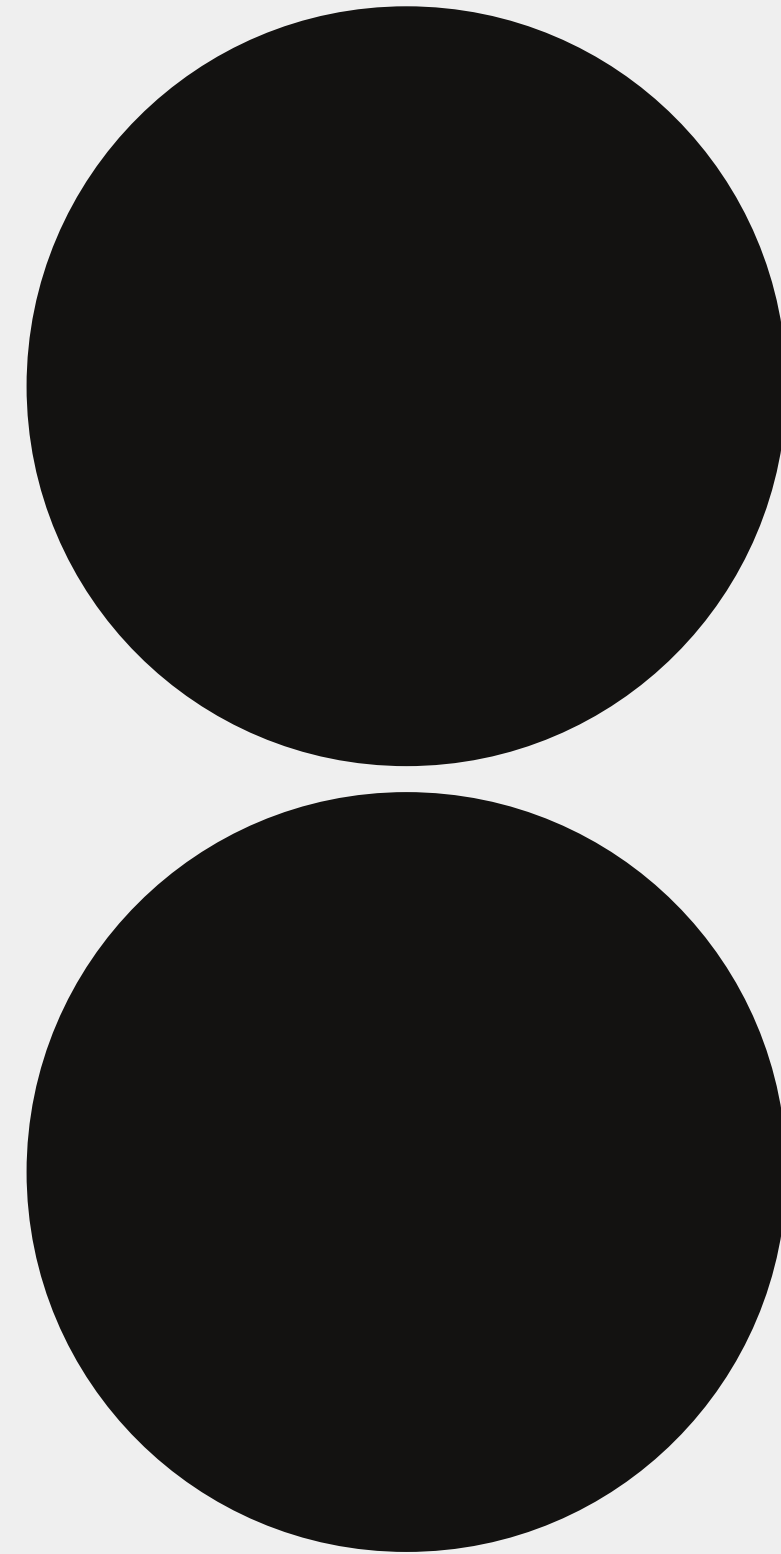
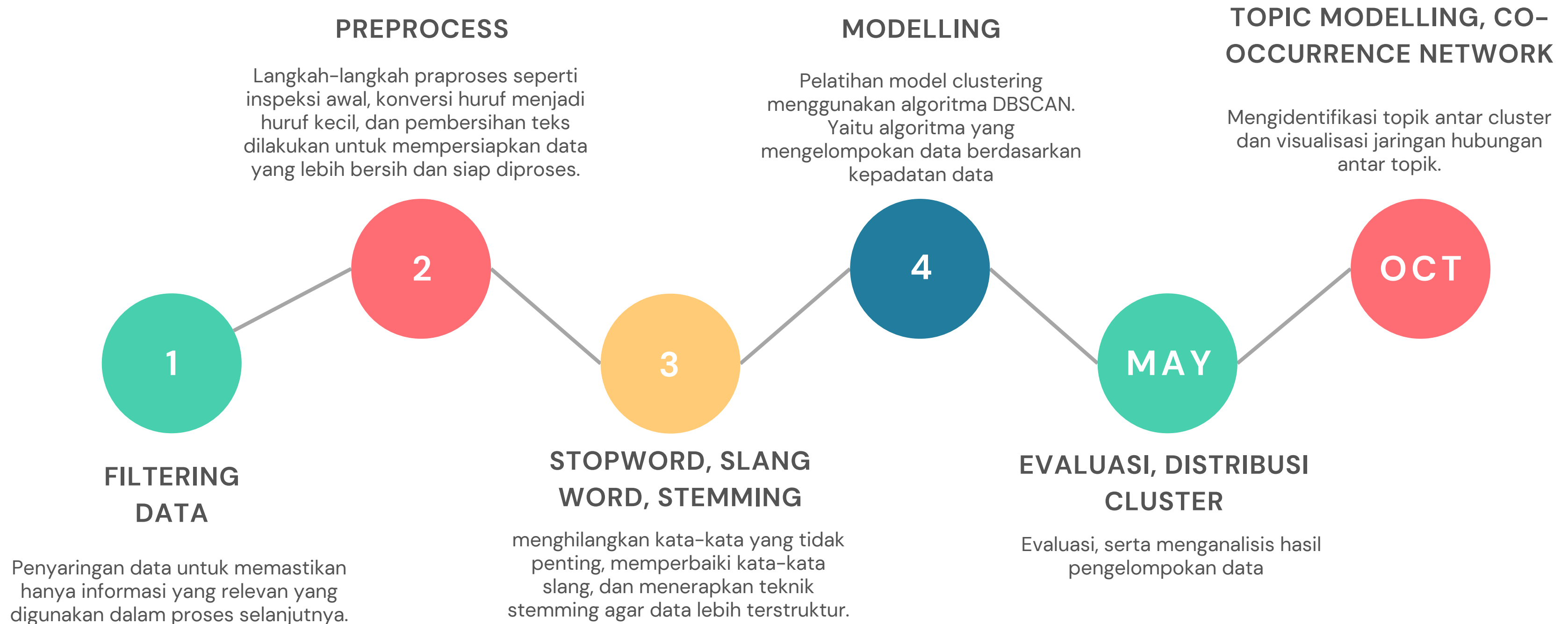


