

# TWITTER (X) SENTIMENT ANALYSIS AND TOPIC MODELLING OF THE 2024 INDONESIAN PRESIDENTIAL ELECTION

KELOMPOK 9  
DATA MINING II SD-A2



# GROUP MEMBERS



NIKHLA ISFA KHURAIYA (049)

SHOFIA ISHMA NAJIYYA (087)

JASMINE TAJ ARIVA (063)

RICHARD HASANNAIN M. (093)

FARRAS ALFIYYAH (086)



# PROBLEM STATEMENT

## Latar Belakang

Untuk mengetahui dan menganalisis opini publik di Twitter (X) terkait pemilihan presiden 2024. Dalam era media sosial yang meluas, kami tertarik menganalisis sentimen dan isu-isu krusial di platform ini, dengan tujuan mendukung strategi kampanye dan pengambilan keputusan politik yang lebih terarah dan informasional.

## Kenapa pilih topik ini?

Karena kami melihat permasalahan pada pemahaman opini publik di Twitter (X) terkait pemilihan presiden 2024. Keriuhan informasi menciptakan tantangan dalam merinci sentimen dan isu-isu krusial. Kami ingin memberikan pandangan secara menyeluruh untuk mendukung strategi kampanye dan pengambilan keputusan politik yang informasional menjelang pemilihan presiden.

## Apa tujuan dari topik ini?

Untuk menganalisis opini publik di Twitter (X) terkait pemilihan presiden 2024. Melalui analisis sentimen dan isu-isu krusial di media sosial, kami bertujuan mendukung penyusunan strategi kampanye yang lebih terarah dan responsif, serta memberikan informasi relevan untuk pengambilan keputusan politik.

