

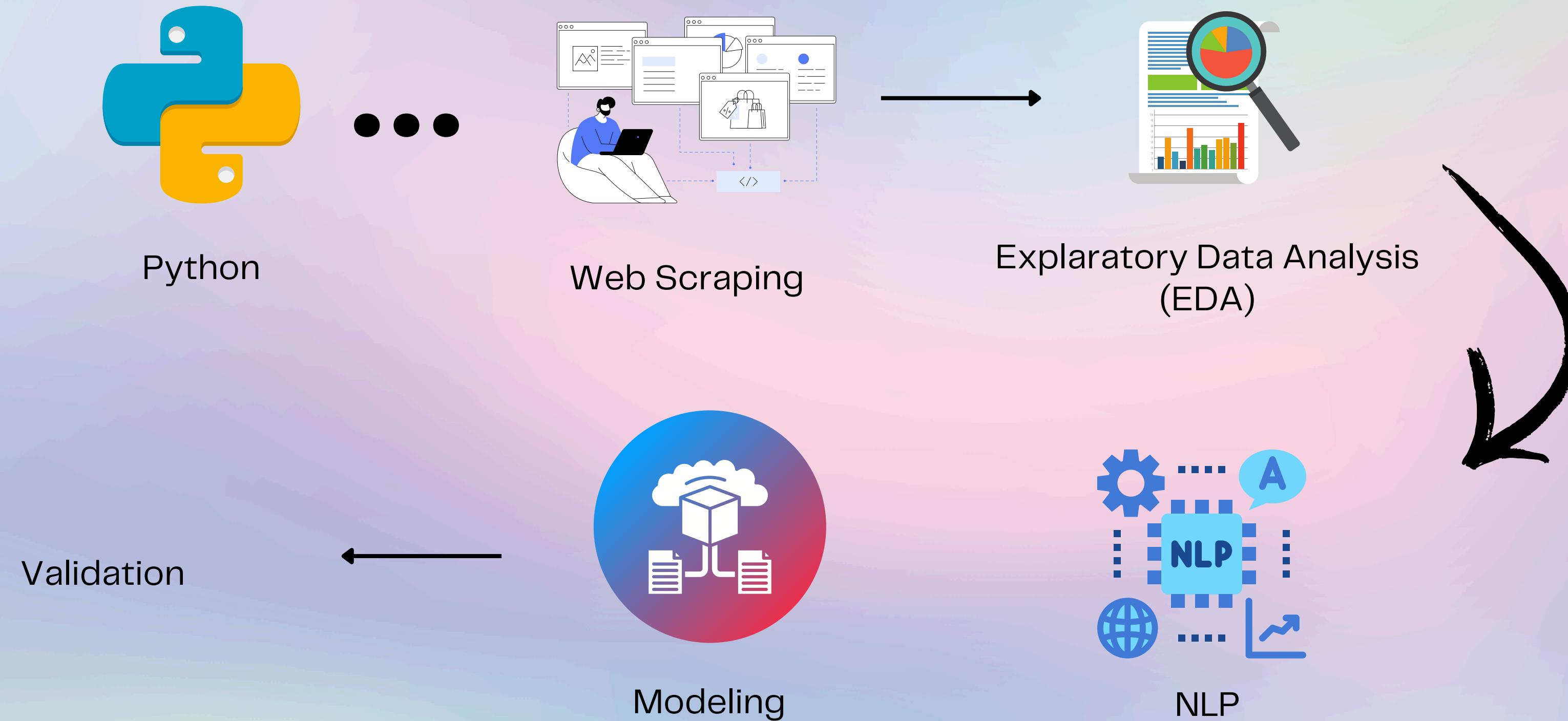


NATURAL LANGUAGE PROCESSING UNTUK SENTIMEN ANALISIS TERHADAP INSTANSI BEA CUKAI MENGGUNAKAN LONG SHORT-TERM MEMORY

○ ○ ○ ○

- E1E122064 _ LA ODE MUHAMMAD YUDHY PRAYITNO
- E1E122093 _ DIKI ARDIKA WIRATAMA
- 2255201019 _ PUTRI RAHAYU