DIRTY VOTE CLUSTERING

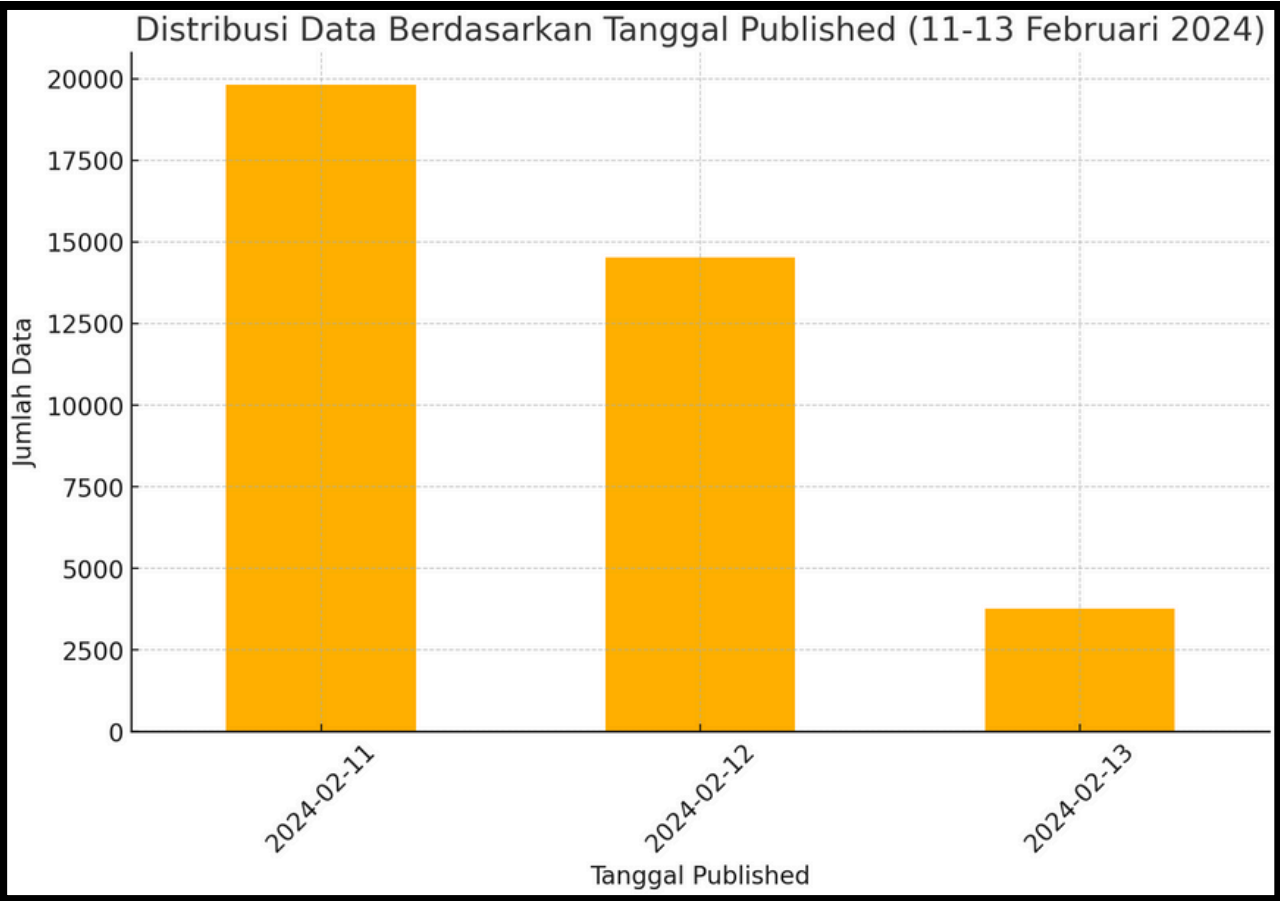


PROJECT TIMELINE



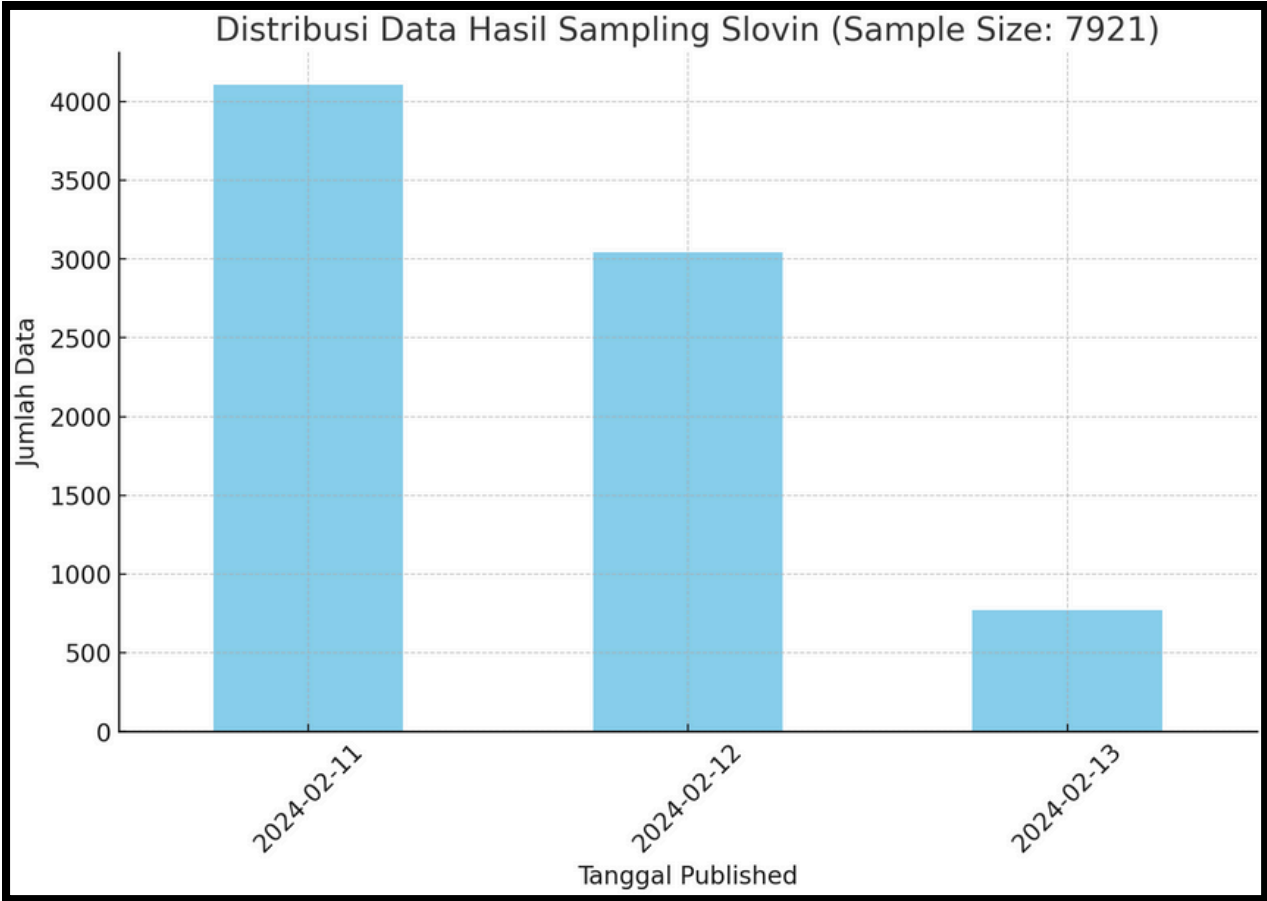
FILTERING DATA

38103 BARIS DATA



SLOVIN SAMPLING
-->

7921 BARIS DATA



PREPROCESS

FIRST INSPECTION

