Doğal dil projesi rapor

Projede, en sık geçen 20 kelimeyi collections kütüphanesinin counter fonksiyonuyla belirledim. Ayrıca, Word2Vec kullanarak bu 20 kelimeye benzer 5 kelimeyi saptadım. Aşağıda belirtilen kütüphaneleri kullanarak rastgele seçtiğim 5 cümleye benzer 3 cümle bulma adımında ise TfidfVectorizer ve cosine_similarity metriklerini kullanmayı tercih ettim.

En çok geçen 20 kelime:

covid: 1803 kez
new: 529 kez
people: 526 kez
cases: 506 kez
deaths: 356 kez
us: 354 kez
pandemic: 341 kez
vaccine: 308 kez
health: 288 kez
get: 270 kez
like: 257 kez
dont: 217 kez
still: 215 kez
one: 213 kez
many: 182 kez
vaccines: 178 kez
total: 174 kez
vaccinated: 171 kez
would: 167 kez
time: 165 kez

En benzer 5 farklı kelimeler

kelime_analizi.py dosyasının çıktısı aşağıdaki gibidir: covid

kelimesine benzer kelimeler:
people: 0.9980438351631165
us: 0.9980133175849915
pandemic: 0.9977877140045166

like: 0.9977747201919556
one: 0.9976652264595032

new kelimesine benzer kelimeler:
day: 0.9907505512237549
deaths: 0.9874858260154724
today: 0.9846656322479248
covid: 0.984281599521637
reported: 0.9828744530677795

people kelimesine benzer kelimeler:
covid: 0.9980437755584717
pandemic: 0.9975391626358032
would: 0.9972959160804749
us: 0.9971497654914856
health: 0.9970716834068298

cases kelimesine benzer kelimeler:
deaths: 0.9868499636650085
new: 0.9824338555335999
day: 0.9814876317977905
today: 0.9760556221008301
reported: 0.9733312726020813

deaths kelimesine benzer kelimeler:
new: 0.9874858260154724
day: 0.987414538860321
cases: 0.9868499040603638
pandemic: 0.9853182435035706
today: 0.985115647315979

us kelimesine benzer kelimeler:
covid: 0.9980132579803467
people: 0.9971497058868408
pandemic: 0.9969032406806946
one: 0.9966102242469788
like: 0.9965512156486511

pandemic kelimesine benzer kelimeler:
covid: 0.997787594795227
people: 0.997539222240448
us: 0.9969032406806946

one: 0.9966307282447815
like: 0.9965562224388123

vaccine kelimesine benzer kelimeler:
covid: 0.9973947405815125
like: 0.9969877004623413
people: 0.9967899322509766
time: 0.9967790842056274
many: 0.9963980913162231

health kelimesine benzer kelimeler:
covid: 0.9971373081207275
people: 0.9970716238021851
us: 0.9963895082473755
like: 0.9963625073432922
get: 0.9959485530853271

get kelimesine benzer kelimeler:
covid: 0.9970259070396423
people: 0.9965485334396362
like: 0.9963463544845581
dont: 0.996117115020752
would: 0.996097207069397

like kelimesine benzer kelimeler:
covid: 0.997774600982666
still: 0.9971526265144348
vaccine: 0.996987521648407
people: 0.9969576597213745
one: 0.9967014789581299

dont kelimesine benzer kelimeler:
people: 0.996996283531189
covid: 0.9967563152313232
like: 0.9966878890991211
would: 0.9963042140007019
get: 0.9961171746253967

still kelimesine benzer kelimeler:
covid: 0.9976348876953125
like: 0.9971527457237244
people: 0.9969191551208496

even: 0.9962649345397949
pandemic: 0.9961618185043335

one kelimesine benzer kelimeler:
covid: 0.9976649880409241
people: 0.9967107772827148
like: 0.9967015385627747
pandemic: 0.9966306686401367
us: 0.9966101050376892

many kelimesine benzer kelimeler:
covid: 0.997198760509491
people: 0.9965128898620605
vaccine: 0.9963980317115784
pandemic: 0.9962214231491089
like: 0.9962148666381836

vaccines kelimesine benzer kelimeler:
covid: 0.9963565468788147
vaccine: 0.9957613945007324
people: 0.9953745603561401
like: 0.9951885938644409
us: 0.9948962330818176

total kelimesine benzer kelimeler:
new: 0.9788808226585388
reported: 0.9738929271697998
deaths: 0.9711862802505493
today: 0.9676193594932556
day: 0.9666181802749634

vaccinated kelimesine benzer kelimeler:
many: 0.9895797371864319
vaccine: 0.9891900420188904
like: 0.988166093826294
people: 0.9879657030105591
would: 0.9876643419265747

would kelimesine benzer kelimeler:
people: 0.9972959756851196
covid: 0.9971023201942444
one: 0.9965130090713501

dont: 0.9963041543960571
like: 0.9962629675865173

time kelimesine benzer kelimeler:
covid: 0.9969623684883118
vaccine: 0.9967789649963379
like: 0.9963994026184082
people: 0.9963671565055847
pandemic: 0.9963542819023132

