim kategorijama:

```
time_row <- ocekivanje['file.time', , drop = FALSE]
time_df <- as.data.frame(t(time_row))
colnames(time_df) <- "file.time"
ordered_df <- time_df[order(time_df$file.time), , drop = FALSE]
top_five_configs <- head(ordered_df, 5)
print(top_five_configs)
```

```
##           file.time
## config_6   29.95767
## config_15  30.71633
## config_24  30.71867
## config_12  39.13400
## config_3   39.73000
```

```
get_top_configurations <- function(metric) {
  metric_row <- ocekivanje[metric, , drop = FALSE]
  metric_df <- as.data.frame(t(metric_row))
  colnames(metric_df) <- metric
  ordered_df <- metric_df[order(-metric_df[[metric]]), , drop = FALSE]
  top_five_configs <- head(ordered_df, 5)
  print(top_five_configs)
}

get_top_configurations("file.evaluation.bleu_score")
```

```
##           file.evaluation.bleu_score
## config_16                  0.1781632
## config_17                  0.1684152
## config_10                  0.1536526
## config_7                   0.1534384
## config_8                   0.1527259
```

```
get_top_configurations("f1")
```

```
##                       f1
## config_14      0.2787334
## config_16      0.2698516
## config_8       0.2684531
## config_parent1 0.2661522
## config_17      0.2650807
```

```
get_top_configurations("expression_and_logic")
```

```
##           expression_and_logic
## config_1              4.766667
## config_13             4.700000
## config_18             4.700000
## config_7              4.700000
## config_9              4.700000
```

```
get_top_configurations("accuracy_and_relevance")
```

```
##           accuracy_and_relevance
## config_13               4.233333
## config_10               4.200000
## config_11               4.166667
## config_14               4.133333
## config_1                4.100000
```

```
get_top_configurations("overall_quality_and_engagement")
```

```
##           overall_quality_and_engagement
## config_1                        4.133333
## config_10                       4.100000
## config_14                       4.066667
## config_13                       4.033333
## config_5                        3.933333
```

Ispisujemo medijane svih kategorija:

```
#trazenje medijana
medijan <- lapply(config_list, function(x) {
  sapply(x[,-1], function(y) median(as.numeric(y)))
})
```

```r
medijan <- as.data.frame(medijan)
colnames(medijan) <- c("config_1","config_10","config_11","config_12","config_13",
                       "config_14","config_15","config_16","config_17","config_18","config_19",
                       "config_2","config_20","config_21","config_22","config_23","config_24",
                       "config_25","config_26","config_27","config_3",
                       "config_4","config_5","config_6","config_7","config_8","config_9", "config_pa
                       "config_parent2")
medijan
```

```
##                                    config_1   config_10    config_11
## file.time                       59.42500000 78.97500000 131.74000000
## file.evaluation.bleu_score       0.02542243  0.02226057   0.05731327
## precision                        0.24283204  0.25892857   0.26973684
## recall                           0.28062040  0.32452769   0.30510030
## f1                               0.22249381  0.21864184   0.24386571
## file.evaluation.diversity        0.98503192  0.98947140   0.99629630
## expression_and_logic             5.00000000  5.00000000   5.00000000
## accuracy_and_relevance           5.00000000  5.00000000   5.00000000
## overall_quality_and_engagement   5.00000000  5.00000000   4.00000000
##                                    config_12    config_13    config_14
## file.time                       3.759000e+01 6.186000e+01 104.17000000
## file.evaluation.bleu_score      2.221376e-78 1.919537e-78   0.07325613
## precision                       2.400000e-01 2.378205e-01   0.26562500
## recall                          1.904355e-01 2.664682e-01   0.36068852
## f1                              1.888074e-01 1.927390e-01   0.24558458
## file.evaluation.diversity       1.000000e+00 1.000000e+00   1.00000000
## expression_and_logic            5.000000e+00 5.000000e+00   5.00000000
## accuracy_and_relevance          4.000000e+00 5.000000e+00   5.00000000
## overall_quality_and_engagement  4.000000e+00 4.000000e+00   4.00000000
##                                    config_15   config_16    config_17
## file.time                       29.22500000 84.79500000 251.11500000
## file.evaluation.bleu_score       0.01048091  0.07960346   0.02154752
## precision                        0.23158602  0.23569648   0.25933442
## recall                           0.21691973  0.30949520   0.27546387
## f1                               