Tahap ini sekedar mengecek jumlah data kolom, tipe data kolom, dan hal-hal basic lainnya

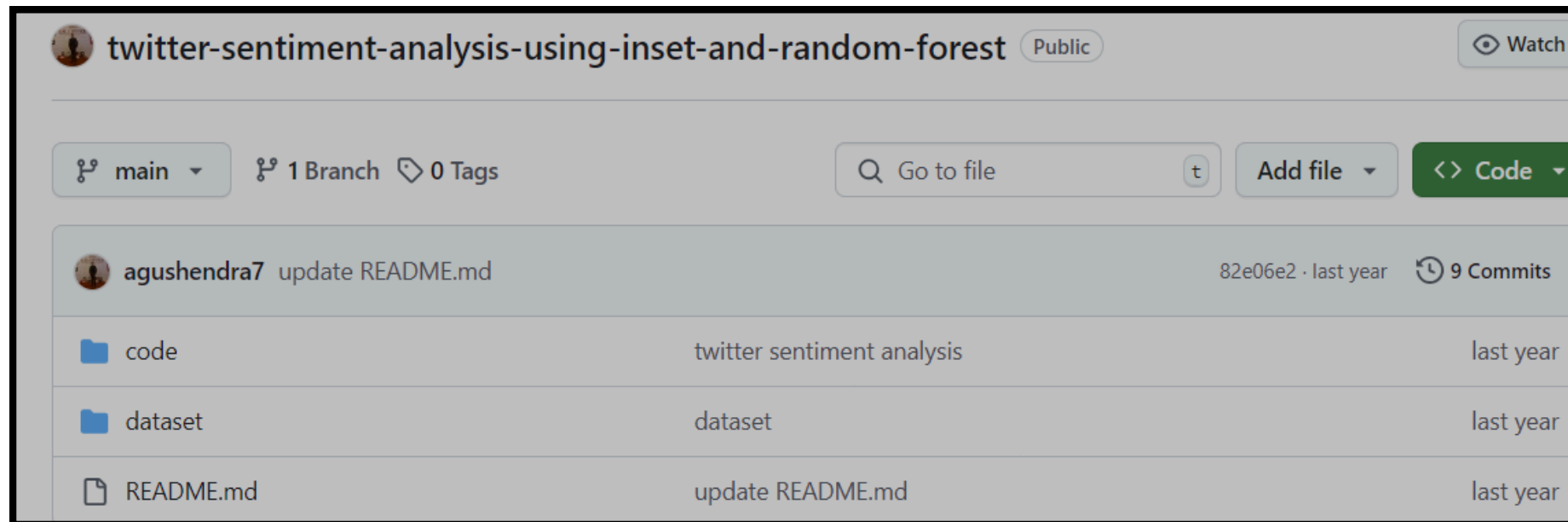
LOWERCASING

Tahap ini membuat semua teks menjadi bentuk non kapital, bertujuan untuk memudahkan pada saat di tahap text cleansing nantinya

