



Puspresnas
Pusat Prestasi Nasional

BPTI
Balai Pengembangan Talenta Indonesia

Telkom
University



Kampus
Merdeka
INDONESIA JAYA

Supervised Learning untuk Analisis Sentimen Tweet Pilpres 2024:

Klasifikasi Sentiment, Topik, Kandidat dengan GPT-4 API dan XLM-RoBERTa

SD2024040000054

Big Data Challenge



Puspresnas
Pusat Prestasi Nasional

BPTI
Balai Pengembangan Talenta Indonesia

Telkom
University



Kampus
Merdeka
INDONESIA JAYA

Problem

“

Tahun 2024 merupakan tahun politik bagi Indonesia dengan diadakannya pemilihan umum serentak termasuk pemilihan presiden dan wakil presiden. berdasarkan keputusan KPU RI No 21 Tahun 2022.

“

Pemilihan Umum adalah sebuah kompetisi untuk mengisi jabatan-jabatan politik di pemerintahan yang didasarkan pada pilihan formal dari warga negara yang memenuhi syarat.

(Defreates, 2023).

”



ANIES-MUHAIMIN



PRABOWO-GIBRAN



GANJAR-MAHFUD



Puspresnas
Pusat Prestasi Nasional

BPTI
Balai Pengembangan Talenta Indonesia

Telkom
University



Kampus
Merdeka
INDONESIA JAYA

Problem



sumber: www.rri.co.id

Kampanye merupakan serangkaian tindakan komunikasi yang terencana dengan tujuan untuk menciptakan efek tertentu pada sejumlah besar khalayak yang dilakukan secara berkelanjutan pada kurun waktu tertentu.

(Wibowo, 2023).



sumber: www.dafideff.com

Media sosial telah mengubah paradigma komunikasi politik di era modern karena media sosial telah menjadi sarana komunikasi yang kuat dan luas digunakan oleh individu, partai politik, dan kandidat dalam kampanye politik.

(Farid, 2023).



Puspresnas
Pusat Prestasi Nasional

BPTI
Balai Pengembangan Talenta Indonesia

Telkom
University



Kampus
Merdeka
INDONESIA JAYA

Problem



“ Selama masa kampanye pilpres, aktivitas pengguna dalam platform X ini menjadi sangat padat. Hal ini menghasilkan besarnya data yang tersimpan dalam platform X, sehingga data yang tersimpan ini memiliki potensi mendapatkan wawasan mendalam tentang pandangan dan preferensi publik terhadap calon presiden dan wakil presiden. ”



Problem



Tujuan



Mengidentifikasi **sentimen**, **topik**, dan **kandidat** yang terkandung pada tweet antar pengguna media sosial X selama masa kampanye pemilu 2024.









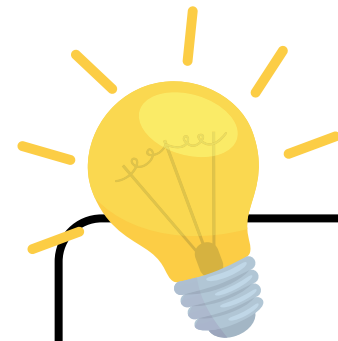
Menemukan **informasi penting** dan **trend** berdasarkan sentimen, topik dan kandidat pada media sosial X selama masa kampanye pemilu 2024.





Model Background

| | XLM-RoBERTa | GPT-4o |
|------------------|--|---|
| Cost |  Open-source | Pay as you go  |
| Inference speed | Personal hardware dependency | OpenAI server dependency |
| Language support |  Multilanguage | Multilanguage  |
| Scalability | Easier to scale, adaptable to your hardware | Scalable with cloud, High resource requirement |
| Accuracy |  Decent Accuracy, Excellent for text classification | Outstanding Accuracy, Excellent for complex task  |



SOLUSI YANG DIANJURKAN



Pendekatan Supervised Learning untuk menentukan 3 fitur penting



Menggunakan GPT-4o Batch API untuk pre-labeled dataset



Base model *pre-trained* XLM-RoBERTa dengan data *pre-labeled* oleh GPT-4o



Deploy model yang telah di-fine-tune untuk memahami isu pilpres 2024.



Sentimen



Topik



Kandidat



XLM-RoBERTa

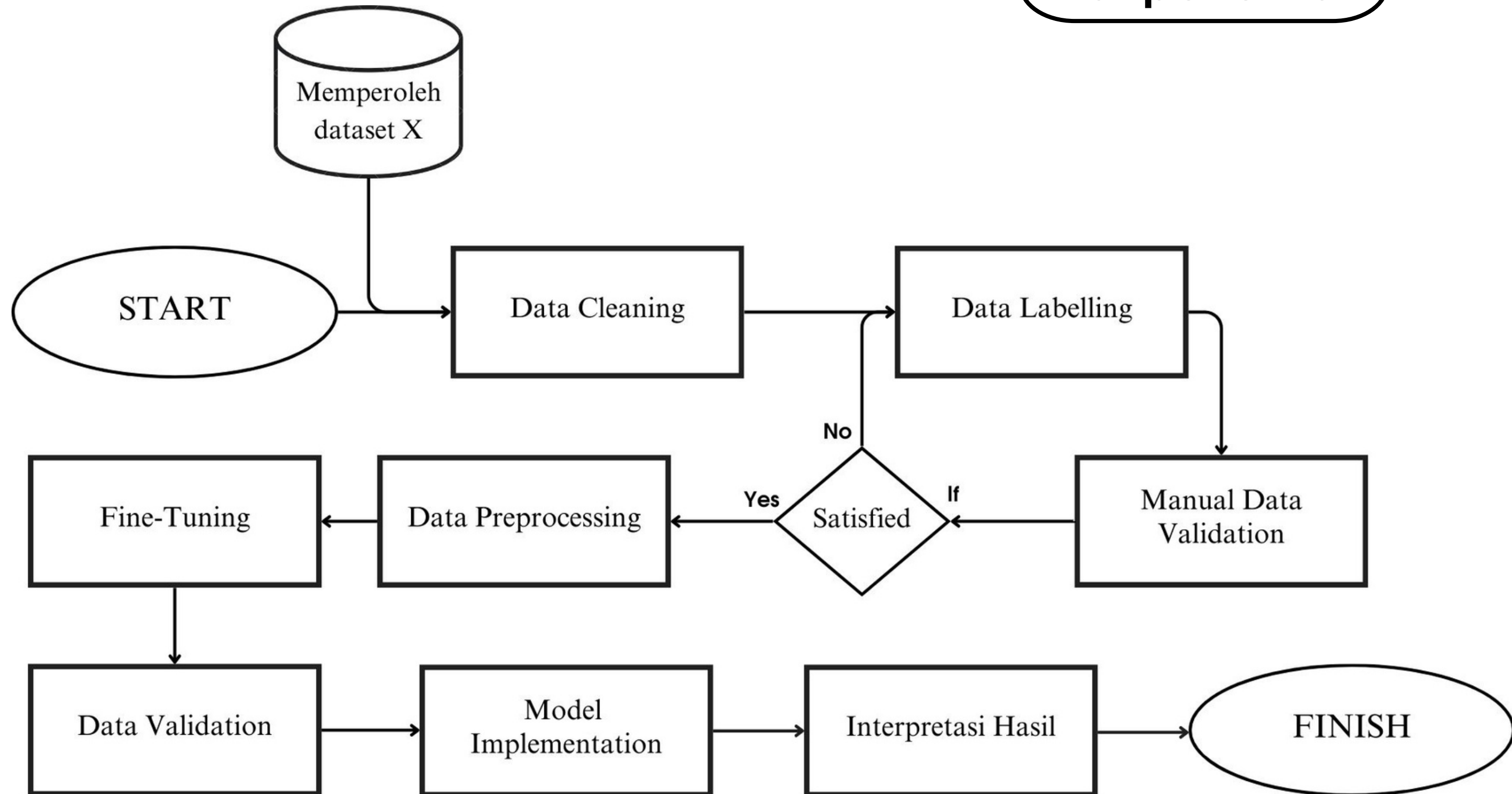


GPT-4o



Alur penelitian

Metodologi





Puspresnas
Pusat Prestasi Nasional

BPTI
Badai Pengembangan Talenta Indonesia

Telkom
University



Kampus
Merdeka
INDONESIA JAYA

Metodologi



Location Cleaning



Cached

Return
"Missing"

yes

Return
"Geocode"

yes

if

Missing

No

if

Geocode

No

Lowercase

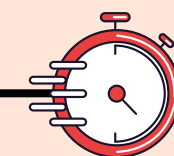
Standarisasi lokasi
dengan Fuzzy
Matching Library

Standarisasi
spasi

Ganti koma dengan
spasi

Remove special
char, kecuali koma

Evaluasi speed



Time taken for preprocessing 9817355 samples: 42.89 seconds
Preprocessing speed: 228907.84 rows/second

