

BAB IV

ANALISA DAN HASIL

Bagian ini memberikan penjelasan mengenai hasil penelitian tersebut bersumber pada formulasi dan tujuan penelitian. Tujuan penelitian adalah untuk 1) mengumpulkan data dari hasilnya analisis log dataset dengan KMeans Clustering untuk cyber profiling; dan 2) mengidentifikasi ciri-ciri pengguna saat melakukan kegiatan penelusuran web.

4.1 Hasil

4.1.1 Log Access

Log access pengunjung dimaksudkan untuk studi ini diperoleh dari Politeknik Negeri Lampung. Log berasal dari data access pada web server pada website polinela.ac.id.

Pada data ini terdapat 30344 record, gambar 4.1 di atas menunjukkan bahwa pembersihan data, juga dikenal sebagai preprocessing, masih perlu dilakukan.

	A	B	C	D	E	F	G	H	I	J
1	IP Address			Waktu		Request	Kode Status HTTP	Ukuran Respons	User-Agent:	Browser
2	185.191.34.185	-	-	[01/Jun/2022:00:11:41	+0700]	GET /153-mahasiswa-polinela-ikuti-wisuda	200	164641	-	Mozilla/5.0 (Windows NT 10.0;
3	185.191.34.185	-	-	[01/Jun/2022:00:11:46	+0700]	GET /wp-comments-post.php HTTP/1.1	301	251	-	Mozilla/5.0 (Windows NT 10.0;
4	185.191.34.185	-	-	[01/Jun/2022:00:11:47	+0700]	GET / HTTP/1.1	301	231	-	Mozilla/5.0 (Windows NT 10.0;
5	185.191.34.185	-	-	[01/Jun/2022:00:11:50	+0700]	GET /153-mahasiswa-polinela-ikuti-wisuda	301	-	-	Mozilla/5.0 (Windows NT 10.0;
6	185.191.34.185	-	-	[01/Jun/2022:00:11:53	+0700]	GET /153-mahasiswa-polinela-ikuti-wisuda	200	165563	-	Mozilla/5.0 (Windows NT 10.0;
7	66.220.149.20	-	-	[01/Jun/2022:00:19:09	+0700]	GET /templates/jb-polinela2016-25/image	404	132863	-	facebookexternalhit/1.1 (+http
8	66.220.149.120	-	-	[01/Jun/2022:00:21:18	+0700]	GET /publikasi/pengumuman/504-hasil-te	404	132863	-	facebookexternalhit/1.1 (+http
9	216.244.66.241	-	-	[01/Jun/2022:00:22:12	+0700]	GET /robots.txt HTTP/1.1	200	67	-	Mozilla/5.0 (compatible; DotBc
10	216.244.66.241	-	-	[01/Jun/2022:00:22:13	+0700]	GET /robots.txt HTTP/1.1	301	-	-	Mozilla/5.0 (compatible; DotBc
11	66.220.149.18	-	-	[01/Jun/2022:00:22:13	+0700]	GET /publikasi/berita/496-serah-terima-m	404	132863	-	facebookexternalhit/1.1 (+http
12	216.244.66.241	-	-	[01/Jun/2022:00:22:14	+0700]	GET /robots.txt HTTP/1.1	200	67	-	Mozilla/5.0 (compatible; DotBc
13	66.220.149.24	-	-	[01/Jun/2022:00:22:14	+0700]	GET /publikasi/berita/491-polinela-sabet-	404	132863	-	facebookexternalhit/1.1 (+http
14	66.220.149.11	-	-	[01/Jun/2022:00:22:18	+0700]	GET /images/berita/LSP1.jpg HTTP/1.1	301	-	-	facebookexternalhit/1.1 (+http
15	66.220.149.11	-	-	[01/Jun/2022:00:22:19	+0700]	GET /images/berita/LSP1.jpg HTTP/1.1	404	132863	-	facebookexternalhit/1.1 (+http
16	66.220.149.1	-	-	[01/Jun/2022:00:22:36	+0700]	GET /images/banners/um_polinela_2018.j	301	-	-	facebookexternalhit/1.1 (+http
17	66.220.149.1	-	-	[01/Jun/2022:00:22:37	+0700]	GET /images/banners/um_polinela_2018.j	404	132863	-	facebookexternalhit/1.1 (+http
18	103.151.63.12	-	-	[01/Jun/2022:00:22:39	+0700]	GET /index.php?format=feed&type=rss HT	301	-	-	WordPress/6.0; https://pascas
19	173.252.95.11	-	-	[01/Jun/2022:00:22:47	+0700]	GET /templates/jb-polinela2016-25/image	301	-	-	facebookexternalhit/1.1 (+http
20	173.252.95.11	-	-	[01/Jun/2022:00:22:48	+0700]	GET /templates/jb-polinela2016-25/image	404	132863	-	facebookexternalhit/1.1 (+http

Gambar 4.1 potongan access log

Beberapa field, seperti Kode Status HTTP, Ukuran Respons, User-Agent, dan Referer, dibuang karena tidak penting untuk proses selanjutnya. Selain itu, ada cell yang perlu diubah menjadi.

Proses preprocessing pertama akan membuat Gambar 4.2

seperti dibawah ini :

IP Adress	Waktu	Requet Halaman
185.191.34.185	[01/Jun/2022:00:11:41	GET /153-mahasiswa-polinela-ikuti-wisuda-daring/ HTTP/1.1
185.191.34.185	[01/Jun/2022:00:11:46	GET /wp-comments-post.php HTTP/1.1
185.191.34.185	[01/Jun/2022:00:11:47	GET / HTTP/1.1
185.191.34.185	[01/Jun/2022:00:11:50	GET /153-mahasiswa-polinela-ikuti-wisuda-daring HTTP/1.1
185.191.34.185	[01/Jun/2022:00:11:53	GET /153-mahasiswa-polinela-ikuti-wisuda-daring/ HTTP/1.1
66.220.149.20	[01/Jun/2022:00:19:09	GET /templates/jb-polinela2016-25/images/politekniknegerilampung.png HTTP/1.1
66.220.149.120	[01/Jun/2022:00:21:18	GET /publikasi/pengumuman/504-hasil-tes-tulis-um.html HTTP/1.1
216.244.66.241	[01/Jun/2022:00:22:12	GET /robots.txt HTTP/1.1
216.244.66.241	[01/Jun/2022:00:22:13	GET /robots.txt HTTP/1.1
66.220.149.18	[01/Jun/2022:00:22:13	GET /publikasi/berita/496-serah-terima-mahasiswa-pkn-politeknik-negeri-lampung-di-kab-lambar-lampung-barat.html HTTP/1.1
216.244.66.241	[01/Jun/2022:00:22:14	GET /robots.txt HTTP/1.1

Gambar 4.2 potongan setelah preprocessing

Dengan 18 kategori yang dikunjungi oleh pengguna berdasarkan data log akses pada website, dan total kunjungan sebanyak 7574 kali, Anda memiliki bahan yang bagus untuk menerapkan teknik clustering seperti K-Means untuk menganalisis pola kunjungan.