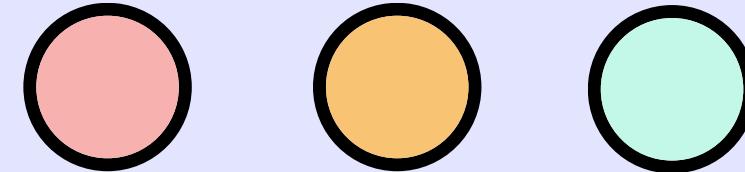
Metodologi



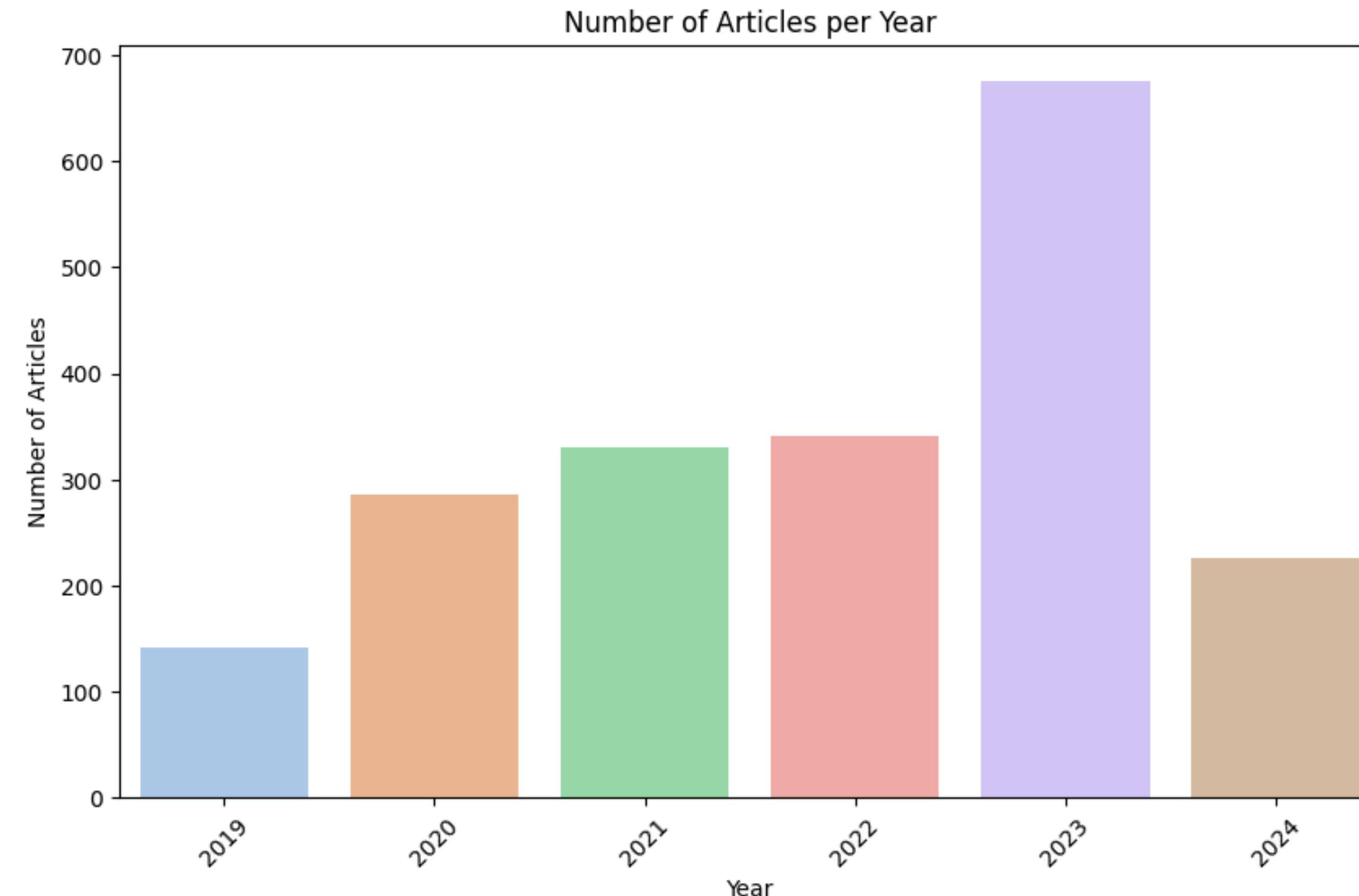
Exploratory Data Analysis

Features

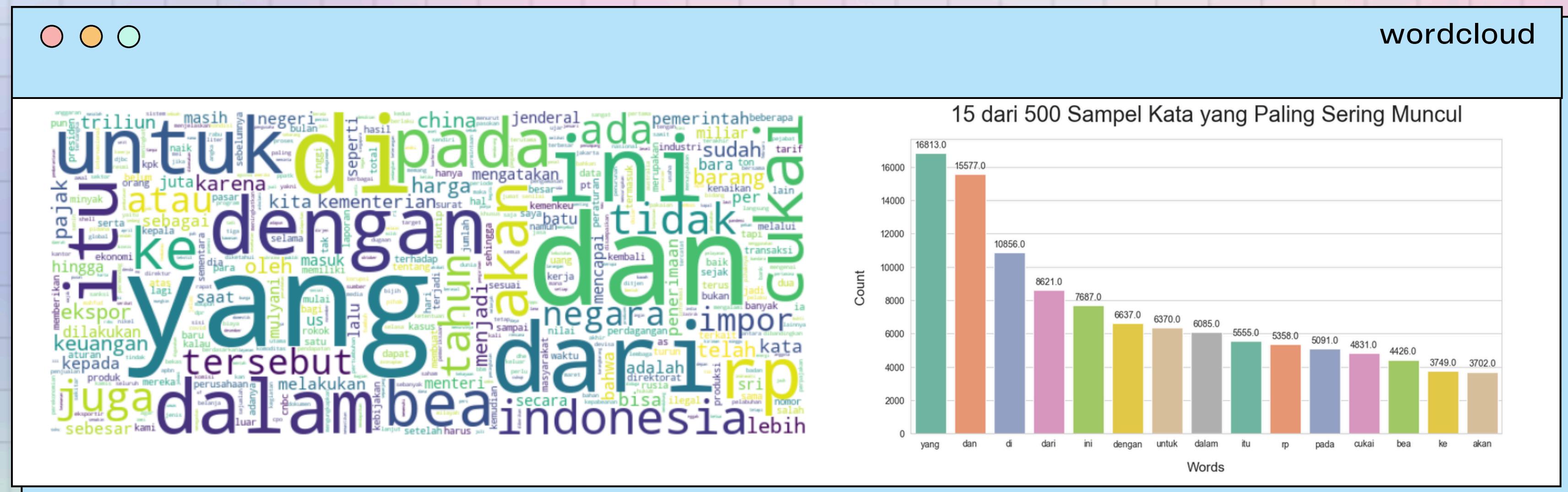
- headline
- date
- link
- content
- Token
- Stop_article
- Stem_article
- Perpos_article
- polarity_score
- polarity