Time taken for inference 9817355 samples: 86.86 seconds
Inference speed: 113031.07 rows/second
Total time taken for cleaning and normalizing 9817355 samples: 129.74 seconds
Total cleaning speed: 75667.61 rows/second



Metodologi

Preview hasil

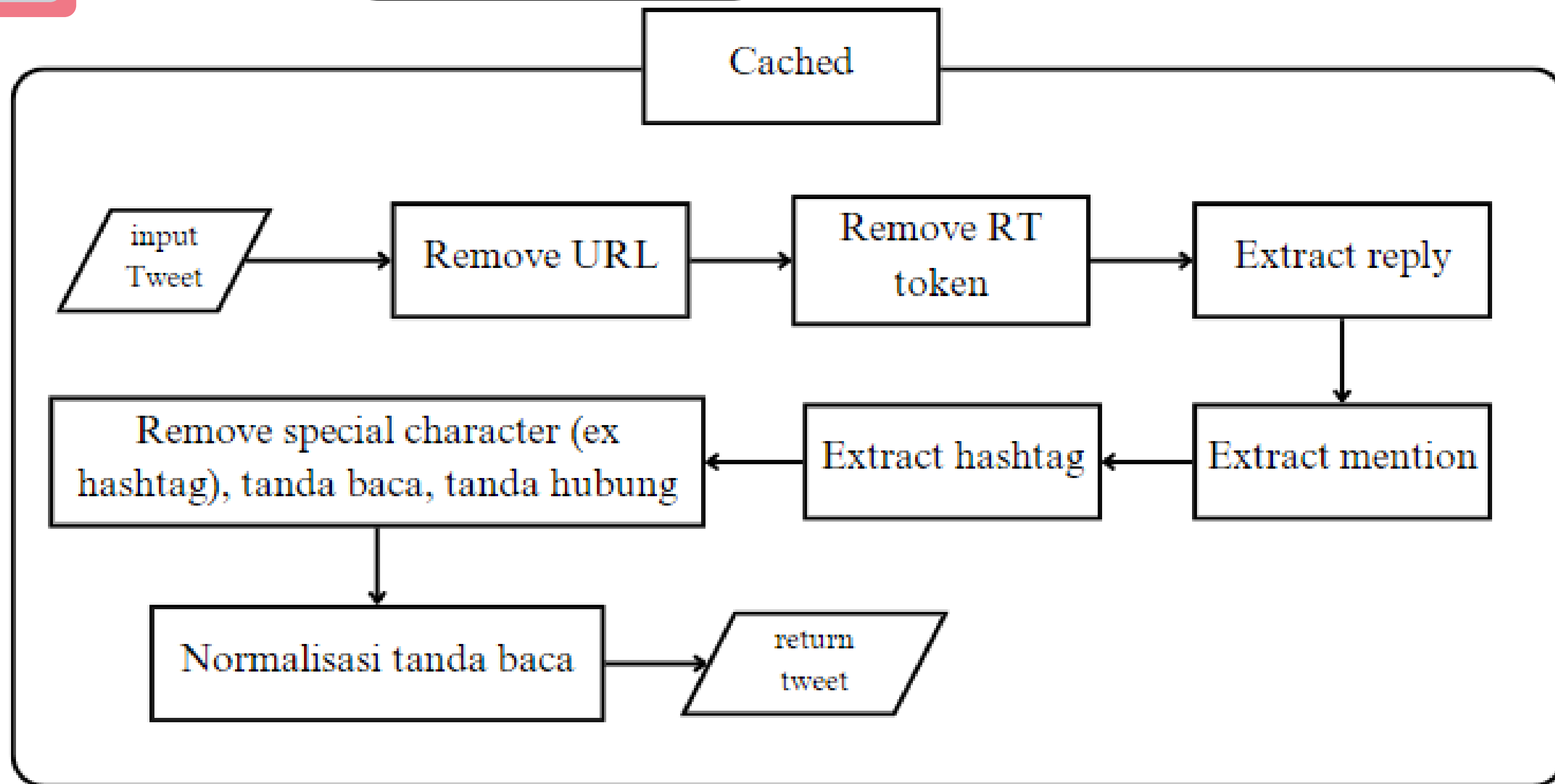


Location Cleaning



| loc | cleaned_loc | standardized_loc |
|--------------------|-------------------|------------------------------|
| Jakarta, Indonesia | jakarta indonesia | Jakarta, Indonesia |
| | Missing | Missing |
| | Missing | Missing |
| Denpasar | denpasar | Denpasar, Bali, Indonesia |
| Indonesia | indonesia | Indonesia |
| Garut | garut | Garut, Jawa Barat, Indonesia |
| Republic of Korea | republic of korea | Unknown Location |
| Jakarta | jakarta | Jakarta, Indonesia |
| Republic of Korea | republic of korea | Unknown Location |

Metodologi





Metodologi



Tweet Cleaning



Evaluasi speed

clean_text_pipeline took 334.5838 seconds, 29341.99 rows/sec

Preview

| content | lang | clean_content | reply_tokens | mentions | hashtags |
|--|------|--|----------------|---|--|
| @cPicLYULRezE8aoQWNW5OUehsJSE3HvtYDTaITid3sM= Sungguh mulianya Prabowo Gibran membagikan bansos kesemua warga | id | Sungguh mulianya Prabowo Gibran membagikan bansos kesemua warga | | ['@cPicLYULRezE8aoQWNW5OUehsJSE3HvtYDTaITid3sM='] | [] |
| RT Selama 1 bulan berarti minimal ada 4 x Jum'at, alias 4.000 kotak. Kalo per kotak 15 rb, berarti sebulan keluar duit 60 jt. Kalo gitu, kita AMIN-kan aja, Gaes. Kalo Prabowo kalah, ada 4.000 orang yg tersantuni makannya. #AMINAJaDulu [RE greypunch] | id | Selama 1 bulan berarti minimal ada 4 x Jumat, alias 4.000 kotak. Kalo per kotak 15 rb, berarti sebulan keluar duit 60 jt. Kalo gitu, kita AMIN-kan aja, Gaes. Kalo Prabowo kalah, ada 4.000 orang yg tersantuni makannya. #AMINAJaDulu | [re greypunch] | [] | ['AMINAJaDulu'] |
| RT Lucu dan Gemes!!! Pak Prabowo masak nasi goreng nih!!! #prabowo #gibran #prabowogibran #masaknasigoreng #masak #nasigoreng #viral #lucu #gemes #indonesiamaju #bersamaindonesiamaju #dekade08 #mendingprabowo #terusmajubersamaprabowo https://t.co/ EKpxfjmvOM [RE dekade_08] | id | Lucu dan Gemes! Pak Prabowo masak nasi goreng nih! #prabowo #gibran #prabowogibran #masaknasigoreng #masak #nasigoreng #viral #lucu #gemes #indonesiamaju #bersamaindonesiamaju #dekade08 #mendingprabowo #terusmajubersamaprabowo | [re dekade_08] | [] | ['prabowo', 'gibran', 'prabowogibran', 'masaknasigoreng', 'masak', 'nasigoreng', 'viral', 'lucu', 'gemes', 'indonesiamaju', 'bersamaindonesiamaju', 'dekade08', 'mendingprabowo', 'terusmajubersamaprabowo'] |