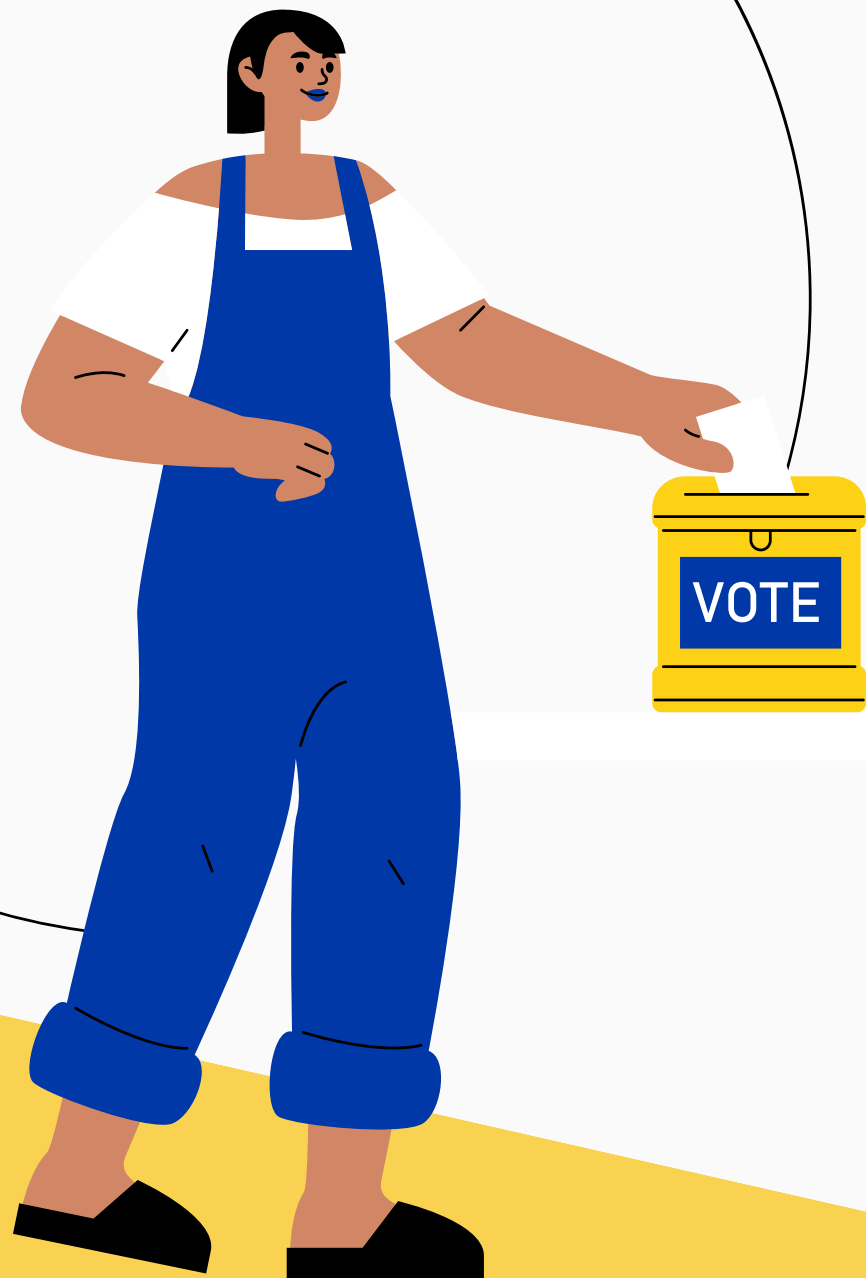


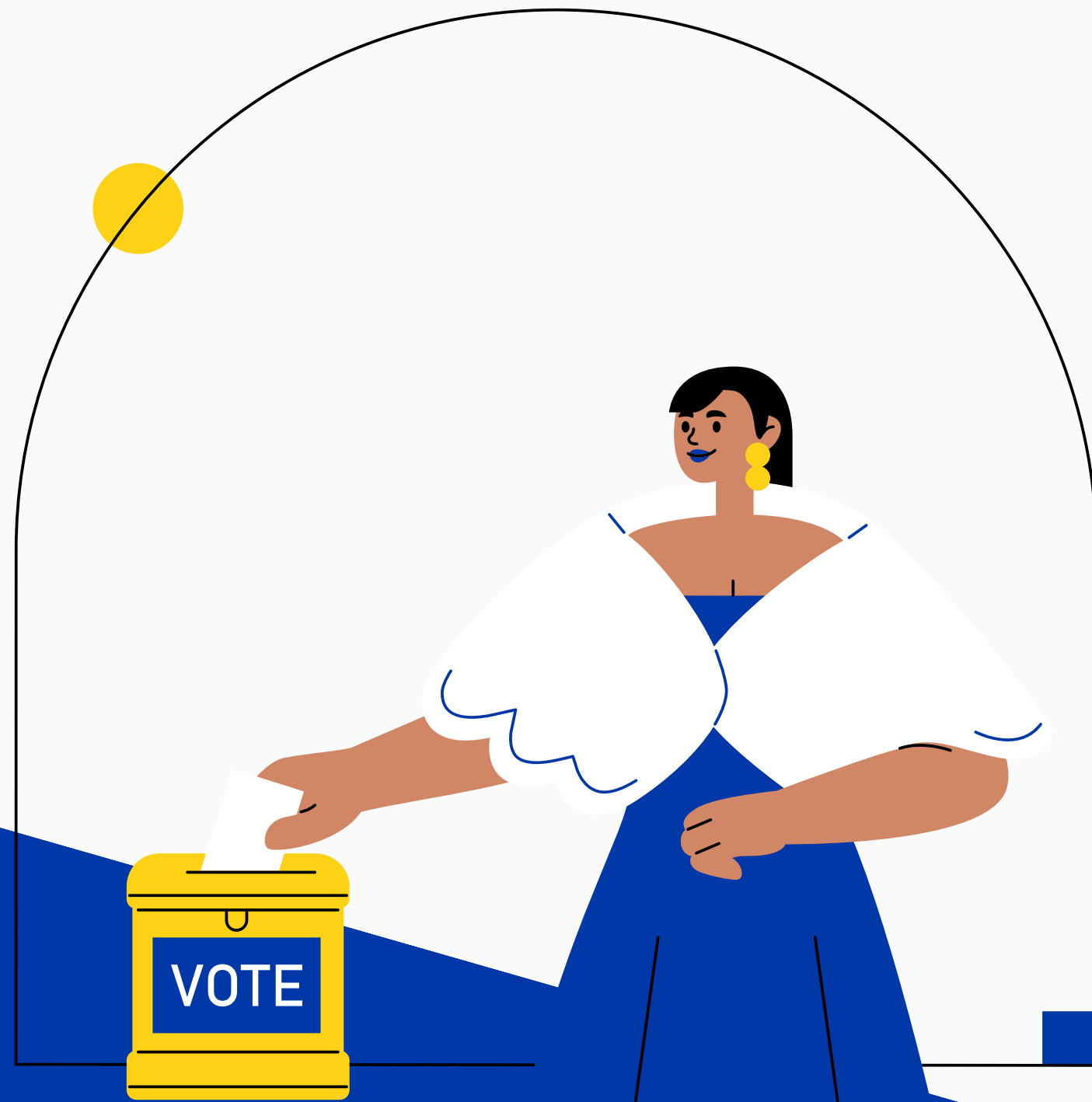
# Pilpres

Pemilihan Umum (Pemilu) adalah sarana kedaulatan rakyat untuk memilih anggota Dewan Perwakilan Rakyat (DPR), anggota Dewan Perwakilan Daerah (DPD), Presiden dan Wakil Presiden, anggota Dewan Perwakilan Rakyat Daerah (DPRD), yang dilaksanakan secara langsung, umum, bebas, rahasia, jujur, dan adil dalam Negara Kesatuan Republik Indonesia berdasarkan Pancasila dan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945. Pemilihan Umum (atau disingkat sebagai pemilu) merupakan salah satu alat penting dalam sistem perwakilan di negara demokrasi.

## Analisis Sentimen

Analisis sentimen adalah suatu proses otomatis untuk mengekstrak, memproses, dan memahami teks yang tidak terstruktur dengan tujuan menarik informasi mengenai sentimen yang terdapat dalam sebuah kalimat, pendapat, atau opini. Analisis sentimen bertujuan untuk mengevaluasi pendapat serta kecondongan suatu opini terhadap suatu topik, baik itu bersifat negatif maupun positif.





# Topic Modelling

*Topic modelling* merupakan suatu teknik yang digunakan untuk mengungkap, menemukan, dan menjelaskan konsep utama dari sebuah teks. *Output* yang diberikan dari *topic modelling* berupa sekumpulan topik dengan beberapa kluster kata yang terdapat dalam dokumen. Dalam *topic modelling* terdapat beberapa metode yang dapat digunakan, salah satunya adalah BERTopic. BERTopic sendiri merupakan metode yang memanfaatkan BERT atau transformer sebagai alat untuk melakukan *topic modelling* itu sendiri

## Wordcloud

*Wordcloud* merupakan salah satu visualisasi yang digunakan untuk memvisualisasi bahasa atau teks pada data. Dalam *wordcloud*, kata yang lebih sering muncul dari kata yang lain akan tercetak dengan font lebih besar juga daripada kata lainnya. *Wordcloud* ini merupakan visualisasi yang populer karena mudah dipahami. *Wordcloud* didapatkan dari nilai paling signifikan dari setiap *vector eigen* dalam sebuah *cluster*

# Sumber Data

Data yang digunakan dalam penelitian ini diperoleh melalui proses pengumpulan *tweet* menggunakan kode *tweet-harvest* yang ditulis dengan bahasa pemrograman *Python* menggunakan *library Tweepy*. Untuk keperluan penelitian ini, kami menggunakan API Twitter (X) untuk mengakses aliran *tweet* yang berkaitan dengan topik pemilihan presiden 2024. Dengan parameter yang kami gunakan sebagai berikut:

Pasangan Calon Presiden	Keywords	Language	Date	Data Count
Paslon 01 - Anies & Muhaimin	“anies imin”, “amin”	Indonesia	11 November 2023 - 13 Desember 2023	483
Paslon 02 - Prabowo & Gibran	“prabowo gibran”	Indonesia	11 November 2023 - 13 Desember 2023	365
Paslon 03 - Ganjar & Mahfud	“ganjar mahfud”	Indonesia	11 November 2023 - 13 Desember 2023	490