0.18503453  0.21955134   0.21867023
## file.evaluation.diversity        1.00000000  0.98093682   0.98233696
## expression_and_logic             5.00000000  5.00000000   5.00000000
## accuracy_and_relevance           4.00000000  5.00000000   4.50000000
## overall_quality_and_engagement   4.00000000  4.00000000   4.00000000
##                                    config_18    config_19     config_2
## file.time                       49.070000000 64.410000000 192.74000000
## file.evaluation.bleu_score       0.006084812  0.007398061   0.05748891
## precision                        0.250000000  0.222222222   0.23495475
## recall                           0.255018598  0.240250928   0.28110203
## f1                               0.193059497  0.192315415   0.21344584
## file.evaluation.diversity        1.000000000  1.000000000   0.97560976
## expression_and_logic             5.000000000  5.000000000   5.00000000
## accuracy_and_relevance           5.000000000  4.000000000   4.00000000
## overall_quality_and_engagement   4.000000000  4.000000000   4.00000000
##                                    config_20    config_21    config_22
## file.time                       138.25500000 4.307500e+01 5.170500e+01
## file.evaluation.bleu_score        0.09933124 2.426020e-78 2.857509e-78
## precision                         0.29365796 2.613636e-01 2.424242e-01
```

```
## recall                              0.28950294 2.469737e-01 2.249784e-01
## f1                                  0.24723033 1.785567e-01 1.905129e-01
## file.evaluation.diversity           1.00000000 1.000000e+00 1.000000e+00
## expression_and_logic                5.00000000 5.000000e+00 5.000000e+00
## accuracy_and_relevance              5.00000000 5.000000e+00 4.500000e+00
## overall_quality_and_engagement      4.00000000 4.000000e+00 4.000000e+00
##                                      config_23    config_24   config_25
## file.time                          103.87000000 2.961000e+01 94.61000000
## file.evaluation.bleu_score           0.06766269 1.556302e-78  0.04197998
## precision                            0.32094671 2.277778e-01  0.24550000
## recall                               0.26027894 1.959096e-01  0.28719362
## f1                                   0.23138126 1.737314e-01  0.23083858
## file.evaluation.diversity            1.00000000 1.000000e+00  0.98651322
## expression_and_logic                 5.00000000 5.000000e+00  5.00000000
## accuracy_and_relevance               4.00000000 4.500000e+00  5.00000000
## overall_quality_and_engagement       4.00000000 4.000000e+00  4.00000000
##                                      config_26    config_27     config_3
## file.time                          1.944750e+02 4.784500e+01 38.950000000
## file.evaluation.bleu_score         1.919563e-78 2.533925e-78  0.006598912
## precision                          2.439103e-01 2.462121e-01  0.244680851
## recall                             2.596723e-01 2.445052e-01  0.246248019
## f1                                 2.095059e-01 1.845848e-01  0.184226045
## file.evaluation.diversity          1.000000e+00 1.000000e+00  0.988504228
## expression_and_logic               5.000000e+00 5.000000e+00  5.000000000
## accuracy_and_relevance             5.000000e+00 4.000000e+00  5.000000000
## overall_quality_and_engagement     4.000000e+00 4.000000e+00  4.500000000
##                                       config_4     config_5     config_6
## file.time                          5.177000e+01 100.13500000 2.833000e+01
## file.evaluation.bleu_score         2.241527e-78   0.04857035 2.781091e-78
## precision                          2.500000e-01   0.25000000 2.288082e-01
## recall                             2.342549e-01   0.24515909 1.802214e-01
## f1                                 2.074510e-01   0.21226002 1.717370e-01
## file.evaluation.diversity          1.000000e+00   1.00000000 1.000000e+00
## expression_and_logic               5.000000e+00   5.00000000 5.000000e+00
## accuracy_and_relevance             4.500000e+00   5.00000000 4.000000e+00
## overall_quality_and_engagement     4.000000e+00   4.00000000 4.000000e+00
##                                       config_7     config_8     config_9
## file.time                          87.29000000 186.44500000 61.180000000
## file.evaluation.bleu_score          0.01069184   0.02154752  0.007663644
## precision                           0.25351406   0.25935673  0.217592593
## recall                              0.26644739   0.28393457  0.199773756
## f1                                  0.20238129   0.21783012  0.192459061
## file.evaluation.diversity           0.98488499   0.97644124  1.000000000
## expression_and_logic                