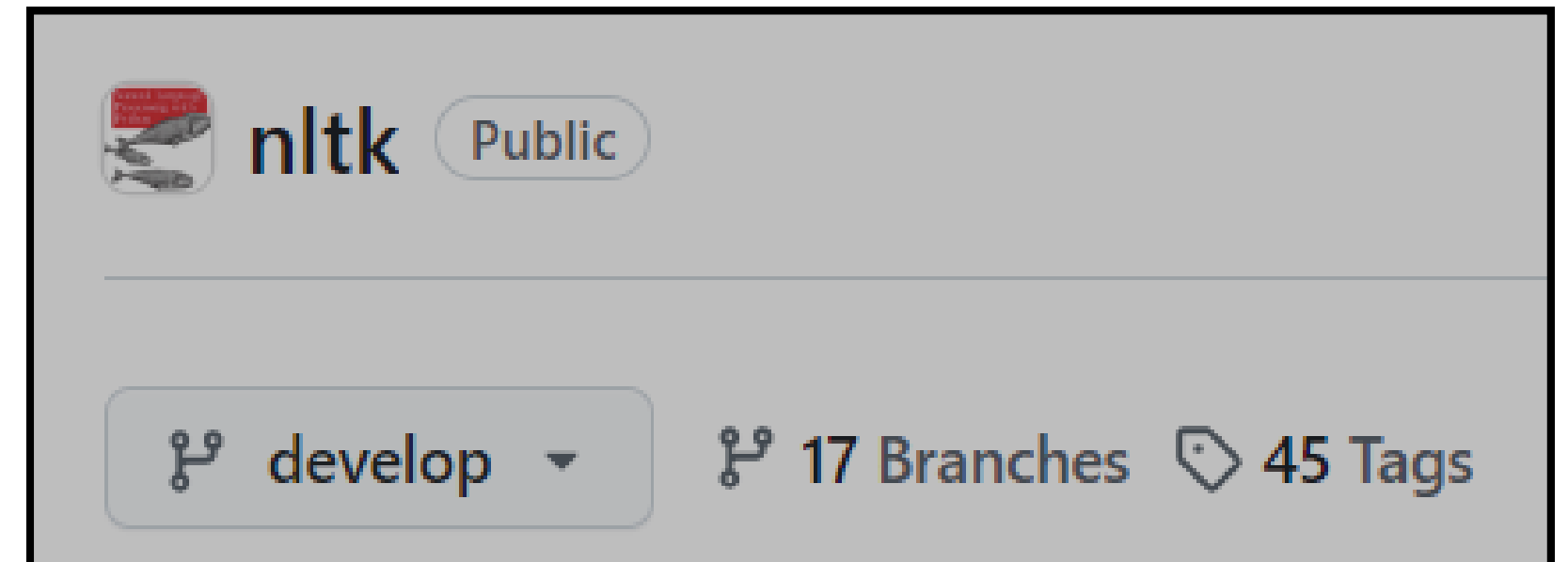
TEXT CLEANSING

Tahap menghilangkan noise dari data teks, seperti tanda baca, emoji, dan elemen lain yang tidak memiliki kontribusi dalam makna dari suatu teks.

PREPROCESS



SLANG WORD



STOPWORD

MODELLING WITH KMEANS ++

K-MEANS ++ MERUPAKAN
VARIASI LAIN DARI
K-MEANS

ASAL MULA

Some methods for classification and analysis
of multivariate observations

J. MacQueen

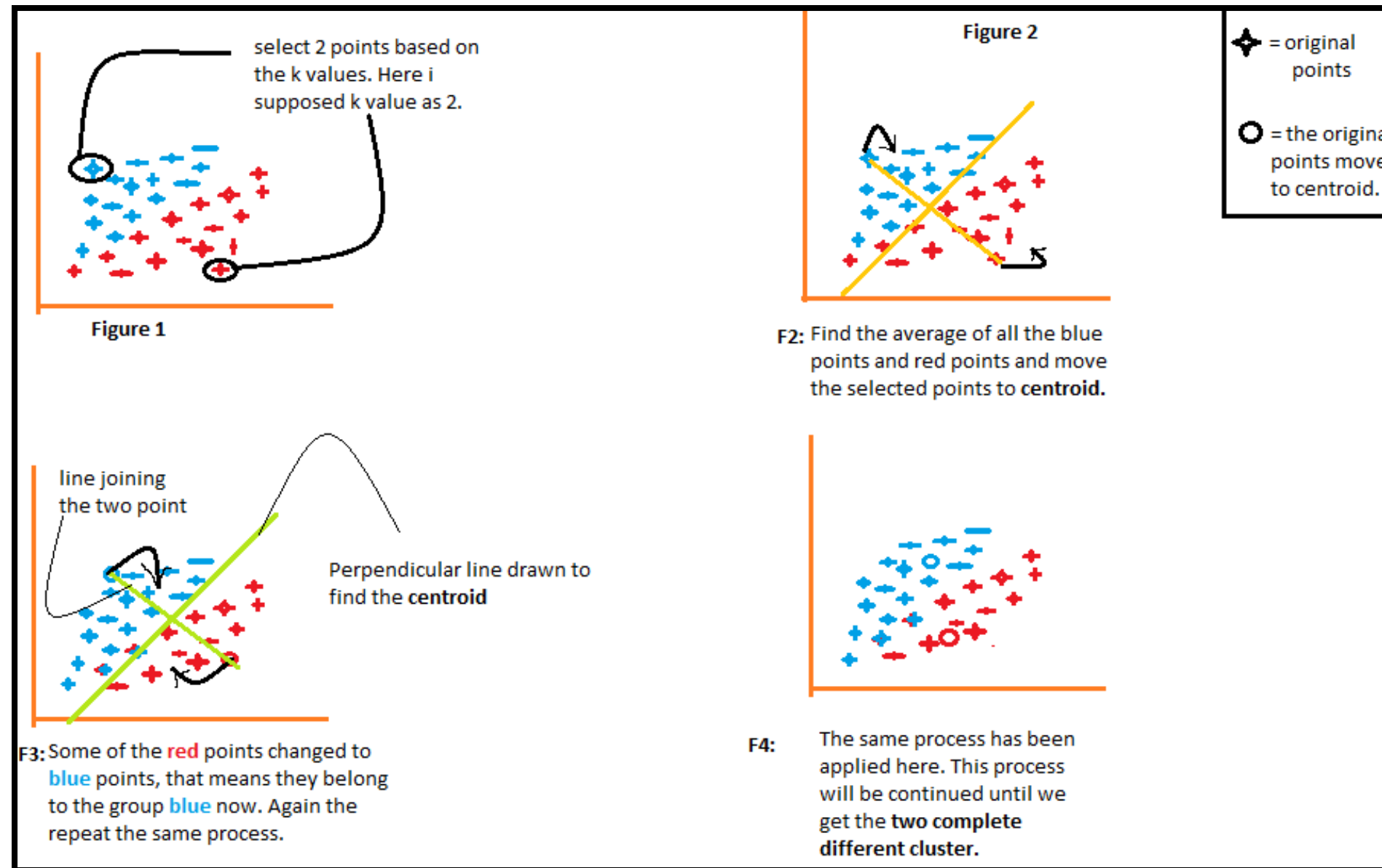
Editor(s) Lucien M. Le Cam, Jerzy Neyman

Berkeley Symp. on Math. Statist. and Prob., 1967: 281-297 (1967)

DEFINISI

Mengelompokkan data berdasarkan kedekatan ke centroid (titik pusat data), Hasil clustering bergantung pada jumlah cluster yang ingin dibentuk.