Puspresnas
Pusat Prestasi Nasional

BPTI
Balai Pengembangan Talenta Indonesia

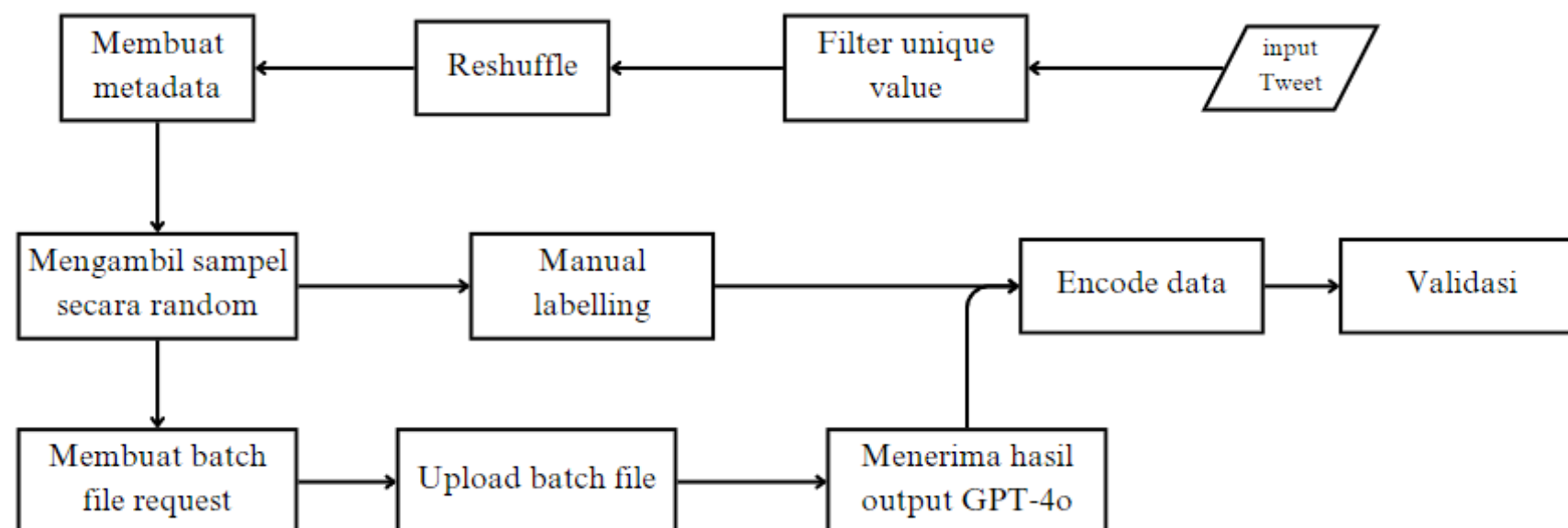
Telkom
University



Kampus
Merdeka
INDONESIA JAYA

Metodologi

Alur Pelabelan data dan validasi



preview using label-studio

#45 jihan.nafie313 #45 21 days ago

Kebersamaan kita adalah kunci keberhasilan. Mari bergandengan tangan memilih Capres Ganjar Pranowo , pemimpin yang akan membawa kita ke puncak kemajuan.JN19BX #JNK #GanjarMahfudRebound #GanjarPranowoPilihanUmat

Choose text sentiment

☒ Positive^[1]

☐ Neutral^[2]

☐ Negative^[3]

☐ Comparative^[4]

Candidate Mention (Multiple Choice)

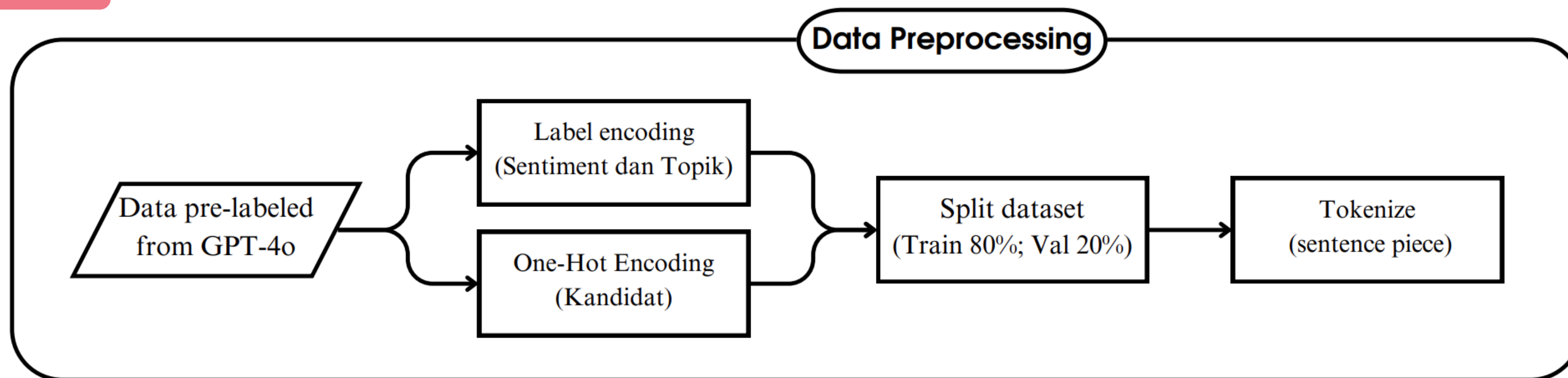
☐ Anies^[5] ☐ Prabowo^[6] ☒ Ganjar^[7] ☐ None^[8]

batch file request .json

```
{
  "custom_id": "request-1",
  "method": "POST",
  "url": "/v1/chat/completions",
  "body": {
    "model": "gpt-4o",
    "messages": [
      {
        "role": "system",
        "content": "You will be provided with a tweet. Classify its sentiment, candidate mentioned,"
      },
      {
        "role": "user",
        "content": "Tweet 1: Rakyat Jabar, senangkan Prabowo untuk meraih prestasi luar biasa. #Ja"
      }
    ],
    "max_tokens": 4000,
    "temperature": 0.1
  }
}
```



Metodologi



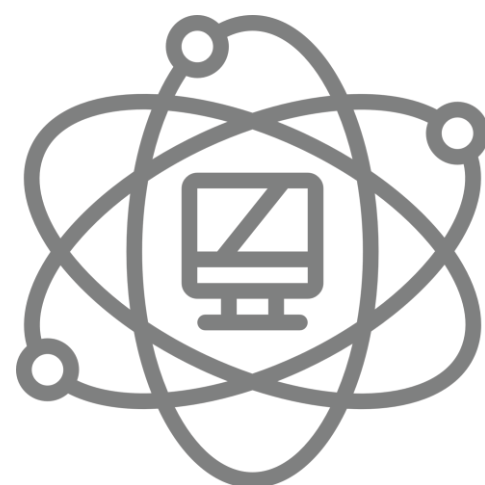
Note

An XLM-RoBERTa tokenizer using **SentencePiece** subword segmentation.

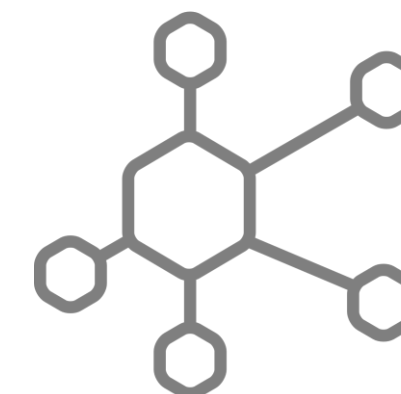
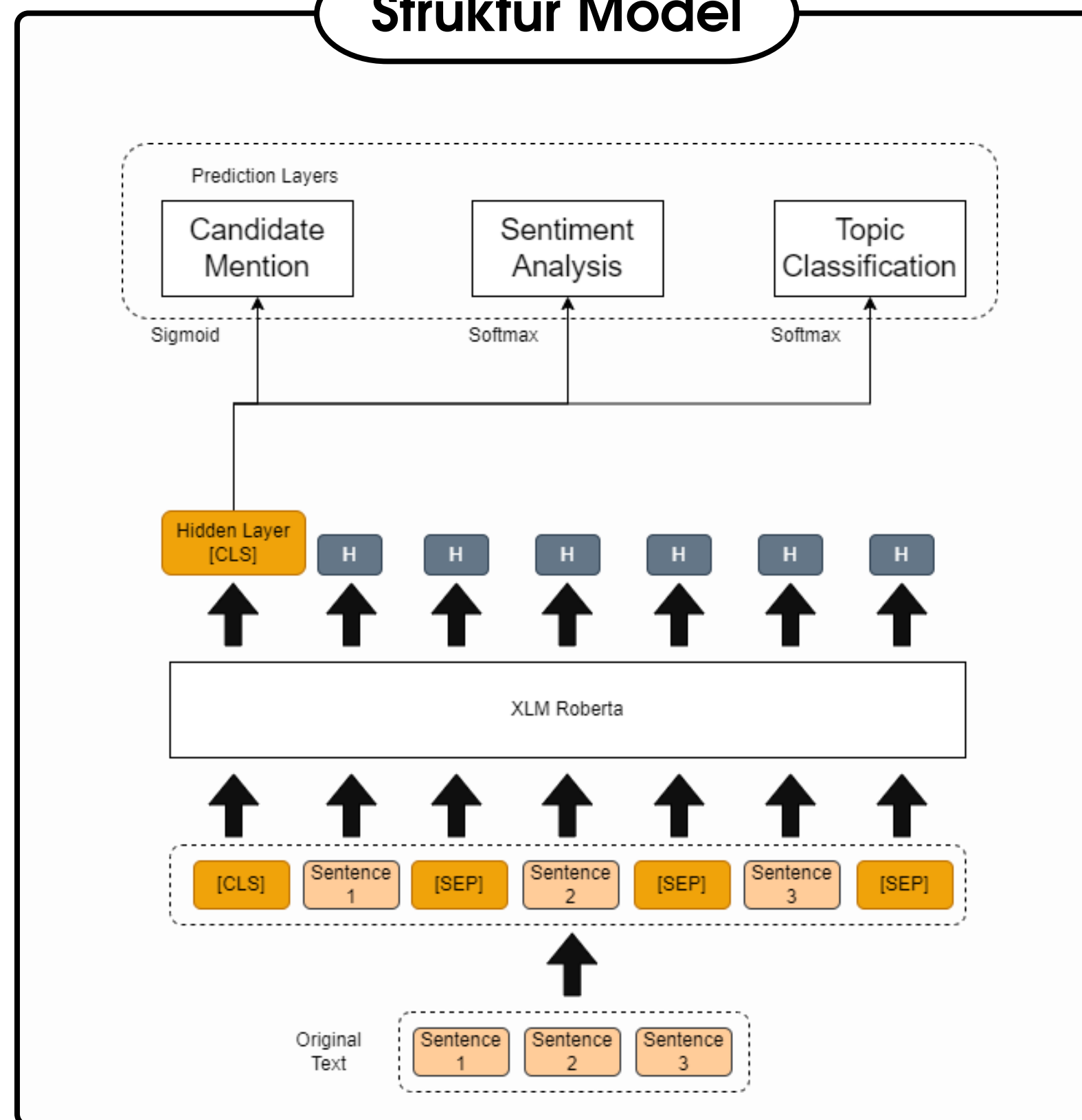
Sumber: Huggingface