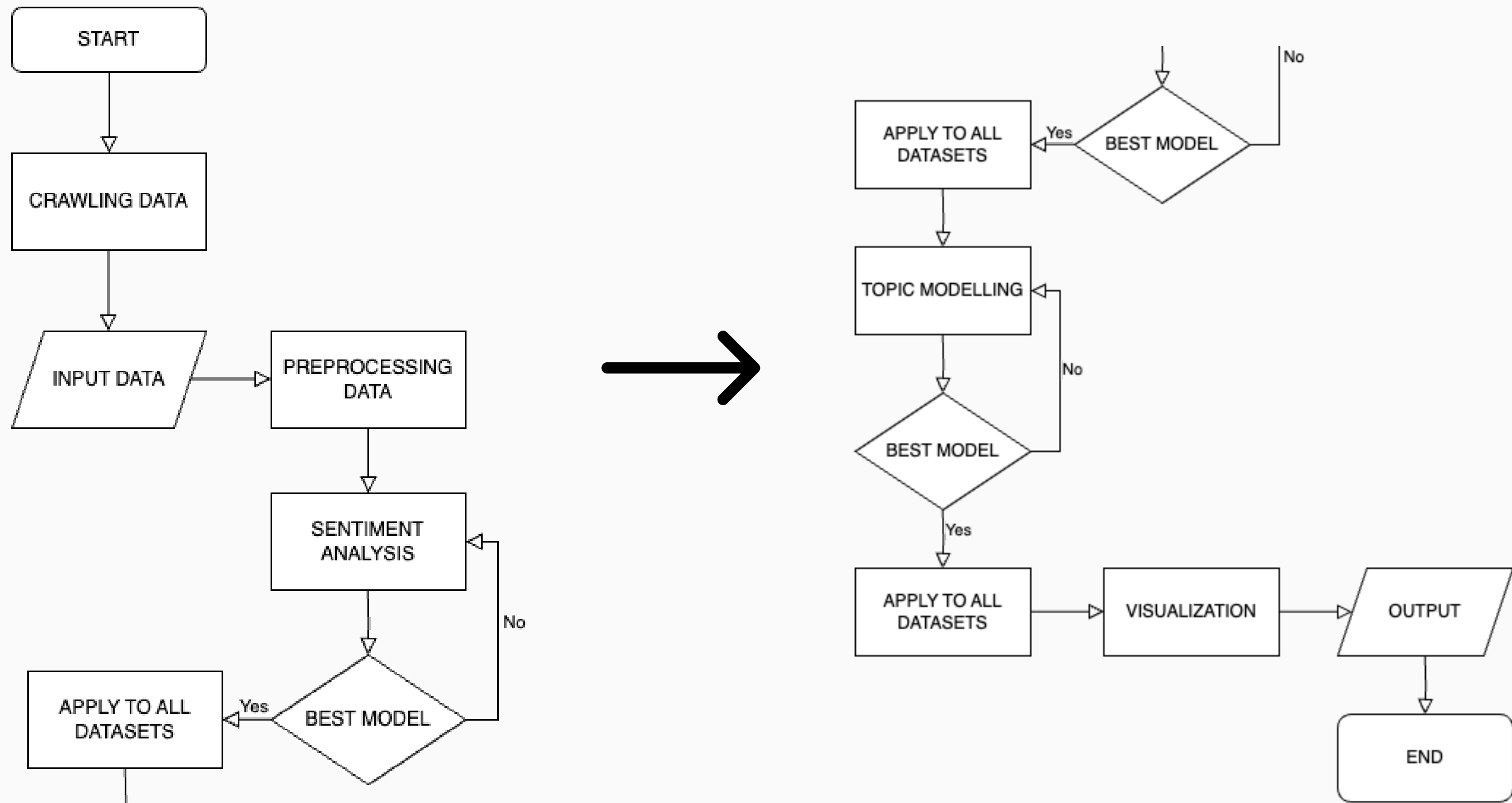
# Penyimpanan Data

Data yang diperoleh dari proses pengumpulan tweet kemudian disimpan dalam format CSV. Berikut adalah variabel yang digunakan dari data masing-masing paslon:

Variabel	Keterangan	Skala Data
full_text	cuitan yang ditulis user	<i>string</i>
no_paslon	nomor paslon yang sedang diperbincangkan	kategorik
label	sentimen positif atau negatif dari cuitan	kategorik
retweet_count	jumlah retweet dalam cuitan	<i>string</i>
username	nama akun user	<i>string</i>

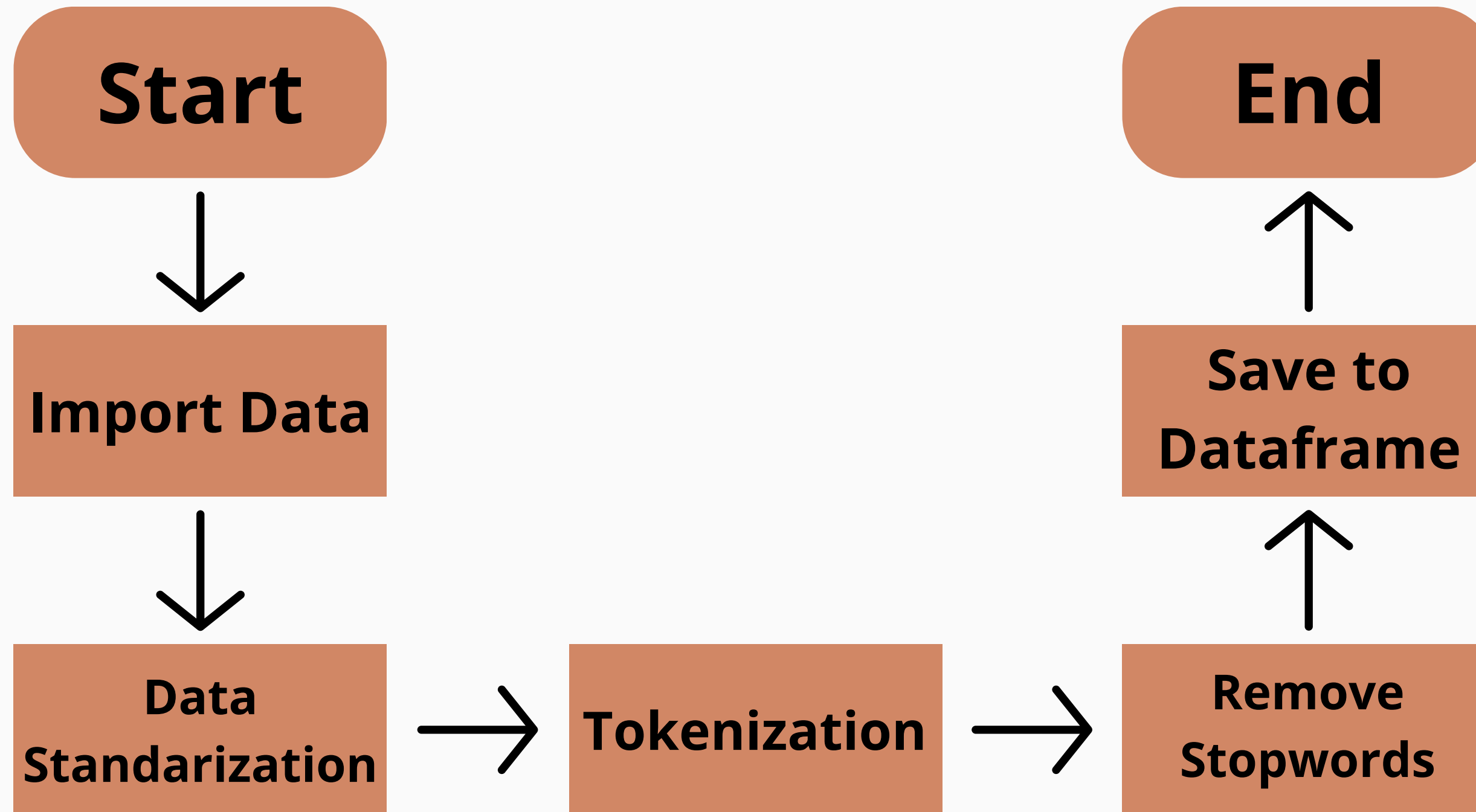
Didapatkan data sebanyak 1338 yang dibagi menjadi data train sebanyak 300 dan lainnya sebagai test data. Data test yang kami ujikan dengan model yang telah dibentuk akan digabungkan kembali dengan data train untuk dilakukan *topic modelling*

# Langkah Penelitian





# Langkah Preprocessing



# MOST RETWEETED: ANIES-IMIN



1. Most Retweeted Paslon 1



2. Most Retweeted Paslon 1



3. Most Retweeted Paslon 1



# MOST RETWEETED: PRABOWO-GIBRAN



← pos



Satya Azyumar

@satyaazyumar

...

ANDA SEMUA MENERTAWAKAN PRABOWO YANG DIMASAK OLEH ANIES DAN GANJAR.. NAMUN ANDA LUPA BAHWA MASIH ADA CAWAPRES.. SEORANG GIBRAN.. MENGHADAPI MAHFUD DAN IMIN.. DAMN..  
[Terjemahkan postingan](#)

22:38 · 12 Desember 2023 · **1,5 juta** Dimainkan

470

5 ribu

32 ribu

1000

1. Most Retweeted Paslon 2

← pos



Ini

@\_Pertama

...

