

PEMANFAATAN DATA TWITTER DAN BERITA ONLINE UNTUK MENDUKUNG OFFICIAL STATISTICS



Jakarta, 19/08/22



Oleh: Wahyu Calvin Frans Mariel



OUTLINE

PENGANTAR

METODE PENELITIAN

IMPLEMENTASI

PENGEMBANGAN





1

Pengantar

TENTANG BPS

Badan Pusat Statistik (BPS) adalah Lembaga Pemerintah Non Kementerian yang bertanggung jawab langsung kepada Presiden.

BPS bertugas dalam melaksanakan pemerintahan di bidang statistik sesuai peraturan perundang-undangan. Hasil statistik yang diselenggarakan oleh BPS diumumkan dalam Berita Resmi Statistik (BRS) secara teratur dan transparan agar masyarakat dengan mudah mengetahui dan mendapatkan data yang diperlukan.





“Official statistics are statistics published by government agencies or other public bodies such as international organizations as a public good. They provide quantitative or qualitative information on all major areas of citizens' lives, such as economic and social development, living conditions, health, education, and the environment.”

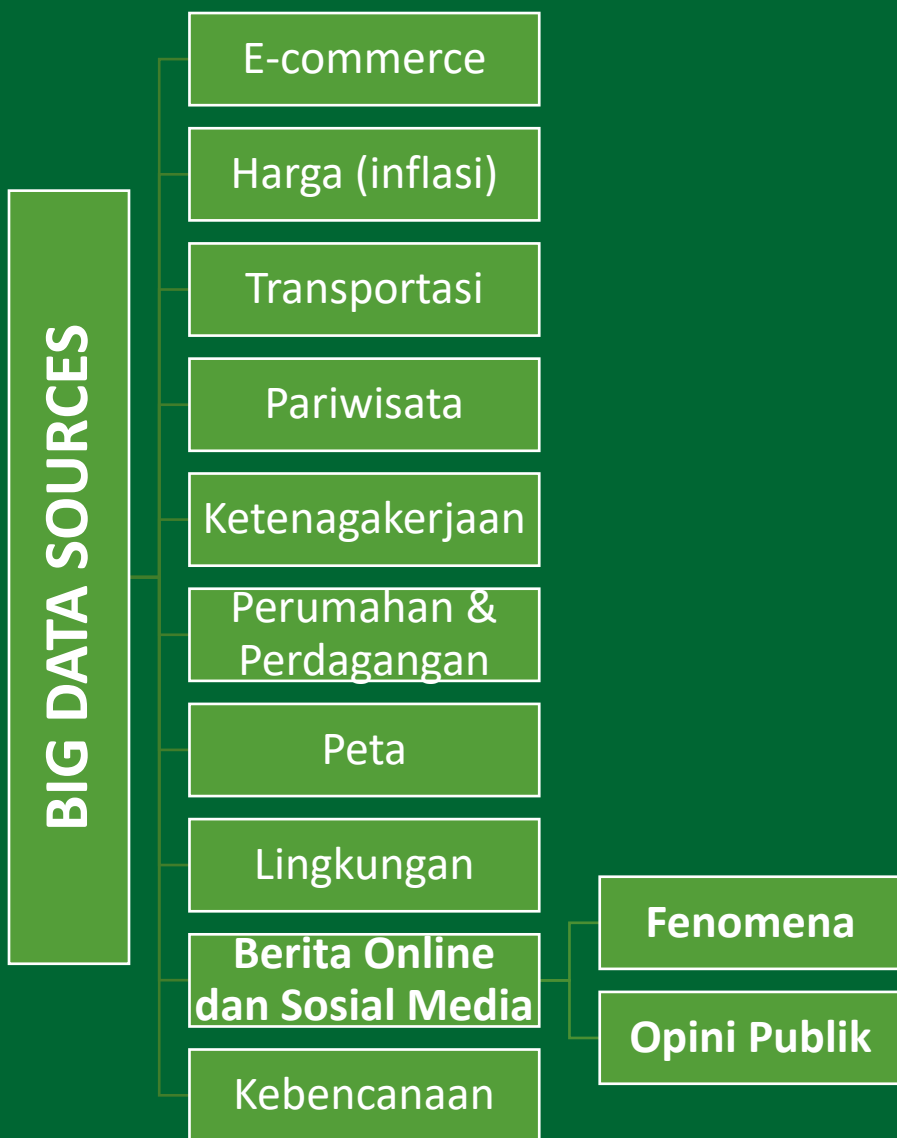


Data BPS



*lihat selengkapnya di www.bps.go.id

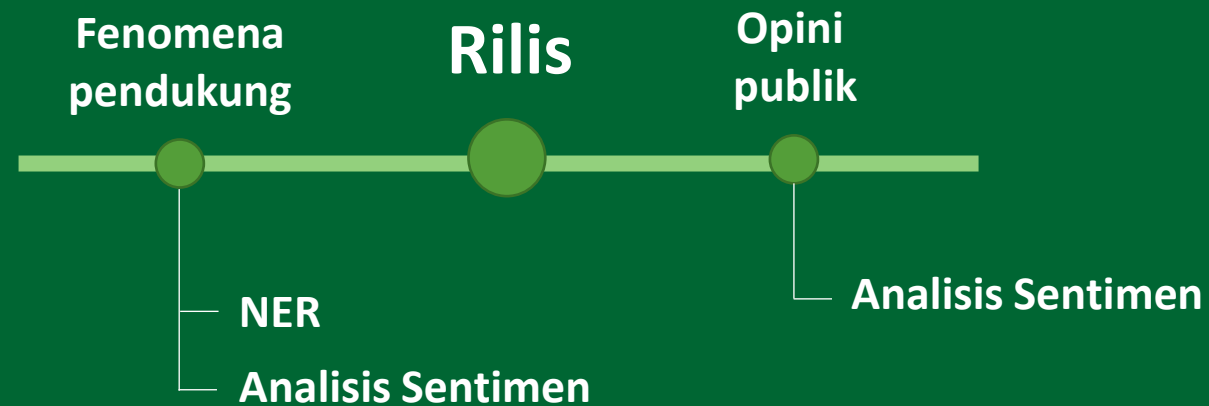
PEMANFAATAN BIG DATA DI BPS



PEMANFAATAN DATA BERITA ONLINE & TWITTER



RILIS Berita Resmi Statistik (BRS)