Metodologi

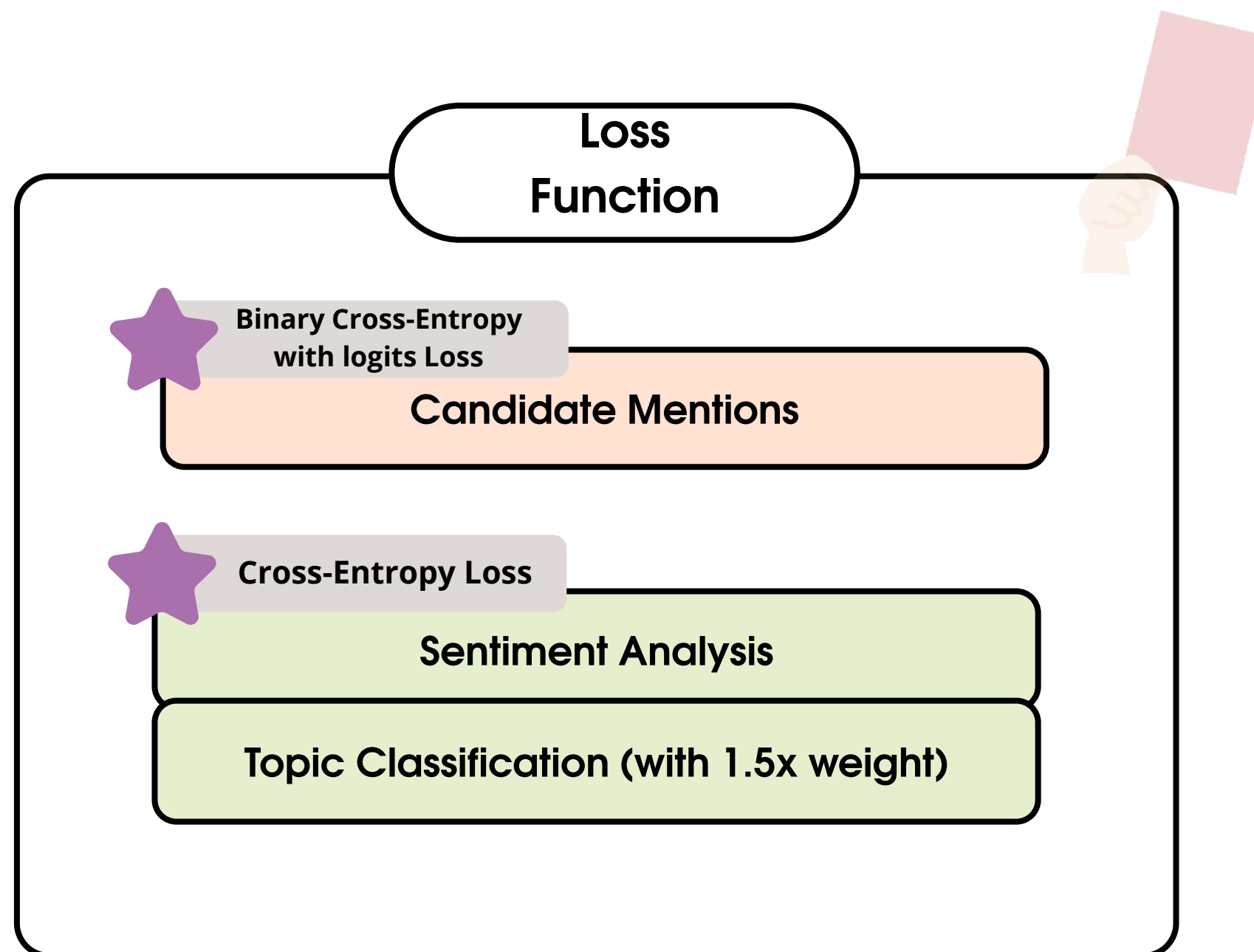


Struktur Model





Metodologi



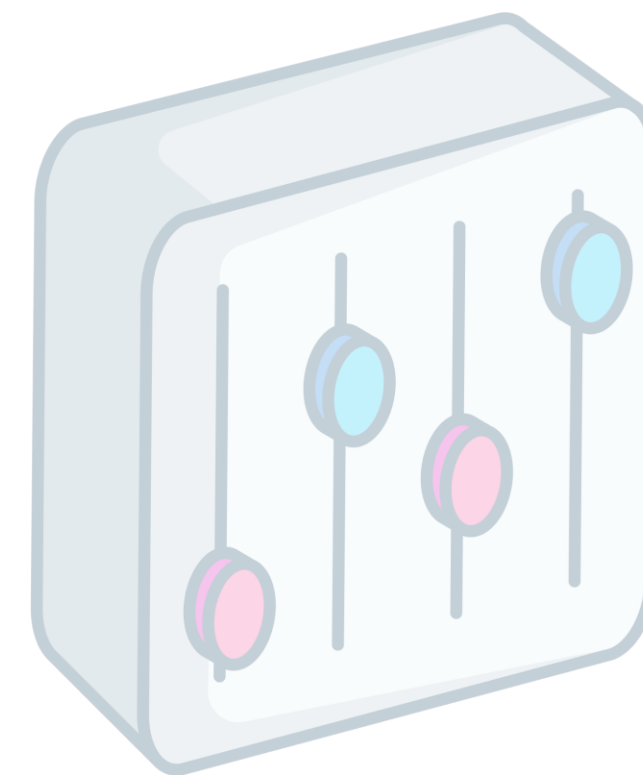
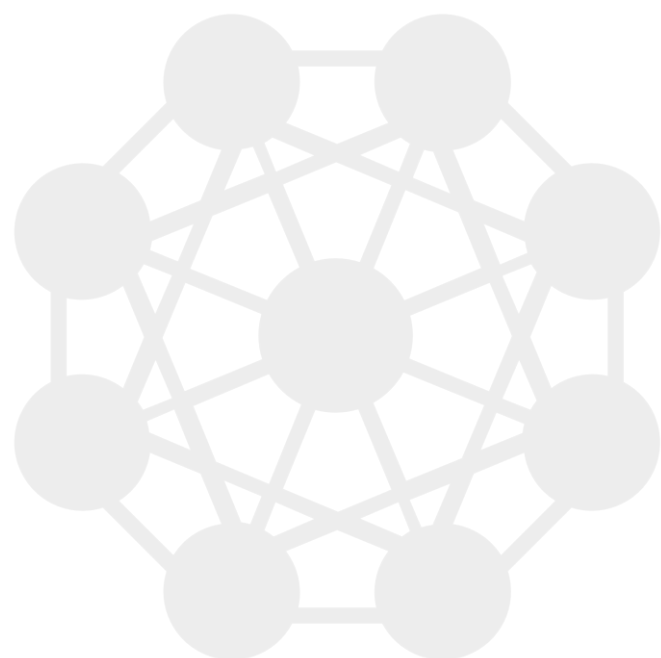


Metodologi

Training Parameters



```
learning rate = 1.5e-5  
per device train batch size = 16  
per device eval batch size = 16  
num train epochs = 4  
weight decay = 0.01  
metric for best model = "eval_loss"  
EarlyStoppingCallback(early_stopping_patience=2)
```