Data Per Year

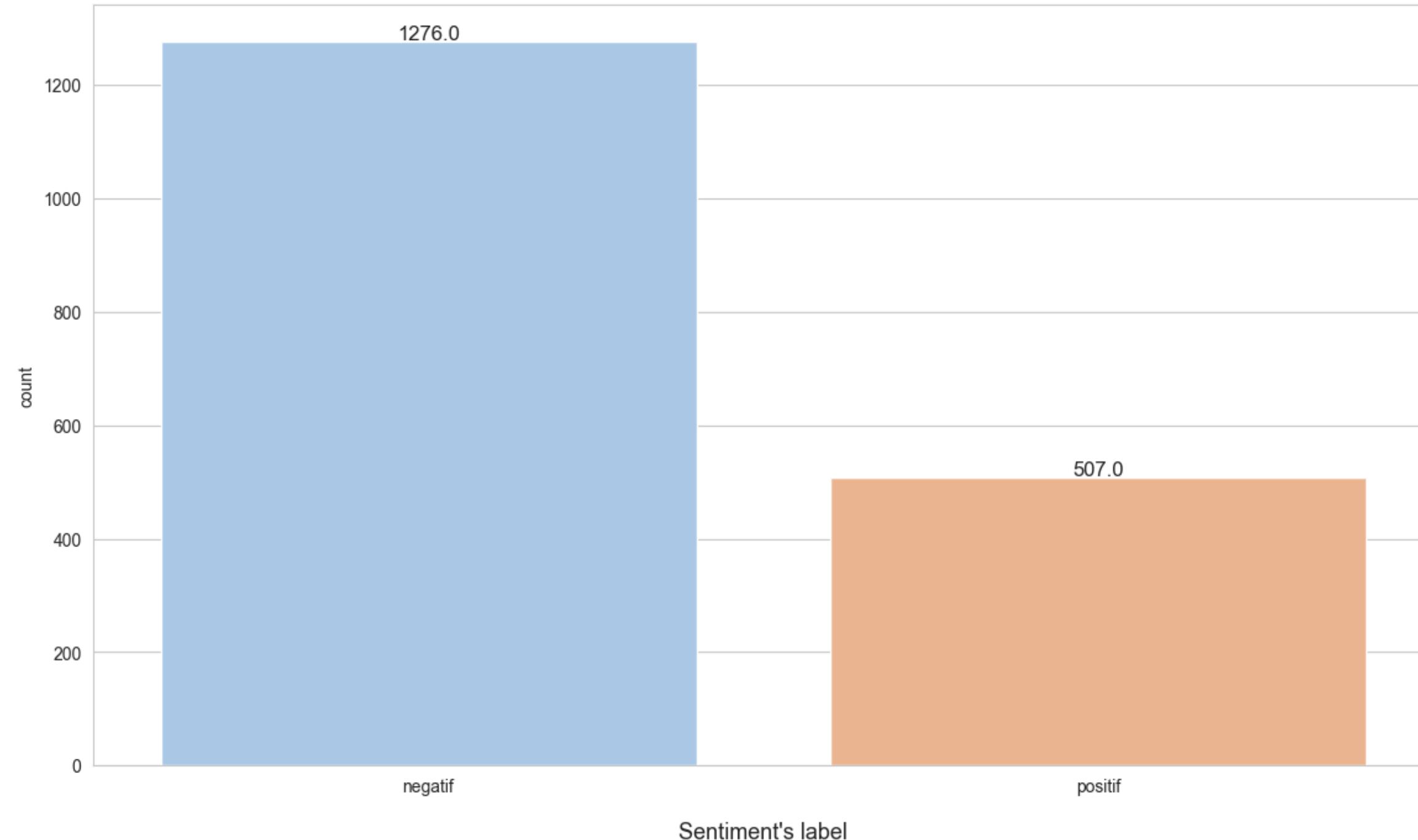


Common Article



Distribusi Sentimen

Jumlah Label Positif dan Negatif.

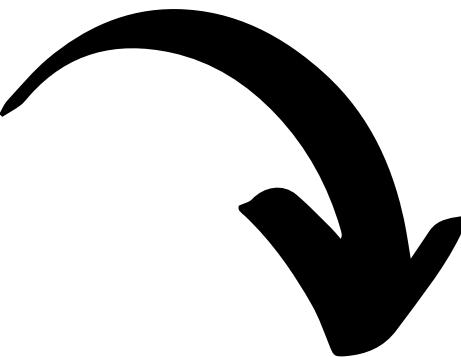


Pre-processing

o o o o

1. Case Folding
2. Remove URL
3. Remove Punctuation dan Numbers
4. Remove Specific Phrases
5. Remove Duplicate

Infak Masjid Raya Sumbar dan sej
Rp799,1 juta.\n\nTerdakwa ment
ah Masjid Raya Sumbar tahun 2013
lebih dahulu ke rekening Masjid
an untuk kepentingan pribadi.\n\n
n Inspektorat Sumbar Nomor:11INS
gkap jabatan bendahara yang diem
19.\n\nKemudian menjabat sebagai
M Arief Iskandar\nCOPYRIGHT