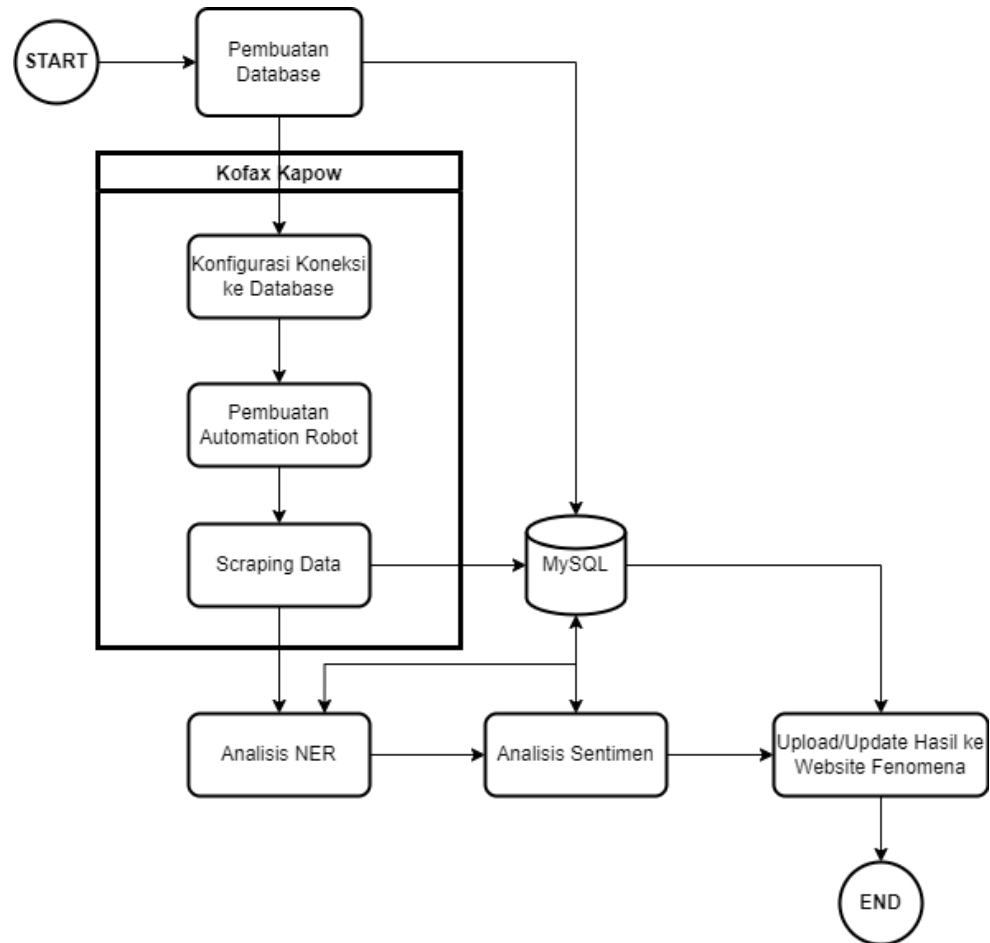


2

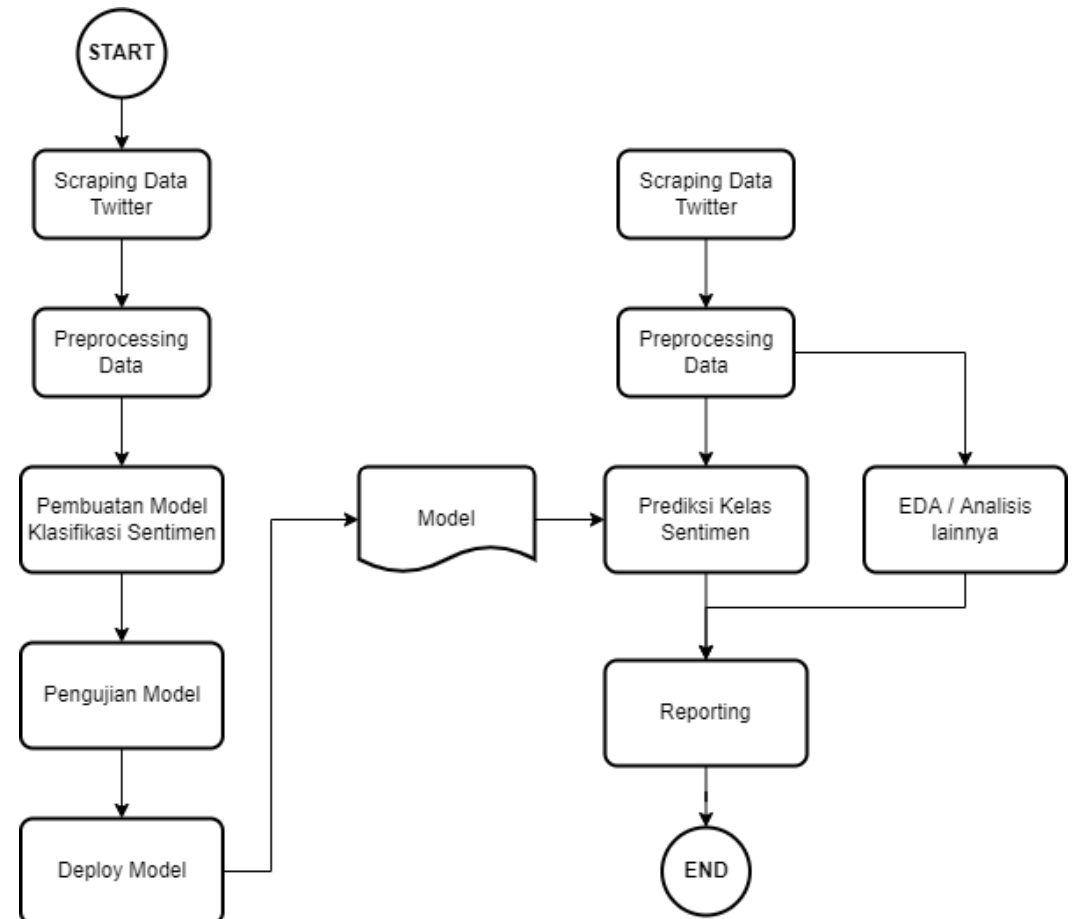
Metode Penelitian

ALUR PELAKSANAAN

BERITA ONLINE

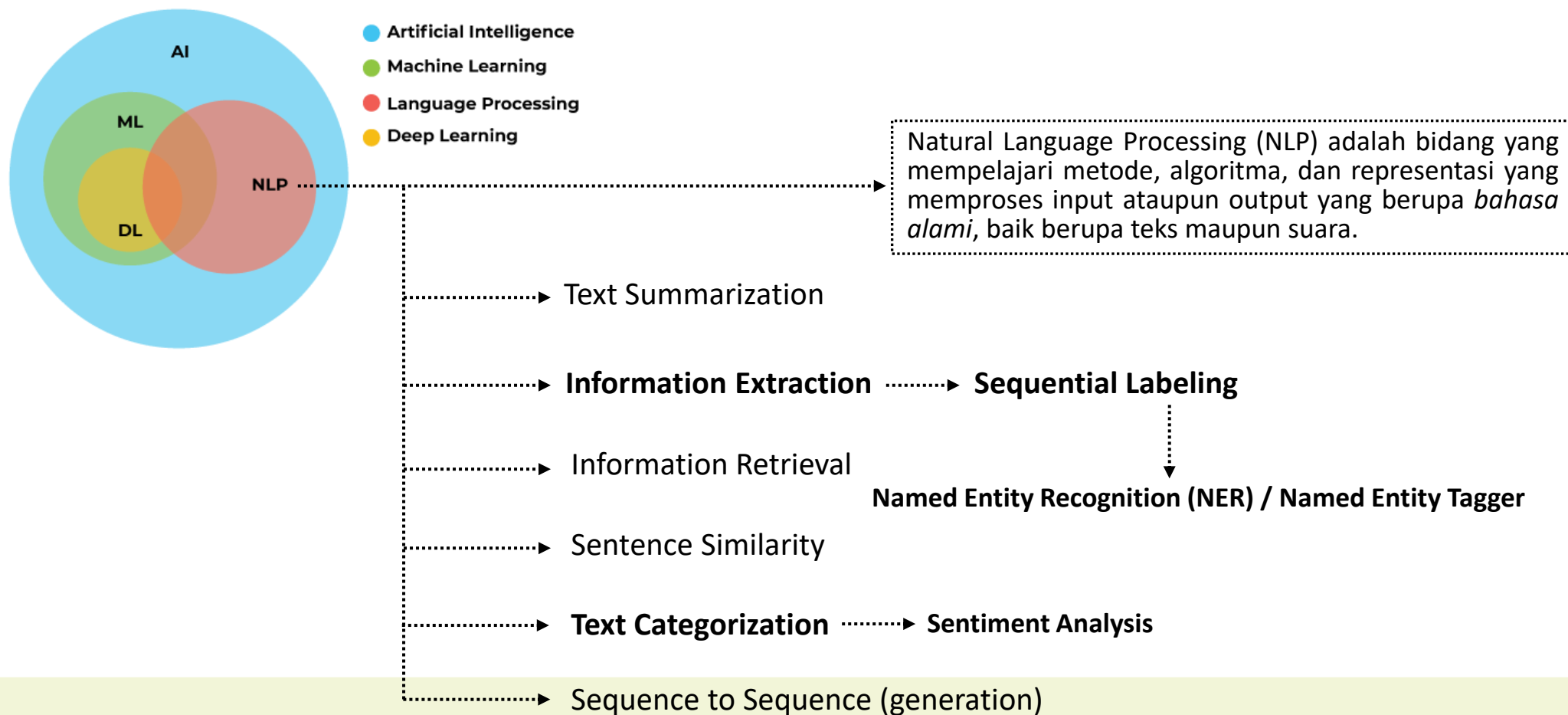


TWITTER





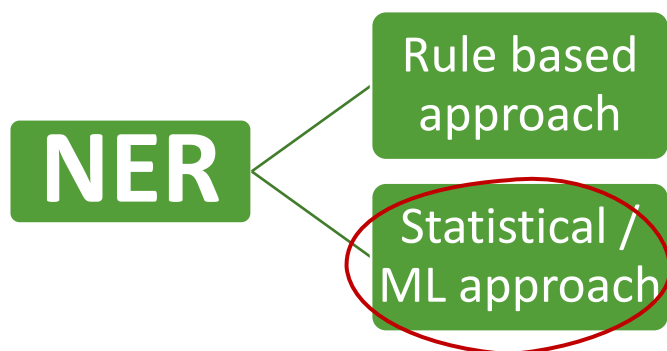
METODE ANALISIS





METODE ANALISIS: NER

NER merupakan teknik ekstraksi informasi yang berfungsi untuk mengenali dan mengklasifikasikan entitas yang terdapat pada suatu teks tak terstruktur ke dalam kategori-kategori yang telah didefinisikan sebelumnya. Entitas tersebut bisa berupa nama seseorang, lokasi, organisasi, waktu, dan sebagainya. NER termasuk ke dalam teknik *Sequence Labelling*.



Contoh

“Badan Pusat Statistik (BPS) melaporkan total nilai ekspor Indonesia mencapai US\$20,46 miliar pada Februari 2022. Secara bulanan, nilai ekspor tersebut mengalami pertumbuhan sebesar 6,73 persen”

Badan Pusat Statistik (BPS) **ORGANIZATION** melaporkan total nilai ekspor Indonesia **LOCATION**
mencapai US\$20,46 miliar **QUANTITY** pada Februari 2022 **TIME**. Secara bulanan, nilai ekspor
tersebut mengalami pertumbuhan sebesar 6,73 persen **QUANTITY**





METODE ANALISIS: NER(2)

PEMBANGUNAN MODEL NER



Entitas:

- | | |
|-----------------|--------------|
| 1) Person | 4) Location |
| 2) Position | 5) Indicator |
| 3) Organization | 6) Quote |