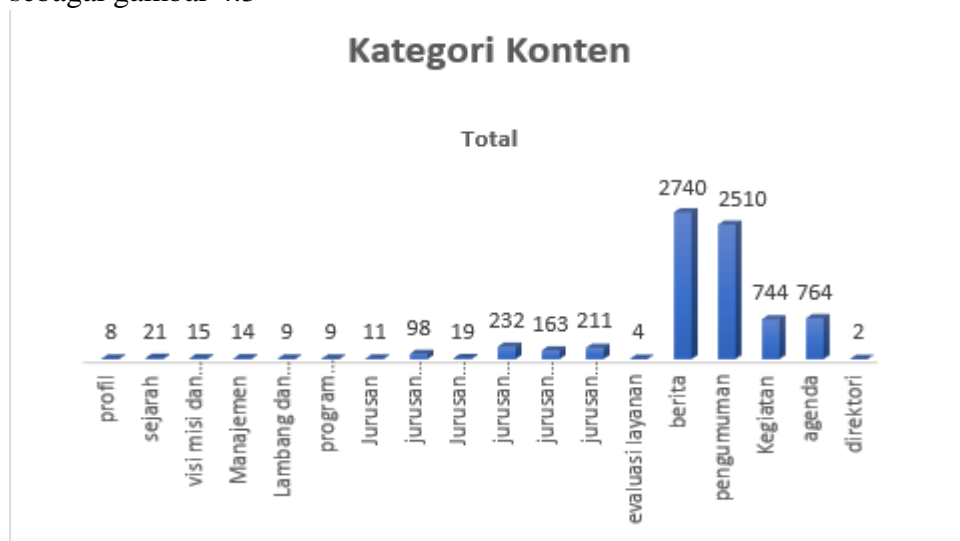
4.1.2 Jenis Data Berdasarkan Konten

Pengkategorian website berdasarkan menu dan submenu dapat membantu memahami struktur dan konten website dengan lebih baik. Berdasarkan 7574 akses konten yang didapat, ada 18 jenis akses konten yang dapat dikategorikan berdasarkan Menu dan Sub Menu pada website polinela.ac.id. Pada Tabel 4.1 merupakan contoh jenis akses konten yang berhasil dikategorikan.

Pada table 4.1 contoh jenis data berdasarkan konten.

Kategori Konten	Jumlah Kunjungan
Agenda	764
Evaluasi layanan	4
Berita	2740
Direktori	2
Jurusan	11
Jurusan budidaya tanaman pangan	98
Jurusan Budidaya tanaman perkebunan	19

Menunjukkan contoh pengkategorian data yang didasarkan pada konten web sebagai gambar 4.3



Gambar 4.3 Kategori Konten

Dari total 7574 akses konten, Sebagai bagian dari proses cyber profiling, hasil pengkategorian yang ada di situs web ini dirancang untuk mengidentifikasi kategori apa yang biasanya dapat diakses oleh pengunjung website. Ini membantu dalam penarikan kesimpulan yang berkaitan dengan proses cyber profiling.

4.1.3 Waktu Akses untuk Kategori Data

Untuk menjelaskan Waktu Akses untuk Kategori Data, bagian ini membaginya menjadi empat kategori: pagi-siang, siang-sore, sore-malam, dan malam-pagi. Tabel 4.2 menunjukkan ilustrasi kategori waktu akses berdasarkan waktu akses.

Tabel 4.2 Contoh kategori data menggunakan waktu akses

Akses	Jumlah Pengunjung	Kategori
Berita	1283	Pagi-Siang
Pengumuman	891	Siang-Sore
Sejarah	1457	Sore-Malam
Jurusan	6	Pagi-Siang
Visi Misi dan Tujuan	4	Siang-Sore

Pada kategori waktu dari pagi-siang, terdapat 3534 kunjungan web yang dapat diakses oleh 264.

Situs pencarian saat ini seperti sub menu berita adalah salah satu yang paling sering diperhatikan oleh pengguna, diikuti oleh kunjungan sub menu pengumuman juga memiliki banyak pengunjung dan paling sedikit dikunjungi sub menu profil pada kategori waktu pagi-siang. Ini adalah kategori yang mencakup sedikit anggota dibandingkan kategori waktu lainnya dalam melakukan aktifitas kunjungan ke website.

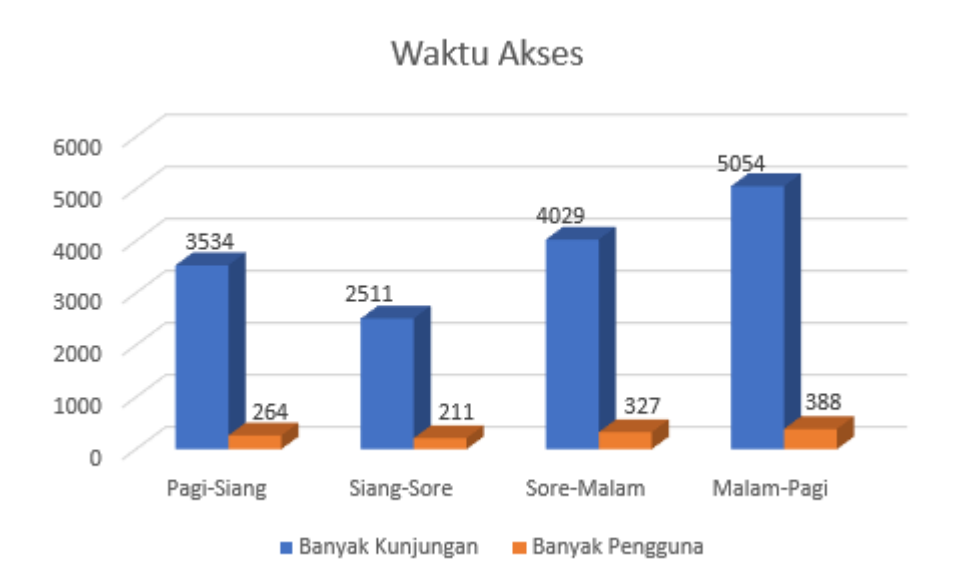
Pada kategori waktu siang-sore, terdapat 2511 kunjungan web yang diakses oleh pengguna dengan total pengguna sebanyak 211. Situs di kategori ini seperti submenu berita merupakan salah satu konten yang paling banyak dikunjungi, disusul secara berurutan dengan sub menu pengumuman, dan sub menu profil pencarian yang paling sedikit. Kategori ini merupakan kategori paling sedikit pengunjung web, walaupun waktu akses masih dalam jam produktif pengguna tidak mengakses website di waktu tersebut sehingga adanya penurunan kunjungan website.