En benzer 3 farklı cümle

cümle.py dosyasının çıktısı:

Orjinal Cümle 1: "test distinguish sarscov influenza b respiratory syncytial virus" (222. cümle)
En Benzer 3 Cümle:

1. (2370. cümle) decided start counting influenza infections influenza clue precovid influenza influenza infection
2. (9104. cümle) protective role combined polyphenols micronutrients influenza virus sarscov infection vitro
3. (9133. cümle) integrated respiratory diseases dashboard tracking respiratory virus activity updated

Orjinal Cümle 2: vaccine attitudes covid vaccine intention among parents children kidney disease primary hypertension(295. cümle)
En Benzer 3 Cümle:

1. (273. cümle) vaccine opinions survey september studied attitudes people indicated hesitancy uncertainty towards getting refused get covid vaccine
2. (7089. cümle) however covid vaccine actual vaccine
3. (455. cümle) would like acknowledge patience parents worked hard protect children waiting vaccine rollout

Orjinal Cümle 3: thought pandemic wasnt real covid trick played pharma(368. cümle)
En Benzer 3 Cümle:

1. (3105. cümle) wasnt funny thought would
2. (2630. cümle) vaccines manufacturers played havoc
3. (8490. cümle) wasnt covid list

Orjinal Cümle 4: efficacy safety prophylaxis systematic review metaanalysis randomized trials(582. cümle)

En Benzer 3 Cümle:

1. (101. cümle) results consistent analysis randomized trials hospitalized patients
2. (6916. cümle) parenchymal lung abnormalities following hospitalisation covid viral pneumonitis systematic review metaanalysis thoraxbmj
3. (2928. cümle) taking time ensure safety efficacy quality control

Orjinal Cümle 5: ask health care provider treatment(7558. cümle)
En Benzer 3 Cümle:

1. (1650. cümle) test positive covid call health care provider ask treatment options
2. (5179. cümle) soon feel sick talk health care provider visit
3. (2638. cümle) dont provider call covid

Veri setindeki ilk 100 cümlenin duygu analizi:

Text: 1 mark calendars
Sentiment: 2, Confidence: 0.2685595452785492

Text: 2 covidian cult world like plague threatening everyone
Sentiment: 0, Confidence: 0.37734878063201904

Text: 3 status id
Sentiment: 2, Confidence: 0.274911105632782

Text: 4 reboot post pandemic
Sentiment: 1, Confidence: 0.2432985156774521

Text: 5 plot admissions diagnoses vs deaths
Sentiment: 0, Confidence: 0.2706371247768402

Text: 6 well anyone paying attention also
Sentiment: 1, Confidence: 0.2937726676464081

Text: 7 remember told wed well keep covid deaths k today found k died march june care homes following unlawful policy im sure still defend govt deny
Sentiment: 0, Confidence: 0.6868210434913635

Text: 8 equaling severe sickness
Sentiment: 1, Confidence: 0.2764244079589844

Text: 9 debt deficits tripled b covid thanks u treasurer followed dunce kooyong
Sentiment: 0, Confidence: 0.582563579082489

Text: 10 one week rotation months december last year company lower literacy sites need physical support guys
Sentiment: 0, Confidence: 0.45860156416893005

Text: 11 stuck bed ill spending days

Sentiment: 1, Confidence: 0.3503963053226471

Text: 12 people biohazard suits due covid
Sentiment: 0, Confidence: 0.3977408110954285

Text: 13 new tool cojac early detection genomic surveillance sarscov including low frequency potentially emerging variants
Sentiment: 3, Confidence: 0.3284643590450287

Text: 14 biden deliberately overwhelmed us million disease ridden illegals refugees
Sentiment: 0, Confidence: 0.6987576484680176

Text: 15 essay one difficult things ever written
Sentiment: 2, Confidence: 0.23944076895713806

Text: 16 wheres wreath kristinoem
Sentiment: 1, Confidence: 0.2650350332260132

Text: 17 take care
Sentiment: 2, Confidence: 0.29167118668556213

Text: 18 objective evaluations smell taste acuity guide appropriate management treatment post patients persistent sensory distortion according new data
Sentiment: 3, Confidence: 0.29742398858070374

Text: 19 isnt invisible menace face
Sentiment: 1, Confidence: 0.32815787196159363