5.00000000   5.00000000  5.000000000
## accuracy_and_relevance              5.00000000   5.00000000  4.500000000
## overall_quality_and_engagement      4.00000000   4.00000000  4.000000000
##                                  config_parent1 config_parent2
## file.time                          76.55500000   9.159000e+01
## file.evaluation.bleu_score          0.03424114   2.950621e-78
## precision                           0.28203914   2.550000e-01
## recall                              0.28436099   2.589330e-01
## f1                                  0.24653551   2.081764e-01
## file.evaluation.diversity           0.98731884   9.846244e-01
## expression_and_logic                5.00000000   5.000000e+00
```

```
## accuracy_and_relevance             5.00000000   5.000000e+00
## overall_quality_and_engagement     4.50000000   4.000000e+00
```

```
get_top_configurations <- function(metric) {
  metric_row <- medijan[metric, , drop = FALSE]
  metric_df <- as.data.frame(t(metric_row))
  colnames(metric_df) <- metric
  ordered_df <- metric_df[order(-metric_df[[metric]]), , drop = FALSE]
  top_five_configs <- head(ordered_df, 5)
  print(top_five_configs)
}
get_top_configurations("file.evaluation.bleu_score")
```

```
##           file.evaluation.bleu_score
## config_20                  0.09933124
## config_16                  0.07960346
## config_14                  0.07325613
## config_23                  0.06766269
## config_2                   0.05748891
```

```
get_top_configurations("f1")
```

```
##                       f1
## config_20      0.2472303
## config_parent1 0.2465355
## config_14      0.2455846
## config_11      0.2438657
## config_23      0.2313813
```

```
get_top_configurations("expression_and_logic")
```

```
##           expression_and_logic
## config_1                     5
## config_10                    5
## config_11                    5
## config_12                    5
## config_13                    5
```

```
get_top_configurations("accuracy_and_relevance")
```

```
##           accuracy_and_relevance
## config_1                       5
## config_10                      5
## config_11                      5
## config_13                      5
## config_14                      5
```

```
get_top_configurations("overall_quality_and_engagement")
```

```
##                overall_quality_and_engagement
## config_1                                  5.0
## config_10                                 5.0
## config_3                                  4.5
## config_parent1                            4.5
## config_11                                 4.0
```

Ispisujemo devijaciju svih kategorija:

```
devijacija <- lapply(config_list, function(x) {
  sapply(x[,-1], function(y) sd(as.numeric(y)))
})
devijacija <- as.data.frame(devijacija)
colnames(devijacija) <- c("config_1","config_10","config_11","config_12","config_13",
                          "config_14","config_15","config_16","config_17","config_18","config_19",
                          "config_2","config_20","config_21","config_22","config_23","config_24",
                          "config_25","config_26","config_27","config_3",
                          "config_4","config_5","config_6","config_7","config_8","config_9", "config_pa
                          "config_parent2")
devijacija
```

```
##                                config_1  config_10  config_11 config_12
## file.time                     24.3703592 22.3506044 24.2520521 8.9517502
## file.evaluation.bleu_score     0.2148906  0.2340392  0.2131319 0.1801609
## precision                      0.1531531  0.1557025  0.1780753 0.1924494
## recall                         0.1426345  0.1739770  0.1556890 0.1580710
## f1                             0.1312359  0.1441396  0.1554100 0.1298674
## file.evaluation.diversity      0.0294613  0.1813786  0.1807555 0.2504123
## expression_and_logic           0.6789106  0.8502873  0.9732042 0.8200084
## accuracy_and_relevance         1.2134306  1.3235272  1.2887667 1.5387710
## overall_quality_and_engagement 1.1366416  1.2689936  1.2793677 1.4259500
##                                config_13   config_14 config_15  config_16
## file.time                     15.7208833 28.73788641 7.0380971 19.6054410
## file.evaluation.bleu_score     0.2102221  0.20180401 0.2419133  0.2547911
## precision                      0.1886987  0.17847523 0.1821236  0.1778661
## recall                         0.1583113  0.16481596 0.1683783  0.1592349
## f1                             0.1448081  0.15864204 0.1479024  0.1475895
## file.evaluation.diversity      0.2506364  0.06888484 0.2511672  0.1806625
## expression_and_logic           0.8366600  0.93526074 0.7580980  0.9732042