METODE ANALISIS: SENTIMENT ANALYSIS

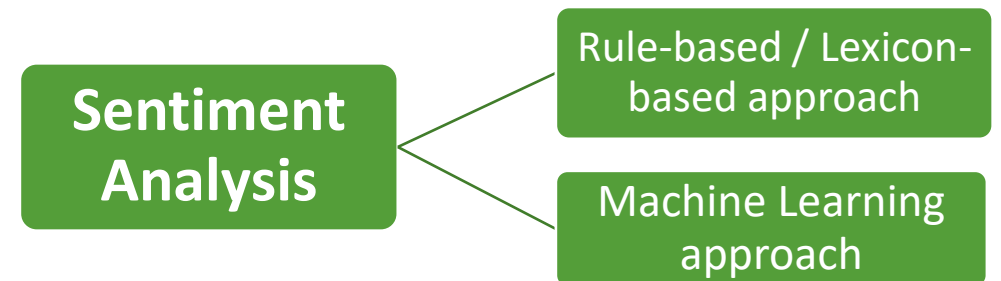
Analisis Sentimen (*sentiment analysis*) merupakan analisis untuk menentukan nada emosional dari penulisnya, atau juga dapat didefinisikan sebagai proses memahami dan mengelompokkan emosi (biasanya dikelompokkan sebagai positif, negatif, atau netral).

Contoh

Alhamdulillah kenaikan pertumbuhan ekonomi mencapai 7 persen tahun 2021 ini, mantap bapak presiden dan tim.
(Positif)

Hari ini kepala BPS mengumumkan pertumbuhan ekonomi kuartal II 2021 Indonesia sebesar 7.07%.
(Netral)

Pertumbuhan ekonomi kalah jauh dari pemerintahan sebelumnya. Hutang dan korupsi semakin merajalela.
(Negatif)



SENTIMENT ANALYSIS

LEXICON BASED

KAMU KEREN
+4

POSITIF

KAMU TIDAK
KEREN -4

NEGATIF

KAMU SANGAT
KEREN +6

POSITIF

KAMU KEREN TAPI
JELEK -4

NETRAL

KAMU SANGAT JELEK
-6

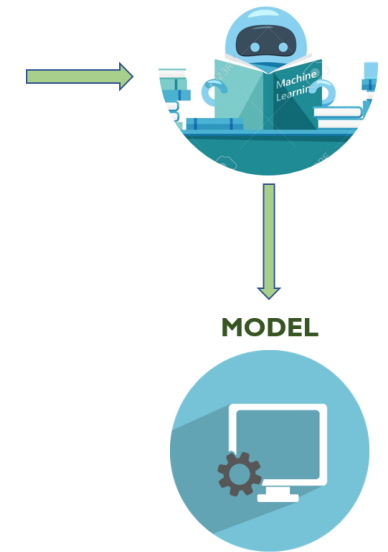
NEGATIF

- **SENTIMEN POSITIF**
Skor positif terbesar > Skor negatif terbesar
- **SENTIMEN NEGATIF**
Skor negative terbesar > Skor positif terbesar
- **SENTIMEN NETRAL**
Skor positif terbesar = Skor negatif terbesar

MACHINE LEARNING BASED

Sentiment	Text Tweet
negative	Jelek filmnya... apalagi si ernest gak mutu bgt actingnya... film sampah
negative	Film king Arthur ini film paling jelek dari seluruh cerita King Arthur
negative	@beexkuanlin Sepanjang film gwa berkata kasar terus pada bapaknya
negative	Ane ga suka fast and furious..menurutku kok jelek ya tu film
positive	10. Walaupun mau bikin film horror, film Danur bisa jadi rujukan .. ringan tapi menakutkan
positive	Banyak moral yg bisa dipetik dari film wonder woman. Salah satunya adalah menjaga auratmu dr cowok hidung belang ?
positive	Berikut 5 film horor Indonesia berpendapatan tinggi dalam 10 tahun terakhir. Dapatkan tiket @DanurMovie di http://bit.ly/belidanur
positive	Trust me. #GetOut itu tipe film paling nyaman dinikmati saat kita gak tau sama sekali filmnya mau kemana. Kalau sudah baca sana-sini ya beda

TRAINING



SENTIMENT ANALYSIS

LEXICON BASED

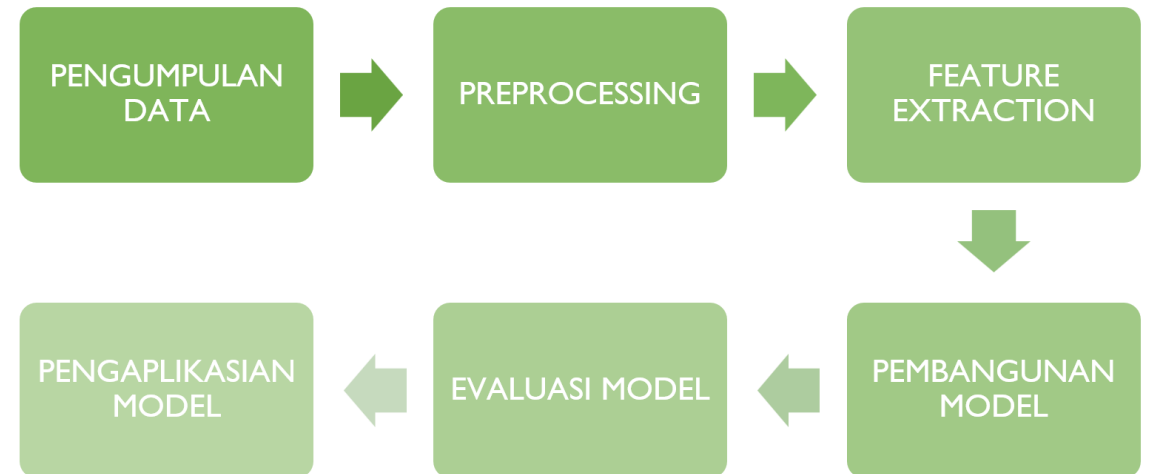
PENDEKATAN LEKSIKON SENTISTRENGTH



Algoritma : SentiStrength Indonesia
Penelitian Terkait: <https://doi.org/10.22146/ijccs.16625>

MACHINE LEARNING BASED

PEMBANGUNAN MODEL ANALISIS SENTIMEN



Algoritma : Neural Network (*Deep Learning*)
Penelitian Terkait: <https://doi.org/10.13106/jidb.2022.vol13.no6.9>