Text: 20 university students faculty staff members campus community tested positive july march
Sentiment: 0, Confidence: 0.46876564621925354

Text: 21 information call
Sentiment: 2, Confidence: 0.26259884238243103

Text: 22 hats volunteers read names todays week covid remembrance action
Sentiment: 4, Confidence: 0.39964088797569275

Text: 23 people start wearing crowded areas especially deal new wave infections fuelled new disease health minister brigitte bourguignon said monday
Sentiment: 0, Confidence: 0.6796907782554626

Text: 24 dont take care health experiencing malnutrition fact health deteriorate
Sentiment: 0, Confidence: 0.45077359676361084

Text: 25 part musk demands f leech public property remove freedom others personal access every resource needed function productively
Sentiment: 4, Confidence: 0.293049693107605

Text: 26 cases fall presence ever decreasing mask wearing find great humor watching seeing tweets still trying somehow justify use public health interventions ridicule removal mandated use
Sentiment: 0, Confidence: 0.5884349346160889

Text: 27 click buy
Sentiment: 2, Confidence: 0.2446298748254776

Text: 28 remember get vaccinated stay home family sick get tested help slow spread
Sentiment: 0, Confidence: 0.4996935725212097

Text: 29 barrington coasts destination marketing launched
Sentiment: 0, Confidence: 0.2648793160915375

Text: 30 finland seems best employment rate years shelves full shops
Sentiment: 4, Confidence: 0.23262052237987518

Text: 31 covid show continues relevant interesting
Sentiment: 3, Confidence: 0.2852588891983032

Text: 32 parliamentarians time thieves
Sentiment: 1, Confidence: 0.27127909660339355

Text: 33 information call wellness coalition
Sentiment: 2, Confidence: 0.27435675263404846

Text: 34 listen jag club madness interview tv informer
Sentiment: 0, Confidence: 0.2519243657588959

Text: 35 good day
Sentiment: 2, Confidence: 0.28512969613075256

Text: 36 mild covid liberal tune destroying sacrifices since
Sentiment: 0, Confidence: 0.32173049449920654

Text: 37 omicron surges contact tracers face unexpected problem much less work wsj rpts
Sentiment: 0, Confidence: 0.47017550468444824

Text: 38 public health failure people think covid wont kill someone love
Sentiment: 0, Confidence: 0.5558201670646667

Text: 39 team working diligently remain goto trusted provider vaccinations boosters
Sentiment: 4, Confidence: 0.40548738837242126

Text: 40 everyone right highquality masks w enough disposable income purchase
Sentiment: 4, Confidence: 0.31555110216140747

Text: 41 seems long way pecking order
Sentiment: 1, Confidence: 0.32554957270622253

Text: 42 sup educ grants abbvie janssenus ferring takedapharma lilly
Sentiment: 3, Confidence: 0.266622394323349

Text: 43 greghuntmp healthcaregov victoriancho said take vaccination holy grail opening
economy without restrictions reduction infections
Sentiment: 0, Confidence: 0.35420310497283936

Text: 44 clifford hofferd chofferd
Sentiment: 1, Confidence: 0.2565097510814667

Text: 45 still running health workers firefighters military administrative refused experimental injection longer paid forbidden work rights days
Sentiment: 0, Confidence: 0.6068729162216187

Text: 46 loser forced believe science get vaccinated
Sentiment: 0, Confidence: 0.5232416987419128

Text: 47 cumulative number cases
Sentiment: 2, Confidence: 0.2991209924221039

Text: 48 must every single mutation ever
Sentiment: 4, Confidence: 0.33287855982780457

Text: 49 okay affidavit saying covid vaccines contain key ingredient used antifreeze considered toxic humans
Sentiment: 2, Confidence: 0.3671635091304779

Text: 50 new study shows unvaccinated covid patients less infectious vaxxed
Sentiment: 0, Confidence: 0.34188440442085266

Text: 51 united states statistics
Sentiment: 2, Confidence: 0.3067503273487091

Text: 52 may ons covid infection survey
Sentiment: 1, Confidence: 0.3102215528488159

Text: 53 face masks fascism
Sentiment: 1, Confidence: 0.282294899225235

Text: 54 total active cases
Sentiment: 1, Confidence: 0.30087926983833313

Text: 55 people nasty
Sentiment: 1, Confidence: 0.35324037075042725

Text: 56 new cases
Sentiment: 1, Confidence: 0.28282710909843445

Text: 57 like let covid older folks die
Sentiment: 1, Confidence: 0.26729846000671387