Hasil analisis

Peforma
XLM-RoBERTa

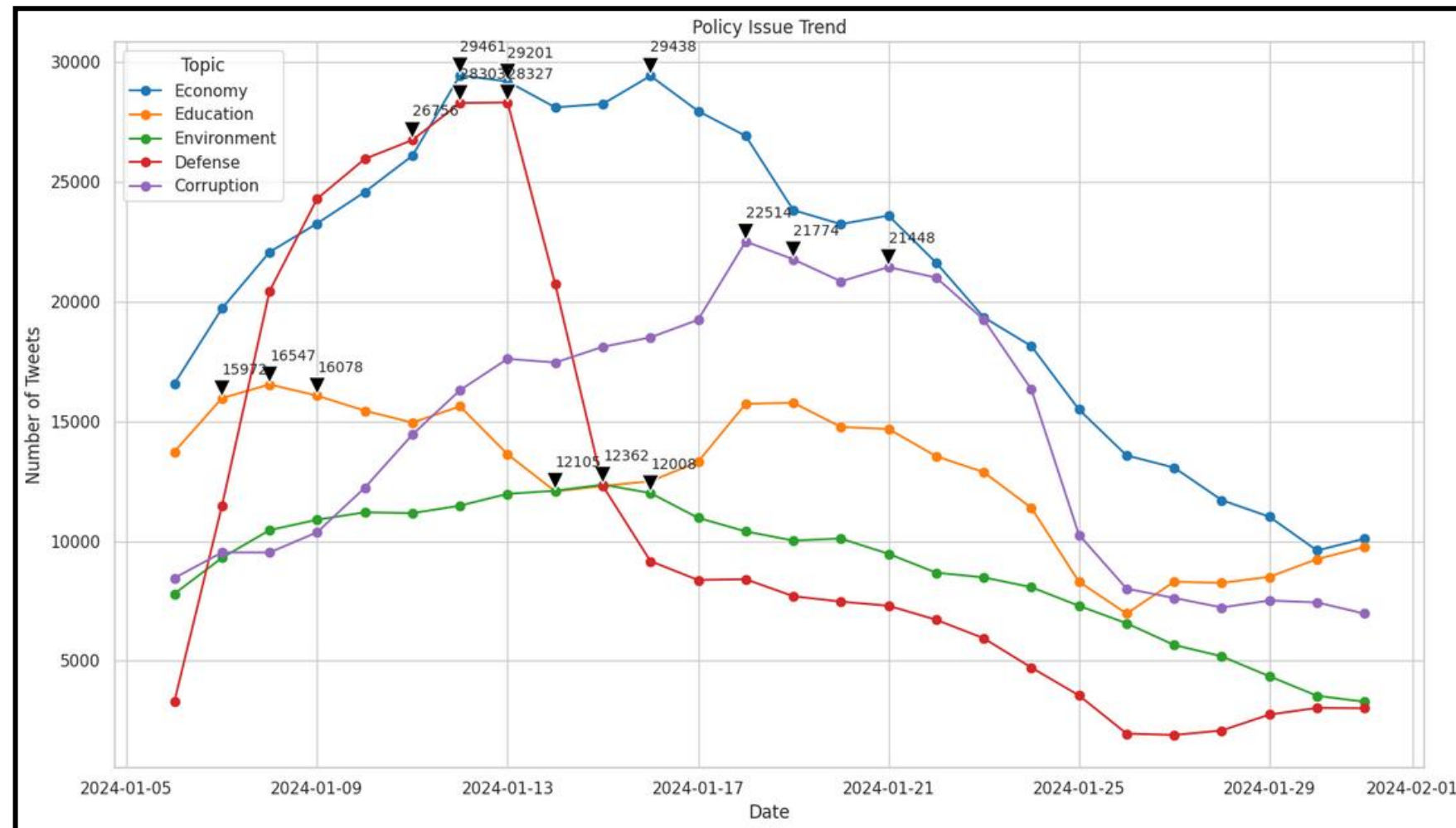
| Epoch | Training Loss | Validation Loss | Candidate F1-Score | Sentiment Accuracy | Topic Accuracy |
|-------|---------------|-----------------|--------------------|--------------------|----------------|
| 1 | 2.165700 | 2.112343 | 0.845313 | 0.805367 | 0.614680 |
| 2 | 1.781800 | 1.897998 | 0.861078 | 0.829992 | 0.654933 |
| 3 | 1.583300 | 1.870082 | 0.864221 | 0.838516 | 0.664720 |
| 4 | 1.410700 | 1.899978 | 0.864740 | 0.834886 | 0.663773 |

