

BAB IV

IMPLEMENTASI SISTEM DAN ANALISIS HASIL

Bagian ini berisi tentang pelaksanaan dan pengujian hasil dari kerangka kerja dengan perhitungan yang telah digunakan dan rencana antarmuka.

4.1 Identifikasi Masalah

Berdasarkan latar belakang masalah pada Bab 1 maka dapat diidentifikasi permasalahan yang diteliti yaitu:

- a. Bagaimana menghasilkan tiap-tiap data jama'ah umroh, terendah dan tertinggi dari jumlah jama'ah dari jenis Kelamin, daerah, Pekerjaan, dan Paket
- b. Bagaimana menghasilkan data baru dari clustering Daerah dan Paket (Bulan keberangkatan), sehingga menjadi acuan promosi pada bulan berapa dan daerah yang tepat untuk target marketing.

4.2 Studi Literatur

Step awal dari penelitian ini adalah dengan mencari dan mempelajari masalah yang ada pada data Jama'ah Umroh yang diperoleh dari PT. Nur Haramain Mulia Probolinggo, Jawa Timur. Data yang digunakan ialah data tahun 2019. Kemudian menganalisa masalah yang akan ditulis dilatar belakang penelitian, membuat tujuan pada penelitian menyampaikan manfaat penelitian dan membuat batasan pada penelitian. Adapun langkah langkahnya adalah sebagai berikut:

- a. Melakukan identifikasi masalah yang terjadi sehingga hasilnya bisa di implementasikan pada latar belakang masalah.
- b. Melakukan pengkajian pada latar belakang masalah agar bisa menghasilkan tujuan dan manfaat penelitian.

4.3 Pengelohan Data

Dengan melakukan pengelohan data sehingga data yang telah tertera pada bab akan kita ubah menjadi data seperti yang diatas, untuk awal pengelolahan data kita menggunakan cara manua sehingga data-data tersebut telah diubah menjadi numerik. Menentukan Data yang akan di olah pada tahap ini data yang akan di olah hanyadata. Data Alamat dan Paket tujuannya untuk menghasilkan data terbaru pada bulan dan daerah mana Jama'ah terbanyak.

L/P	ALAMAT	PEKERJAA	PAKET
2	15	1	19
1	3	2	19
2	3	2	19
1	8	9	19
2	8	4	19
2	15	4	19
1	8	5	19
2	8	7	19
2	30	14	19
1	8	1	19
2	8	4	19
1	15	5	19

Gambar 4. 1 Data Set Pekerjaan Jama'ah Umroh csv

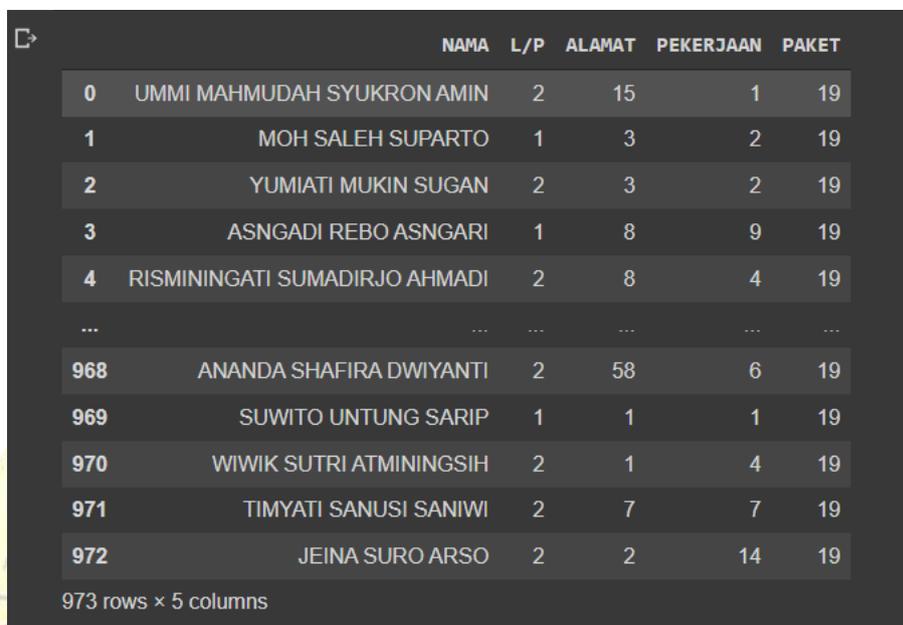
Dengan melakukan pengelohan data sehingga data yang telah tertera pada bab akan kita ubah menjadi data seperti yang diatas, untuk awal pengelolahan data kita menggunakan cara manua sehingga data-data tersebut telah diubah menjadi numerik.

