

P1

Bayu Rahayudi, S.T., M.M. (I)
Nanang Yudi Setiawan, S.T., M.Kom. (II)



ANALISIS PENGELOMPOKAN ULASAN PENGGUNA MENGUNAKAN K-MEANS CLUSTERING UNTUK EVALUASI APLIKASI MY SAPK BKN

Oleh: M Ighfar Maulana
205150400111055

Pemetaan Masalah ISHIKAWA DIAGRAM

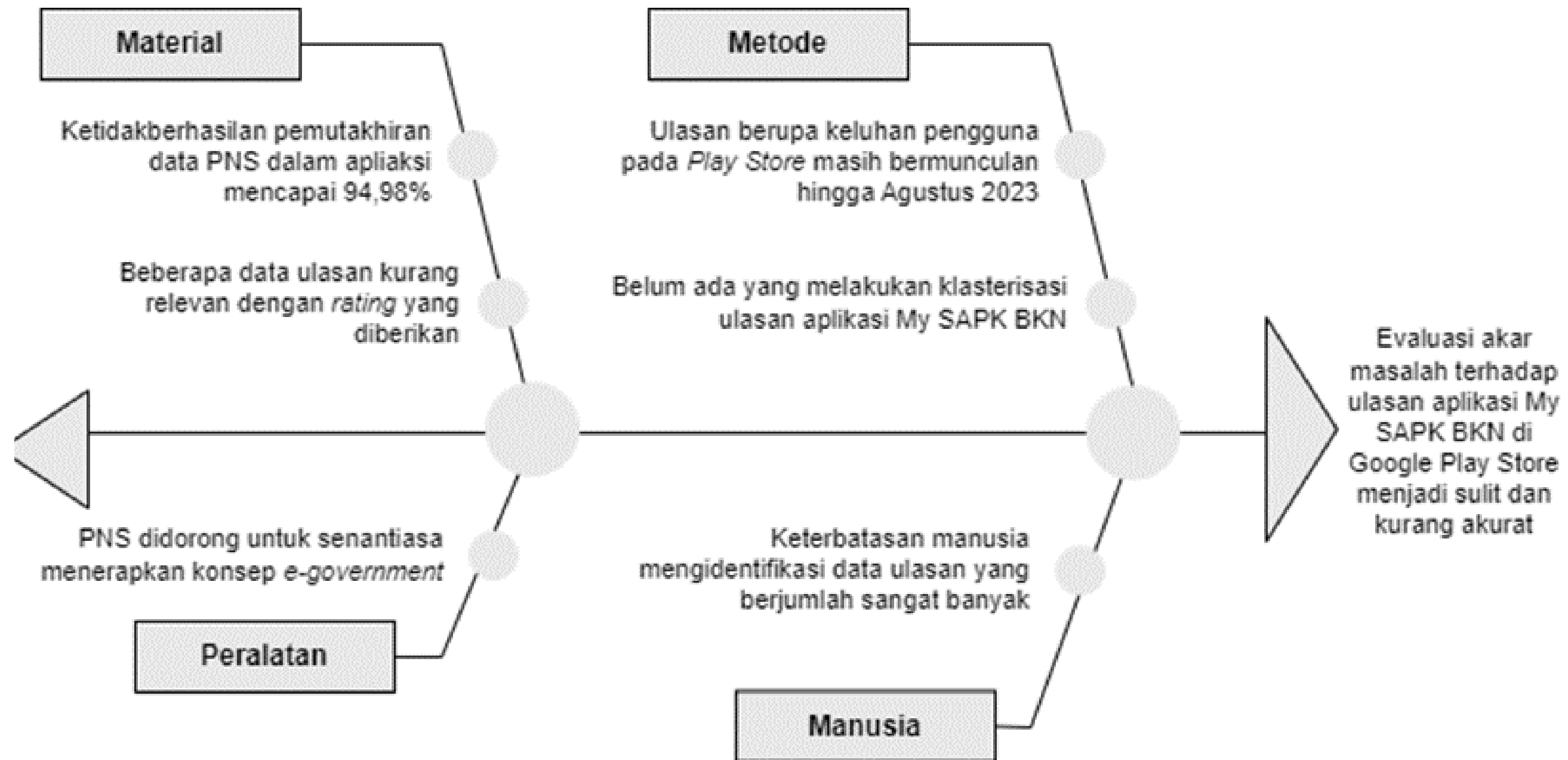
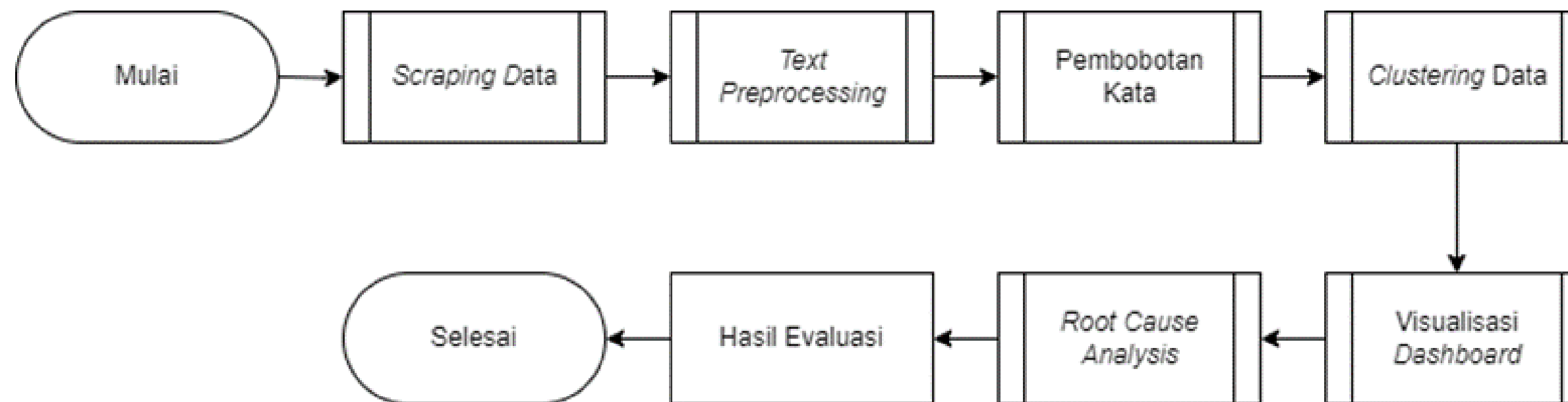