Prabowo bilang "kalo tidak suka prabowo gibran jangan pilih kami, kami tidak takut tidak punya jabatan" Tapi ikut pilpres mulu ga kelar kelar.  
[Terjemahkan postingan](#)

21:06 · 12 Desember 2023 · **1,2 juta** Dimainkan

490

4000

27 ribu

635

2. Most Retweeted Paslon 2

← pos



222 Gunung Untuk Palestina

@kafiradikalis

...

...Ini mengerikan. Terlalu mengerikan. Vulgar banget. Video ini adalah resolusi paling sempurna yang menggambarkan betapa buruknya sosok [@Jokowi](#)  
  
Terkuak sudah kenapa ngotot mencalonkan Gibran sebagai cawapres. Jelas yah alasan kenapa haram memilih Prabowo di pilpres 2024... (``)  
[Terjemahkan postingan](#)



Raja Purwa

@BosPurwa · 1 Desember 2023

DHUAARR! Cerita Eks Ketua KPK Agus Rahardjo Dimarahi Jokowi terkait Kasus E-KTP: Beliau Teriak 'Hentikan' ... Selengkapnya di ...



10:02 · 1 Desember 2023 · **466,8 ribu** Dimainkan

216

2000

5 ribu

267

3. Most Retweeted Paslon 2

11

# MOST RETWEETED: GANJAR-MAHFUD ●

← pos



Satya Azyumar

@satyaazyumar

...

ANDA SEMUA MENERTAWAKAN PRABOWO YANG DIMASAK OLEH ANIES DAN GANJAR.. NAMUN ANDA LUPA BAHWA MASIH ADA CAWAPRES.. SEORANG GIBRAN.. MENGHADAPI MAHFUD DAN IMIN.. DAMN..

[Terjemahkan postingan](#)

22:38 · 12 Desember 2023 · 1,5 juta Dimainkan

470

5 ribu

32 ribu

1000

1. Most Retweeted Paslon 3

← Post



Extra Time Indonesia

@idextratime

...

Klasemen

1. Anies Muhaimin

2. Prabowo Gibran

3. Ganjar Mahfud

4. Manchester United

5:15 a. m. · 13 dic. 2023 · 213,3 mil Reproducciones

156

1 mil

4 mil

36

2. Most Retweeted Paslon 3



NitNot

@Leonita\_Lestari · 12 dic. 2023

...

KEMARIN BILANG TUNGGU HASIL LITBANG KOMPAS |katanya survei dari lembaga ini lebih bisa dipercaya dibanding yang lain.

Jadi gimana pendapat anda saat sadar bahwa keterpilihan Ganjar lebih rendah daripada keterpilihan Mahfud?

Apakah ada terbersit ide bahwa Mahfud sebaiknya memang...

[Mostrar más](#)

Survei Kepemimpinan Nasional  
Elektabilitas Calon Presiden  
Edisi Desember 2023