## accuracy_and_relevance         1.2228664  1.25212463 1.3493293  1.4793599
## overall_quality_and_engagement 1.1591713  1.14269293 1.3308886  1.4546793
##                                config_17 config_18   config_19    config_2
## file.time                     66.49522397 9.6091487 16.39378316 40.93448358
## file.evaluation.bleu_score     0.25431810 0.2046779  0.19665712  0.21943539
## precision                      0.21209563 0.1579069  0.16215796  0.20258588
## recall                         0.16509128 0.1541593  0.14261272  0.14808121
## f1                             0.16284627 0.1390158  0.13271732  0.14823304
## file.evaluation.diversity      0.02963121 0.2519935  0.03475271  0.06553105
## expression_and_logic           1.12444111 0.8769068  0.92785750  1.07478947
## accuracy_and_relevance         1.44794742 1.3390681  1.42393336  1.52413533
## overall_quality_and_engagement 1.38298361 1.2958821  1.26854066  1.45467933
##                                config_20  config_21  config_22   config_23
## file.time                     27.6924346 14.6594494 16.1157383 27.92106229
```

```
## file.evaluation.bleu_score      0.1778147  0.1447709  0.1950939  0.16540329
## precision                       0.1577645  0.1639074  0.1725175  0.16792908
## recall                          0.1366686  0.1553548  0.1658585  0.14667749
## f1                              0.1301802  0.1321097  0.1433744  0.14477487
## file.evaluation.diversity       0.1812621  0.2505034  0.2506501  0.02552553
## expression_and_logic            1.1626367  1.1064077  0.8441823  1.22286643
## accuracy_and_relevance          1.5070715  1.6220464  1.4076964  1.41624402
## overall_quality_and_engagement  1.3060425  1.4323841  1.2972119  1.33735027
##                                config_24  config_25   config_26  config_27
## file.time                      9.1118209 18.7066113 55.03366021 11.1293603
## file.evaluation.bleu_score     0.1581011  0.2177971  0.24984941  0.1622356
## precision                      0.1571146  0.1734162  0.20202564  0.1636947
## recall                         0.1537714  0.1461392  0.16368349  0.1563671
## f1                             0.1241443  0.1278260  0.15707447  0.1363080
## file.evaluation.diversity      0.2524506  0.1811336  0.03064871  0.2508710
## expression_and_logic           1.1064077  0.9321832  0.92785750  1.1043280
## accuracy_and_relevance         1.5241353  1.4735211  1.51049965  1.5465943
## overall_quality_and_engagement 1.5222488  1.3879614  1.39786372  1.4223180
##                                 config_3   config_4   config_5  config_6
## file.time                      9.2329085 17.6738121 27.8235585 7.6973021
## file.evaluation.bleu_score     0.1681985  0.1975667  0.1976047 0.1302931
## precision                      0.1957022  0.1782532  0.1775566 0.1693184
## recall                         0.1569257  0.1655716  0.1752388 0.1534862
## f1                             0.1290885  0.1444430  0.1632104 0.1084044
## file.evaluation.diversity      0.2505410  0.2505976  0.1812851 0.1814686
## expression_and_logic           0.8944272  0.8502873  0.9222661 0.9278575
## accuracy_and_relevance         1.5742176  1.2594471  1.2847469 1.5196037
## overall_quality_and_engagement 1.5421287  1.2058288  1.2298958 1.4562588
##                                  config_7    config_8  config_9 config_parent1
## file.time                      17.8831131 34.55049372 9.7356109     39.6819484
## file.evaluation.bleu_score      0.2503266  0.24677513 0.2014979      0.1966194
## precision                       0.1658059  0.20927287 0.1957350      0.2073107
## recall                          0.1567481  0.15785186 0.1626002      0.1402709
## f1                              0.1423094  0.16145074 0.1319328      0.1395609
## file.evaluation.diversity       0.1805313  0.02408152 0.2510172      0.1811236
## expression_and_logic            0.9153857  0.96846840 0.8366600      1.0283342
## accuracy_and_relevance          1.2958821  1.56873183 1.5687318      1.4383899
## overall_quality_and_engagement  1.3480510  1.44635888 1.4700066      1.3829836
##                                config_parent2
## file.time                         28.9972866
## file.evaluation.bleu_score         0.2230883
## precision                          0.1913444
## recall                             0.1702784
## f1                                 0.1598770
## file.evaluation.diversity          0.2509054
## expression_and_logic               0.9321832
## accuracy_and_relevance             1.4936648
## overall_quality_and_engagement     1.4840144
```

```r
mean_deviation <- apply(devijacija[, -1], 2, mean)

mean_deviation_df <- data.frame(Config = colnames(devijacija)[-1], Mean_Deviation = mean_deviation)