Pada kategori waktu sore-malam, terdapat 4029 kunjungan web yang diakses oleh semua pengguna sebanyak 327. Pada kategori ini situs pencarian seperti submenu berita merupakan salah satu konten yang paling sering dilihat, disusul secara berurutan dengan sub menu pengumuman, sub menu agenda dan sub menu kegiatan. Kategori ini merupakan kategori sedang pengunjung web, walaupun waktu akses menunjukkan jam istirahat pengguna lebih banyak mengakses website sehingga adanya peningkatan kunjungan website.

Pada kategori waktu malam-pagi, terdapat 5054 kunjungan menu web yang dapat diakses pengguna melalui jumlah jumlah total pengguna sebesar 388. Situs pencarian dalam kategori ini seperti submenu berita adalah salah satu dari konten yang paling sering dilihat, diikuti oleh sub menu pengumuman, sub menu agenda dan sub menu kegiatan . Kategori ini merupakan kategori paling banyak pengunjung web, pada waktu akses website mengalami peningkatan kunjungan website yang sangat banyak.

Website pencarian konten berita memiliki jumlah pengunjung tertinggi berdasarkan kategori waktu akses, seperti yang ditunjukkan oleh submenu berita sebagai kunjungan konten teratas berdasarkan jumlah pengunjung.

Secara keseluruhan, pengkategorian data berdasarkan waktu akses digambarkan sebagai gambar 4.4:



Gambar 4.4 Waktu Akses untuk Kategori Data

4.1.4 Jenis Data Berdasarkan Jumlah Pengunjung

Dengan menggunakan algoritma K-MEANS, kami akan menjelaskan kategori data berdasarkan banyak pengunjung web di bagian ini. Banyak kunjungan web digunakan sebagai atribut untuk kategori ini. Website paling banyak yang dikunjungi, paling sedikit yang dikunjungi, dan rata-rata yang dikunjungi digunakan untuk menghitung tingkat kunjungan ini. Tabel 4.3 berikut menunjukkan bagaimana pusat cluster awal diaktifkan pada proses klustering ini:

Tabel 4.3 Pusat Awal Klaster Dimulai

	Klaster		
	1	2	3
Jumlah Pengunjung	2	232	2740

Nilai Terbanyak, tengah, dan terkecil dari data digunakan untuk menentukan nilai di atas. Nilai-nilai ini berfungsi pusat cluster awal yang akan diikuti oleh metode K-MEANS.

4.1.5 Implementasi Tahap K-MEANS

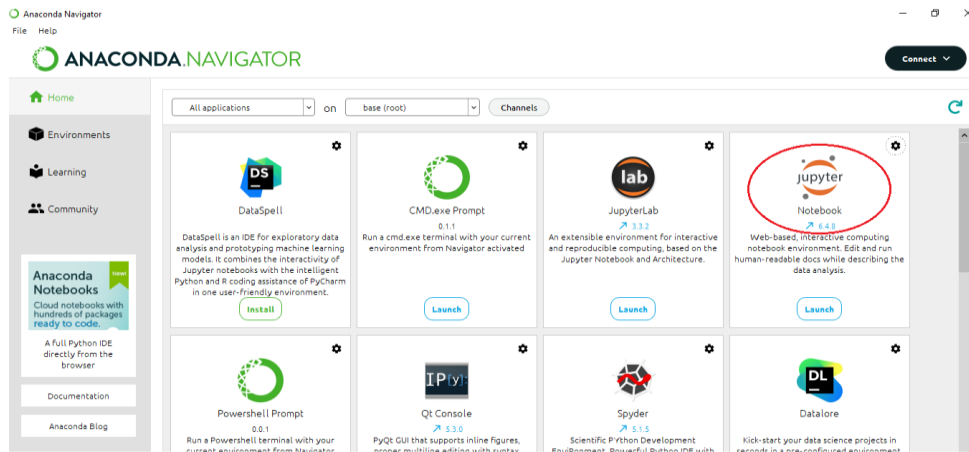
Untuk mengidentifikasi proses cyber profiling pada pola kunjungan website polinela.ac.id, penulis menggunakan program atau aplikasi pendukung, seperti anaconda navigator dan Jupyter Notebook, yang dikodekan dalam bahasa pemrograman Python. Program ini dirancang untuk mengelompokkan data jumlah kunjungan website dalam proses cyber profiling. Untuk memulai pengolahan program clustering K-MEANS, pengguna pertama kali membuat dataset menggunakan Microsoft Excel dalam format.csv, yang kemudian telah diproses sebagai gambar 4.5:

	A	B	C	D	E	F	G
1	Kategori	data ke	jumlah pengunjung(X)	Pagi-siang	Siang-sore	Sore-malam	Malam-Pagi
2	agenda	1	764	365	270	399	494
3	evaluasi layanan	2	4	1	0	0	0
4	berita	3	2740	1283	891	1457	1849
5	direktori	4	2	1	2	0	0
6	jurusan	5	11	6	4	5	7
7	jurusan budidaya tanaman pangan	6	98	48	29	50	69
8	Jurusan Budidaya tanaman perkebunan	7	19	10	6	9	13
9	jurusan ekonomi dan bisnis	8	211	101	67	110	144
10	jurusan peternakan	9	163	80	56	83	107
11	jurusan teknologi pertanian	10	232	110	78	122	154
12	kegiatan	11	744	341	261	403	483
13	lambang dan makna	12	9	6	0	0	6
14	manajemen	13	14	6	5	8	9
15	pengumuman	14	2510	1160	827	1350	1683
16	profil	15	8	3	2	5	6
17	program kerjasama	16	9	0	5	7	4
18	sejarah	17	21	9	5	12	16
19	visi misi dan tujuan	18	15	6	5	9	10

Gambar 4.5 Dataset di Microsoft Excel

Data jumlah kunjungan untuk setiap kategori yang telah dibuat, bersama dengan jumlah data kunjungan keseluruhan dan waktu, akan ditampilkan dalam file keluaran dari Microsoft Excel dalam format.csv, seperti yang ditunjukkan pada gambar 4.5 dari dataset di atas. Data yang akan digunakan untuk pengelompokan atau pengklasteran jumlah kunjungan terbanyak, sedang dan sedikit.