17,4%

39,7%

18,0%

Anies Baswedan

Prabowo Subianto

Ganjar Pranowo



Belum menentukan pilihan: 24,9%

12,7%

37,3%

21,6%

Abdul Muhaimin Iskandar

Gibran Rakabuming Raka

M. Mahfud MD



Belum menentukan pilihan: 28,4%

394

811

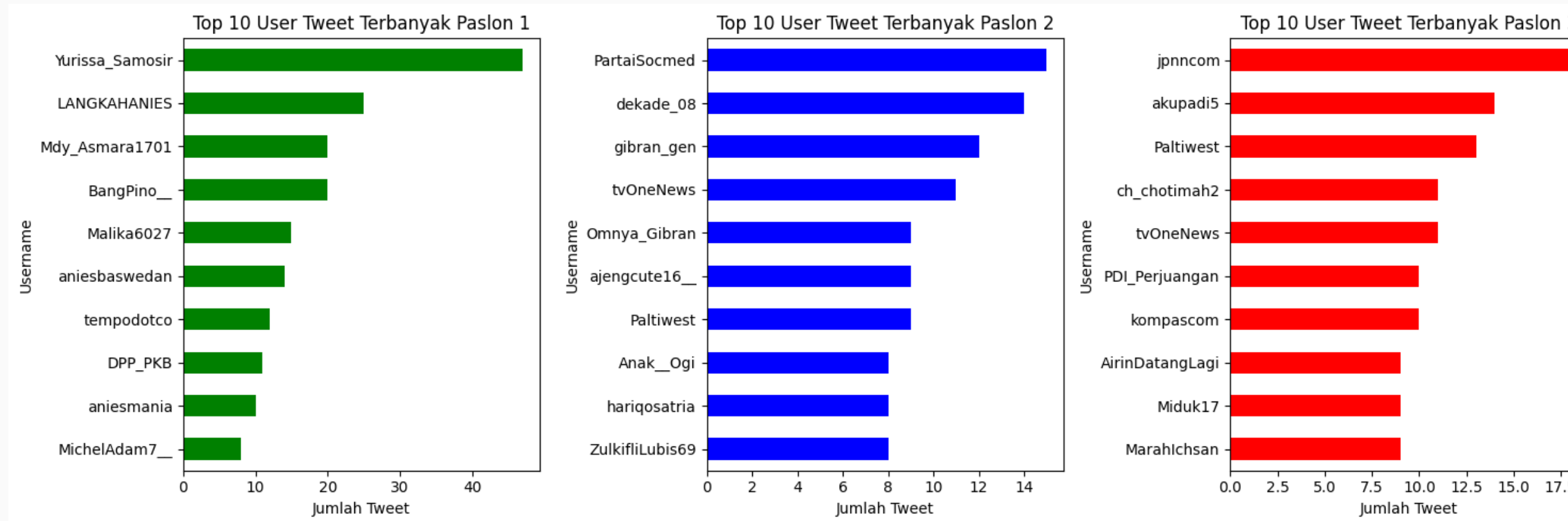
1 mil

398 mil

3. Most Retweeted Paslon 3

12

# Analisis Terhadap Pengguna Twitter (X)

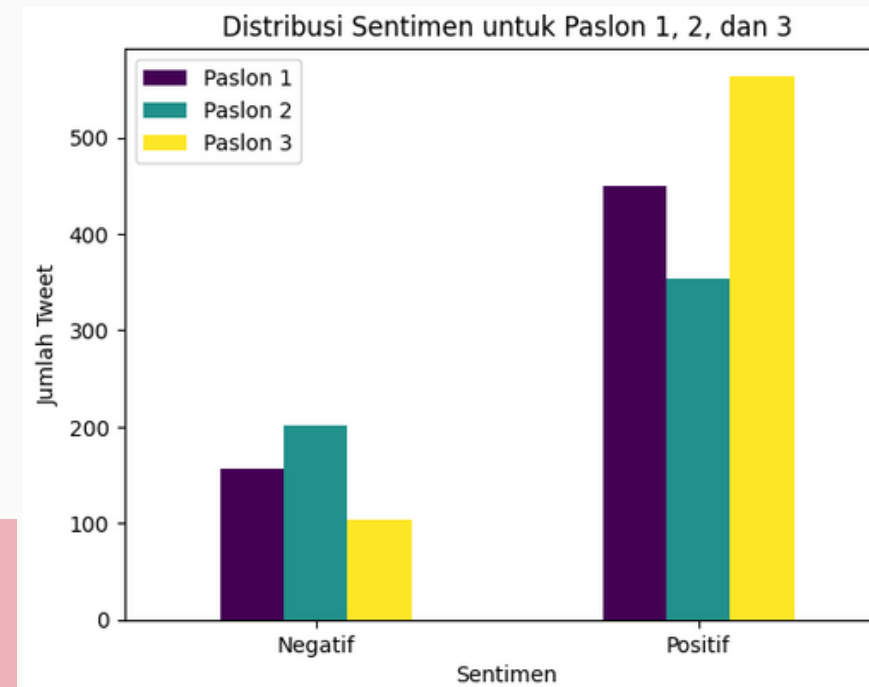


Pada analisis ini menghasilkan visualisasi diatas, Yaitu Top 10 User *Tweet* Terbanyak dari tiap paslon dengan menampilkan 10 *user* yang berbeda dari tiap paslon, seperti ‘Yurissa\_Samosir’ memposting *tweet* pada paslon 1 sebanyak 40 lebih, lebih banyak dari ‘PartaiSocmed’ yang hanya memposting *tweet* pada paslon 2 sebanyak 14 dan lebih sedikit daripada ‘jpnncom’ yang memposting *tweet* pada paslon 3 sebanyak 17.

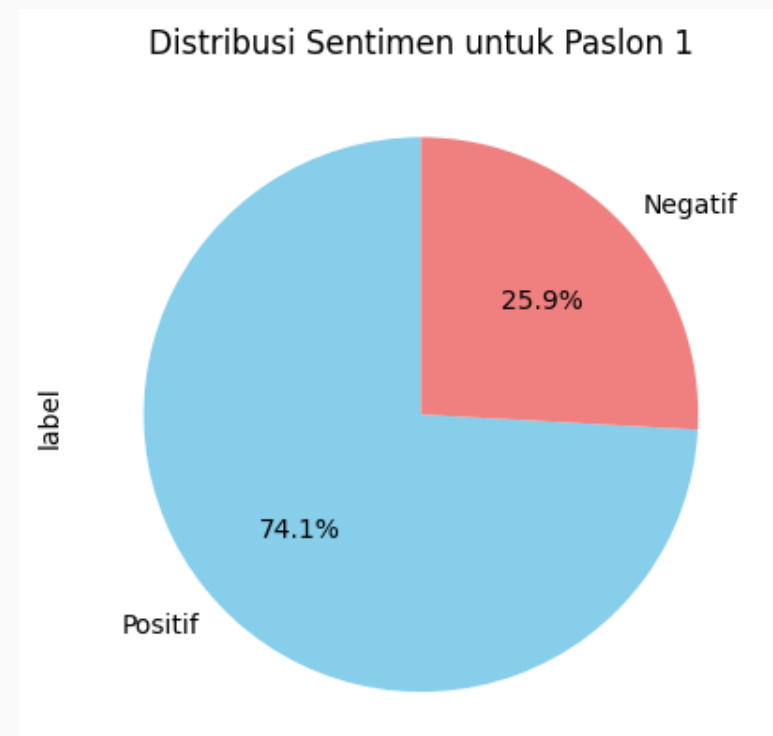
# SENTIMENT ANALYSIS

Akurasi	Skor Akurasi
Precision	0.84
Recall	0.72
F1 Score	0.77

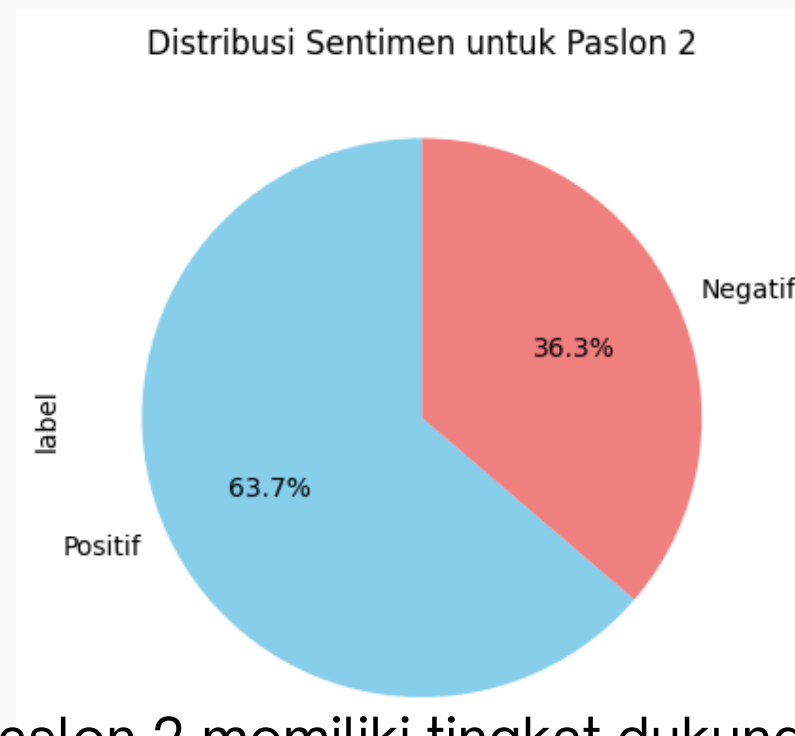
Dari model sentiment analysis yang dibuat dengan metode LSTM didapatkan akurasi sebesar 0.84



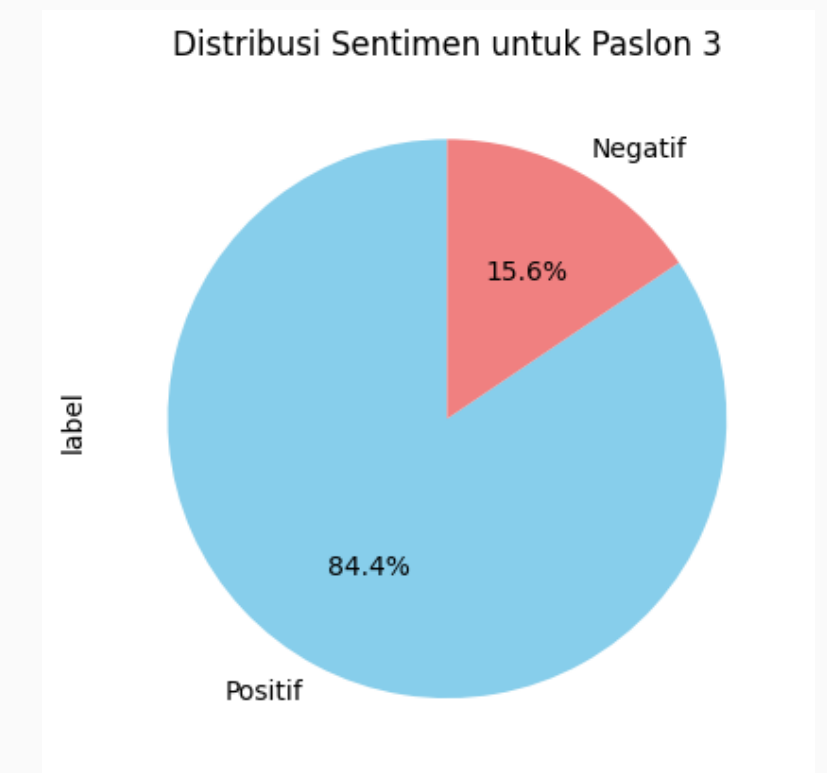
perbedaan signifikan terlihat pada distribusi sentimen antara ketiga paslon. Paslon 3 mencapai nilai positif tertinggi, sementara Paslon 2 memiliki nilai negatif yang paling tinggi. Perbedaan ini dapat disebabkan oleh faktor-faktor seperti konten dari *buzzer* atau murni dari pendapat masyarakat.