3

Implementasi



FENOMENA: BERITA ONLINE

Kategori	Jumlah_Subkategori
Industri Pengolahan	16
Transportasi dan Pergudangan	6
Pertanian, Kehutanan, dan Perikanan	6
Pertambangan dan Penggalian	4
Jasa Perantara Keuangan dan Asuransi	4
Perdagangan Besar dan Eceran; Reparasi Mobil dan S...	2
Pengadaan Listrik dan Gas	2
Penyediaan Akomodasi dan Makan Minum	2
Aktivitas di Masa Pandemi COVID-19	1
Kepatuhan Terhadap Protokol Kesehatan	1
Vaksinasi COVID-19	1
Jasa Kesehatan dan Kegiatan Sosial Swasta	1
Informasi dan Komunikasi	1
Konstruksi	1
Pengadaan Air, Pengelolaan Sampah, Limbah dan Daur...	1
Jasa Pendidikan Swasta	1
Jasa Kesenian, Hiburan dan Rekreasi Swasta	1
Real Estat	1
Aktivitas Jasa Lainnya Swasta	1
Jasa Perusahaan	1

```
# Load model NER masing-masing entitas
Person      = spacy.load('Person')
Position    = spacy.load('Position')
Organization= spacy.load('Organization')
Location    = spacy.load('Location')
Indicator   = spacy.load('Indicator')
Quote       = spacy.load('Quote')
```

```
# Config Sentistrength
config = dict()
config["negation"] = True
config["booster"] = True
config["ungkapan"] = True
config["consecutive"] = True
config["repeated"] = True
config["emoticon"] = True
config["question"] = True
config["exclamation"] = True
config["punctuation"] = True
senti = sentistrength(config)
skor = 0
```



WEBSITE FENOMENA (1)

FENOMENA 2.0

Dashboard

DAFTAR MENU

Daftar Fenomena

NER Analisis

Sentimen Analisis

Search for...



TOTAL BERITA PDB
14585



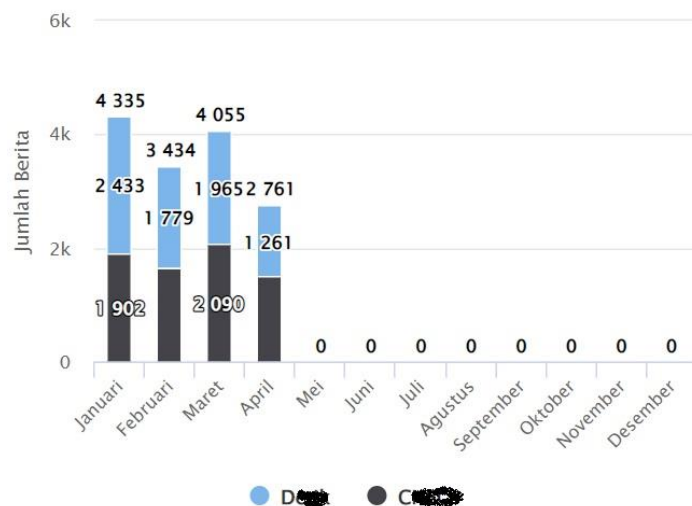
SUMBER BERITA
2



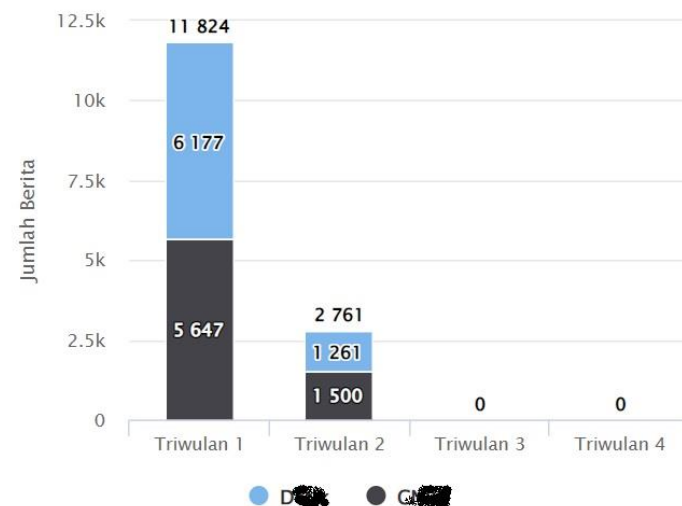
PROVINSI TERCAKUP
34



Jumlah Berita menurut Sumber Berita per Bulan



Jumlah Berita menurut Sumber Berita per Triwulan





WEBSITE FENOMENA (2)

NerOutput #0017c5e6-2c

×

Berikut adalah hasil analisis NER dari berita yang dipilih:

KATEGORI LAPANGAN USAHA

Pertanian, Kehutanan, dan Perikanan (A)



SUBKATEGORI LAPANGAN USAHA

Pertanian, Peternakan, Perburuan dan Jasa
Pertanian; Tanaman Pangan (A1.a)



BERITA - Tak Semua Salah Babi China, Ini Penyebab Tempe-Tahu RI Langka

▼

NER

▼

Nama Tokoh	['Muhamad Lutfi', 'Brazil']
Jabatan	['Menteri Perdagangan', 'Menteri BUMN Erick Thohir dan', 'Bupati Purworejo meresmikan Pasar Purworejo']
Organisasi	['BUMN', 'Kementerian Perdagangan (Kemendag)', 'Administrasi Bea dan Cukai China']
Lokasi	['Jakarta', 'Indonesia', 'Lutfi', 'China', 'Cina', 'Amerika Serikat', 'Amerika Selatan', 'Argentina', 'Ukraina']
Kutipan	["Masalah kedelai ini masalah internasional. Pada 2021, kita mengimpor 2,5 juta ton kedelai, sementara produksi dalam negeri tidak lebih dari 300 ribu ton. Jadi kita bergantung pada masalah internasional,", "Permintaan tidak sebaik yang diharapkan; sejumlah besar gandum dan beras digunakan untuk pakan, yang akan menggantikan banyak kedelai,", "Karena cuaca kering yang panjang,"]

SENTIMENT

▼


Sentimen Berita	Negatif
Sentimen Kutipan	Negatif

Close





WEBSITE FENOMENA (3)

**FENOMENA 2.0**

Dashboard

DAFTAR MENU

Daftar Fenomena

NER Analisis >

Sentimen Analisis >

Search for...