Selanjutnya, buka Anaconda Navigator dan pilih jupyter notebook, yang dibuat menggunakan bahasa pemrograman python, yang ditunjukkan pada gambar 4.6:



Gambar 4.6 Navigator Anaconda View

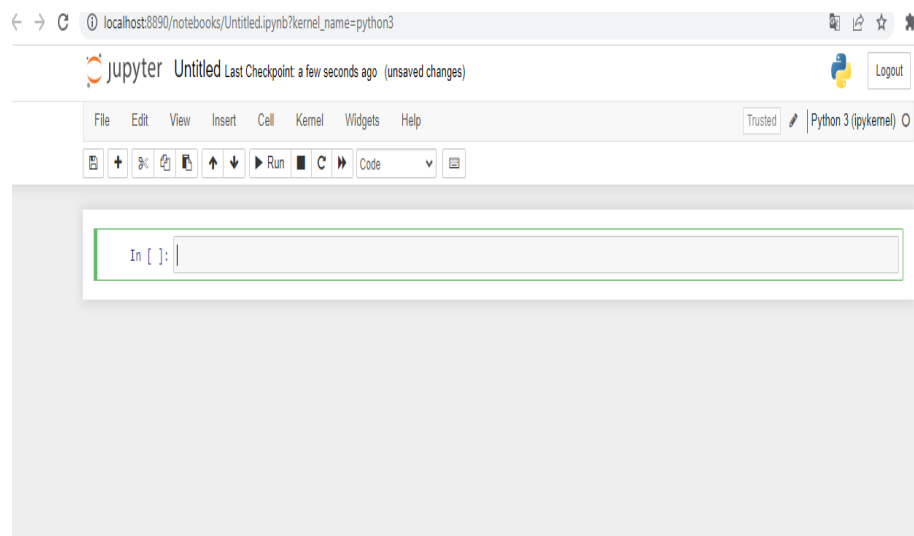
Pengguna dapat secara interaktif mengintegrasikan kode dan output dalam satu dokumen dengan Jupyter Notebook, alat pengolahan data Python yang populer. Setelah memulai, pengguna akan dibawa ke tampilan browser untuk melakukan tahap pemrograman atau pengolahan data. Seperti yang ditunjukkan pada Gambar 4.7:



Gambar 4.7 Menunjukkan tampilan notebook jupyter.

Setelah memilih lokasi penyimpanan atau direktori, pengguna harus menuju menu baru disebelah kanan pojok atas dan memilih python3.

Kemudian dibawa ke tampilan jupyter notebook, di mana program diproses. seperti yang ditunjukkan pada gambar 4.8:



Gambar 4.8 Visualisasi jupyter notebook di workspace

Untuk mendukung pengolahan program clustering K-MEANS, ada banyak library, module, dan paket yang tersedia sesuai kebutuhan pengguna dan program yang ingin diolah. Tahap pertama adalah memasukkan sejumlah buku bacaan setelah Visualisasi jupyter notebook terbuka. Beberapa library yang digunakan termasuk numpy, matplotlib adalah program yang membantu Anda melakukan komputasi data numerik dengan pandas yang memungkinkan Anda membuat tabel, mengubah ukuran data, dan melakukan pengecekan. pyplot memungkinkan Anda membuat berbagai jenis plot untuk memvisualisasikan data Anda dan mendapatkan wawasan darinya, dan library sklearn. Cluster melakukan pengelompokan K-MEANS pada data Anda dan mendapatkan tugas klaster dan pusat klaster untuk analisis.

Selanjutnya, tekan tombol shift dan enter untuk masuk ke sel baru. Selanjutnya adalah mengimport kumpulan data yang telah dimodifikasi ke microsoft office excel.

Untuk memanggil dataset dan menampilkan dataset, pengguna menggunakan perintah yang ada di gambar 4.9:

```
[2]: # Load data from CSV
file_path = 'data2024.csv' # Ganti dengan path file CSV Anda
data = pd.read_csv(file_path)
#Menampilkan Dataset yang telah ditambahkan
data.head()
```

```
[2]:
```

	Kategori	data ke	jumlah pengunjung(X)	Pagi-siang	Siang-sore	Sore-malam	Malam-Pagi
0	agenda	1	764	365	270	399	494
1	evaluasi layanan	2	4	0	0	0	0
2	berita	3	2740	1283	891	1457	1849
3	direktori	4	2	0	0	0	2
4	jurusan	5	11	6	4	5	7

Gambar 4.9 Bentuk Dataframe dari Dataset

Dapat dilihat nilai jumlah, rata-rata, standar deviasi, minimum, 25%, 50%, 75%, dan maksimum dari masing-masing variable. Selanjutnya, seperti yang ditunjukkan pada gambar 4.10, penulis ingin melihat nomor index dan tipe data dari dataset tersebut:

```
#Describing Data
data.describe()
```

	data ke	jumlah pengunjung(X)	Pagi-siang	Siang-sore	Sore-malam	Malam-Pagi
count	19.000000	19.000000	19.000000	19.000000	19.000000	19.000000
mean	10.421053	417.368421	195.368421	138.842105	221.421053	278.157895
std	6.067361	813.020583	378.521276	267.020653	434.866942	546.321056
min	1.000000	2.000000	0.000000	0.000000	0.000000	0.000000
25%	5.500000	10.000000	6.000000	4.500000	6.000000	6.500000
50%	10.000000	21.000000	10.000000	6.000000	12.000000	16.000000
75%	15.500000	294.000000	144.000000	102.500000	150.000000	191.500000
max	20.000000	2740.000000	1283.000000	891.000000	1457.000000	1849.000000

Gambar 4.10 Describing data

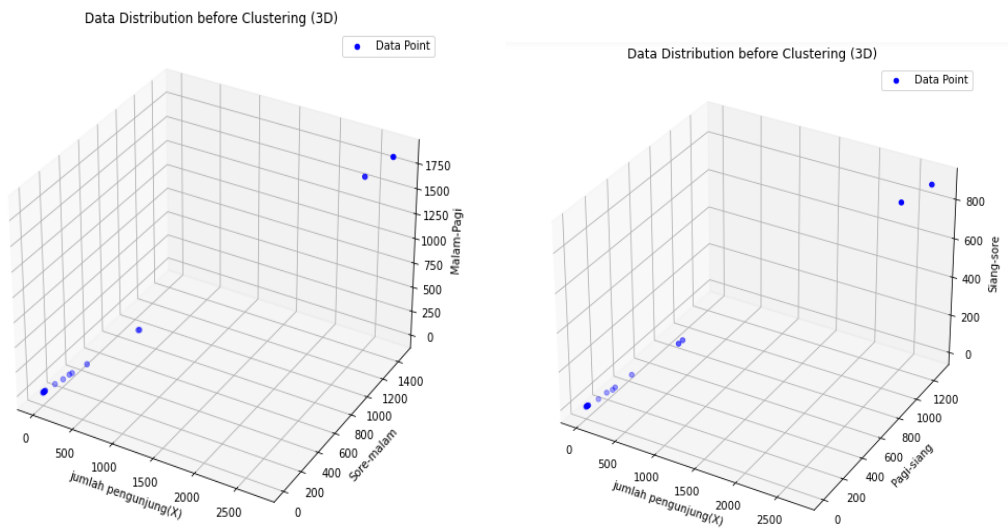
Setelah informasi tentang masing-masing variabel dan jenis datanya ditampilkan, penulis dapat melihat kolom pada dataframe dan banyaknya data dari masing-masing variabel, sebagai yang digambarkan pada gambar 4.11:

```
#Data Info
data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 19 entries, 0 to 18
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Kategori              19 non-null    object
1   data ke               19 non-null    int64
2   jumlah pengunjung(X)  19 non-null    int64
3   Pagi-siang           19 non-null    int64
4   Siang-sore           19 non-null    int64
5   Sore-malam           19 non-null    int64
6   Malam-Pagi           19 non-null    int64
dtypes: int64(6), object(1)
memory usage: 1.2+ KB
```