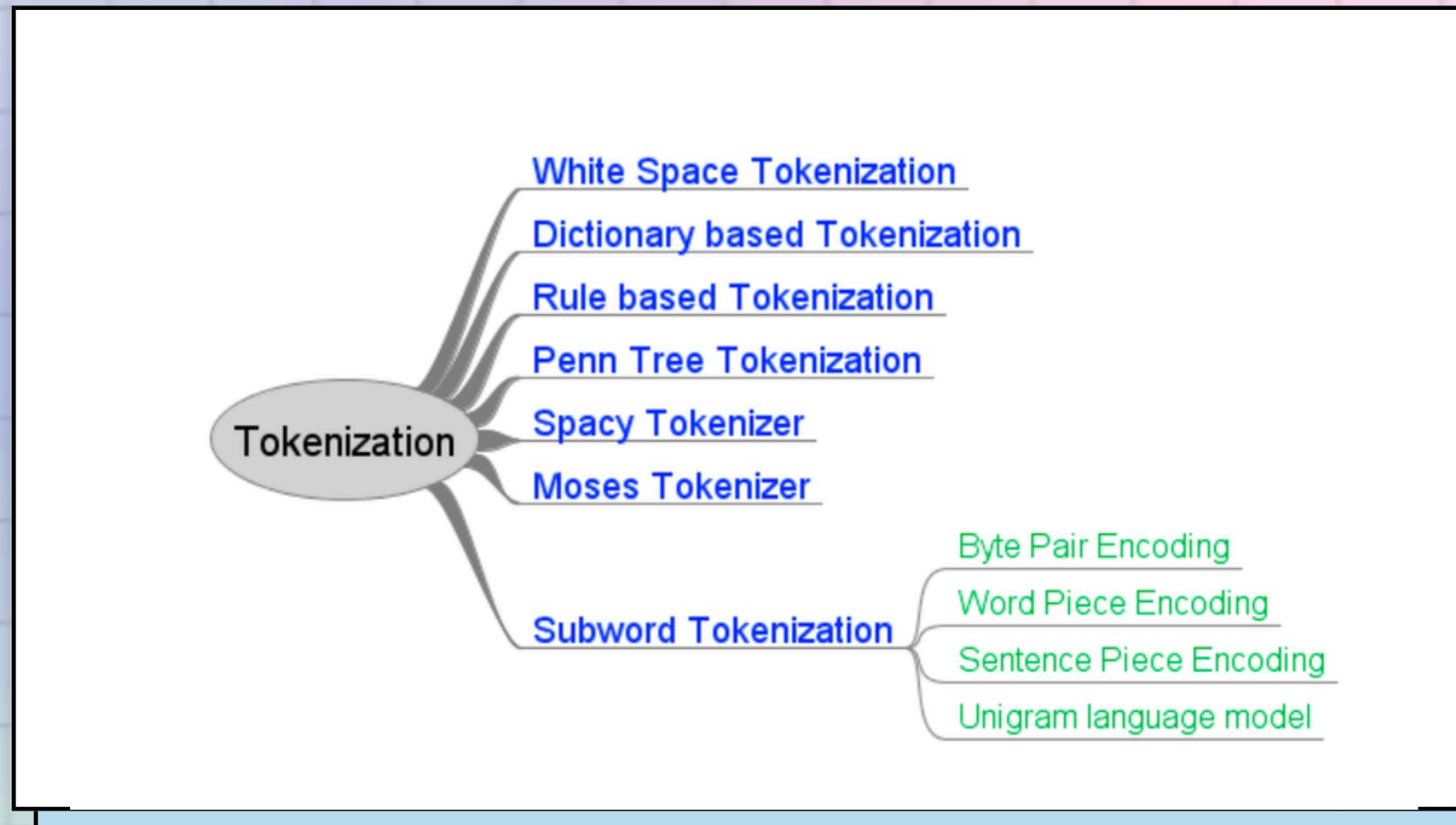


ro ke sejumlah rekening seolaholah u
adalah uang pada unit pengumpulan za
sisa dana peringatan hari besar isla
rp berdasarkan laporan hasil pengh
a yang menyelewengkan sejumlah angga
ovinsi sumbar sejak januari hingga
ud editor m arief iskandar copyright

MDM Company

Tokenizing

- o o o o





Stopwords

o o o o

Stop Words

These words include:

- a
- of
- on
- I
- for
- with
- the
- at
- from
- in
- to

Stopwords seperti "dan", "di", "yang", dan kata-kata lain yang tidak memberikan kontribusi signifikan terhadap analisis sentimen dihapus. Langkah ini bertujuan untuk memfokuskan analisis pada kata-kata yang membawa makna sentimen yang lebih kuat dalam konteks artikel berita.

**Natural Language Processing
with Python!**