Diagram Alir Sistem



Scrapping

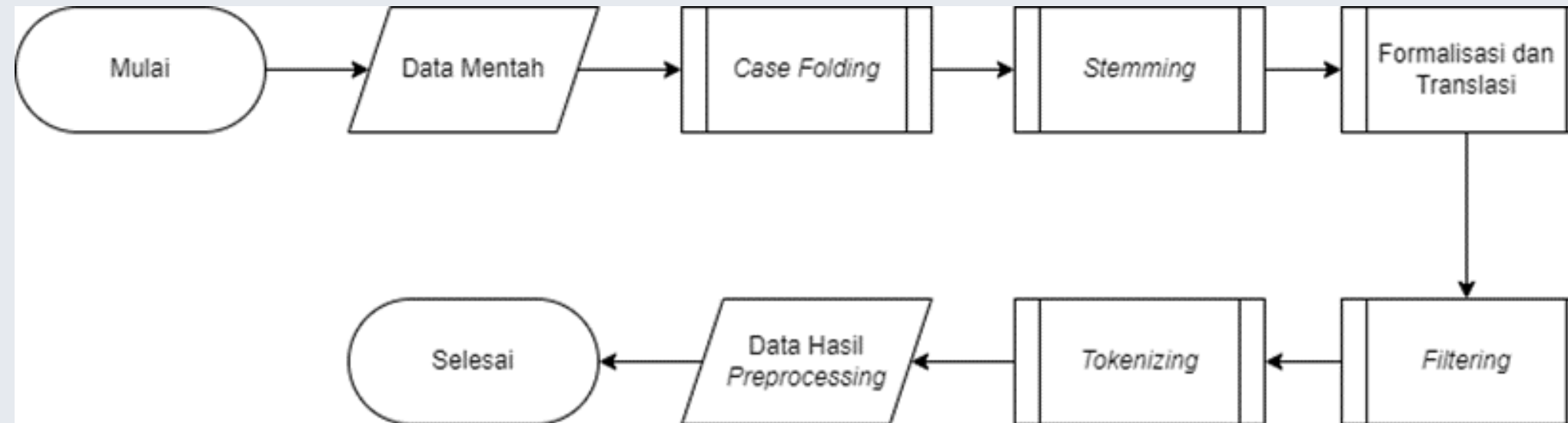


No.	Variabel	Tipe Data	Keterangan
1.	<i>userName</i>	String	Nama dari pengguna aplikasi My SAPK BKN
2.	<i>score</i>	String	Besar angka bintang yang diberikan oleh pengguna aplikasi My SAPK BKN
3.	<i>at</i>	String	Waktu dari pemberian ulasan oleh pengguna aplikasi My SAPK BKN
4.	<i>content</i>	String	Ulasan yang diberikan pengguna

Scraper google-play-scraper
Rentang 17 Juli 2021 - 31 Agustus 2023
3321 Data Ulasan (Raw)

<i>userName</i>	<i>score</i>	<i>at</i>	<i>content</i>
Ane Lubu	1	2023-08-18 16:52:30	Selalu gagal buat log in. Padahal username dan password nya benar. Dicoba berkali kali tetap saja ga bs. Banyak komplain masuk <u>mas</u> mbak nya. Diresppn doongg
Arkan Faisal Adillah	1	2023-06-14 10:03:27	Untuk karpeg virtual nya loading mulu ngak mau kebuka, tolong apk nya di tingkatkan lagi.
Bb Br	2	2023-04-08 20:15:05	Cek update kepangkatan, bulan lalu di cek pangkat udah naik, di cek lagi bulan ini kok balik ke pangkat sebelumnya Gimana nih apps nya?

Data Preprocessing



ID	Data Mentah	Data Hasil <i>Preprocessing</i>
D1	Untuk proses lupa pasword tidak terintegrasi dengan versi web. Selalu gagal ketika ubah pasword.	['proses', 'lupa', 'sandī', 'integrasi', 'versi', 'web', 'gagal', 'ubah', 'sandī']
D2	Sangat di sayangkan Server selalu eror untuk aplikasinya, dan Versi Web pun saya tak bisa akses, dan menjadi kendala untuk kinerja saya	['sayang', 'server', 'eror', 'aplikasi', 'versi', 'web', 'akses', 'kendala', 'kerja']
D3	Mantap,,, jadi pemantauan kepegawaian update	['pantau', 'pegawai', 'update']
D4	Apl. Bagus memudahkan ASN dalam melakukan pendataan 🗂️	['aplikasi', 'asn', 'laku', 'data']
D5	Kenapa kartu Asn virtual tidak bisa keluar	['kartu', 'asn', 'virtual']

Stemming-> Library Sastrawi

Formalisasi & Translasi->

- Rama Prakoso (kbba_updated.txt)
- Louise Owen (slang_words.txt)

Filtering->

- Rama Prakoso (combined_stop_words.txt)
- Natural Language Kit (Bahasa Indonesia)

Term Frequency - Inverse Document Frequency (TF-IDF)