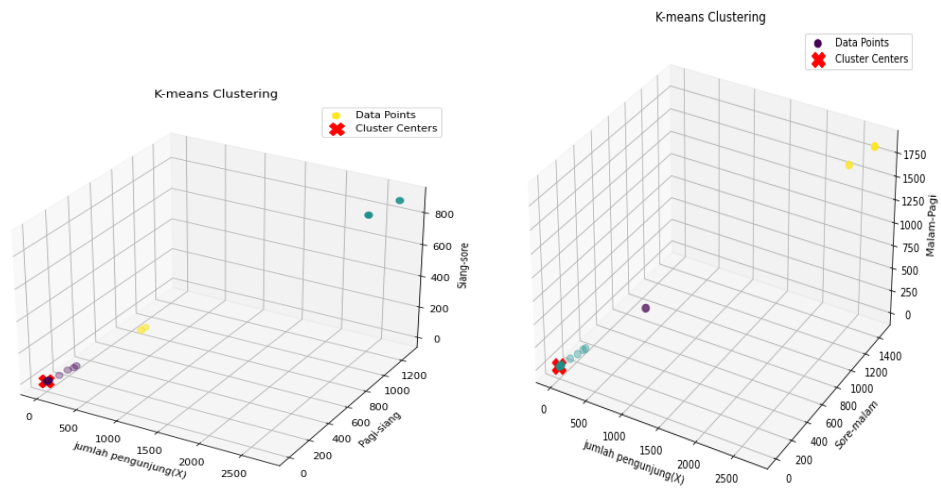
Gambar 4.11 Tampilan info dari Dataset

Penulis ingin menganalisa data kunjungan log sebelum di clustering supaya terlihat perbedaannya ditunjukkan pada gambar 4.12:



Gambar 4.12 Data kunjungan sebelum di clustering

Hasil grafik scatterplot yang telah dibuat menunjukkan ketiga kelompok dan anggota-anggotanya secara jelas seperti pada gambar 4.13:



Gambar 4.13 Grafik Scatter plot hasil akhir

Terdapat 18 data kategori dengan masing-masing cluster terletak di antara cluster 0, cluster 1, dan cluster 2. Hasil dari cluster ini dapat ditampilkan dalam bentuk frame data dengan mengubah kolom baru untuk cluster tersebut. seperti pada gambar 4.14:

```

Cluster 1 data points:
    Kategori data ke jumlah pengunjung(X) \
1     evaluasi layanan      2           4
3     direktori              4           2
4     jurusan                5          11
5     jurusan budidaya tanaman pangan      6          98
6     Jurusan Budidaya tanaman perkebunan  7          19
7     jurusan ekonomi dan bisnis          8         211
8     jurusan peternakan                  9         163
9     jurusan teknologi pertanian         10        232
11    lambang dan makna                   12          9
12    manajemen                           13         14
14    profil                              15          8
15    program kerjasama                   16          9
16    sejarah                             17         21
17    visi misi dan tujuan                18         15

    Pagi-siang Siang-sore Sore-malam Malam-Pagi Cluster
1         0         0         0         0         0
3         0         0         0         2         0
4         6         4         5         7         0
5        48        29        50        69         0
7        10         6         9        13         0
8       101        67        110       144         0
9        80         56         83       107         0
11       110        78        122       154         0
12        6         0         0         6         0
14        6         5         8         9         0
15        3         2         5         6         0
16        0         5         7         4         0
17        9         5        12        16         0
17        6         5         9        10         0

Cluster 2 data points:
    Kategori data ke jumlah pengunjung(X) Pagi-siang Siang-sore \
2     berita      3           2740       1283       891
13  pengumuman  14           2510       1160       827

    Sore-malam Malam-Pagi Cluster
2         1457       1849       1
13        1350       1683       1

Cluster 3 data points:
    Kategori data ke jumlah pengunjung(X) Pagi-siang Siang-sore \
0     agenda      1           764        365       270
10    kegiatan   11           744        341       261

    Sore-malam Malam-Pagi Cluster
0         399        494       2
10        403        483       2

```

Gambar 4.14 Menampilkan Cluster

4.1.6 Pembahasan

Perhitungan yang dilakukan secara manual untuk mengolah clustering K-MEANS dan data kunjungan halaman web akan dijelaskan dalam pembahasan proses pengolahan clustering K-MEANS. Bagian dari proses pengolahan clustering K-MEANS secara manual menggunakan aplikasi microsoft excel pada data akses log website polinela.ac.id pasti memiliki beberapa perhitungan dan persamaan.

Untuk melakukan perhitungan clustering K-MEANS, langkah-langkah berikut harus diikuti:

- 1) Ada studi kasus atau data yang ingin dikelompokkan, seperti yang ditunjukkan pada gambar 4.15:

A	B	C	D	E	F
Kategori	jumlah pengunjung	Pagi-siang	Siang-sore	Sore-malam	Malam-pagi
agenda	764	365	270	399	494
evaluasi layanan	4	1	0	0	0
berita	2740	1283	891	1457	1849
direktori	2	0	1	0	0
jurusan	11	6	4	5	7
jurusan budidaya tanaman pangan	98	48	29	50	69
Jurusan Budidaya tanaman perkebunan	19	10	6	9	13
jurusan ekonomi dan bisnis	211	101	67	110	144
jurusan peternakan	163	80	56	83	107
jurusan teknologi pertanian	232	110	78	122	154
kegiatan	744	341	261	403	483
lambang dan makna	9	6	0	0	6
manajemen	14	6	5	8	9
pengumuman	2510	1160	827	1350	1683
profil	8	2	1	5	6
program kerjasama	9	0	5	7	4
sejarah	21	9	5	12	16
visi misi dan tujuan	15	6	5	9	10

Gambar 4.15 Data awal excel

Data awal atau data yang diinginkan untuk pengklasteran harus ada, seperti yang dijelaskan pada gambar 4.15. Penulis menggunakan data konten sebanyak 18 kategori yang memiliki 5 variable yang akan digunakan yaitu “jumlah pengunjung” dan waktu akses terdiri dari “Pagi-siang”, ”Siang-Sore”, ”Sore-malam”, Malam-pagi” untuk dilakukan clustering K-MEANS.