Tabel 1. Hasil *fine-tuning* model

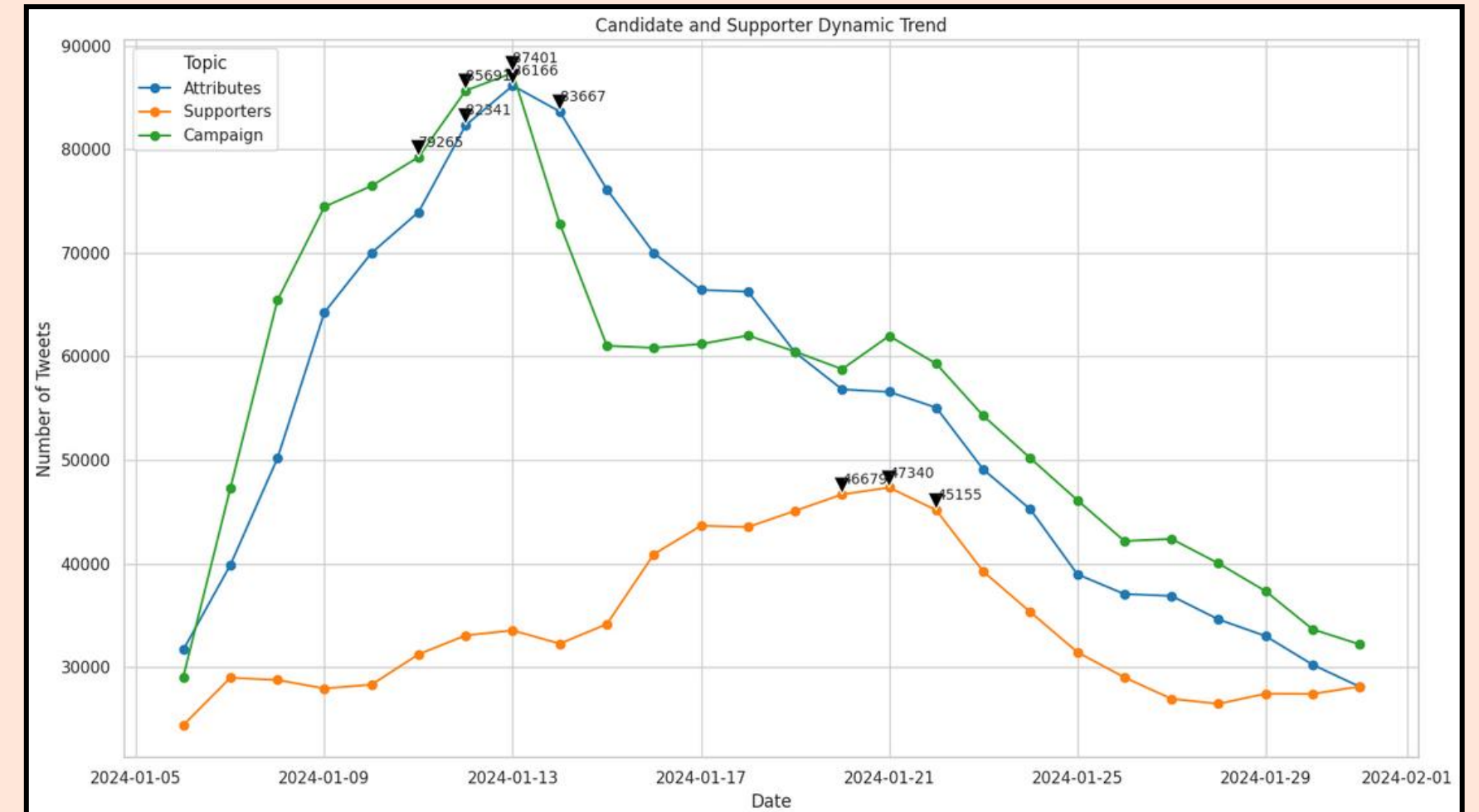


Hasil analisis

Tren Topik



Gambar 1. Line chart tren isu kebijakan politik



Gambar 2. Line chart tren dinamika kandidat dan pendukung

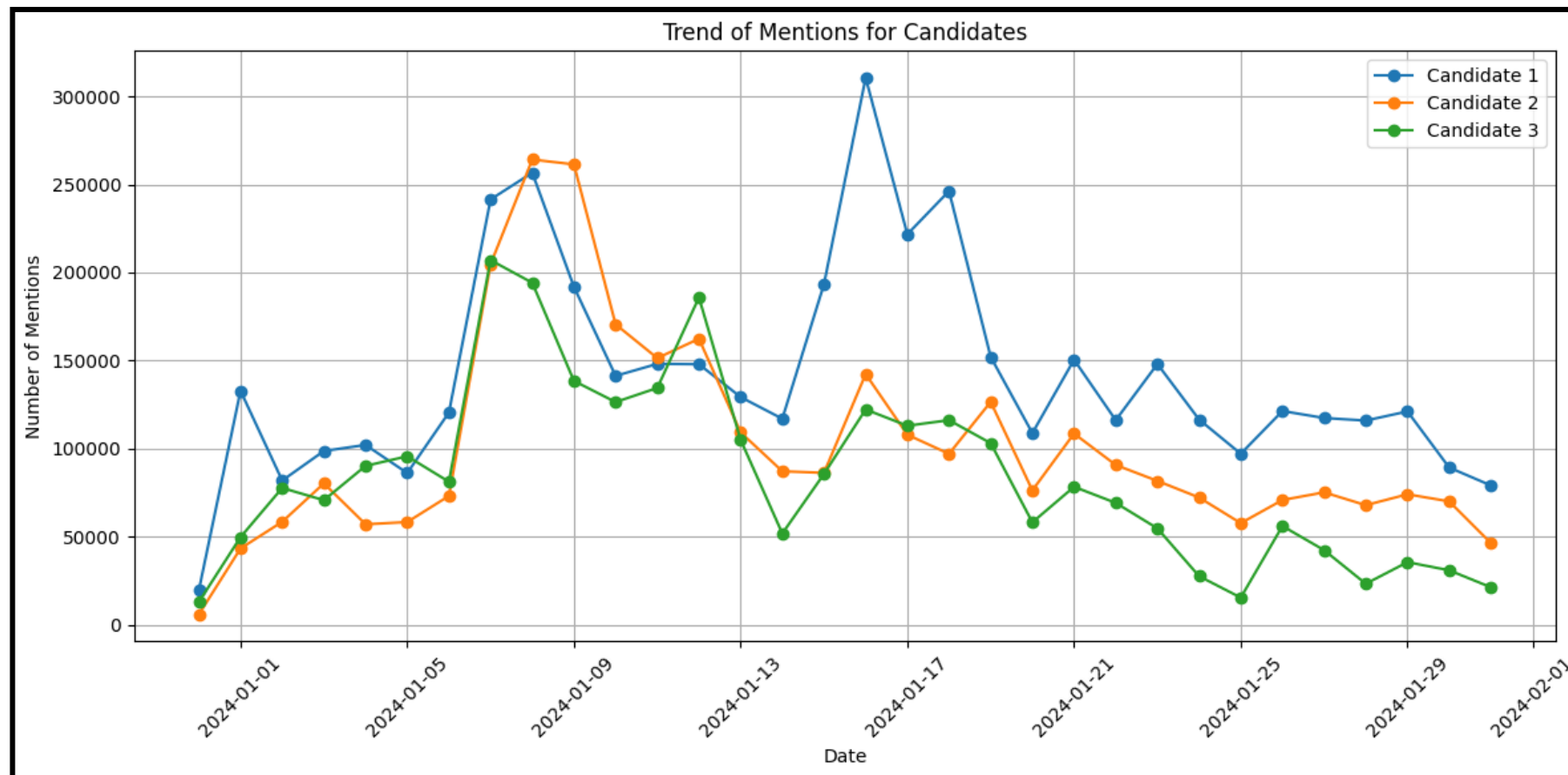
Insight

Publik lebih tertarik membicarakan tentang atribut kandidat dan pendukung dibandingkan dengan kebijakan politik.

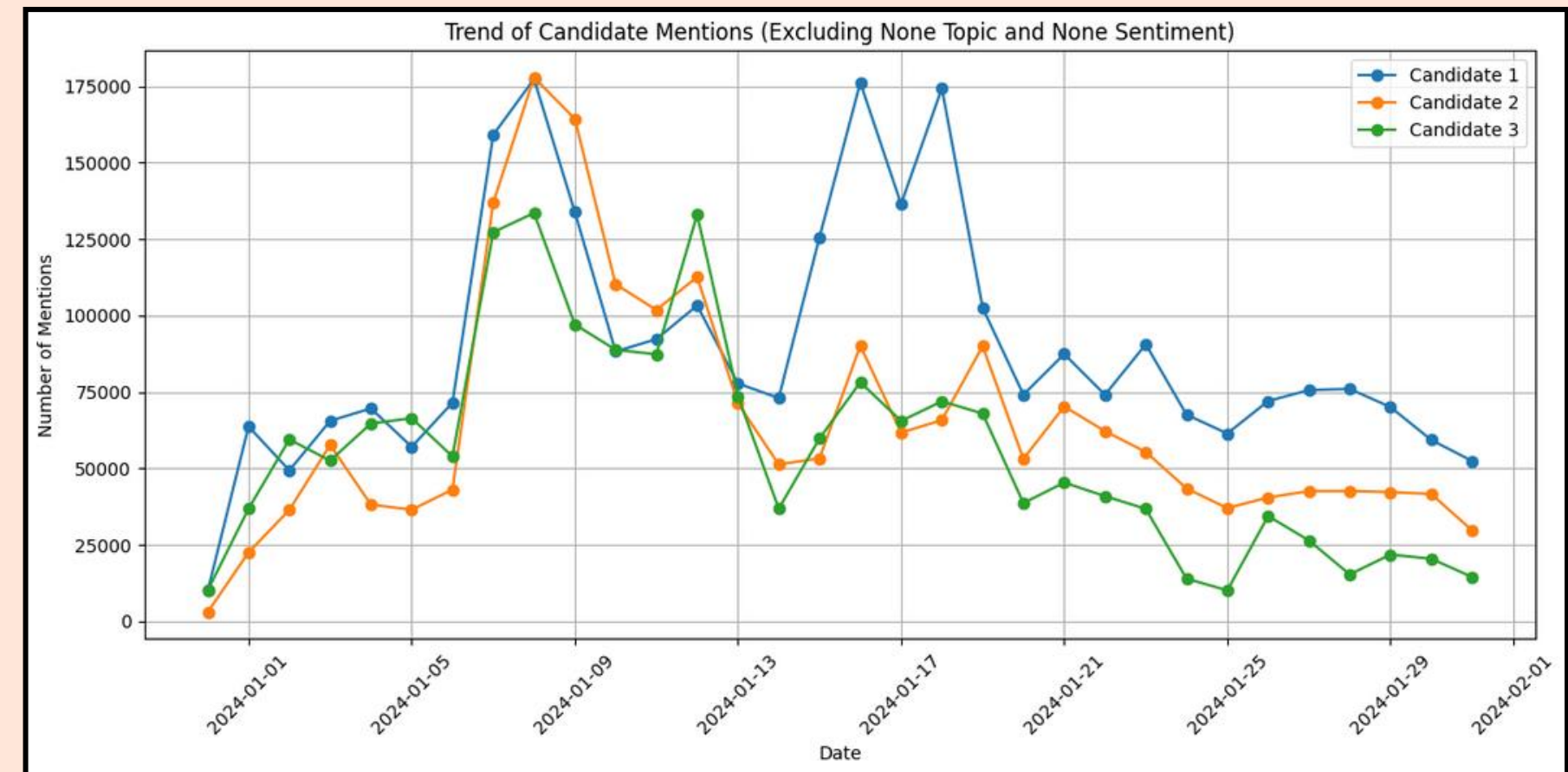


Hasil analisis

Candidate mention trend



Gambar 3. Line chart Candidate mention trend with none topic and none sentiment (neutral)



Gambar 4. Line chart Candidate mention trend without none topic and none sentiment (neutral)