Paslon 1 memiliki dukungan positif yang tinggi (74.1%), dengan tingkat negatif yang lebih rendah (25.9%).



Paslon 2 memiliki tingkat dukungan positif yang lebih rendah (63.7%) dibandingkan dengan Paslon 1, dan tingkat negatif yang lebih tinggi (36.3%).



Paslon 3 mendapatkan tingkat dukungan positif tertinggi (84.4%), dengan tingkat negatif yang paling rendah (15.6%).

# TOPIC MODELLING

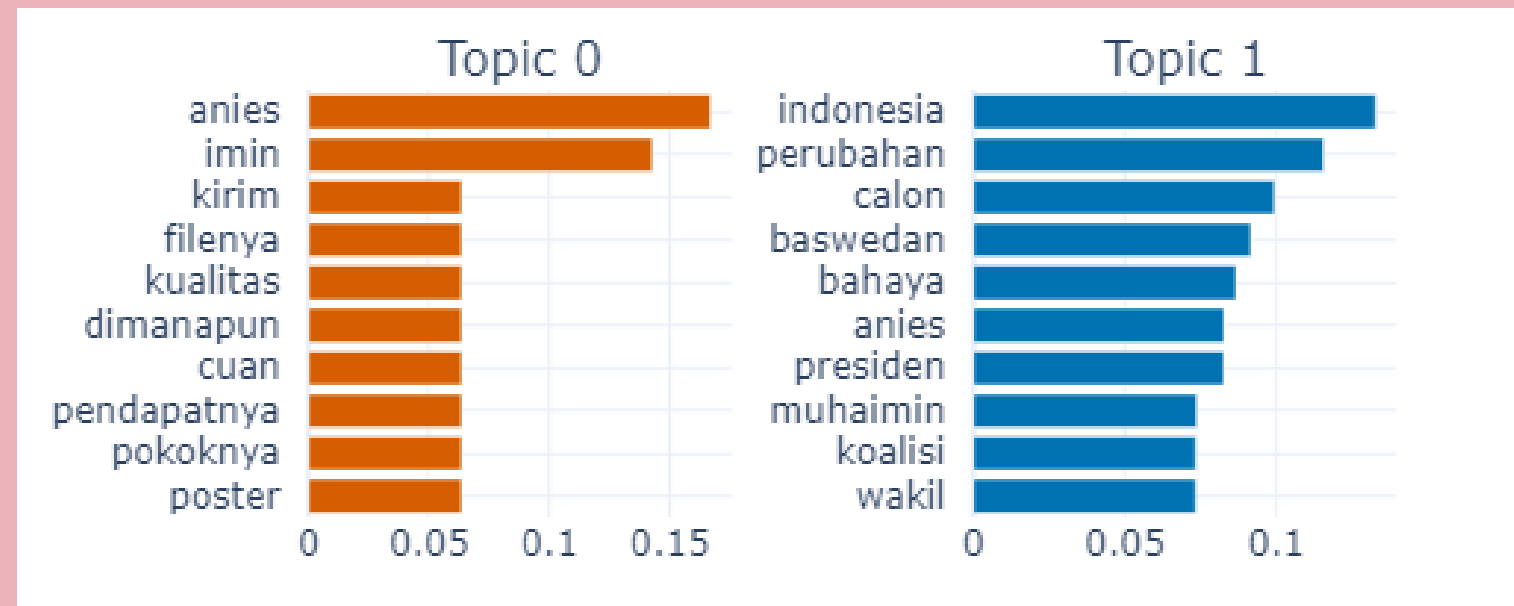
Paslon	Sentimen	Skor Perplexity
1	Negatif	1.067
	Positif	1.178
2	Negatif	1.223
	Positif	1.204
3	Negatif	Topik tidak terbentuk
	Positif	1.400

Skor Model



# HASIL TOPIC MODELLING TIAP PASLON

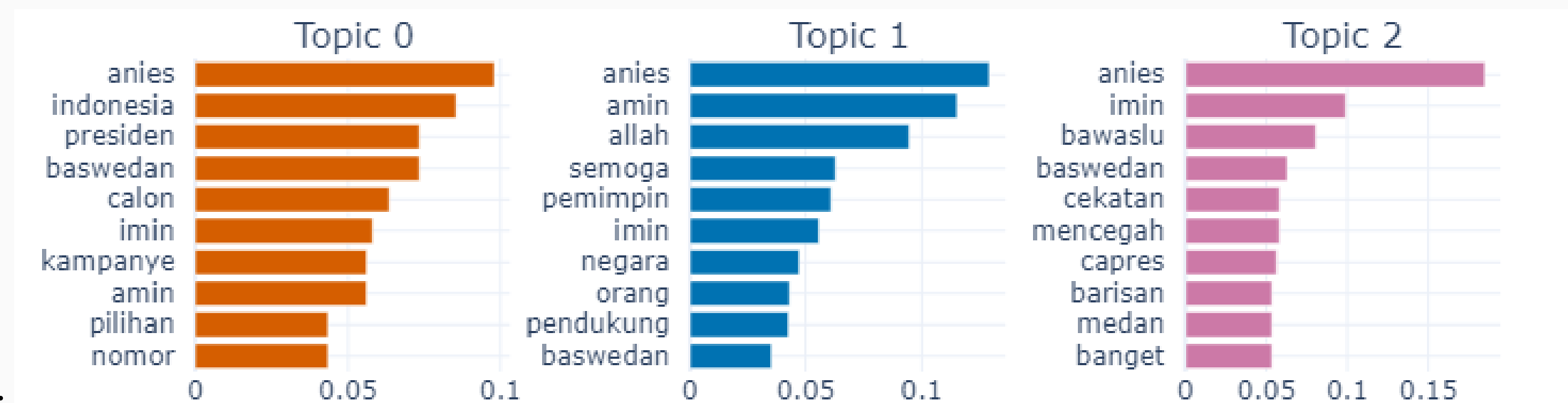
## Paslon 1



### Topik Negatif

Terdapat keterkaitan paslon 1 dengan sebuah file, poster, dan cuan atau uang dalam konteks negatif. Topik yang lain adalah terdapat keterkaitan paslon 1 dengan perubahan Indonesia namun perubahan yang membahayakan bangsa. Dua kata dengan skor topik tertinggi pada topik 1 adalah 'Anies' dan 'Imin' yang merupakan anggota dari paslon 1, sedangkan pada topik 2 adalah 'Indonesia' dan 'perubahan'.