Q

Daftar Nama Tokoh Publik

Showing 1-20 of 118 items.

* Anda dapat mengubah ukuran kolom tabel dengan menyeret tepi kolom.

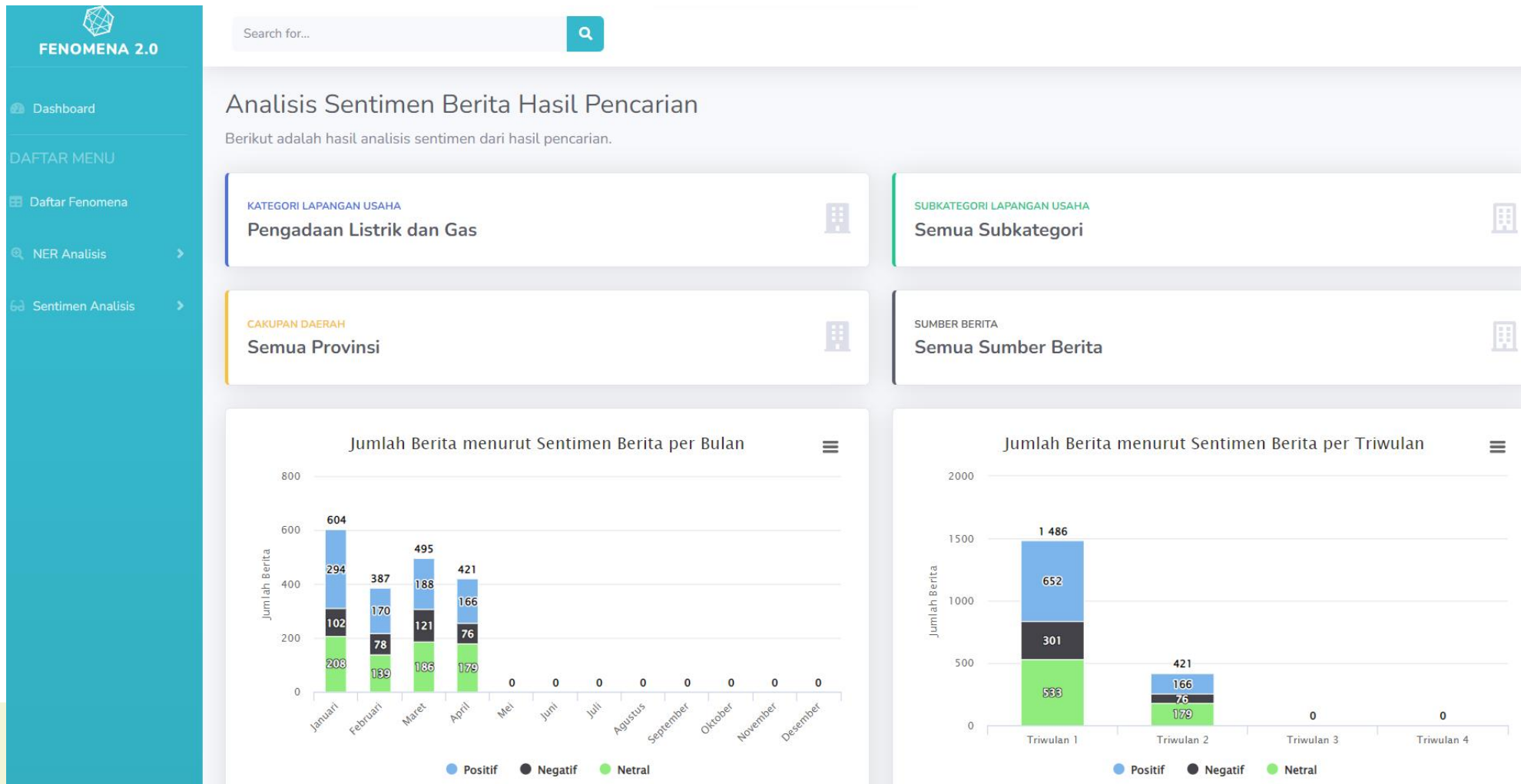
All

Nama Tokoh	Jumlah Berita ↓↑	Pilihan
Joko Widodo (Jokowi)	780	
Risbiani Fardaniah	687	
Sri Mulyani	634	
Faisal Yuniarto	617	
Agung	532	
Luhut	514	
Airlangga	478	
Budi Suyanto	447	
Airlangga Hartarto	432	



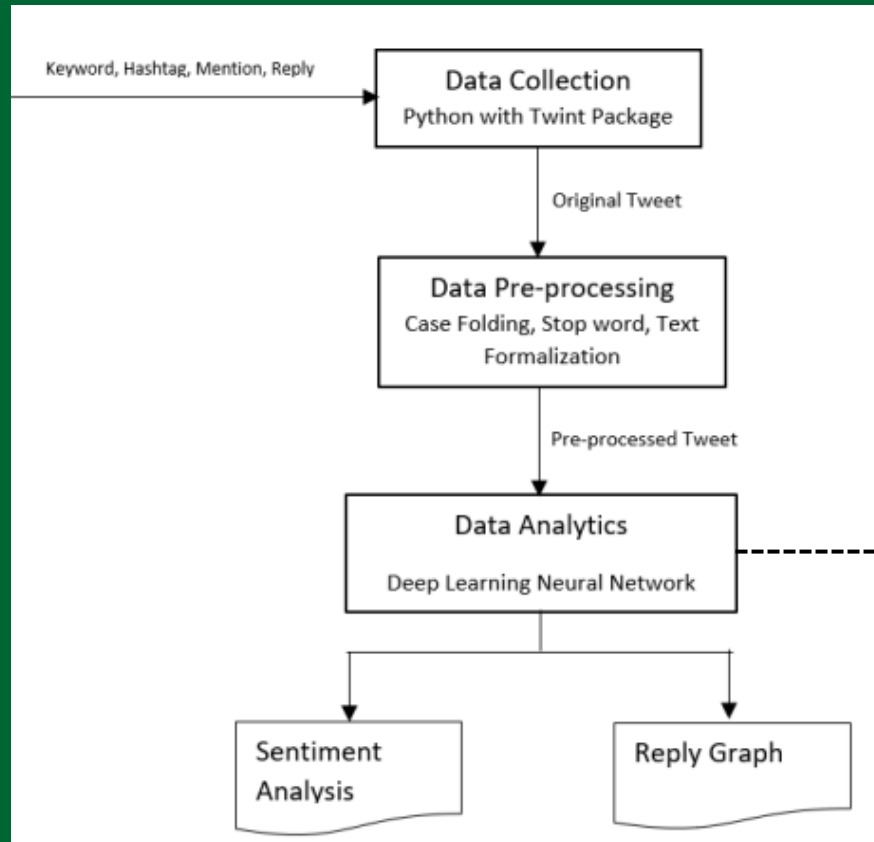


WEBSITE FENOMENA (4)