4.4 Data Selection

Penentuan informasi atau pilihan informasi yang berlaku untuk pemeriksaan yang akan diperoleh dari indeks informasi saat ini. Berikutnya adalah cara yang paling umum untuk mentransfer dataset. Ilustrasi eksekusi dapat ditemukan pada gambar 4.1 berikut.

```
data = pd.read_csv("DATA_NEW.csv")
data.head(973)
```

Segment Program 4. 1 Data seleksi melalui google colaboratory



	NAMA	L/P	ALAMAT	PEKERJAAN	PAKET
0	UMMI MAHMUDAH SYUKRON AMIN	2	15	1	19
1	MOH SALEH SUPARTO	1	3	2	19
2	YUMIATI MUKIN SUGAN	2	3	2	19
3	ASNGADI REBO ASNGARI	1	8	9	19
4	RISMININGATI SUMADIRJO AHMADI	2	8	4	19
...
968	ANANDA SHAFIRA DWIYANTI	2	58	6	19
969	SUWITO UNTUNG SARIP	1	1	1	19
970	WIWIK SUTRI ATMININGSIH	2	1	4	19
971	TIMYATI SANUSI SANIWI	2	7	7	19
972	JEINA SURO ARSO	2	2	14	19

973 rows x 5 columns

Gambar 4. 2 Hasil Dataset

4.5 K-Means Clustering

Algoritma Clustering melakukan pengelompokan pada cluster berdasarkan kesamaan data yang memiliki kemiripan karakter, Data-data yang memiliki karakteristik (*similarity*) antara satu data dengan data yang lain pada dataset umroh. Proses pengelompokan dengan menggunakan metode *K-Means Clustering* dengan tiga *cluster*. Sehingga didapatkan hasil yang tidak seimbang atau memiliki selisih jumlah data yang sangat jauh pada setiap *cluster*.

```
km = KMeans(n_clusters=3)
km
```

Segment Program 4. 2 Kmeans 3 clusters

Proses clustering menggunakan 3 *clusters*, KM sama dengan data pada 3 *cluster* maka hasil dari sebuah table dapat di ambil dari Km setelah dianalisi menggunakan *elbow*.

	NAMA	L/P	ALAMAT	PEKERJAAN	PAKET	cluster
0	UMMI MAHMUDAH SYUKRON AMIN	2	15	1	19	0
1	MOH SALEH SUPARTO	1	3	2	19	1
2	YUMIATI MUKIN SUGAN	2	3	2	19	1
3	ASNGADI REBO ASNGARI	1	8	9	19	1
4	RISMININGATI SUMADIRJO AHMADI	2	8	4	19	1
...
968	ANANDA SHAFIRA DWIYANTI	2	58	6	19	2
969	SUWITO UNTUNG SARIP	1	1	1	19	1
970	WIWIK SUTRI ATMININGSIH	2	1	4	19	1
971	TIMYATI SANUSI SANIWI	2	7	7	19	1
972	JEINA SURO ARSO	2	2	14	19	1

973 rows x 6 columns

Gambar 4.3 Implementasi - *K-Means* dengan tiga cluster

Tabel 4.1 Jumlah Cluster

Cluster	Jumlah
0	307
1	499
2	167

4.6 Hasil Analisis

Implementasi yang telah dijelaskan di bab 4 membantu analisis terhadap pengelompokan untuk memetakan potensi lokasi daerah analisis dilakukan terhadap 972 data pada tahun 2019. Dalam pengelompokan ini, yang menjadi atribut adalah jumlah daerah, paket dan pekerjaan. Data tersebut akan diolah menggunakan K-Means clustering. Hasil penelitian menggunakan data produksi di daerah kab probolinggo dan luar kota.

a. Analisa sistem lama

Setelah melakukan observasi dan wawancara pada bagian bidang Marketing data yang ada pada kantor tidak digunakan untuk melakukan pemanfaatan di bidang marketing selama ini yang masih bersifat manual sehingga petugas kantor sangat kesulitan dalam menemukan lokasi yang tepat untuk

Adapun kekurangan sistem yang bersifat manual antara lain :

- a. Data hanya sebagai view dan tidak dikelola menjadi informasi
- b. Proses Promosi tanpa menggunakan data
- c. Marketing yang berjalan dari semua jama'ah