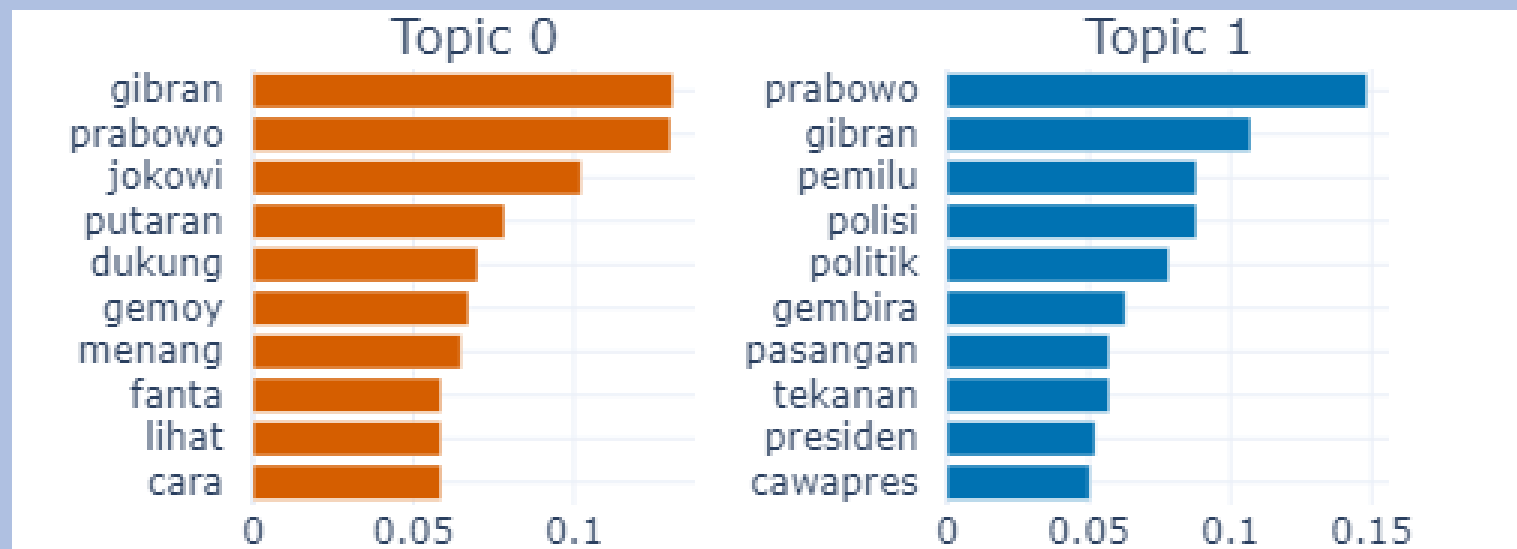
Topik pertama menggambarkan paslon 1 sebagai calon presiden pilihan, dengan skor untuk kata 'indonesia' dan 'presiden' tertinggi kedua dan ketiga. Topik kedua membahas doa untuk paslon 1 menjadi pemimpin negara dengan dua kata tertinggi setelah nama paslon adalah 'semoga' dan 'pemimpin'. Topik ketiga membahas paslon 1 sebagai calon yang cekatan dan berkaitan dengan mencegah. Skor tertinggi setelah nama paslon pada topik ini adalah 'bawaslu', 'cekatan', dan 'mencegah'.



### Topik Positif

# HASIL TOPIC MODELLING TIAP PASLON

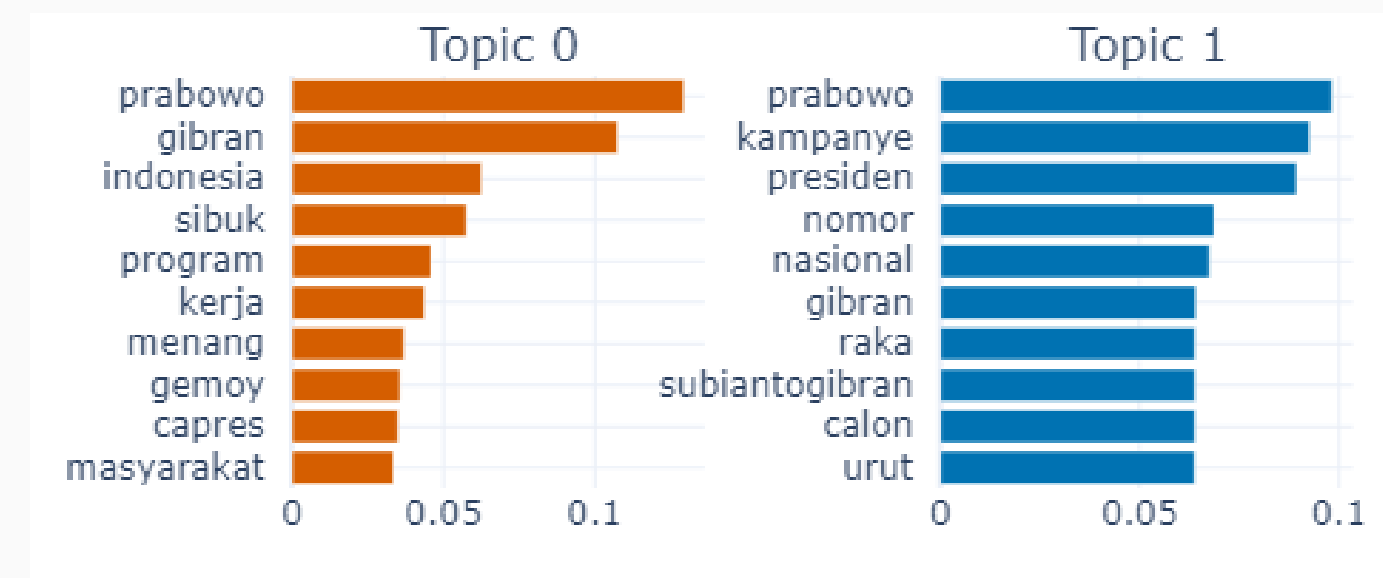
## Paslon 2



### Topik Negatif

Topik pertama memuat kata ‘sibuk’, ‘program’, dan ‘kerja’ sebagai bagian dari topik. Adapun kata ‘gemoy’ yang sebenarnya telah menjadi topik pada sentimen negatif, yang menandakan branding tersebut bermakna positif maupun negatif dari sudut pandang masyarakat. Topik kedua tidak menggambarkan sebuah topik secara jelas, namun dapat diinterpretasikan bahwa topik sentimen positif masyarakat pada paslon 2 membahas kampanye dan nomor urut paslon 2.