best_config <- mean_deviation_df[which.min(mean_deviation_df$Mean_Deviation), ]
```

```
print("Best Configuration:")
```

```
## [1] "Best Configuration:"
```

```
print(best_config)
```

```
##              Config Mean_Deviation
## config_15 config_15       1.274211
```

Ispisujemo razdiobu svih ocjena ljudske evaluacije:

```
human_eval <- lapply(config_list, function(x) {
  df <- data.frame(x$expression_and_logic, x$accuracy_and_relevance,
                   x$overall_quality_and_engagement)
  colnames(df) <- c("expression_and_logic", "accuracy_and_relevance",
                    "overall_quality_and_engagement")
  row.names(df) <- 1:30
  df
})
raspodjela <- lapply(human_eval, function(x) {
  sapply(x, function(raspodjela_ocjena) table(raspodjela_ocjena))
})
names(raspodjela) <- c("config_1","config_10","config_11","config_12","config_13",
                       "config_14","config_15","config_16","config_17","config_18","config_19",
                       "config_2","config_20","config_21","config_22","config_23","config_24",
                       "config_25","config_26","config_27","config_3",
                       "config_4","config_5","config_6","config_7","config_8","config_9", "config_pa
                       "config_parent2")
raspodjela
```

```
## $config_1
## $config_1$expression_and_logic
## raspodjela_ocjena
##  2  3  4  5
##  1  1  2 26
##
## $config_1$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  1  3  5  4 17
##
## $config_1$overall_quality_and_engagement
## raspodjela_ocjena
##  2  3  4  5
##  4  5  4 17
##
##
## $config_10
## $config_10$expression_and_logic
## raspodjela_ocjena
##  2  3  4  5
```

```
##  2  1  3 24
##
## $config_10$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  2  3  2  3 20
##
## $config_10$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  1  5  1  6 17
##
##
## $config_11
##   expression_and_logic accuracy_and_relevance overall_quality_and_engagement
## 1                    1                      2                               2
## 2                    1                      2                               3
## 3                    1                      4                               5
## 4                    5                      3                               7
## 5                   22                     19                              13
##
## $config_12
## $config_12$expression_and_logic
## raspodjela_ocjena
##  2  3  4  5
##  1  3  6 20
##
## $config_12$accuracy_and_relevance
## raspodjela_ocjena
##  1  3  4  5
##  6  5  6 13
##
## $config_12$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  4  3  4  8 11
##
##
## $config_13
## $config_13$expression_and_logic
## raspodjela_ocjena
##  1  3  4  5
##  1  1  3 25
##
## $config_13$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  2  1  4  4 19
##
## $config_13$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  1  3  4  8 14
##
```

```
##
## $config_14
## $config_14$expression_and_logic
## raspodjela_ocjena
##  1  2  4  5
##  1  1  6 22
##
## $config_14$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  1  4  3  4 18
##
## $config_14$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  1  3  3  9 14
##
##
## $config_15
## $config_15$expression_and_logic
## raspodjela_ocjena
##  2  3  4  5
##  1  2  3 24
##
## $config_15$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  3  2  6  6 13
##
## $config_15$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  2  5  3  8 12
##
##
## $config_16
##   expression_and_logic accuracy_and_relevance overall_quality_and_engagement
## 1                    1                      3                               3
## 2                    1                      4                               5
## 3                    1                      4                               2
## 4                    5                      2                               6
## 5                   22                     17                              14
##
## $config_17
##   expression_and_logic accuracy_and_relevance overall_quality_and_engagement
## 1                    1                      3                               4
## 2                    2                      4                               3
## 3                    3                      4                               5
## 4                    4                      4                               9
## 5                   20                     15                               9
##
## $config_18
## $config_18$expression_and_logic
## raspodjela_ocjena
```

```