Stop Words

[“This”, “is”, “a”, “test”]





Stemming

o o o o

Stemming in NLP

Stemming mengubah kata-kata menjadi bentuk dasarnya dengan menghapus afiks seperti imbuhan dan sufiks. Proses ini membantu dalam mengurangi variasi kata dan memastikan konsistensi dalam analisis sentimen.

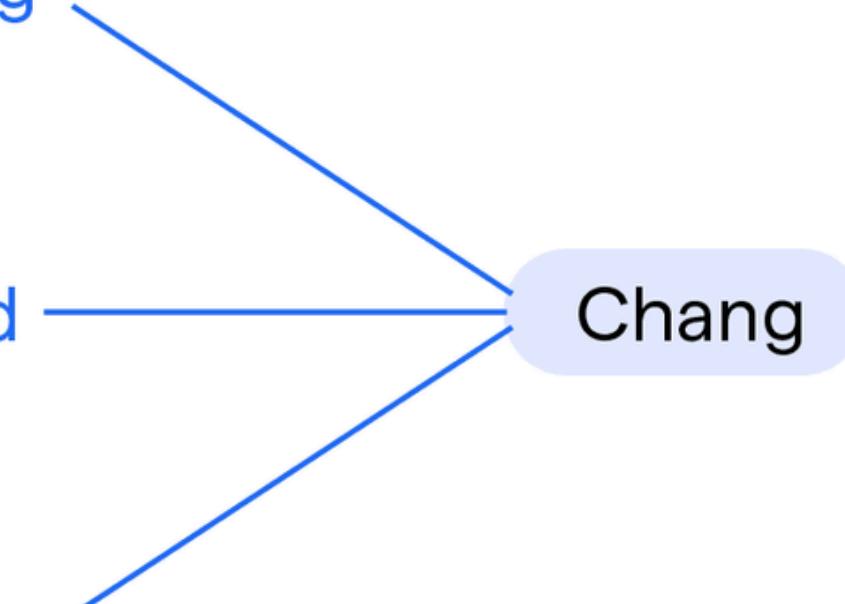
Stemming

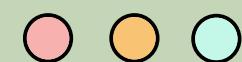
Chang^{ing}

Changed

Change

Chang





Sentiment Scoring



Pelabelan Data Set

Proses pelabelan data pada penelitian ini dilakukan dengan sentiment scoring menggunakan Indonesian Sentiment Lexicon (InSet). InSet menyediakan bobot sentimen untuk setiap kata, dengan rentang nilai antara 1 hingga 5 untuk sentimen positif dan -1 hingga -5 untuk sentimen negatif.