Menghasilkan 1297 rows jenis kata, yang kemudian disebut dimensi

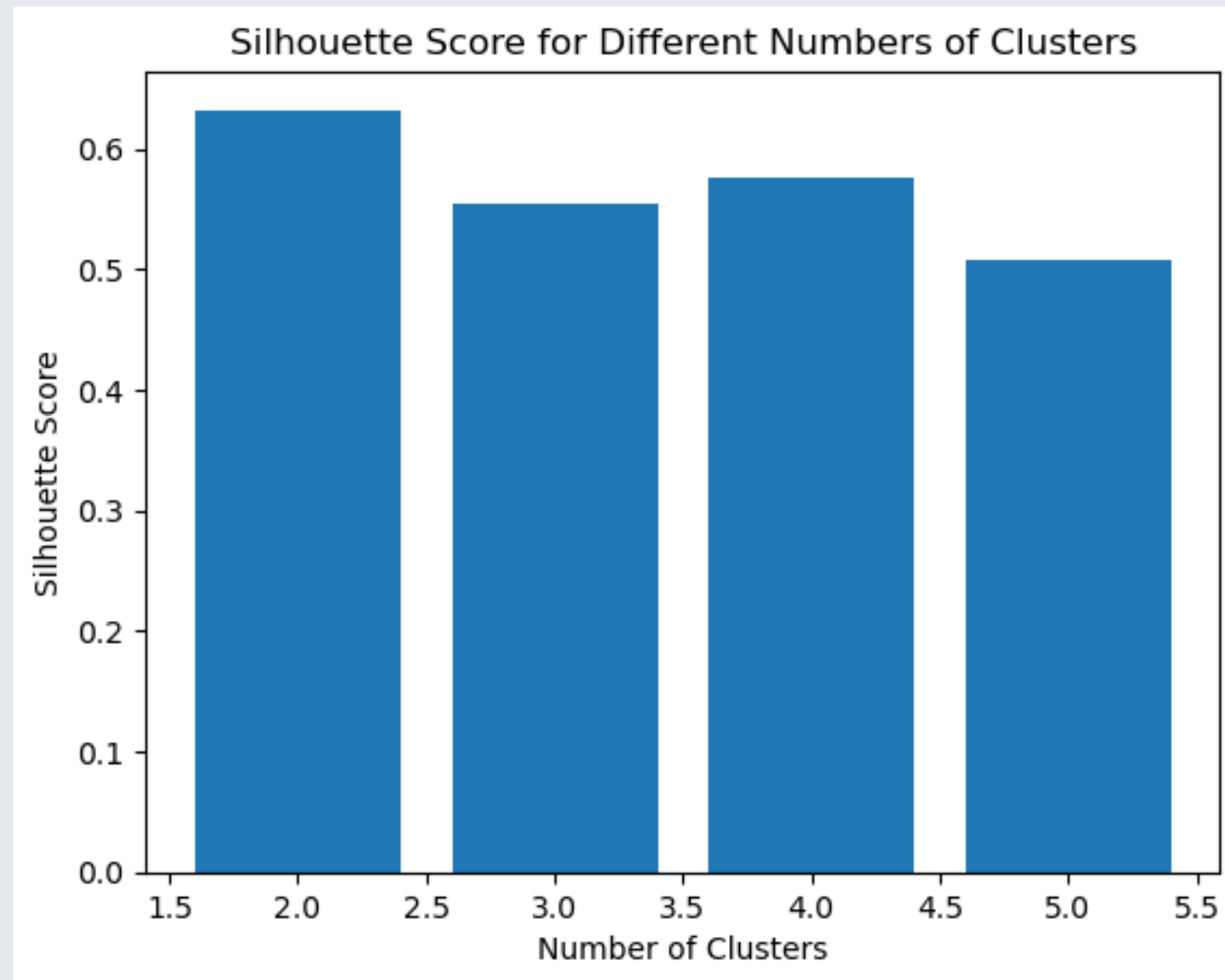
Untuk menghindari *noise* pada dataset dan kepentingan visualisasi, data hasil TF-IDF dilakukan reduksi dimensi dengan *Principal Component Analysis (PCO)*

	0	1	2	3	4	5	6	7	8	9	...	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266
abad	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
abai	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
abal	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
abang	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
abdi	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
...
wifi	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
wilayah	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
youtube	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
zaman	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
zenfone	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

1297 rows × 2267 columns

PENENTUAN JUMLAH CLUSTER

Silhouette Coefficient



Silhouette Coefficient:

- 2 Cluster-> 0,63
- 3 Cluster-> 0,55
- 4 Cluster-> 0,58
- 5 Cluster-> 0,51

Skor **di atas 0,5** dianggap bahwa hasil cluster berkategori **baik**. (Layton, et al., 2013)