##  1  3  4  5
##  1  2  1 26
##
## $config_18$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  3  1  5  5 16
##
## $config_18$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  2  3  5  6 14
##
##
## $config_19
## $config_19$expression_and_logic
## raspodjela_ocjena
##  1  2  4  5
##  1  1  4 24
##
## $config_19$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  3  4  3  6 14
##
## $config_19$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  1  7  3  9 10
##
##
## $config_2
##   expression_and_logic accuracy_and_relevance overall_quality_and_engagement
## 1                    1                     5                               4
## 2                    2                     3                               4
## 3                    1                     4                               4
## 4                    3                     6                               7
## 5                   23                    12                              11
##
## $config_20
##   expression_and_logic accuracy_and_relevance overall_quality_and_engagement
## 1                    2                     4                               3
## 2                    1                     3                               5
## 3                    1                     1                               2
## 4                    5                     5                              13
## 5                   21                    17                               7
##
## $config_21
## $config_21$expression_and_logic
## raspodjela_ocjena
##  1  3  4  5
##  2  2  3 23
##
## $config_21$accuracy_and_relevance
```

```
## raspodjela_ocjena
##  1  2  3  4  5
##  6  1  5  2 16
##
## $config_21$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  4  4  5  7 10
##
##
## $config_22
## $config_22$expression_and_logic
## raspodjela_ocjena
##  1  3  4  5
##  1  1  4 24
##
## $config_22$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  3  3  4  5 15
##
## $config_22$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  1  6  4  6 13
##
##
## $config_23
## $config_23$expression_and_logic
## raspodjela_ocjena
##  1  4  5
##  3  5 22
##
## $config_23$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  3  4  2  7 14
##
## $config_23$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  4  1  4 11 10
##
##
## $config_24
## $config_24$expression_and_logic
## raspodjela_ocjena
##  1  3  4  5
##  2  2  3 23
##
## $config_24$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  5  1  5  4 15
```

```
##
## $config_24$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  4  5  3  5 13
##
##
## $config_25
## $config_25$expression_and_logic
## raspodjela_ocjena
##  1  2  4  5
##  1  1  5 23
##
## $config_25$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  3  4  2  3 18
##
## $config_25$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  2  6  3  6 13
##
##
## $config_26
## $config_26$expression_and_logic
## raspodjela_ocjena
##  1  2  4  5
##  1  1  4 24
##
## $config_26$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  3  5  3  2 17
##
## $config_26$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  4  2  5  8 11
##
##
## $config_27
## $config_27$expression_and_logic
## raspodjela_ocjena
##  1  3  4  5
##  2  2  1 25
##
## $config_27$accuracy_and_relevance
## raspodjela_ocjena
##  1  3  4  5
##  6  3  7 14
##
## $config_27$overall_quality_and_engagement
## raspodjela_ocjena
```

```
##  1  2  3  4  5
##  4  3  3  9 11
##
##
## $config_3
## $config_3$expression_and_logic
## raspodjela_ocjena
##  1  3  4  5
##  1  2  4 23
##
## $config_3$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  5  2  5  2 16
##
## $config_3$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  4  4  6  1 15
##
##
## $config_4
## $config_4$expression_and_logic
## raspodjela_ocjena
##  1  3  4  5
##  1  1  5 23
##
## $config_4$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  2  2  5  6 15
##
## $config_4$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  1  4  6  7 12
##
##
## $config_5
## $config_5$expression_and_logic
## raspodjela_ocjena
##  1  2  4  5
##  1  1  3 25
##
## $config_5$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  1  5  2  5 17
##
## $config_5$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  1  5  2  9 13
##
```

```