Artikel dengan skor lebih besar dari 0 diberi label sentimen positif (1), sementara artikel dengan skor kurang dari 0 diberi label sentimen negatif (0)



Lexicon
In NLP

Data Set Analysis

o o o o

| | Content | Token | Stop_article | Stem_article | Prepos_article | polarity_score | polarity |
|---|---|---|---|---|---|----------------|----------|
| 0 | pemerintah tengah bersiap merealisasikan ekspo... | ['pemerintah', 'tengah', 'bersiap', 'merealisasi... | ['pemerintah', 'merealisasikan', 'ekspor', 'pa... | perintah realisasi ekspor pasir laut presiden ... | ['perintah', 'realisasi', 'ekspor', 'pasir', '... | -12 | -1 |
| 1 | direktur jenderal bea cukai kementerian keuang... | ['direktur', 'jenderal', 'bea', 'cukai', 'keme... | ['direktur', 'jenderal', 'bea', 'cukai', 'keme... | direktur jenderal bea cukai menteri uang askol... | ['direktur', 'jenderal', 'bea', 'cukai', 'ment... | -39 | -1 |
| 2 | ketua umum ikatan pengusaha konveksi berkarya ... | ['jakarta', 'cnbc', 'indonesia', 'ketua', 'umu... | ['jakarta', 'cnbc', 'indonesia', 'ketua', 'ika... | ketua ikat usaha konveksi karya ipkb nandi her... | ['ketua', 'ikat', 'usaha', 'konveksi', 'karya'... | -58 | -1 |
| 3 | pengusaha tekstil di dalam negeri menyoroti pe... | ['pengusaha', 'tekstil', 'di', 'dalam', 'neger... | ['pengusaha', 'tekstil', 'negeri', 'menyoroti'... | usaha tekstil negeri sorot nyata menteri uang ... | ['usaha', 'tekstil', 'negeri', 'sorot', 'nyata... | -85 | -1 |
| 4 | presiden joko widodo jokowi memanggil beberapa... | ['presiden', 'joko', 'widodo', 'jokowi', 'mem... | ['presiden', 'joko', 'widodo', 'jokowi', 'mema... | presiden joko widodo jokowi panggil menteri ra... | ['presiden', 'joko', 'widodo', 'jokowi', 'pang... | 15 | 1 |

Word Embedding

o o o o

Word embedding memetakan kata-kata ke dalam ruang vektor yang memungkinkan komputer untuk memahami hubungan semantik antara kata-kata tersebut. Teknik ini membantu meningkatkan kualitas analisis sentimen dengan menyandikan makna kata-kata dalam artikel berita ke dalam representasi matematis.

| | | | | | | | | | | | | |
|------|-----|------|------|------|------|------|------|------|------|------|------|------|
| 15 | 235 | 10 | 231 | 80 | 375 | 402 | 96 | 18 | 18 | 15 | 548 | 68 |
| 22 | 231 | 18 | 535 | 1164 | 147 | 231 | 482 | 118 | 32 | 529 | 1022 | 712 |
| 254 | 5 | 23 | 1031 | 1080 | 5 | 231 | 555 | 1301 | 35 | 102 | 30 | 3 |
| 5 | 6 | 251 | 1070 | 452 | 518 | 16 | 4 | 678 | 658 | 477 | 954 | 529 |
| 174 | 87 | 125 | 153 | 5 | 23 | 1031 | 1080 | 836 | 231 | 10 | 69 | 290 |
| 252 | 147 | 407 | 96 | 678 | 76 | 27 | 535 | 571 | 568 | 5 | 10 | 20 |
| 22 | 168 | 26 | 932 | 1670 | 98 | 168 | 870 | 932 | 258 | 645 | 369 | 1607 |
| 932 | 100 | 1671 | 390 | 932 | 754 | 390 | 820 | 20 | 933 | 369 | 955 | 390 |
| 369 | 407 | 18 | 663 | 15 | 1155 | 18 | 1412 | 891 | 1608 | 777 | 95 | 46 |
| 231 | 965 | 121 | 19 | 188 | 975 | 777 | 145 | 184 | 206 | 1609 | 249 | 1610 |
| 231 | 555 | 1301 | 188 | 231 | 19 | 464 | 1637 | 4 | 976 | 1032 | 976 | 1032 |
| 1965 | 249 | 976 | 1032 | 177 | 1056 | 1194 | 10 | 231 | 348 | 1314 | 231 | 4 |
| 947 | 249 | 489 | 231 | 194 | 175 | 1920 | 927 | 490 | 548 | 96 | 257 | 217 |