- 2) Menentukan berapa banyak cluster dan kategori nilai pusat yang ada
 Tahap kedua, menentukan jumlah kelompok, digambarkan pada gambar 4.16. Penulis memilih tiga cluster: "Banyak", "Sedang", dan "Sedikit" untuk studi kasus tugas akhir untuk menentukan banyaknya kunjungan untuk proses profiling. Kemudian, mereka menentukan nilai pusat pada data awal berdasarkan jumlah cluster tersebut. Nilai pusat dapat dipilih secara acak atau acak. Di sini, penulis menggunakan nilai pusat cluster 1 "2", nilai pusat cluster 2 "232", dan nilai pusat cluster 3 "2740".

Kategori	jumlah pengunjung	Pagi-siang	Siang-sore	Sore-malam	Malam-pagi	
agenda	764	365	270	399	494	
evaluasi layanan	4	1	0	0	0	
direktori	2	0	1	0	0	C1
jurusan	11	6	4	5	7	
jurusan budidaya tanaman pangan	98	48	29	50	69	
Jurusan Budidaya tanaman perkebunan	19	10	6	9	13	
jurusan ekonomi dan bisnis	211	101	67	110	144	
jurusan peternakan	163	80	56	83	107	
jurusan teknologi pertanian	232	110	78	122	154	C2
kegiatan	744	341	261	403	483	
lambang dan makna	9	6	0	0	6	
manajemen	14	6	5	8	9	
pengumuman	2510	1160	827	1350	1683	
profil	8	2	1	5	6	
program kerjasama	9	0	5	7	4	
sejarah	21	9	5	12	16	
visi misi dan tujuan	15	6	5	9	10	
berita	2740	1283	891	1457	1849	C3

Gambar 4.16 Menentukan jumlah cluster

3) Berapa jauh jarak terdekat

Pada titik ini, proses perhitungan dilakukan untuk menentukan nilai centroid

1, centroid 2 dan centroid 3, seperti yang ditunjukkan pada gambar 4.17:

Data ke	C1	C2	C3	Minimum	cluster
1	609585	300390	4185730	300390	1
2	4	54453	7981791	4	0
3	0	54519	7982577	0	0
4	128	49571	7921495	128	0
5	10445	17448	7418960	10445	0
6	392	45887	7874075	392	0
7	47602	327	6800063	327	1
8	27924	4503	7069451	4503	1
9	56859	100	6694656	100	1
10	580321	280284	4263132	280284	1
11	80	51571	7946839	80	0
12	209	48204	7903652	209	0
13	6685373	5536264	58460	58460	2
14	71	51170	7940746	71	0
15	88	51177	7940341	88	0
16	516	44785	7858705	516	0
17	246	47707	7897075	246	0
18	7982577	6721456	0	0	2

Gambar 4.17 Nilai jarak terdekat

Proses perhitungan pertama ditunjukkan pada gambar 4.17, di mana rumus euclidean distance pada persamaan 3 digunakan untuk menentukan jarak minimum data terhadap centroid:

- Situs yang paling sedikit dikunjungi (Sedikit) untuk C1:

$$= \sqrt{(764 - 2)^2 + (365 - 0)^2 + (270 - 1)^2 + (399 - 0)^2 + (494 - 0)^2} = 609585$$

Formula Excel : =SQRT((764- 2)^2)+ (365- 0)^2)+ (270- 1)^2)+ (399- 0)^2)+ (494- 0)^2))

- Situs yang paling sedikit dikunjungi (Sedikit) untuk C2:

$$= \sqrt{(764 - 232)^2 + (365 - 110)^2 + (270 - 78)^2 + (399 - 112)^2 + (494 - 154)^2} = 300390$$

Formula Excel : =SQRT((764- 232)^2)+ (365-11 0)^2)+ (270- 78)^2)+ (399- 112)^2)+ (494- 154)^2))

- Situs yang paling sedikit dikunjungi (Sedikit) untuk C3:

$$= \sqrt{(764 - 2740)^2 + (365 - 1283)^2 + (270 - 891)^2 + (399 - 1457)^2 + (494 - 1849)^2} = 4185730$$

Formula Excel : =SQRT((764- 2740)^2)+ (365- 1283)^2)+ (270- 891)^2)+ (399- 1457)^2)+ (494- 1849)^2))

Untuk menemukan nilai centroid terminimum, perhitungan dilakukan sebanyak n data menggunakan rumus jarak geometri. Nilai pusat digunakan untuk menghitung nilai centroid pertama sesuai dengan cluster yang telah ditetapkan sebelumnya; metode ini juga digunakan untuk menghitung nilai centroid kedua dan ketiga, hingga nilai centroid ketiga ditemukan. Setelah memperoleh nilai centroid 0, centroid 1, dan centroid 2, jarak terkecil antara ketiga centroid dapat dilihat untuk menentukan berapa banyak data yang masuk ke cluster.

4) Update nilai centroid

Setelah mendapatkan cluster pada setiap data, langkah selanjutnya adalah mengubah nilai centroid:

$$C1: \frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}$$

$$C2: \frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}$$

$$C3: \frac{\frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}, \frac{\text{Jumlah data}}{\text{Banyak data class}}}{\frac{\text{Jumlah data}}{\text{Banyak data class}}}$$

Pada Perhitungan centroid baru (iterasi) dilakukan oleh algoritma K-MEANS sampai hasil iterasi sebelumnya sama dengan hasil centroid sebelumnya. Dalam studi kasus Analisis Access Log Web Menggunakan K-Means Clustering untuk Cyber Profiling, saya melakukan perhitungan ini secara manual menggunakan aplikasi Microsoft Excel. Saya melakukan ini tiga kali. Adapun nilai centroid akhir yang diperoleh yaitu (2309,9375;1085,375;788,3125;1213,5625;1512,625), (2510;1160;827;1350;1683) dan (2740;1283;891;1457;1849).