##
## $config_6
## $config_6$expression_and_logic
## raspodjela_ocjena
##  1  2  4  5
##  1  1  4 24
##
## $config_6$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  5  2  5  5 13
##
## $config_6$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  4  5  3  8 10
##
##
## $config_7
## $config_7$expression_and_logic
## raspodjela_ocjena
##  1  2  4  5
##  1  1  2 26
##
## $config_7$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  1  5  2  4 18
##
## $config_7$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  2  5  1  8 14
##
##
## $config_8
##   expression_and_logic accuracy_and_relevance overall_quality_and_engagement
## 1                    1                      4                               4
## 2                    1                      4                               2
## 3                    1                      4                               7
## 4                    3                      1                               4
## 5                   24                     17                              13
##
## $config_9
## $config_9$expression_and_logic
## raspodjela_ocjena
##  1  3  4  5
##  1  1  3 25
##
## $config_9$accuracy_and_relevance
## raspodjela_ocjena
##  1  3  4  5
##  6  4  5 15
##
```

```
## $config_9$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  4  4  2  8 12
##
##
## $config_parent1
## $config_parent1$expression_and_logic
## raspodjela_ocjena
##  1  4  5
##  2  2 26
##
## $config_parent1$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  2  5  3  1 19
##
## $config_parent1$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  2  5  3  5 15
##
##
## $config_parent2
## $config_parent2$expression_and_logic
## raspodjela_ocjena
##  1  3  4  5
##  1  3  2 24
##
## $config_parent2$accuracy_and_relevance
## raspodjela_ocjena
##  1  2  3  4  5
##  4  2  4  3 17
##
## $config_parent2$overall_quality_and_engagement
## raspodjela_ocjena
##  1  2  3  4  5
##  5  1  4  7 13
```

Ispisujemo koleraciju izmedu svake ocjenjene ljudske evaluacije i ukupne ocjene te koleraciju izmed vremena i ukupne ocjene:

```
koleracije <- lapply(config_list, function(x) {
  c(cor(as.numeric(x$expression_and_logic), as.numeric(x$overall_quality_and_engagement), method = "spea
    cor(as.numeric(x$accuracy_and_relevance), as.numeric(x$overall_quality_and_engagement), method ="sp
    cor(as.numeric(x$file.time), as.numeric(x$overall_quality_and_engagement), method = "spearman"))
})
koleracije <- as.data.frame(koleracije)
colnames(koleracije) <- c("config_1","config_10","config_11","config_12","config_13",
                          "config_14","config_15","config_16","config_17","config_18","config_19",
                          "config_2","config_20","config_21","config_22","config_23","config_24",
                          "config_25","config_26","config_27","config_3",
                          "config_4","config_5","config_6","config_7","config_8","config_9", "config_pa
                          "config_parent2")
```

```r
rownames(koleracije) <- c("expression_and_logic", "accuracy_and_relevance", "time")
koleracije
```

```
##                          config_1 config_10 config_11 config_12 config_13
## expression_and_logic    0.1958589 0.4646474 0.3216731 0.5450020 0.3270750
## accuracy_and_relevance  0.9386329 0.8965279 0.7869784 0.8224653 0.7587008
## time                    0.3931346 0.1892929 0.3597626 0.1180713 0.4558813
##                         config_14 config_15 config_16 config_17 config_18
## expression_and_logic    0.4125355 0.4869017 0.2728212 0.4057439 0.4231822
## accuracy_and_relevance  0.8061973 0.7780888 0.8206333 0.8909898 0.8057374
## time                    0.1657374 0.2886121 0.1068642 0.5682019 0.3581841
##                         config_19   config_2 config_20 config_21 config_22
## expression_and_logic    0.4230772 0.54884140 0.5163717 0.5399917 0.1006677
## accuracy_and_relevance  0.9081336 0.80923480 0.8141238 0.8540053 0.8685576
## time                    0.3241449 0.02857601 0.2266387 0.2620028 0.5132707
##                         config_23 config_24  config_25 config_26 config_27
## expression_and_logic    0.4026360 0.3966698 0.50301687 0.3857512 0.5133193
## accuracy_and_relevance  0.8183079 0.7409841 0.87351960 0.8581697 0.7898789
## time                    0.3085555 0.1798687 0.03185115 0.3753246 0.2219240
##                          config_3  config_4  config_5  config_6  config_7
## expression_and_logic    0.5297032 0.3939337 0.3695165 0.5147212 0.4690600
## accuracy_and_relevance  0.8486391 0.8906775 0.7636609 0.8169518 0.8194732
## time                    0.2878218 0.4573312 0.2743815 0.2027027 0.1754766
##                          config_8  config_9 config_parent1 config_parent2
## expression_and_logic    0.3623771 0.6195714      0.5101792      0.4880169
## accuracy_and_relevance  0.8689157 0.8280466      0.8717397      0.8601826
## time                    0.4710916 0.1895204      0.2319701      0.0656723
```

```r
correlation_sums <- colSums(abs(koleracije), na.rm = TRUE)

top_configs <- names(sort(correlation_sums, decreasing = TRUE)[1:5])

print(top_configs)
```

```
## [1] "config_17" "config_4"  "config_8"  "config_3"  "config_21"
```