CARA KERJA K-MEANS



INISIASI NILAI K

PENETAPAN CENTROID
SECARA ACAK

BAGI DATA KE KELOMPOK
TERDEKAT

HITUNG ULANG CENTROID
HINGGA STABIL

EVALUASI MODEL

MENGGUNAKAN DAVIES
BOULDIN INDEX & SILHOUTTE
SCORE

SILHOUTTE SCORE

```
from sklearn.metrics import silhouette_score

# Menghitung Silhouette Score
silhouette_avg = silhouette_score(X, kmeans.labels_)
print(f'Silhouette Score: {silhouette_avg}')
```

✓ 2.1s

Silhouette Score: 0.97073363073355

DBI

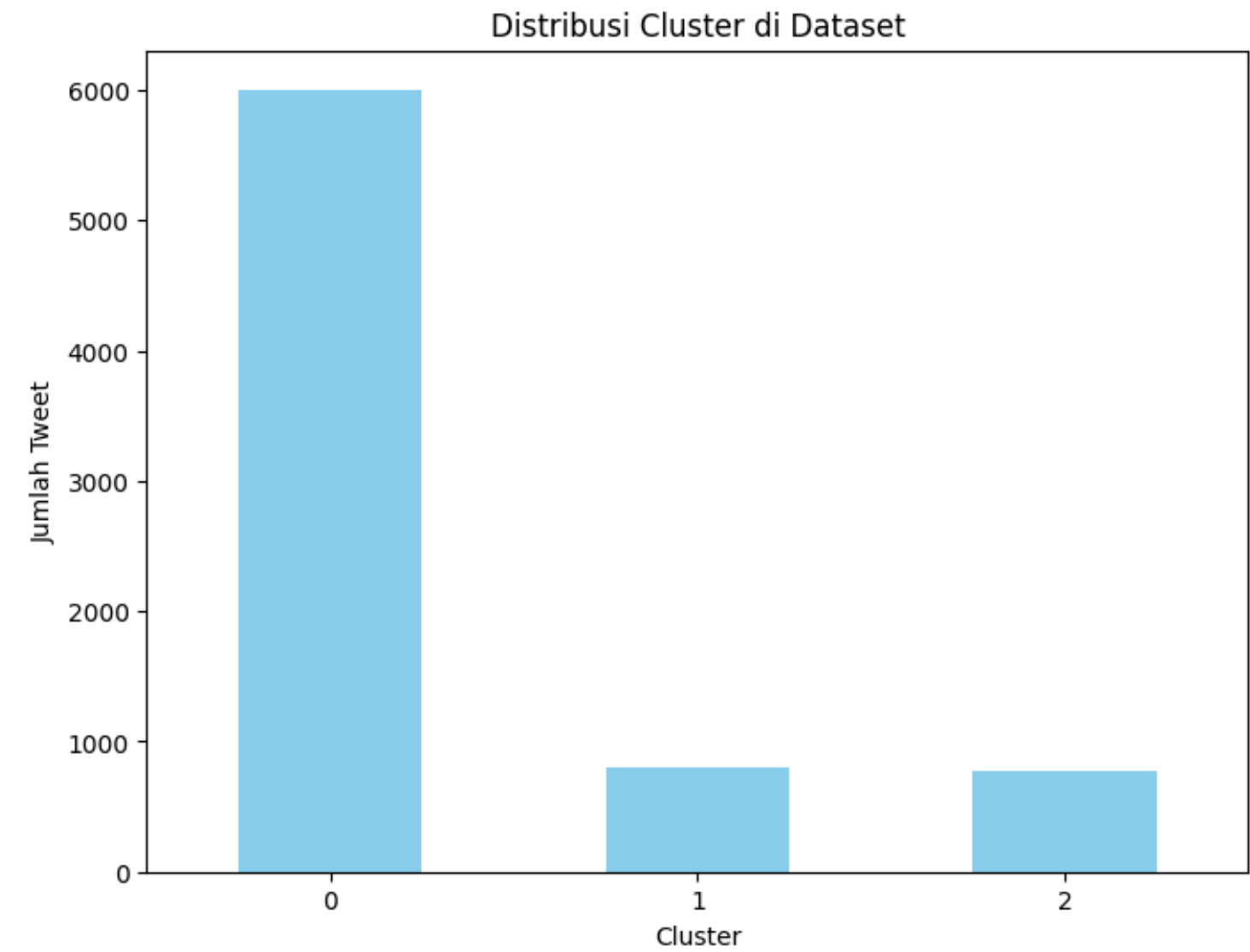
```
# Evaluasi menggunakan DBI
dbi = davies_bouldin_score(X.toarray(), kmeans.labels_)
print(f'Davies-Bouldin Index: {dbi}')
```

✓ 0.0s

Davies-Bouldin Index: 0.21744255726409556

DISTRIBUSI KLUSTER

BAR CHART



```
df['cluster'].value_counts()
```

✓ 0.0s

```
cluster
```

```
0    6003
```

```
1     802
```

```
2     778
```

```
Name: count, dtype: int64
```

TOPIC MODELLING

BY USING TOP WORD

```
# Group data by clusters
clustered_data = df.groupby('cluster')['stem'].apply(lambda x: ' '.join(x)).reset_index()

# Function to get the most common words in each cluster
def get_top_words(text, n=10):
    words = text.split()
    word_counts = Counter(words)
    return word_counts.most_common(n)

# Apply the function to each cluster
clustered_data['top_words'] = clustered_data['stem'].apply(get_top_words)

clustered_data[['cluster', 'top_words']]
```

0.1s

	cluster	top_words
0	0	[(pilih, 866), (prabowo, 753), (gibran, 628), (orang, 614), (indonesia, 602), (negara, 502), (allah, 463), (politik, 422), (curang, 422), (rakyat, 401)]
1	1	[(film, 1006), (pilih, 170), (prabowo, 166), (gibran, 166), (paslon, 152), (orang, 142), (tonton, 133), (politik, 125), (indonesia, 118), (dokumenter, 97)]
2	2	[(moga, 844), (allah, 333), (indonesia, 232), (orang, 217), (lindung, 181), (pilih, 171), (film, 155), (amin, 135), (pimpin, 126), (buka, 120)]