Gambar 4.18 Menunjukkan hasil akhir proses clustering setiap iterasi dengan melihat tidak adanya perubahan iterasi, keterangan “aman” ketika tidak ada perubahan dan ketika masih ada perubahan iterasi diberikan keterangan “berubah”

Data ke	C1	C2	C3	Minimum	cluster	ket
1	2490242,883	3262142	4185730	2490243	0	aman
2	5560380,008	6684705	7981791	5560380	0	aman
3	5560976,133	6685373	7982577	5560976	0	aman
4	5510025,383	6629545	7921495	5510025	0	aman
5	5092845,133	6170756	7418960	5092845	0	aman
6	5470581,133	6586213	7874075	5470581	0	aman
7	4582369,383	5607501	6800063	4582369	0	aman
8	4803160,633	5852253	7069451	4803161	0	aman
9	4495348,758	5511604	6694656	4495349	0	aman
10	2550857,758	3329692	4263132	2550858	0	aman
11	5531440,758	6652975	7946839	5531441	0	aman
12	5495193,883	6613136	7903652	5495194	0	aman
13	54908,50781	0	58460	0	1	aman
14	5526406,508	6647096	7940746	5526407	0	aman
15	5525660,758	6646475	7940341	5525661	0	aman
16	5458089,383	6572107	7858705	5458089	0	aman
17	5489776,508	6607105	7897075	5489777	0	aman
18	222440,3828	58460	0	0	2	aman

Gambar 4.18 Hasil akhir cluster

Ada bukti bahwa, baik dalam pengolahan clustering K-MEANS secara manual menggunakan program Bahasa Python dan aplikasi Microsoft Excel, keakuratan hasil akhir tidak bergantung pada bagaimana clustering K-MEANS diproses, melainkan pada nilai random yang ditetapkan sebelum proses tersebut.

4.1.7 Hasil Clustering K-MEANS

Dengan menggunakan algoritma K-MEANS, hasilnya dipisahkan menjadi tiga kelompok : sedikit, sedang, dan terbanyak. Dari hasil yang diklasifikasikan sebagai output dari notebook jupyter menunjukkan data klaster yang sama. Hasil dari klastering menunjukkan bahwa proses telah berjalan sesuai dengan perkiraan penelitian. Hasil clustering detail akan dijelaskan sebagai berikut:

- Kategori sedikit

Menurut hasil klastering yang telah dilakukan, klaster pertama memiliki 816 record dan memiliki anggota terbanyak. Namun, klaster ini memiliki nilai yang lebih sedikit daripada rata-rata keseluruhan data yang diteliti, dan karena tingkat kunjungan yang sedikit dalam klaster ini, nilai data dalam klaster ini berkisar antara 2-232. Jadi, klaster kesatu dikategorikan pada situs web dengan jumlah kunjungan paling sedikit dibandingkan dengan klaster lainnya. Terdapat beberapa website informasi, seperti profil, direktori, sejarah, dan lainnya, yang dikunjungi hanya beberapa kali. Website-website ini termasuk dalam cluster kesatu, yang memiliki tingkat kunjungan tersedikit.

Pada tabel 4.4 berikut beberapa Contoh data jumlah kunjungan sedikit cluster satu:

Tabel 4.4 Contoh hasil cluster satu dengan banyak kunjungan sedikit

Akses Web	Cluster	Jumlah	Kategori
GET /index.php/profil HTTP/1.1	0	8	Profil
GET/index.php/jurusan.html HTTP/1.1	0	11	Jurusan
GET /index.php/profil/lambang-dan- makna HTTP/1.1	0	9	Lambang dan makna
GET /index.php/profil/sejarah HTTP/1.1	0	21	Sejarah
GET /index.php/profil/visi-misi-dan- tujuan HTTP/1.1	0	15	Visi misi dan tujuan
GET /direktori.html HTTP/1.1	0	2	Direktori

Cluster kesatu terdiri dari beberapa website informasi yang sangat jarang dikunjungi, seperti yang ditunjukkan pada tabel 4.4 di atas. Meskipun termasuk dalam cluster kesatu, informasi yang diperoleh dari website-website ini berdampak pada profil pengguna, seperti yang ditunjukkan oleh beberapa submenu profil, visi misi, dan tujuan. Cluster kesatu, yang mengandung 816 data, mencakup semua kategori konten yang sudah ada sebelumnya.

Pada tabel 4.5 terdapat kategori konten web yang ada pada cluster satu.

Tabel 4.5 Jenis cluster satu

Kategori	Jumlah
Profil	8
Sejarah	21
Visi misi dan tujuan	15
Manajemen	14
Program kerjasama	9
Lambang dan makna	9
Jurusan	11
Jurusan Budidaya tanaman perkebunan	19
Evaluasi layanan	4
Direktori	2
Jurusan teknologi pertanian	232
jurusan peternakan	163
Jurusan budidaya tanaman pangan	98
Jurusan ekonomi dan bisnis	211

- **Kategori Sedang**

Klaster ketiga memiliki 1505 rekor anggota, dengan nilai hasil 744–764. Karena nilainya sedikit lebih terbanyak daripada rata-rata klaster, nilai ini menunjukkan bahwa tingkat kunjungan anggota klaster ketiga sedang. Oleh karena itu, klaster ketiga dianggap memiliki tingkat kunjungan yang sedang.

Beberapa bagian dari cluster ketiga, yang memiliki banyak tingkat kunjungan yang relatif sedang, berisi informasi tentang akses ke submenu kegiatan dan agenda. Tabel 4.6 menunjukkan contoh konten website dari cluster tiga:

Tabel 4.6 Contoh hasil dari cluster kedua yang memiliki jumlah kunjungan sedang

Akses Web	cluster	Jumlah	Kategori
GET /index.php/publikasi/kegiatan.html HTTP/1.1	2	744	Kegiatan
GET /index.php/publikasi/agenda HTTP/1.1	2	764	Agenda

Kategori konten yang ada di cluster ketiga ditunjukkan dalam tabel 4.7:

4.7 Jenis Konten Dalam Cluster Ketiga

Kategori	Jumlah
Kegiatan	744
Agenda	764

- Kategori Terbanyak

Hasil klaster kedua menunjukkan bahwa anggota klaster kedua memiliki 5250 catatan dan memiliki nilai data terbanyak. Nilai yang dihasilkan oleh klaster kedua berada di antara 2510 dan 2740, menunjukkan bahwa klaster kedua memiliki nilai yang jauh melebihi rata-rata. Akibatnya, klaster kedua dianggap memiliki tingkat kunjungan terbanyak, seperti yang ditunjukkan pada tabel 4.8.

Tabel 4.8 Contoh hasil cluster kedua yang memiliki jumlah kunjungan terbanyak

Akses Web	Cluster	Jumlah	Kategori
GET /index.php/publikasi/pengumuman/219-pengumuman-final-seleksi-cpns-polinela-2015 HTTP/1.1	1	2510	Pengumuman

GET /index.php/publikasi/berita/227-jambore-nasional-fkmpi-2015-polinela.html HTTP/1.1	1	2740	Berita
--	---	------	--------

Tabel 4.9 menunjukkan kategori website yang ada di cluster kedua.