Model Architecture

o o o o

```
# Tokenizing and padding sequences
max_features = 2000 # Adjust as necessary
max_length = 300
```

Epoch: 100

Optimizer: Adam

Train, Test: 80:20

Hyperparameter

Loss: binary_crossentropy

Batch: 64

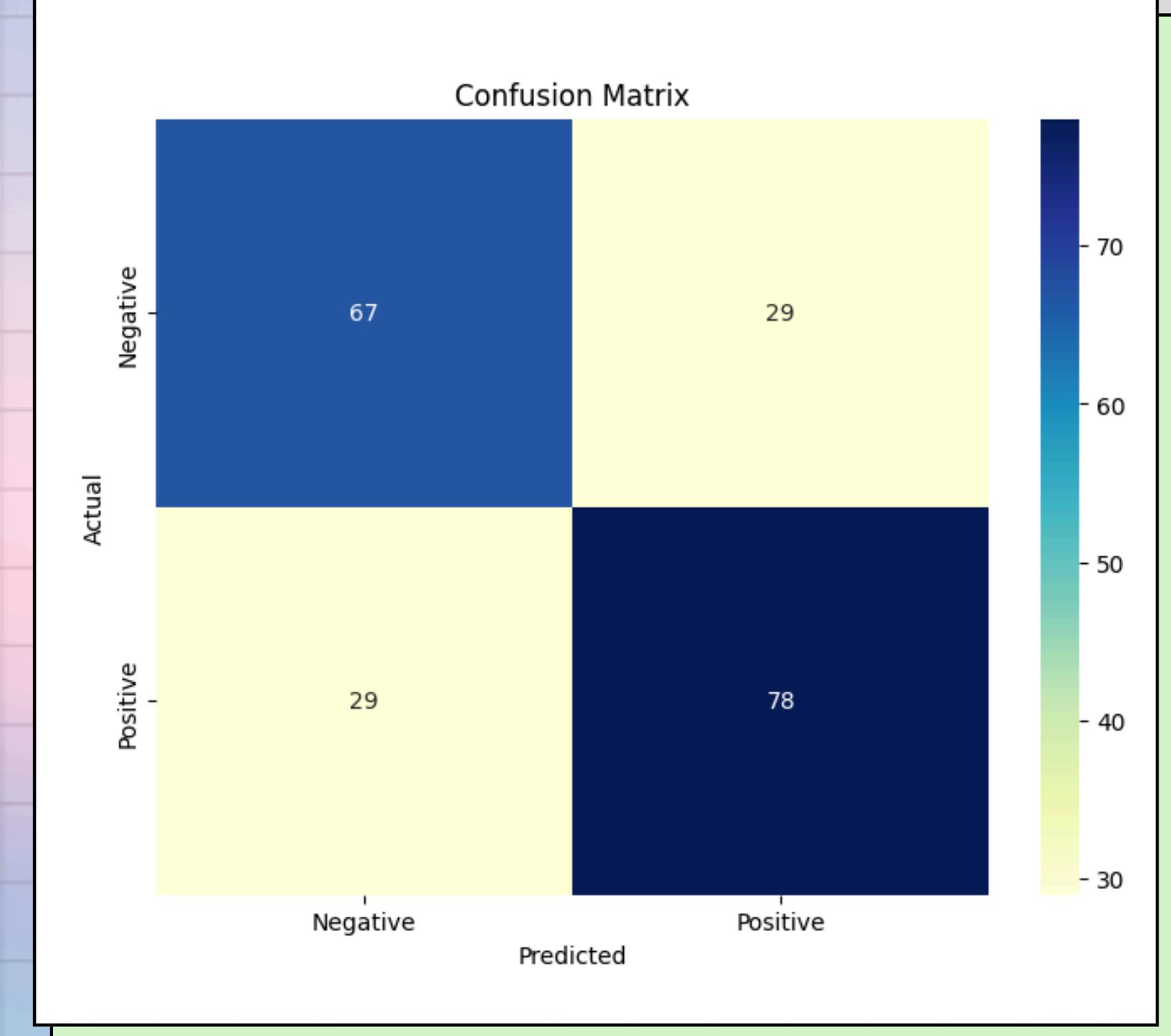
| Layer (type) | Output Shape | Param # |
|---|------------------|---------|
| embedding_18 (Embedding) | (None, 300, 128) | 256,000 |
| spatial_dropout1d_18 (SpatialDropout1D) | (None, 300, 128) | 0 |
| lstm_18 (LSTM) | (None, 100) | 91,600 |
| dense_18 (Dense) | (None, 1) | 101 |