CLUSTER 0 [POLITIK]

Kluster ini menampilkan kata-kata seperti "pilih," "**Prabowo**," "**Gibran**," "orang," "Indonesia," dan "negara," yang menunjukkan bahwa diskusi ini berkaitan dengan proses pemilihan dan kandidat tertentu. Kehadiran kata "**curang**" bersama dengan "**politik**" dan "**rakyat**" memberikan indikasi kuat bahwa percakapan ini mungkin berkaitan dengan kekhawatiran atau tuduhan terkait kecurangan dalam pemilihan—sesuatu yang sangat terkait dengan konsep "dirty vote"

CLUSTER 1 [FILM]

Kluster 1 berpusat pada diskusi tentang film dokumenter Dirty Vote, dengan kata-kata seperti "**film**", "**dokumenter**", "**tonton**" menunjukkan bahwa komentar-komentar ini terkait dengan menyinggung film dokumenter dirty vote itu sendiri. Istilah "**Prabowo**," dan "**Gibran**" mengartikan bahwa sebagian besar paslon yang disinggung dalam kluster ini adalah prabowo gibran, dan kata "**pilih**" menggambarkan kondisi pemilihan umum (pemilu)

CLUSTER 2 [DOA HARAPAN]

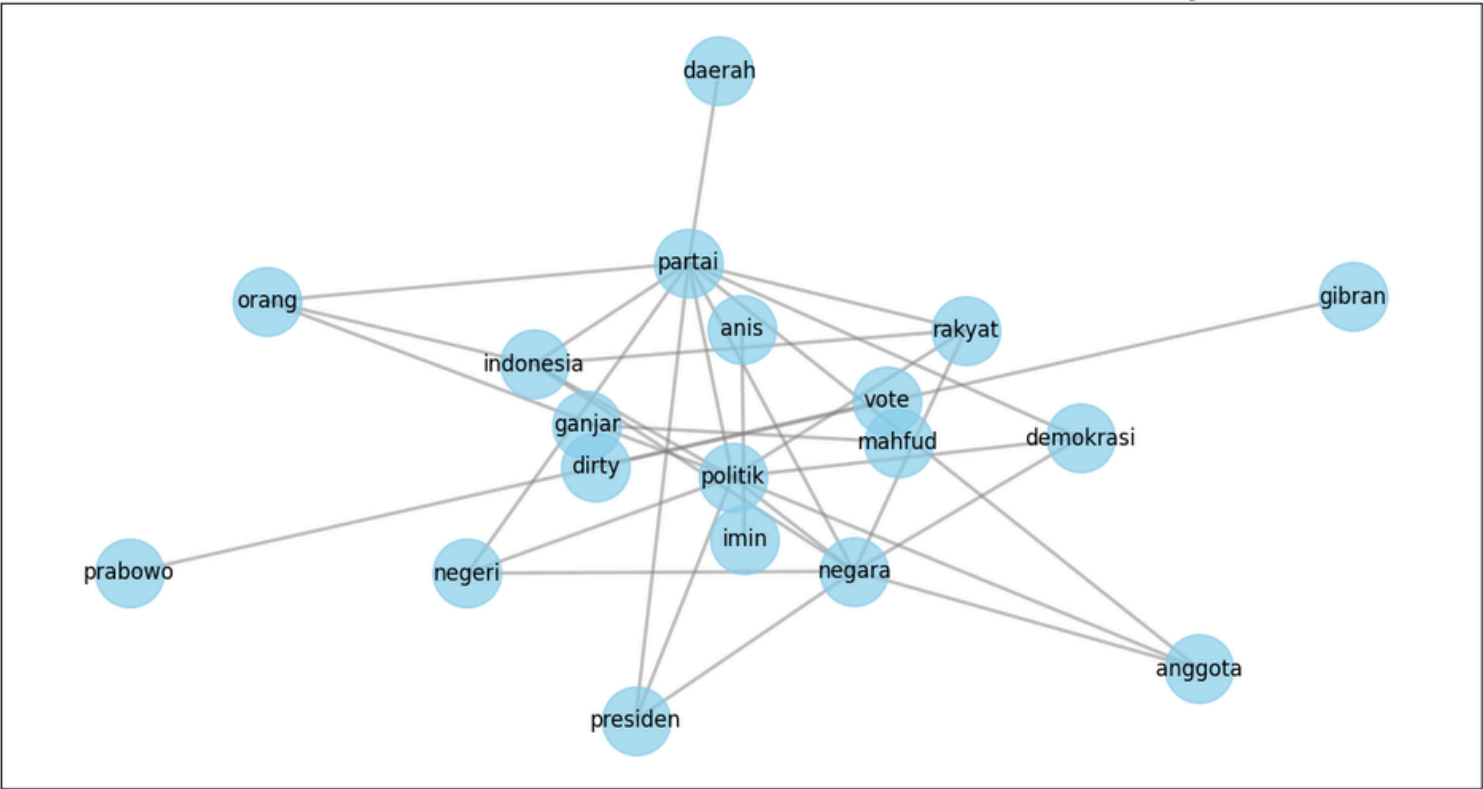
Kluster ini agak unik karena menggabungkan elemen religius ("**Allah**") dengan diskusi politik. Terlihat ada kata harapan seperti "**moga**", "**lindungi**". Lalu kata "**indonesia**", "**pimpin**" yang mengartikan bahwa harapan ini ditujukan ke indonesia dengan lebih spesifik ke pemimpin indonesia

ANALISIS JARINGAN NETWORK

BY KAMADA KAWAI LAYOUT

CLUSTER 0 [POLITIK]

Co-occurrence Network for Cluster 0 (Kamada-Kawai Layout)



SENTRALITAS POLITIK

Kata-kata seperti "partai," "politik," dan "vote" berada di pusat jaringan, menandakan bahwa diskusi dalam kluster ini sangat berfokus pada tema politik dan proses pemilihan.

SENTRALITAS POLITIK

Kehadiran kata "dirty" yang terkait dengan "vote," serta nama-nama seperti tokoh publik menunjukkan adanya diskusi tentang praktik pemilu yang tidak jujur (dirty vote) dengan tokoh-tokoh politik ini sebagai subjek utama.