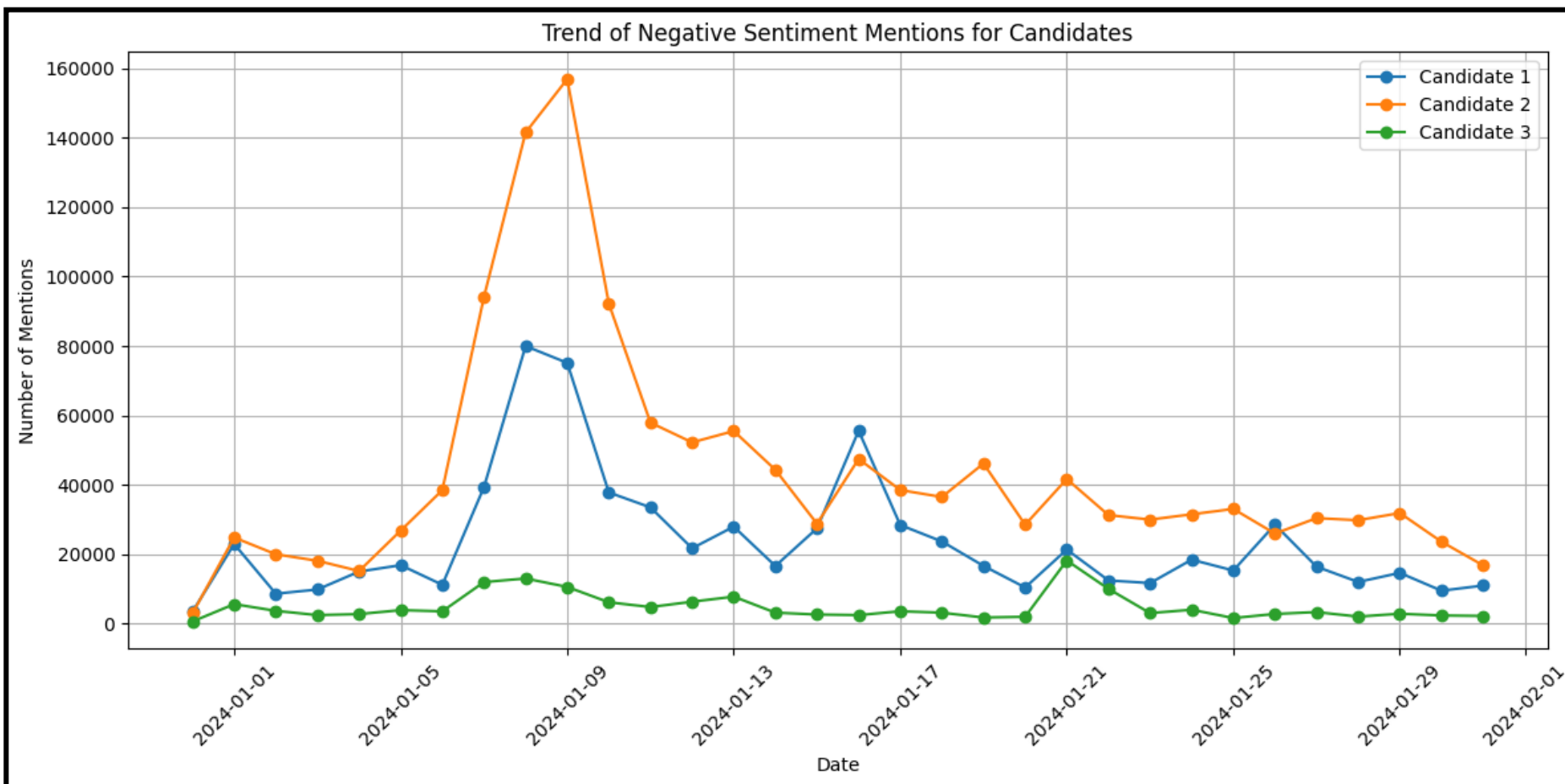
Insight

- pembahasan yang tidak bertopik dan sentimen netral memiliki tren yang mirip dengan penyebutan umum.
- kandidat no 1 memiliki popularitas yang paling unggul dibandingkan dengan 2 kandidat lainnya.

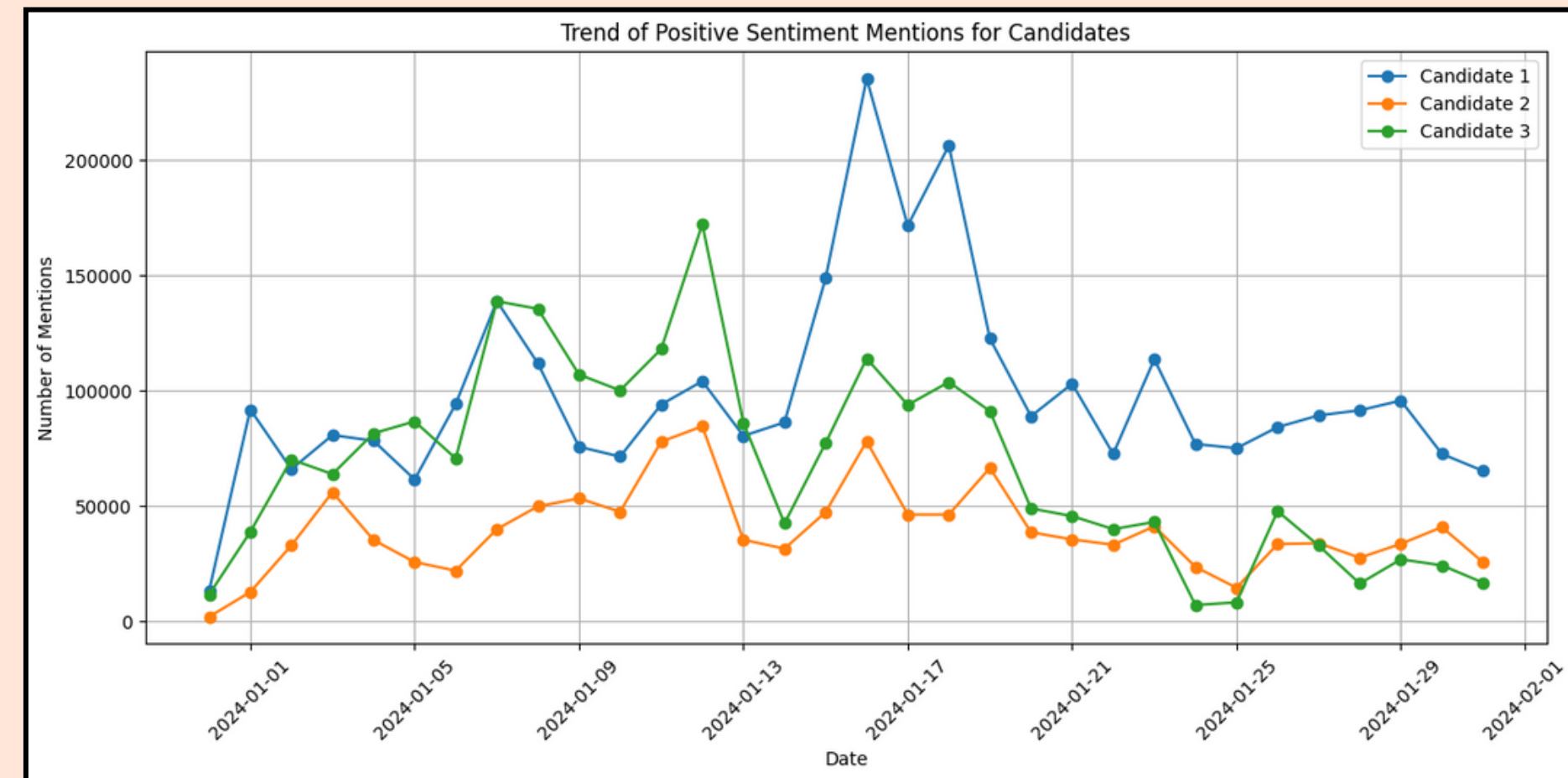


Hasil analisis

Tren Sentimen



Gambar 5. Line chart tren sentimen negatif setiap kandidat



Gambar 6. Line chart tren sentimen positif setiap kandidat

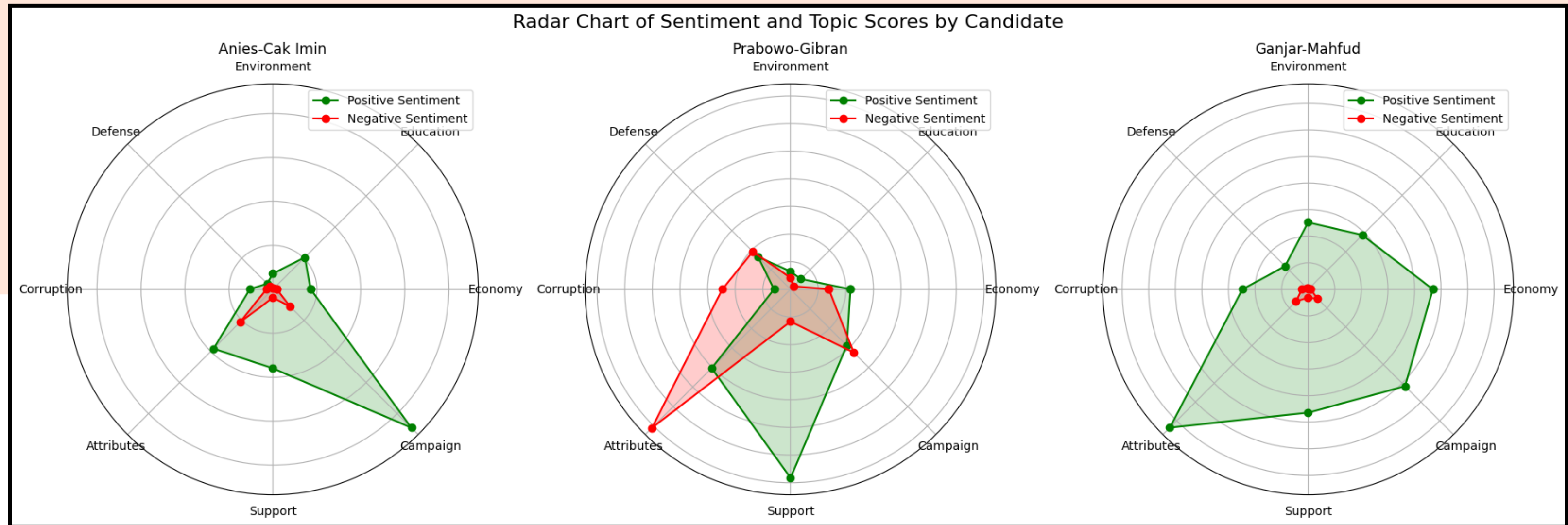
Insight

- Kandidat no 3 menerima sentimen negatif paling rendah dan kandidat no 1 menerima sentiment positif paling tinggi
- kandidat no 2 menerima sentimen negatif paling tinggi dan menerima sentiment positif paling rendah