Tabel 4.9 Jenis konten dalam cluster kedua

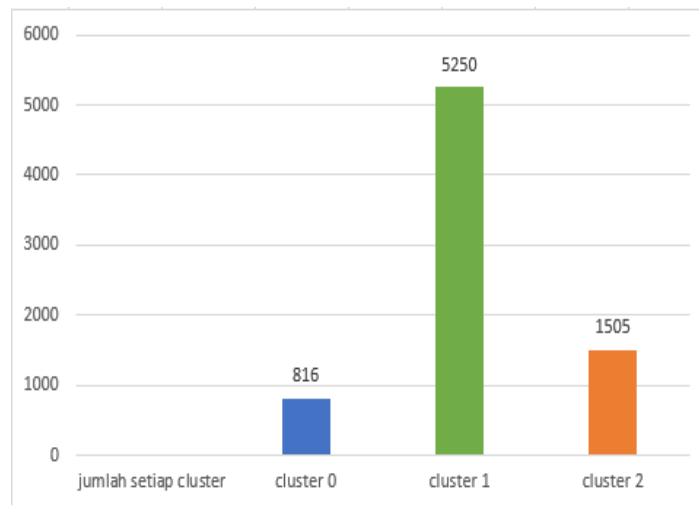
Kategori	Jumlah
Berita	2740
Pengumuman	2510

Tabel 4.10 berikut menunjukkan jumlah setiap cluster yang dibuat dengan algoritma KMeans:

Tabel 4.10 Jumlah setiap kelompok cluster

Jumlah pada	0	1	2	Total
Cluster	816	5250	1505	7574

Data yang digunakan selama proses klastering tidak memiliki nilai yang hilang, seperti yang ditunjukkan dalam hasil tabel 4.10, di mana total data ditunjukkan sesuai dengan data sebelum klastering. Jika ditampilkan dalam bentuk diagram, hasil clustering K-MEANS di atas akan digunakan untuk membantu proses cyber profiling, seperti yang ditunjukkan pada gambar 4.11 berikut.



Gambar 4.11 Menggambarkan hasil cluster

4.2 Hasil Analisis

Hasil di atas menggambarkan algoritma K-MEANS memenuhi harapan dalam penelitian ini, karena disarankan untuk digunakan pada data yang besar Maulana, R., & Dwiyani, N. (2018). Berdasarkan pola banyak pengunjung, algoritma KMeans berhasil mengelompokkan data menjadi kategori terbanyak, sedang, dan sedikit. Kemudian, hasil penelitian sebanding dengan pengelompokan data didasarkan pada waktu akses dan konten website.

Hasil perbandingan tersebut mencapai kesimpulan yang sama, yang menunjukkan bahwa penggunaan algoritma K-MEANS dalam penelitian ini telah berhasil. Kemudian, hasil penelitian tersebut sebanding dengan pengelompokan data berdasarkan waktu akses dan konten website. Hasil perbandingan tersebut mencapai kesimpulan yang sama, yang menunjukkan bahwa penggunaan algoritma K-MEANS dalam penelitian ini telah berhasil.

Hasil K-MEANS Clustering menunjukkan bahwa web dengan submenu berita, ini adalah bagian dari konten yang diakses paling sering oleh pengguna internet. Pencarian submenu direktori, informasi evaluasi layanan, profil, sejarah, visi misi dan tujuan, manajemen, lambing dan makna, program kerjasama, menu jurusan, sub menu Jurusan budidaya tanaman pangan, Jurusan Budidaya tanaman perkebunan, jurusan ekonomi dan bisnis, jurusan peternakan, jurusan teknologi

pertanian. Seperti yang ditunjukkan oleh hasil cluster kesatu, ini adalah kelompok website dengan tingkat kunjungan yang lebih sedikit dalam kategori sedikit. Tingkat kunjungan terbanyak seperti submenu berita dan pengumuman memiliki tingkat kunjungan terbanyak, seperti yang ditunjukkan oleh hasil cluster kedua. Ini menunjukkan bahwa pengguna selalu mencari berita dan pengumuman terbaru. Tingkat kunjungan kategori sedang seperti submenu kegiatan dan agenda memiliki tingkat kunjungan sedang, seperti yang ditunjukkan oleh hasil cluster ketiga.

Berdasarkan waktu akses Hasil kategori waktu akses menunjukkan bahwa penggunaan Internet meningkat pada sore-malam dan malam-pagi. Hal ini disebabkan oleh fakta bahwa waktu tersebut adalah waktu yang paling sering dihabiskan oleh pengguna untuk melakukan aktivitas yang memerlukan akses ke informasi di website Polinela. Pada kategori yang didasarkan pada waktu, website dengan submenu berita menjadi yang paling sering dikunjungi oleh pengguna, diikuti oleh submenu Pengumuman. Karena waktu produktif dan perbedaan waktu istirahat beraktivitas pengguna akses konten menyebabkan sudah mulai berkurang konten yang dikunjungi pada pagi-siang dan siang-sore mulai berkurang. Dalam kategori waktu ini, website dengan submenu direktori menjadi yang paling sedikit dikunjungi oleh pengguna, diikuti oleh submenu evaluasi layanan.

Hasil penelitian Zulfadhilah, M. (2016) dan Aji, RP (2022). Secara psikografis, orang yang menggunakan penelitian ini menganggap internet sebagai kelompok yang terdiri dari individu yang menggunakan Internet untuk tujuan informasi. Dalam konteks ini, pengguna internet dikategorikan dalam kelompok NetWorker, yang secara logis oleh pengguna terhubung ke internet menggunakan jejaring sosial, melakukan transaksi online, dan memperluas jaringannya. Mereka juga dapat termasuk dalam kelompok NetJungki, yang menggunakan Internet untuk berhubungan baik dengan komunitas online mereka.

Untuk menampilkan hasil pengunjung web ini termasuk dalam kategori networker dan netjungki seperti pada tabel 4.11

Tabel 4.11 Networker dan netjungki kategori

Kategori	Jumlah
Profil	8
Sejarah	21
Visi misi dan Tujuan	15
Manajemen	14
Lambang dan makna	9
program kerjasama	9
Jurusan	11
Jurusan budidaya tanaman pangan	98
Jurusan budidaya tanaman perkebunan	19
Jurusan teknologi pertanian	232

Hasil dari anilisa profiling ini dapat membantu pengembang web mengevaluasi perilaku pengguna untuk meningkatkan minat pengunjung website. Namun, mengenai ancaman kejahatan seperti yang disebutkan oleh Aji, R. P. (2022), hasil analisis data menunjukkan bahwa tidak ada ancaman karena data yang dikumpulkan mengenai konten website yang pengguna akses untuk cyber profiling.

Hasil penelitian menunjukkan bahwa istilah "cyber profiling" digunakan dalam penelitian ini untuk mengidentifikasi informasi tentang karakteristik pengguna Internet berdasarkan website yang telah dikunjungi, lokasi pengguna saat mengunjungi website, dan lingkungan tempat pengguna mengakses Internet.