KONTEKS DEMOKRASI DAN NEGARA

Keterkaitan antara "rakyat," "demokrasi," dan "negara" dengan tema utama menekankan kekhawatiran tentang dampak kecurangan pemilu terhadap demokrasi dan kepercayaan publik terhadap negara

Cluster 0 Network Analysis:			
	Degree	Centrality	Clustering Coefficient
partai		0.555556	0.355556
politik		0.500000	0.444444
negara		0.444444	0.500000
indonesia		0.277778	0.800000
rakyat		0.222222	1.000000

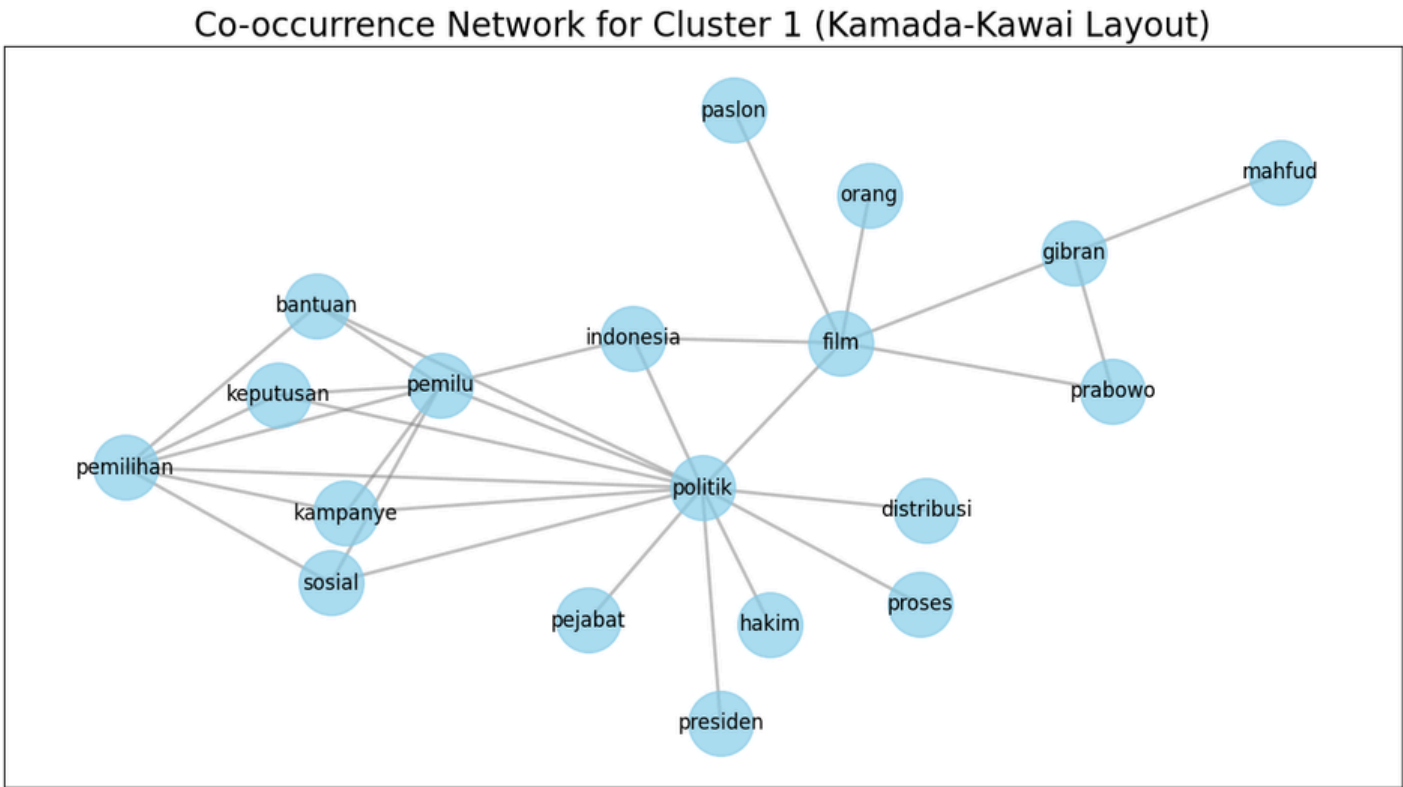
Degree Centrality mengukur seberapa terhubung sebuah node dengan node lain di dalam jaringan. Semakin tinggi degree centrality, semakin banyak koneksi langsung yang dimiliki node tersebut dengan node lainnya.

menunjukkan seberapa erat kata-kata yang terkait dengan sebuah kata saling terhubung, mengindikasikan kohesi atau kekompakan diskusi di sekitar topik tertentu.

ANALISIS JARINGAN NETWORK

BY KAMADA KAWAI LAYOUT

CLUSTER 1 [FILM]



POLITIK SEBAGAI PUSAT DISKUSI

Kata "politik" berada di pusat jaringan, menunjukkan bahwa pembahasan utama dalam kluster ini berpusat pada aspek politik, terutama terkait pemilihan dan kampanye.

PERAN MEDIA DAN FILM

Kata "film" terhubung dengan "politik," "orang," dan nama-nama seperti "Gibran" dan "Prabowo," yang mendukung bahwa media film dokumenter ditrty vote ini, berhasil dalam membentuk persepsi tentang para kandidat dan proses politik.

KETERLIBATAN PEJABAT DAN PROSES HUKUM

Keterhubungan antara "proses," "hakim," dan "presiden" menunjukkan adanya komentar yang meninggung tentang peran pejabat tinggi dan aspek hukum dalam politik. Hal ini menunjukan adanya hubungan pejabat dengan pemilu.