SOCIAL MEDIA: TWITTER



Sequential Model with 3 Dense layers with each activation of relu, relu, and sigmoid

- `model = Sequential ()`
- `model.add(Dense(90, input_dim=X_len, kernel_initializer="normal", activation='relu'))`
- `model.add(Dense(9, kernel_initializer="normal", activation='relu'))`
- `model.add(Dense(3, kernel_initializer="normal", activation='sigmoid'))`

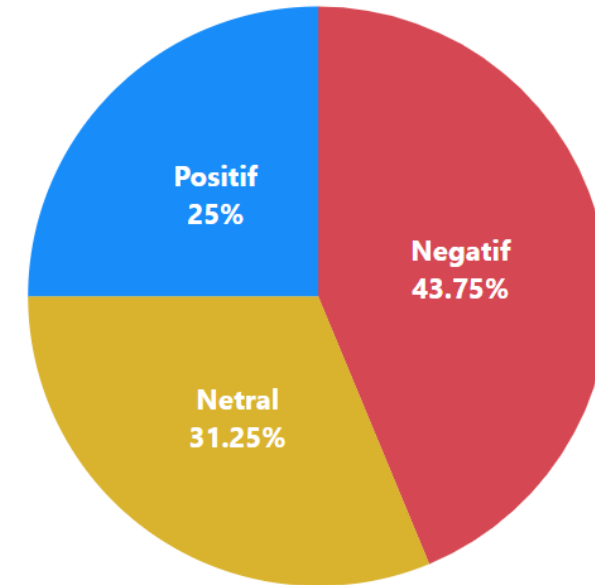
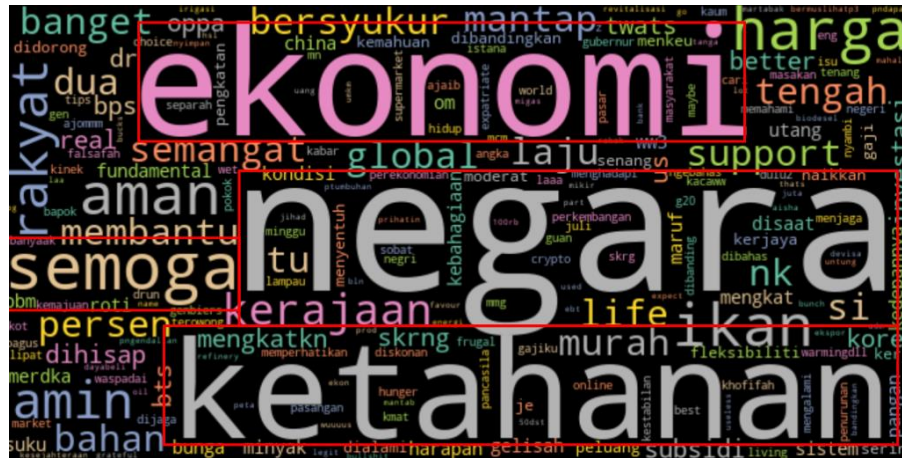
Training:Testing:Validation = 80%:10%:10%

	precision	recall	f1-score
0	0.91	0.86	0.88
1	0.74	0.83	0.78
2	0.79	0.73	0.76
accuracy			0.81
macro avg	0.81	0.81	0.81
weighted avg	0.81	0.81	0.81