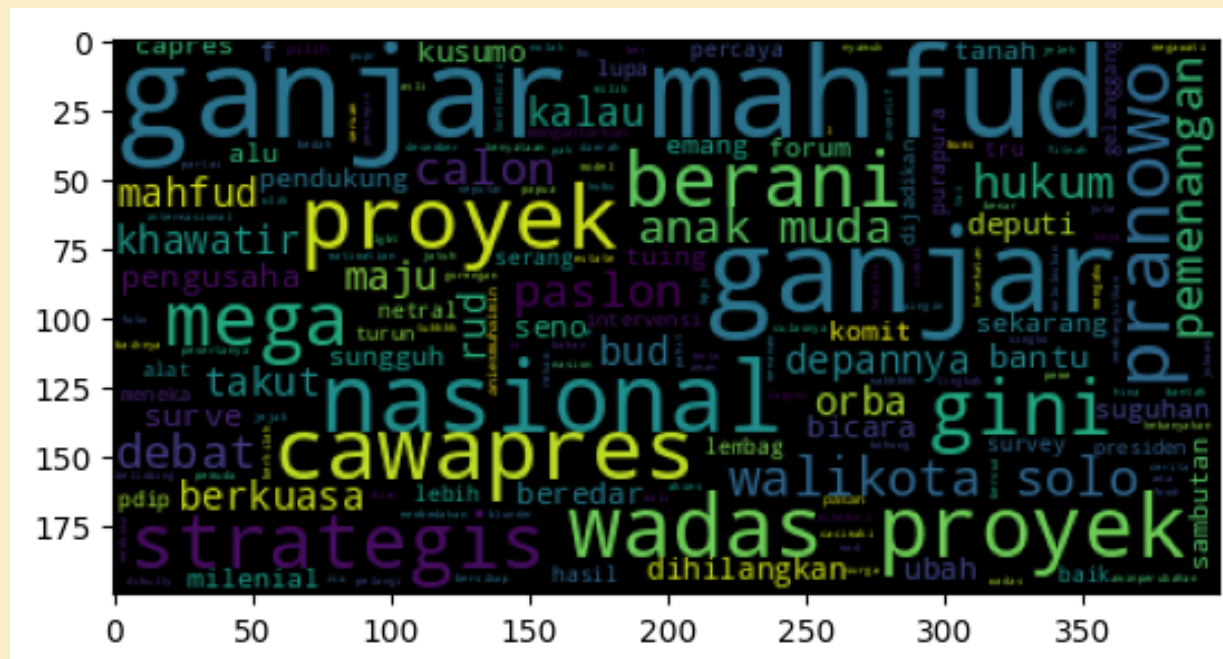
Topik pertama terdapat keterkaitan paslon 2 kepada ‘jokowi’, presiden saat ini yang merupakan ayah dari Gibran dan ‘gemoy’ yang merupakan branding yang melekat pada Prabowo. Topik kedua memuat kedua nama paslon bersama dengan keterkaitan terhadap ‘tekanan’ dan ‘polisi’.



### Topik Positif

# HASIL TOPIC MODELLING TIAP PASLON

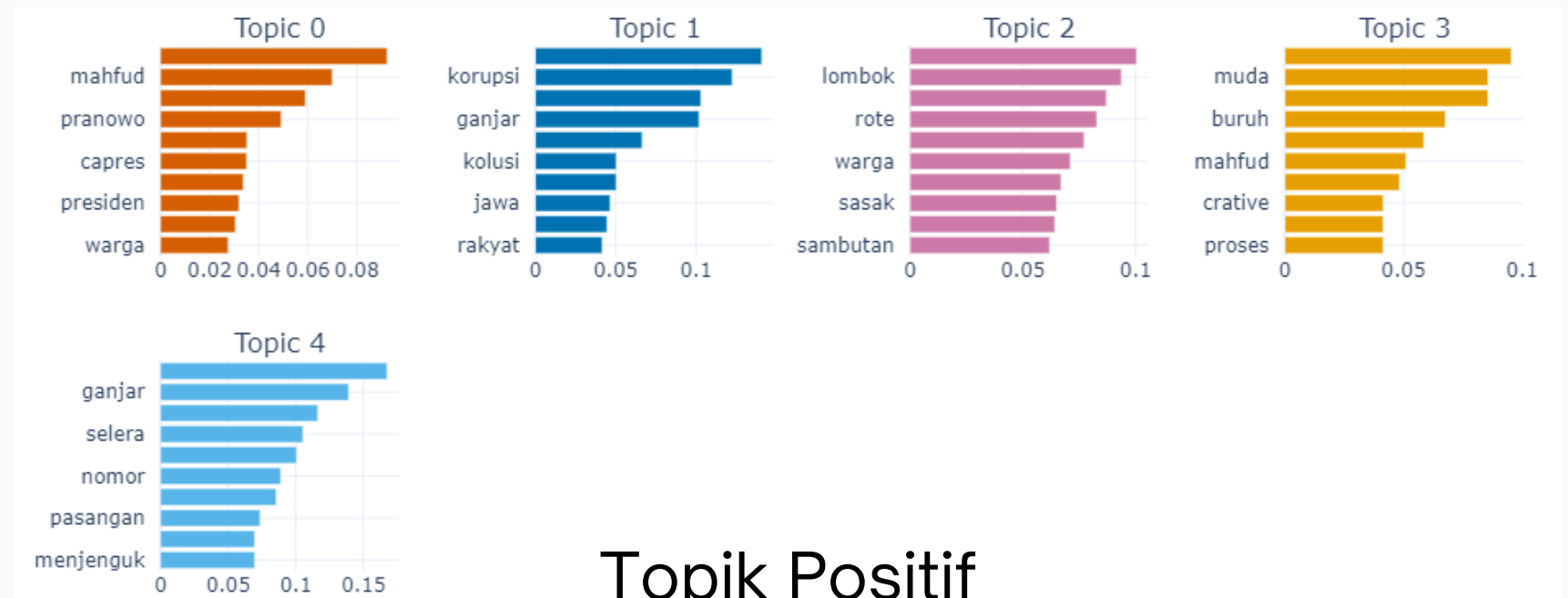
# Paslon 3



# Topik Negatif

Sehingga untuk melakukan evaluasi topik pada sentimen negatif paslon 3 digunakan *wordcloud* untuk melihat frekuensi kata yang paling sering muncul sebagai alternatif. Melalui *wordcloud* dari sentimen negatif paslon 3, diperoleh visualisasi terhadap kata yang sering muncul dari sentimen negatif pada paslon 3. Salah satu kata tersebut adalah ‘proyek’ dan ‘wadas’ yang berkaitan dengan proyek kerja Ganjar selama menjadi gubernur Jawa Tengah.

Diperoleh lima topik yang lebih banyak dari dua paslon sebelumnya. Topik pertama memuat nama paslon 3 dan 'capres', 'presiden', serta 'warga', mengindikasikan pendapat masyarakat terkait paslon 3 sebagai calon presiden warga Indonesia. Topik kedua memuat kata 'korupsi' dan 'kolusi' yang berkaitan dengan prestasi Ganjar selama menjadi gubernur Jawa Tengah. Topik ketiga berkaitan dengan 'sambutan' dari 'warga' 'rote' 'lombok' dan suku 'sasak'. Topik keempat menggambarkan buruh dan bidang kreativitas serta menyinggung 'muda'. Topik kelima menyebutkan nama 'ganjar' berkaitan dengan kata 'menjenguk' dan 'selera'.



# Topik Positif

# THANK YOU FOR LISTENING!