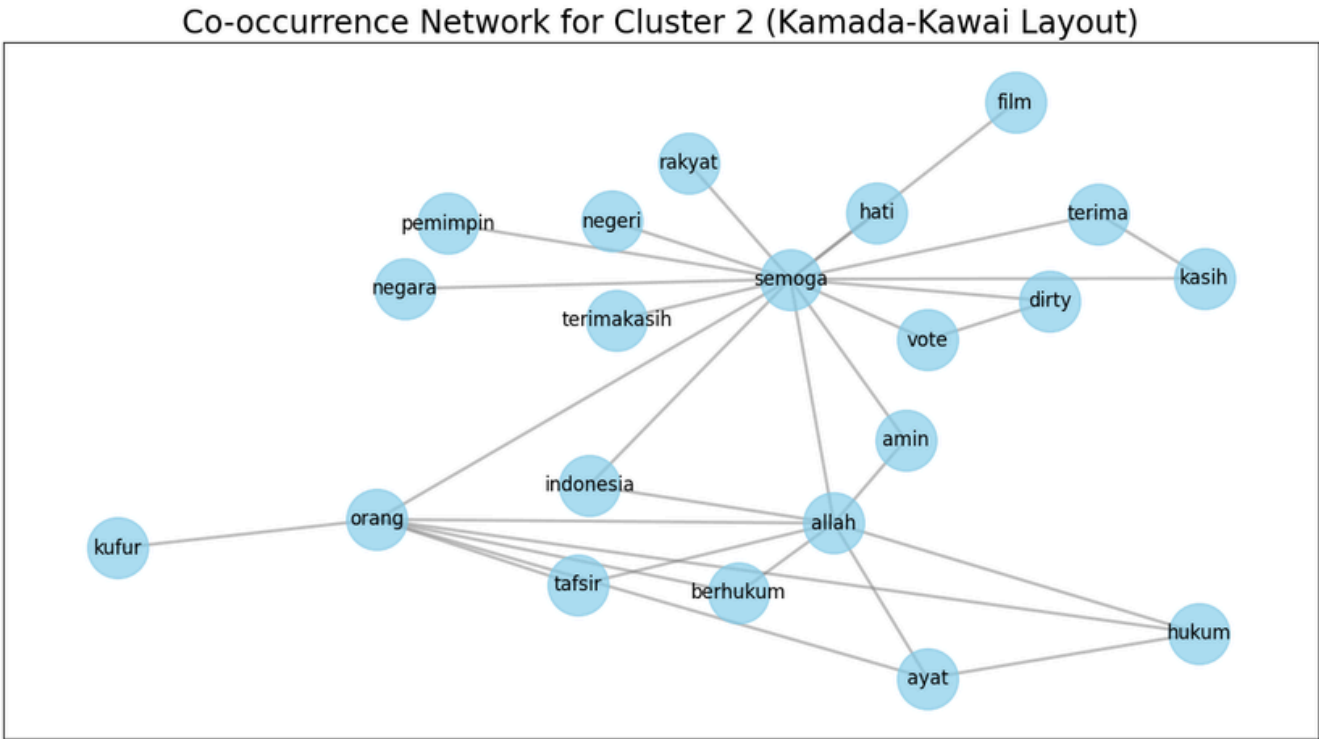
Cluster 1 Network Analysis:

	Degree	Centrality	Clustering Coefficient
politik		0.722222	0.141026
pemilu		0.388889	0.476190
film		0.333333	0.133333
pemilihan		0.333333	0.600000
gibran		0.166667	0.333333
kampanye		0.166667	1.000000

ANALISIS JARINGAN NETWORK

BY KAMADA KAWAI LAYOUT

CLUSTER 2 [HARAPAN]



HARAPAN DAN DOA

Kata "semoga" berada di pusat jaringan, menunjukkan bahwa komentar di kluster ini berkaitan harapan atau doa terkait situasi politik

DIMENSI DAN RELIGIUSM

Kata-kata seperti "Allah," "amin," dan "ayat" menandakan bahwa komentar ini menyinggung isu-isu politik dilihat melalui lensa keagamaan, dengan harapan agar keadilan ditegakkan dalam pemilu

TAFSIRAN HUKUM

Kata "berhukum" dan "tafsir" yang terkait dengan "hukum" dan "ayat" menunjukkan bahwa ada perbincangan tentang bagaimana hukum dan ajaran agama dipahami dan diterapkan dalam konteks pemilu dan pemerintahan

Cluster 2 Network Analysis:		
	Degree Centrality	Clustering Coefficient
semoga	0.75	0.047619
allah	0.40	0.285714
orang	0.35	0.285714
ayat	0.15	1.000000
hukum	0.15	1.000000

OPTIONAL

TOPIC MODELLING

USE LATENT DIRICHLET ALLOCATION

CLUSTER 0 [HARAPAN]

Cluster 0 Topik:

Topik 1: pilih, politik, orang, rakyat, demokrasi, suara, negara, terimakasih, milu, dirty
Topik 2: indonesia, allah, negara, jokowi, kuasa, curang, pimpin, demokrasi, orang, negeri
Topik 3: pilih, tonton, ngeri, curang, video, trending, kuasa, salah, mantap, banget
Topik 4: prabowo, gibran, pilih, anis, imin, video, paslon, kasih, terima, ganjar
Topik 5: orang, keren, anak, malu, buka, gila, tipu, pilih, daya, cerdas

PEMILU DAN
DEMOKRASI

KEKUASAAN DAN
KEPEMIMPINAN
JOKOWI

REAKSI
TERHADAP
KONTEN MEDIA

SENTIMEN
EMOSIONAL

KANDIDAT DAN
PEMILIHAN

CLUSTER 1 [FILM]

Cluster 1 Topik:

Topik 1: film, pilih, prabowo, gibran, tonton, rakyat, anis, imin, indonesia, milu
Topik 2: film, kuasa, prabowo, gibran, curang, cari, kotor, pilih, politik, capres
Topik 3: film, gibran, politik, prabowo, milu, pilih, curang, paslon, salah, dokumenter
Topik 4: film, dirty, vote, paslon, prabowo, gibran, orang, negara, ganjar, kuasa
Topik 5: film, orang, tonton, paslon, buat, indonesia, terima, kasih, dokumenter, pilih

FILM DAN
PENGARUHNYA
PADA PEMILU

Topik ini berfokus pada bagaimana film dokumenter, khususnya yang terkait dengan politik, mempengaruhi persepsi publik terhadap pemilu dan kandidat

NETRALITAS DAN
DAMPAK FILM
PADA KANDIDAT

Topik ini membahas apakah film tersebut netral atau bias, serta dampaknya pada persepsi publik terhadap para kandida

KETERLIBATAN
DAN KONTROVERSI
DALAM POLITIK

mengungkapkan kekhawatiran bahwa film tersebut mungkin digunakan untuk menggiring opini publik menjelang pemilu

KETERLIBATAN
DAN KONTROVERSI
DALAM POLITIK

Komentar-komentar menunjukkan bahwa film tersebut memaparkan "kotornya" proses pemilu, dan ada perdebatan tentang apakah film ini berpihak pada salah satu kandidat atau mengungkap semua kandidat secara merata.