Confusion Matrix

o o o o

Hasil perhitungan overall accuracy menunjukkan tingkat akurasi model algoritma Long Short Term-Memory dalam mengklasifikasikan sentimen terkait instansi bea cukai pada website artikel berita adalah sebesar 71.43%.

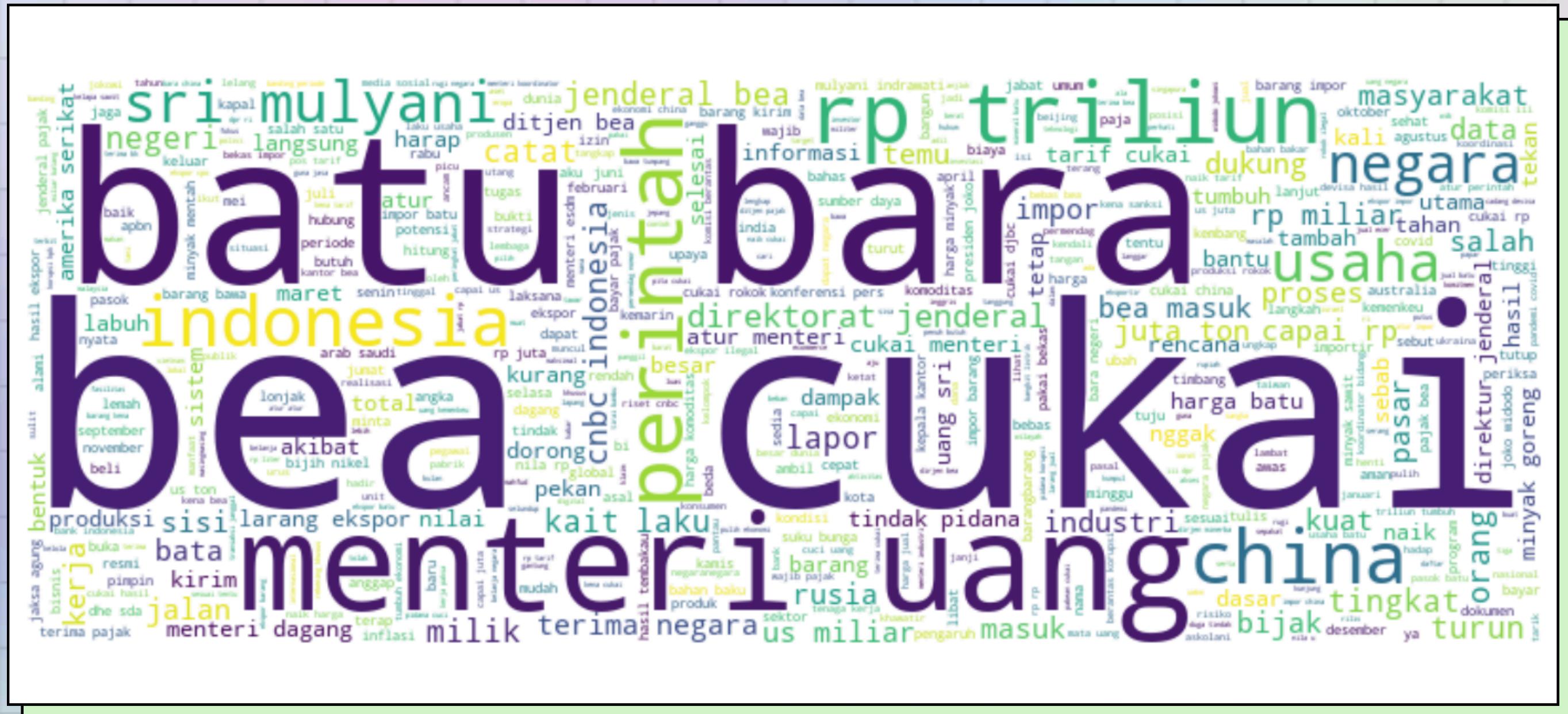
$$Accuracy = \frac{67+78}{67+29+29+78} \times 100\% = 71.43\%$$



Word Cloud Kelas Sentimen Positif



Word Cloud Kelas Sentimen Negatif



**Thank
you!**

o o o o