Text: 58 new death per
Sentiment: 1, Confidence: 0.2969604432582855

Text: 59 didnt tweet much pandemic idea whether effective rona
Sentiment: 0, Confidence: 0.3859463930130005

Text: 60 could help track tweet elon musk said fauci source covid ngo
Sentiment: 0, Confidence: 0.3683551251888275

Text: 61 cancercausing chemical found sunscreen products
Sentiment: 0, Confidence: 0.30129969120025635

Text: 62 data provisional change frequently

Sentiment: 2, Confidence: 0.30510345101356506

Text: 63 yesterday missourians recorded dead
Sentiment: 0, Confidence: 0.3856775164604187

Text: 64 latest report
Sentiment: 2, Confidence: 0.2681654691696167

Text: 65 hi everyone
Sentiment: 1, Confidence: 0.290897011756897

Text: 66 yanjiao town near locked due
Sentiment: 1, Confidence: 0.34521156549453735

Text: 67 canadian occurring pandemic demonstrated disruptive compounding putting
communities increased risk decreased priority
Sentiment: 0, Confidence: 0.4926053583621979

Text: 68 dr pope notes probably going get fifth shot fallstay tuned guess
Sentiment: 0, Confidence: 0.4153420925140381

Text: 69 alexberenson vs pierrekory
Sentiment: 0, Confidence: 0.2737552225589752

Text: 70 canton street
Sentiment: 1, Confidence: 0.2767908275127411

Text: 71 growth us confirmed covid deaths county al covid visualizations give insight covid
thanks data scientists usafacts
Sentiment: 3, Confidence: 0.25705552101135254

Text: 72 still travel standing one place
Sentiment: 1, Confidence: 0.29059427976608276

Text: 73 love twitter
Sentiment: 1, Confidence: 0.27144625782966614

Text: 74 two weeks booster came week misery month lassitude
Sentiment: 0, Confidence: 0.575214684009552

Text: 75 sit nothing making sure youre wearing virus porous mask
Sentiment: 0, Confidence: 0.5210534930229187

Text: 76 thank much covering important topic cynthiamcfadden
Sentiment: 2, Confidence: 0.24772430956363678

Text: 77 richard summerfield explores talent retention age great resignation fwwmagazines
march cover story
Sentiment: 4, Confidence: 0.3152007758617401

Text: 78 also buy hold coviddogs rewards
Sentiment: 4, Confidence: 0.2291925847530365

Text: 79 th day alert level quarantine c body temperature spo bpm pulse bp

Sentiment: 0, Confidence: 0.22402846813201904

Text: 80 presenters able present remotely
Sentiment: 3, Confidence: 0.23969466984272003

Text: 82 munching watermelon tea lays
Sentiment: 1, Confidence: 0.2966792583465576

Text: 83 catching would serious outcome
Sentiment: 1, Confidence: 0.3474313020706177

Text: 84 point im making current mrna covid vaccines designed code response thats spike
proteins
Sentiment: 3, Confidence: 0.27293887734413147

Text: 85 earn bitcoin browsing cryptotab
Sentiment: 0, Confidence: 0.36016929149627686

Text: 86 coercive manipulation
Sentiment: 2, Confidence: 0.24779148399829865

Text: 87 awesome place holds special place vail
Sentiment: 4, Confidence: 0.4365083575248718

Text: 88 still exhausted
Sentiment: 1, Confidence: 0.34603697061538696

Text: 89 fully vaccinated
Sentiment: 1, Confidence: 0.2648468017578125

Text: 90 china pentagon darpa coordination wuhan virology lab come eventually
Sentiment: 2, Confidence: 0.24204958975315094

Text: 91 greatest country earth
Sentiment: 2, Confidence: 0.25193703174591064

Text: 92 rwanda uganda
Sentiment: 1, Confidence: 0.311398446559906

Text: 93 new study published erspublications pulse oximeter measurements blood oxygen levels
may unreliable assess severity across different ethnicities infectdisnews medicineuon
Sentiment: 0, Confidence: 0.3543967306613922

Text: 94 created determines covid
Sentiment: 2, Confidence: 0.271411657333374

Text: 95 returning teach ff spring break rise
Sentiment: 0, Confidence: 0.2994909882545471

Text: 96 pretty much scientists though
Sentiment: 2, Confidence: 0.33965942226360321

Text: 97 years anyone said fauci mandates politically motivated labeled conspiracy theorist
Sentiment: 0, Confidence: 0.6952553987503052

Text: 98 afraid life never
Sentiment: 1, Confidence: 0.2939322888851166

Text: 99 four years term hes still talking train delays
Sentiment: 0, Confidence: 0.316682904958725

Text: 100 work share stories pacific learners impacted pandemic
Sentiment: 0, Confidence: 0.2674310505390167