Sementara disisi lain seiring berjalannya waktu PT. Nur haramain, semakin bertambah jumlah jama'ahnya sehingga harus menggunakan sebuah sistem baru yang terkomputerisasi dengan baik agar pekerjaan semakin mudah dan baik.

b. Analisa sistem baru

Setelah meneliti sistem lama peneliti mengupayakan untuk menanggulangi masalah yang ada dengan membuat sistem baru yaitu "Penerapan Data Mining untuk segmentasi Jama'ah PT. Nurul Haramain Mulia sebagai Rekomendasi Promosi"

Pembuatan sistem baru ini untuk lebih mempermudah bagian marketing PT.Nur Haramain Mulia dalam melakukan proses data jama'ah dalam melakukan pengambilan keputusan promosi. Dimana dalam proses pengelolaan data nantinya akan menghasilkan pengetahuan yang bisa digunakan oleh staff marketing, adapun proses data antara lain adanya data daerah dan data paket keberangkatan pada bula tertentu, input data jumlah keberangkatan. Dengan adanya proses data semoga bisa membantu staff marketing dalam mengambil keputusan daerah dan pada bulan berapa promosi akan diterbitkan.

4.7 Analisis Hasil

Implementasi yang telah dijelaskan di bab 4 membantu analisis terhadap pengelompokan untuk memetakan potensi hasil jama'ah umrah di Kab. Probolinggo dan diluar Kab. Probolinggo, Analisis dilakukan terhadap 900 data pada tahun 2019.

4.8 Dataset Alamat

Hasil Perhitungan pada data alamat total data yang diperoleh sebanyak 862 data dan data paling banyak yaitu 54 data daerah paling banyak adapun daerah paling sedikit yaitu 1 data.

```
data['ALAMAT'].describe()
```

Segment Program 4. 3 Describe Alamat

```
count    973.000000
mean     17.130524
std      15.120722
min       1.000000
25%       6.000000
50%      12.000000
75%      25.000000
max       60.000000
Name: ALAMAT, dtype: float64
```

Gambar 4. 4 Perhitungan Data Rata-Rata Describe Alamat

Tabel 4. 2 Data rata-rata Alamat

Keterangan Data	Hasil Data
Count (menghitung seluruh data)	873 Data
Max (data maksimal)	60 Data
Min (data Minimal)	1 Data
Std (standard Deviation)	15 Data
50 %	12Data

Hasil Alamat terbanyak dan Alamat Terendah pada tahap ini memilih 10 data terbanyak dan 10 data yang sedikit, pada gambar di bawah terdapat dua baris angka yang pertama adalah Inansial Data dan pada sebelah kanan hasil data yang di peroleh.

```
▶ data['ALAMAT'].value_counts()
```

Segment Program 4. 4 Count Alamat

```

8      157
15     73
1      69
30     65
42     58
17     44
5      43
2      40
3      38
14     38
12     35
4      31
6      25
7      24
16     24
9      23
44     23
20     15
21     15
41     15
25     14
24     12
43     11
10     11
45     10
50     9
54     9
53     6
55     5
23     4
11     3
46     3
58     3
57     3
13     3
51     3
60     2
47     2
52     1
49     1
56     1
48     1
59     1
Name: ALAMAT, dtype: int64

```

Gambar 4. 5 Menghitung Dataset Alamat

Tabel 4. 3 Data tertinggi alamat

No	Insalisasi	Alamat	Hasil Data
1	8	Kraksaan	157 data
2	15	Paiton	73 data
3	1	Probolinggo Kota	69 data
4	30	Pulau Madura	65 data
5	42	Situbondo	58 Data
6	17	Pajarakan	44 Data

7	5	Gading	43 Data
8	2	Banyuanyar	40 Data
9	3	Besuk	38 Data
10	14	Maron	38 Data
11	12	Leces	35 Data
12	4	Dringu	31 Data
13	6	Gending	25 data
14	7	Kotaanyar	24 data
15	16	Pakuniran	24 data
16	9	Krejengan	23 data
17	44	Sidoarjo	23 data
18	20	Sumberasih	15 data
19	21	Tegalsiwalan	15 data
20	41	Jember	15 data

