

DDI

Kullandığım Yöntemler

Projede `collections` kütüphanesinin bir fonksiyonu olan `counter` fonksiyonunu kullanarak en sık geçen 20 kelimeyi buldum, ve `Word2Vec` kullanarak en sık geçen 20 kelimeye benzer 5'er kelime buldum.

Aşağıdaki kütüphaneleri kullanarak rastgele seçtiğim 5 cümleye benzer 3 cümle buldum.

```
from sklearn.feature_extraction.text import TfidfVectorizer  
from sklearn.metrics.pairwise import cosine_similarity
```

(projenin başında word2vec yerine fastText kullanmıştım ve daha iyi sonuç almıştım fakat o kütüphanede kelimeler arasında anlama bakmadan benzerliğe bakıyordu. sizin de anlamlara göre kelimeleri istediğinizi varsayarak word2vec kullandım.)

Ayrıca seçtiğim 5 cümle ve onlara benzeyen 3 cümlelerin duygu analizini de önceden eğitilmiş BERT modeliyle yaptım.

En Çok Geçen 20 Kelime

covid: 1990 kez

people: 561 kez

new: 420 kez

pandemic: 393 kez

cases: 388 kez

us: 348 kez

health: 330 kez

vaccine: 319 kez

get: 303 kez
like: 283 kez
deaths: 277 kez
vaccines: 235 kez
one: 214 kez
still: 199 kez
many: 186 kez
dont: 184 kez
virus: 179 kez
years: 173 kez
data: 171 kez
know: 169 kez

Benzer Kelimeler

covid kelimesine benzer kelimeler:

like: 0.9995571970939636
us: 0.999462366104126
vaccine: 0.9994460344314575
virus: 0.9994349479675293
one: 0.9993780255317688

people kelimesine benzer kelimeler:

covid: 0.9992077946662903
would: 0.9990838766098022
pandemic: 0.9990395903587341
many: 0.9990091919898987
like: 0.9989782571792603

new kelimesine benzer kelimeler:

covid: 0.9977774620056152
study: 0.9974275827407837

may: 0.9973515272140503
death: 0.9973382949829102
today: 0.9973227381706238

pandemic kelimesine benzer kelimeler:

covid: 0.9993544816970825
us: 0.9992906451225281
vaccine: 0.999172568321228
like: 0.9991205334663391
many: 0.9990890622138977

cases kelimesine benzer kelimeler:

deaths: 0.9955263137817383
new: 0.9955083131790161
may: 0.9947658777236938
last: 0.9947171211242676
research: 0.9946383833885193

us kelimesine benzer kelimeler:

covid: 0.9994624853134155
pandemic: 0.9992907047271729
world: 0.9992616772651672
like: 0.9992531538009644
vaccine: 0.9991729855537415

health kelimesine benzer kelimeler:

covid: 0.9992683529853821
like: 0.9991322755813599
us: 0.9991204142570496
may: 0.9990966320037842
pandemic: 0.9990630149841309

vaccine kelimesine benzer kelimeler:

covid: 0.9994460940361023

vaccines: 0.9991986751556396

us: 0.9991728663444519

pandemic: 0.9991726279258728

like: 0.9991703629493713

get kelimesine benzer kelimeler:

covid: 0.999017059803009

us: 0.9989002346992493

like: 0.9988899827003479

pandemic: 0.9987879991531372

people: 0.9987666606903076

like kelimesine benzer kelimeler:

covid: 0.9995572566986084

us: 0.9992531538009644

would: 0.9991744756698608

vaccine: 0.9991704225540161

world: 0.9991567134857178

deaths kelimesine benzer kelimeler:

covid: 0.9979841709136963

pandemic: 0.9978737235069275

death: 0.9977339506149292

may: 0.9976344704627991

one: 0.9976094961166382

vaccines kelimesine benzer kelimeler:

covid: 0.9992638230323792

vaccine: 0.9991986155509949

us: 0.9991682767868042

pandemic: 0.9990345239639282

like: 0.9990326762199402

one kelimesine benzer kelimeler:

covid: 0.9993780255317688

like: 0.9990905523300171

dr: 0.9990839958190918

even: 0.9990330934524536

pandemic: 0.999029278755188

still kelimesine benzer kelimeler:

covid: 0.9990882277488708

us: 0.9989975690841675

vaccine: 0.9989560842514038

pandemic: 0.9989472031593323

people: 0.9989078640937805

many kelimesine benzer kelimeler:

pandemic: 0.9990888833999634

covid: 0.9990265369415283

people: 0.9990090727806091

vaccine: 0.9989325404167175

one: 0.9988634586334229

dont kelimesine benzer kelimeler:

would: 0.9986273646354675

covid: 0.9985214471817017

like: 0.9984583854675293

vaccine: 0.9984508752822876

people: 0.9984331130981445

virus kelimesine benzer kelimeler:

covid: 0.9994351267814636

vaccine: 0.9991253018379211

us: 0.9991106986999512

like: 0.9990913271903992

world: 0.9989924430847168

years kelimesine benzer kelimeler:

pandemic: 0.9987602829933167

covid: 0.9987571835517883

vaccine: 0.9986853003501892

us: 0.9986804723739624

many: 0.9986286163330078

data kelimesine benzer kelimeler:

new: 0.9969034194946289

covid: 0.9960290193557739

patients: 0.9956604242324829

study: 0.9956492185592651

vaccine: 0.9955775141716003

know kelimesine benzer kelimeler:

vaccine: 0.998814582824707

many: 0.9986082911491394

covid: 0.9985476732254028

would: 0.9984907507896423

get: 0.9984492063522339

5 farklı cümleye benzer cümleler

Cümle 1: 8329. reported positive result

En Benzer 3 Cümle ve Satır Numaraları:

1. Satır 5164: result covid vaccines

2. Satır 9475: reported close april positive cases deaths

3. Satır 5037: cases reported

Cümle 2: 1557. many souls already lost lives due way covid

En Benzer 3 Cümle ve Satır Numaraları:

1. Satır 829: least canadian lives lost due covid misinformation
2. Satır 9662: texans lost lives leadership
3. Satır 5078: already lost faith confidence cdc

Cümle 3: 10041. million americans died still

En Benzer 3 Cümle ve Satır Numaraları:

1. Satır 7943: million americans died covid
2. Satır 9214: covid million sick million died
3. Satır 6960: meanwhile million americans elsewhere died

Cümle 4: 10058. find free covid vaccine

En Benzer 3 Cümle ve Satır Numaraları:

1. Satır 8526: find vaccine
2. Satır 8253: havent find covid vaccine today
3. Satır 8673: free vaccine sites

Cümle 5: 1636. looks pretty ridiculous people dont collapse dead spot covid

En Benzer 3 Cümle ve Satır Numaraları:

1. Satır 3136: collapse pediatric covid surveillance
2. Satır 9211: get accurate test result spot
3. Satır 4023: pretty clear dont think winstonpeters

Duygu Analizi

Sentiment (1-5):

1 - Çok olumsuz

5 - Çok olumlu

Cümle: reported positive result

Sentiment: 3, Confidence: 0.2633279263973236

Cümle: result covid vaccines

Sentiment: 3, Confidence: 0.25785064697265625

Cümle: reported close april positive cases deaths

Sentiment: 1, Confidence: 0.6540114283561707

Cümle: cases reported

Sentiment: 1, Confidence: 0.3429744839668274

Cümle: many souls already lost lives due way covid

Sentiment: 5, Confidence: 0.21775606274604797

Cümle: least canadian lives lost due covid misinformation

Sentiment: 1, Confidence: 0.6313874125480652

Cümle: texans lost lives leadership

Sentiment: 1, Confidence: 0.4388708472251892

Cümle: already lost faith confidence cdc

Sentiment: 1, Confidence: 0.5514181852340698

Cümle: million americans died still

Sentiment: 1, Confidence: 0.5809951424598694

Cümle: million americans died covid

Sentiment: 1, Confidence: 0.5009440779685974

Cümle: covid million sick million died

Sentiment: 1, Confidence: 0.6133244037628174

Cümle: meanwhile million americans elsewhere died

Sentiment: 1, Confidence: 0.5154972672462463

Cümle: find free covid vaccine

Sentiment: 1, Confidence: 0.23532384634017944

Cümle: find vaccine

Sentiment: 1, Confidence: 0.2495897263288498

Cümle: havent find covid vaccine today

Sentiment: 1, Confidence: 0.25352296233177185

Cümle: free vaccine sites

Sentiment: 4, Confidence: 0.33985140919685364

Cümle: looks pretty ridiculous people dont collapse dead spot covid

Sentiment: 1, Confidence: 0.7471318244934082

Cümle: collapse pediatric covid surveillance

Sentiment: 1, Confidence: 0.5181099772453308

Cümle: get accurate test result spot

Sentiment: 4, Confidence: 0.3495408296585083

Cümle: pretty clear dont think winstonpeters

Sentiment: 3, Confidence: 0.34811851382255554