Hasil analisis

Analisis sentimen terhadap topik setiap kandidat

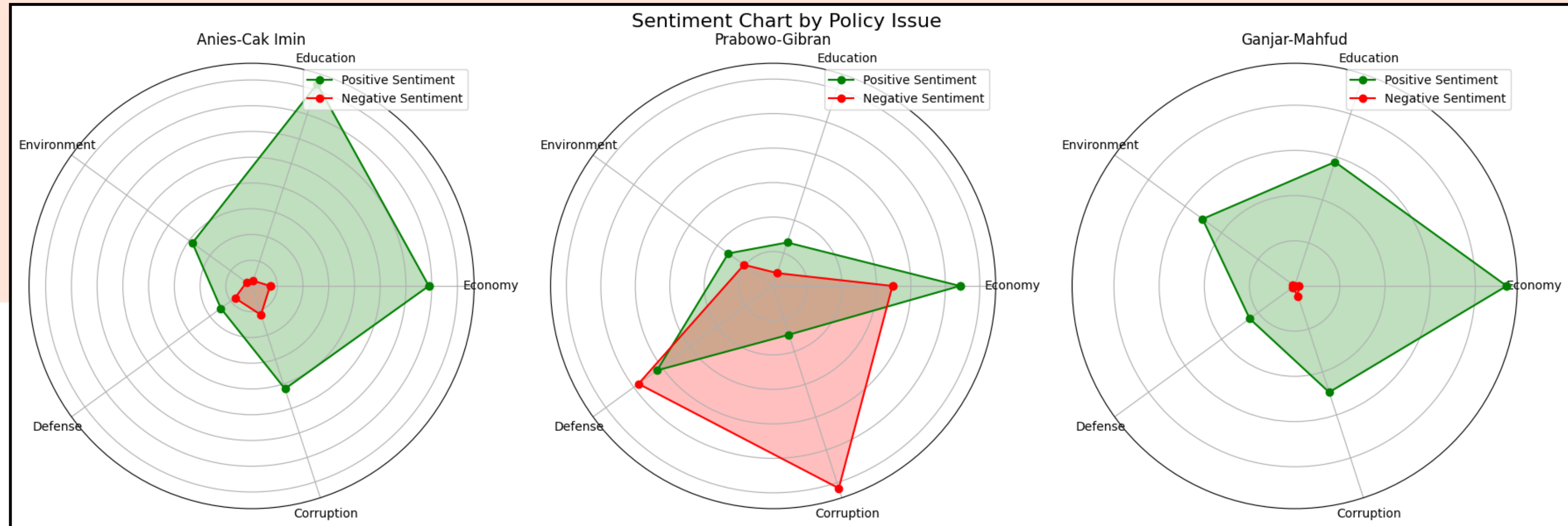


Gambar 7. Radar chart analisis sentimen pada topik setiap kandidat



Hasil analisis

Analisis sentimen terhadap kebijakan setiap kandidat



Gambar 8. Radar chart analisis sentimen pada kebijakan setiap kandidat



Rekomendasi

| | Kandidat 1 | Kandidat 2 | Kandidat 3 |
|-------------------|--|---|--|
| Atribut | <ul style="list-style-type: none">Hindari kesan berbicara tanpa tindakan. | <ul style="list-style-type: none">Perbaiki sikap dan cara bicara agar tidak terkesan emosional.Tekankan jiwa nasionalis. | <ul style="list-style-type: none">Bangun citra positif pemimpin dekat rakyat sebagai kunci utama untuk sentimen positif. |
| Dukungan | <ul style="list-style-type: none">Rekonsiliasi dengan partai politiklawan stigma berkoalisi dengan ormas tertentu | <ul style="list-style-type: none">Manfaatkan dukungan partai terbanyak dan terbesar.Lawan stigma pendukung prabowo pendidikan rendah. | <ul style="list-style-type: none">Pastikan dukungan koalisi memadai dan tidak bergantung pada satu partai |
| Kampanye & Debat | <ul style="list-style-type: none">Tekankan mengkritik program kerja lawan dibanding personalitas | <ul style="list-style-type: none">Kampanye 2 arahAktif debat capres dengan materi dan data yang matang. | <ul style="list-style-type: none">Manfaatkan performa debat capres sentimen negatif terkecil dengan lebih agresif menyerang lawan. |
| Kebijakan Politik | <ul style="list-style-type: none">Komunikasikan kebijakan lebih baik secara realistis disertai dukungan ahli. | <ul style="list-style-type: none">Persiapkan klarifikasi mengenai kepemilikan lahan, alutsista bekas, dan serangan peretasan | <ul style="list-style-type: none">Sentimen positif tertinggi, tetapi tidak terbanyak; fokus tingkatkan perhatian publik |



Kesimpulan

Kesimpulan

Hasil fine-tuning menunjukkan bahwa model memiliki akurasi sentiment sebesar **0.838516**, akurasi topik sebesar **0.664720**, dan F1-Score kandidat sebesar **0.864221**.

Analisis hasil klasifikasi model XLM-RoBERTa yang telah di fine-tuning menunjukkan bahwa,

- topik yang paling banyak dibicarakan oleh publik dalam media sosial X adalah seputar dinamika pendukung dan atribut kandidat, bukan kebijakan politik.
- sentimen publik cenderung menunjukkan pola negatif terhadap kandidat 2. Di sisi lain, kandidat 1 mendapatkan lebih banyak sentimen positif dari publik X.
- pembahasan yang tidak bertopik dan sentimen netral memiliki tren yang mirip dengan penyebutan umum, dengan kandidat 1 yang mendominasi jumlah mention di X.



REFERENSI

REFERENSI



KPU RI. (2022). KEPUTUSAN KOMISI PEMILIHAN UMUM REPLUBLIK INDONESIA NOMOR 21 TAHUN 2022.

Astrianti Defretes, D., & Laga Kleden, K. (2023). EFEKTIVITAS PEMILIHAN UMUM SERENTAK TAHUN 2024. Jurnal Hasil Penelitian, 8, 49–50. <http://jurnal.untag-sby.ac.id/index.php/jhp17>

Wibowo, K. (2023). Kampanye Partisipatif Dalam Pemilihan Umum Tahun 2024. Jurnal Pengawasan Pemilu, 109–118.

Salman Farid, A. (2023). PENGGUNAAN MEDIA SOSIAL DALAM KAMPANYE POLITIK DAN DAMPAKNYA TERHADAP PARTISIPASI POLITIK DAN PERSEPSI PUBLIK (Vol. 4, Issue 1).

Sentimen pascadebat capres dan cawapres 2024. (2024, February 5). [antaranews.com](https://www.antaranews.com).
<https://www.antaranews.com/infografik/3948588/sentimen-pascadebat-capres-dan-cawapres-2024>



Lampiran

Metadata Topik



Topic Discussed - Policy Issues:

- 0 = **None**: Tidak membahas isu kebijakan atau kinerja kandidat.
- 1 = **Economy/Industry/Infrastructure/Technology**: Kebijakan ekonomi, industri, infrastruktur, atau teknologi.
- 2 = **Education/Healthcare/Social Service**: Kebijakan pendidikan, kesehatan, atau layanan sosial.
- 3 = **Agriculture/Environment/Social Culture**: Kebijakan pertanian, lingkungan, atau budaya sosial.
- 4 = **Defense/Security/Foreign Policy**: Kebijakan pertahanan, keamanan, atau luar negeri.
- 5 = **Corruption/Justice/Law**: Korupsi, reformasi sistem peradilan, atau masalah hukum.
- 6 = **Candidate Personal Attributes/Background**: Pengalaman, latar belakang pendidikan, atau skandal kandidat.
- 7 = **Supporter/Stakeholder/Parties**: Dukungan, pemangku kepentingan, atau partai politik.
- 8 = **Rallies/Speeches/Debates/Campaign Performance**: Rallies, pidato, debat, atau kinerja kampanye.



Puspresnas
Pusat Prestasi Nasional

BPTI
Balai Pengembangan Talenta Indonesia

Telkom
University



Kampus
Merdeka
INDONESIA JAYA

Terima Kasih