Penentuan dilanjutkan ke penelusuran WordCloud

Sebaran Kata WordCloud

Percobaan 2 Cluster



Aspek 1: Update Data

Aspek 2: **Server**

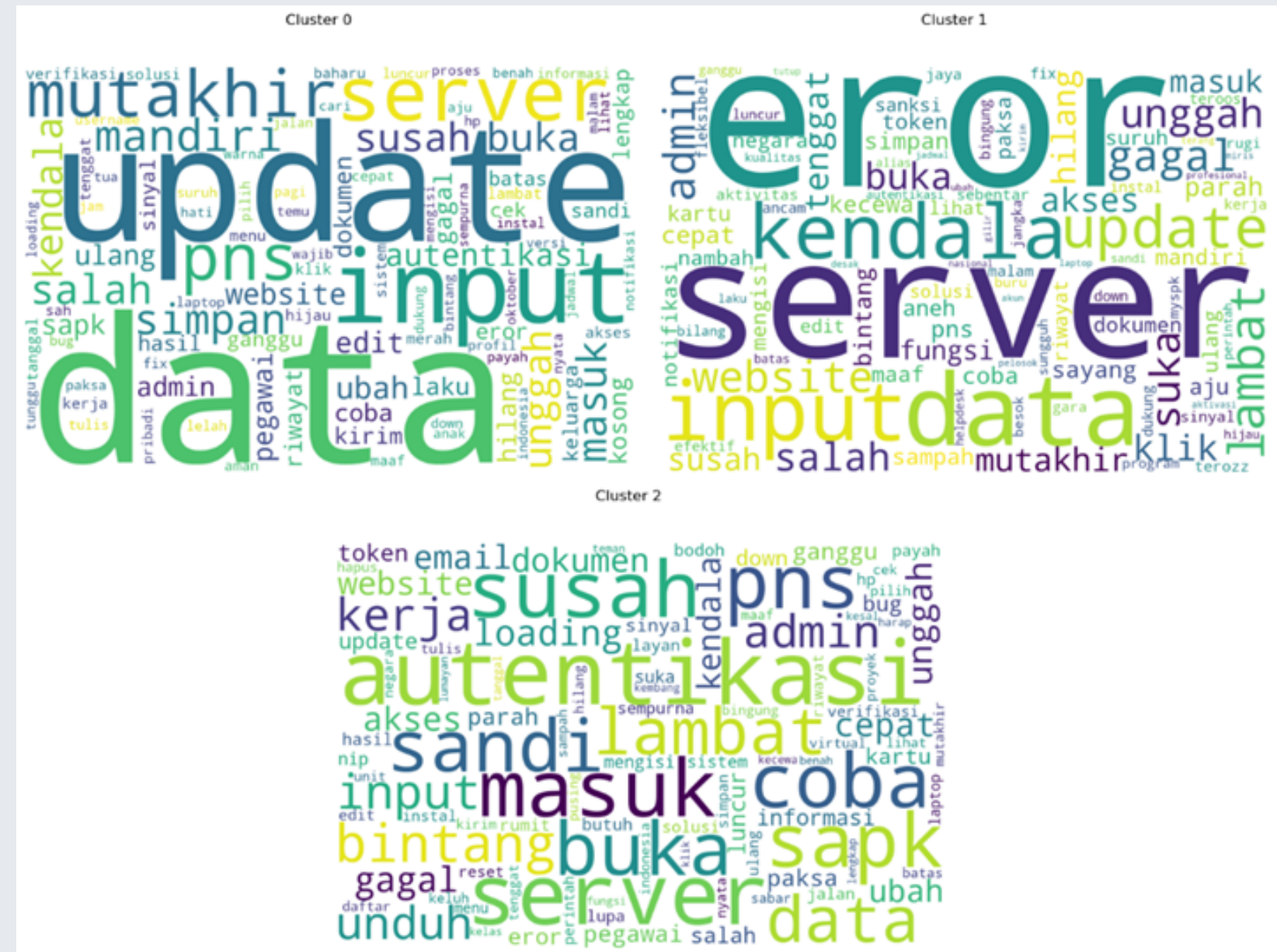
Sebaran Kata WordCloud

Percobaan 3 Cluster

Aspek 1: **Update Data**

Aspek 2: **Server**

Aspek 3: **Autentikasi**



Sebaran Kata WordCloud

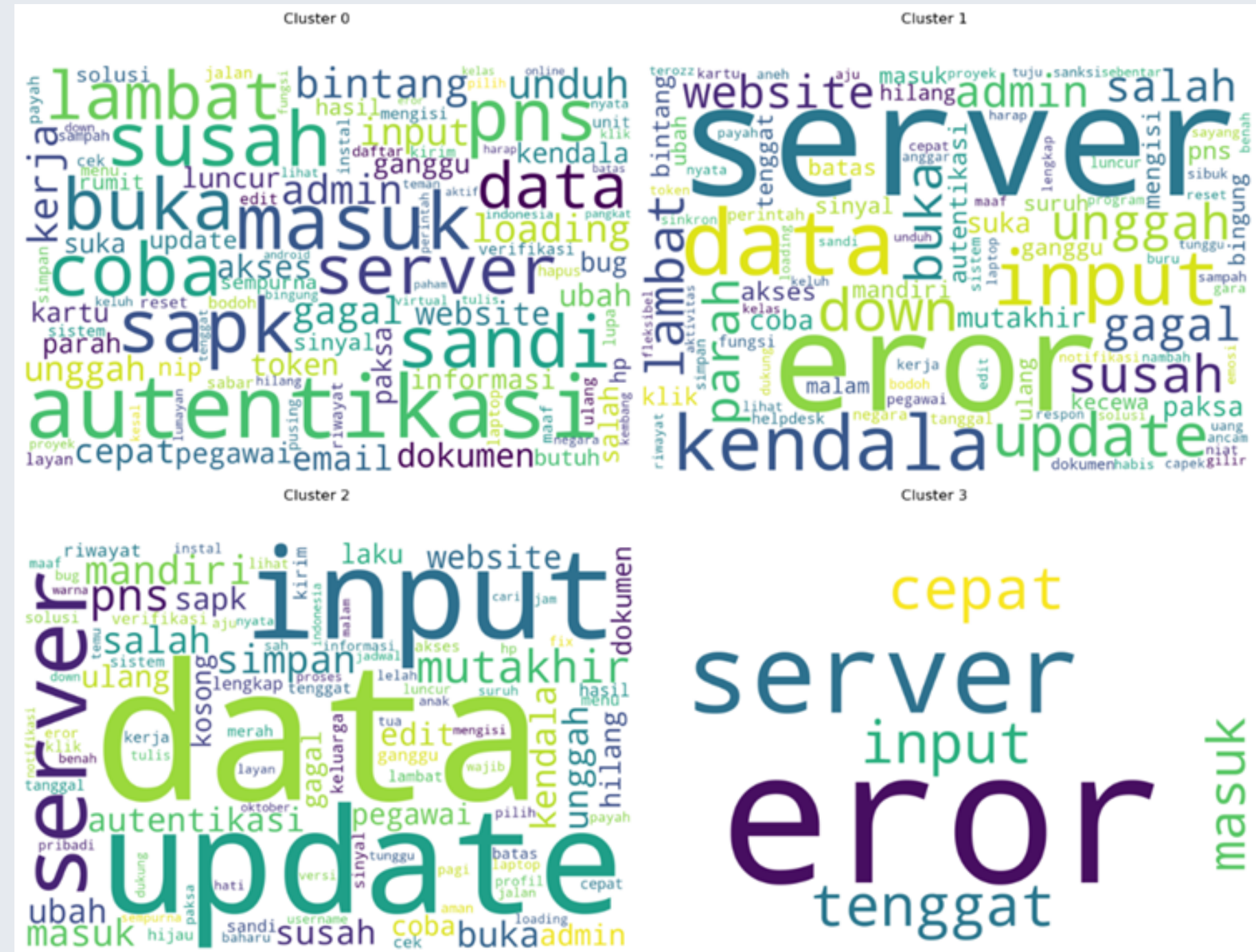
Percobaan 4 Cluster

Aspek 1: **Autentikasi**

Aspek 2: **Server**

Aspek 3: **Update Data**