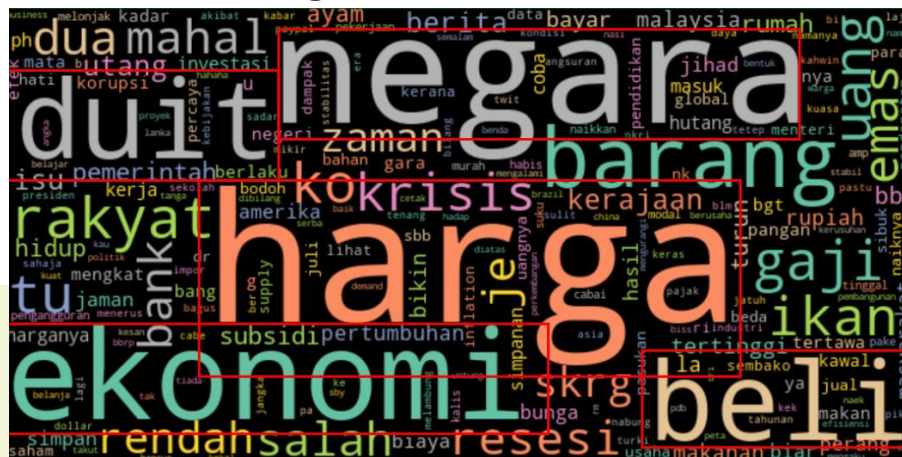


TWITTER : PUBLIC OPINION

Sentimen Positif



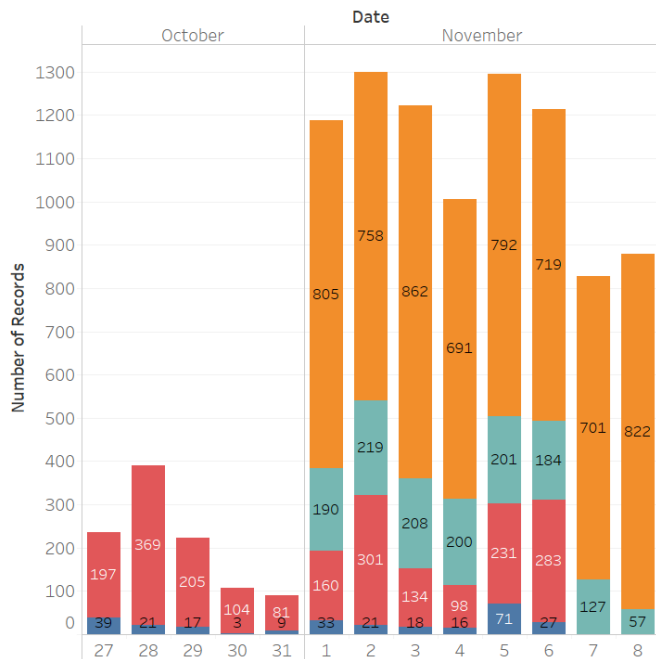
Sentimen Negatif





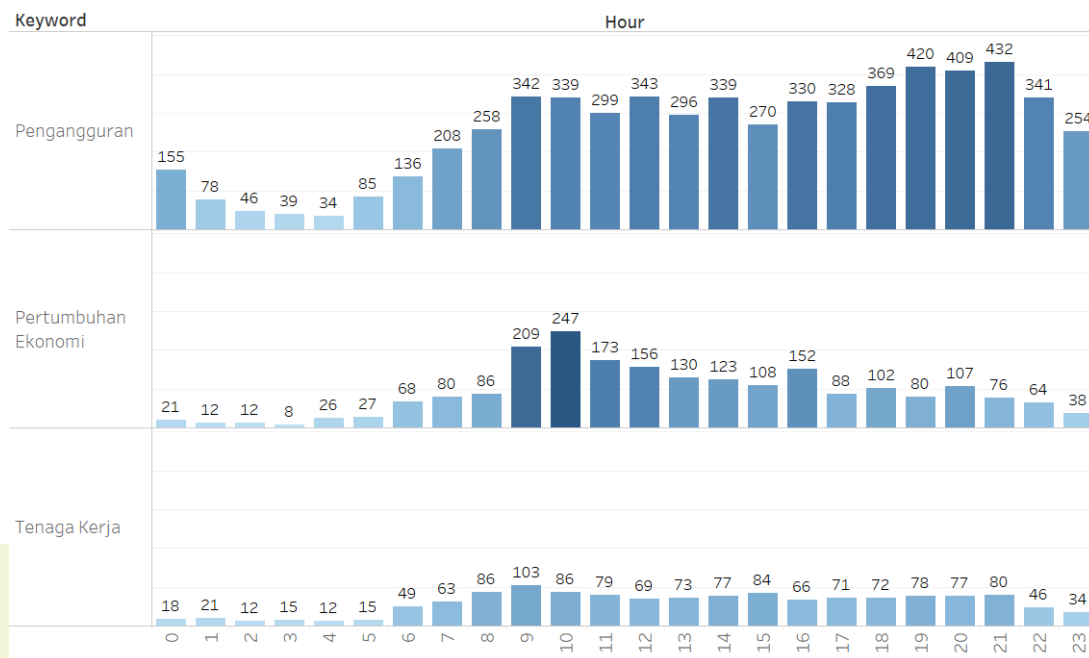
TWITTER : EDA

Jumlah Twit berdasarkan Tanggal

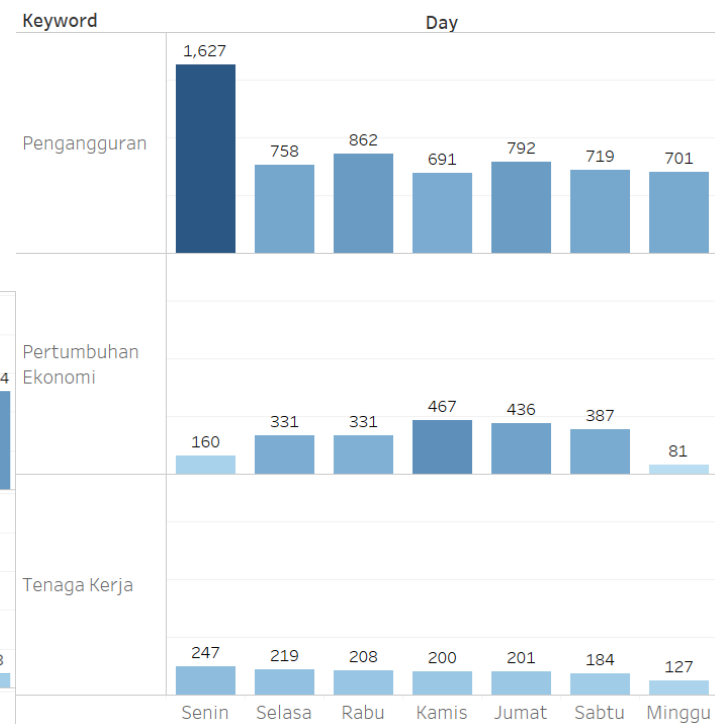


Keyword
Pengangguran
Tenaga Kerja
Pertumbuhan Ekonomi
Badan Pusat Statistik

Jumlah Twit berdasarkan Jam

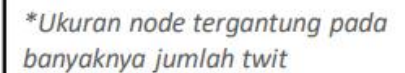


Jumlah Twit berdasarkan Hari





- Periode Waktu : 9 Februari 2020 – 4 Juli 2021





4

Pengembangan



CHALLENGES

- GIGO (Garbage In, Garbage Out)
- Bilingual (misal: Bahasa Indonesia dan Inggris secara bersamaan)
- Penggunaan bahasa daerah
- Penggunaan slang
- Subjektivitas label pada data training





DEVELOPMENT

ONGOING DEVELOPMENT

- Pengumpulan dan perbaikan data set.
- Pengembangan NER menggunakan BERT (IndoNLU/IndoBERT).
- Pengembangan Sentiment Analysis menggunakan pre-trained IndoBERT model.
- Pengembangan analisis teks berita online untuk memprediksi pertumbuhan ekonomi.

Transfer learning

FUTURE DEVELOPMENT

- Optimalisasi model
- Otomatisasi proses secara menyeluruh.
- Mindmap generator.





BADAN PUSAT STATISTIK

ST2023
SENSUS PERTANIAN



TERIMA KASIH

www.bps.go.id