Aspek 4: -



Sebaran Kata WordCloud

Percobaan 5 Cluster

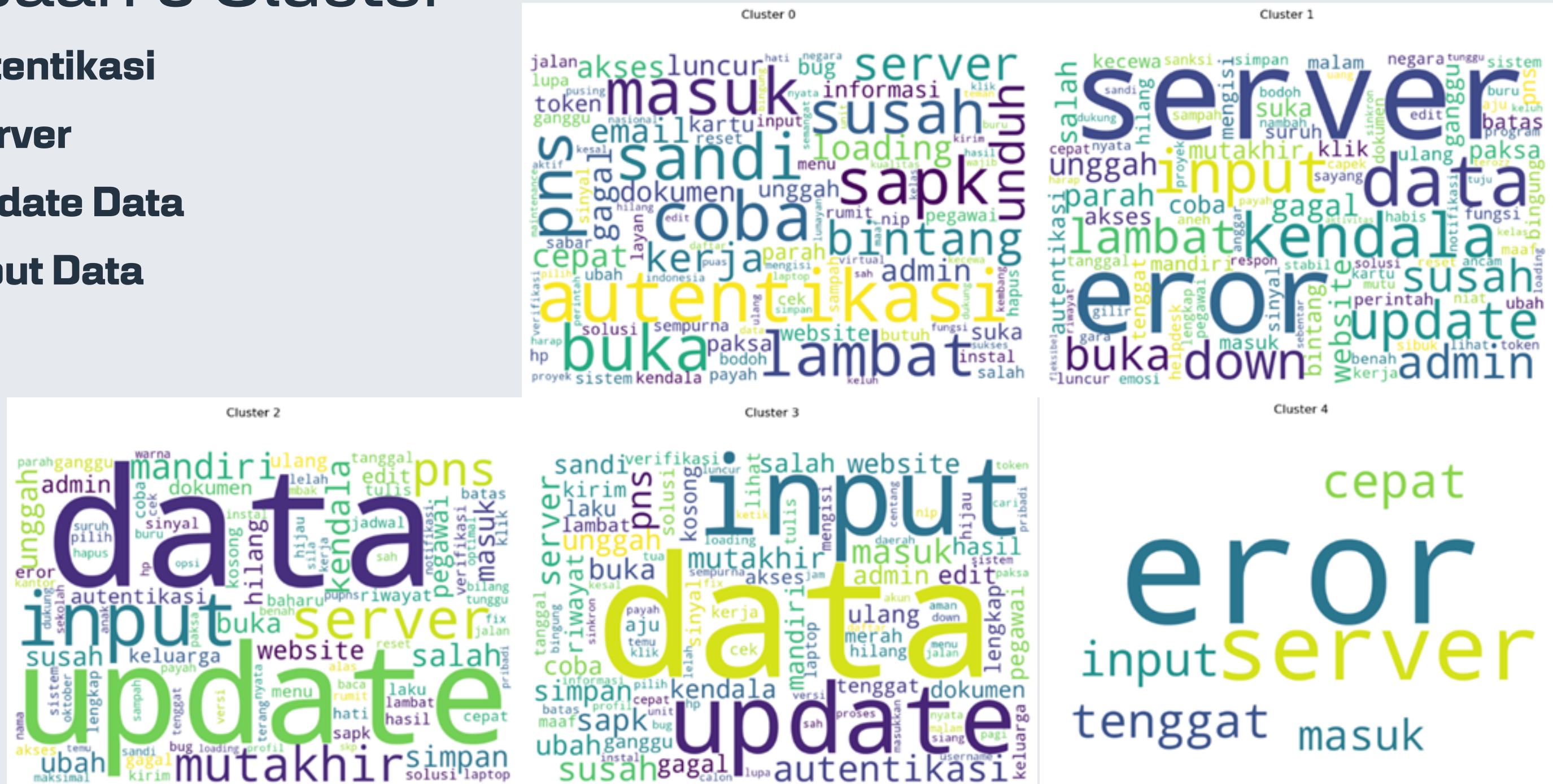
Aspek 1: Autentikasi

Aspek 2: Server

Aspek 3: Update Data

Aspek 4: Input Data

Aspek 5: –



PENENTUAN JUMLAH CLUSTER

Jumlah Klaster	Aspek	Jumlah Data Ulasan
2	Update Data	2.059
	Server	208
3	Update Data	534
	Server	200
	Autentikasi	1.533
4	Autentikasi	1.449
	Server	238
	Update Data	526
	-	54
5	Autentikasi	1.192
	Server	244
	Update Data	229
	Input Data	548
	-	54

2 Cluster-> Selisih jumlah ulasan terlalu besar

3 Cluster-> Memecah aspek Update Data menjadi aspek Autentikasi

4 Cluster-> Tidak terbentuk aspek pada *cluster 3*

5 Cluster-> Tidak terbentuk aspek pada *cluster 2*

Jumlah cluster yang dipilih = **3 cluster**

Hasil Aspek Akhir

No.	Aspek	Jumlah Data Ulasan
1	Update Data	534
2	Server	200
3	Autentikasi	1.533

Nilai *Silhouette Coefficient*:
0,55

Terima Kasih