Tabel 4. 4 Data terendah alamat

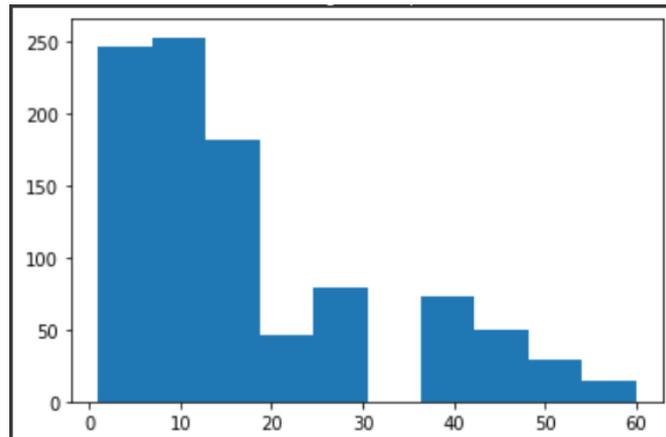
No	Insalisasi	Alamat	Hasil Data
1	25	Bantaran	14 Data
2	24	Wonomerto	12 Data
3	43	Surabaya	11 Data
4	10	Krucil	11 Data
7	45	Bondowoso	10 Data
8	50	Kalimantan	9 Data
9	54	Lumajang	9 Data
10	53	Jakarta	6 Data

11	55	Malang	5 Data
12	23	Tongas	4 Data
13	11	Kuripan	3 Data
14	46	Gresik	3 Data
15	58	Tuban	3 Data
16	57	Trenggalek	3 Data
17	13	Lumbang	3 Data
18	51	Pasuruan	3 Data
19	60	Nganjuk	2 Data
20	47	Blitar	2 Data
21	52	Denspansar	1 Data
22	49	Ambon	1 Data
16	56	Klaten	1 Data
17	48	Bojonegoro	1 Data
18	59	Bantul	1 Data

```
plt.hist(data['ALAMAT'])
(array([246., 253., 182., 46., 79., 0., 73., 50., 29., 15.]),
 array([ 1., 6.9, 12.8, 18.7, 24.6, 30.5, 36.4, 42.3, 48.2, 54.1, 60. ]),
 <a list of 10 Patch objects>)
```

Segment Program 4. 5 Plot data Alamat

Pada hasil list segment program 4.5 menghasilkan data berikut: 10 grafik data, 246 data score 1 253 data score 253 pada grafik data 6 dan 9, 182 data score 12 dan 8, data score 46 inisial 18 dan 7.



Gambar 4. 6 Grafik Dataset Alamat

Tampilan histogram pada data Alamat dari angka 1-19 data meningkat pada 20 data mengalami penurunan data.

4.9 Dataset Paket

Pengelompokan dataset pada paket meperdiksi pada bulan berapa yang akan dilakukan promosi untuk jama' ah umroh, bagi kantor di perlu sulit-sulit untuk menentukan pada bulan berapa akan dilaksanakan promosi besar-besaran.

```
▶ data['PAKET'].value_counts()
```

Gambar 4. 7 Count Paket

```

9      129
12     90
14     82
19     68
11     61
3       60
5       60
2       56
10     53
17     45
16     44
18     41
15     37
4       36
1       33
6       25
13     22
8       17
7       14
Name: PAKET, dtype: int64

```

Gambar 4. 8 count dataset paket

Tabel 4. 5 Data Paket tertinggi

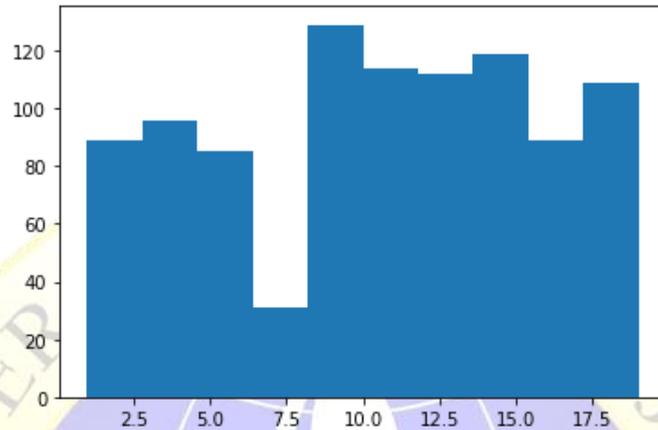
Insalisasi	Paket	Jumlah Data
9	Maulid 21 Hari Madura	129
12	Maulid 13 Hari	90
14	Maulid 10 Hari	82
19	Promo Jan-Feb 10Hari 02 FEB	68
11	Maulid 16 Hari	61
3	Ramadhan I'Tikaf (Tr)	60
5	Rajab 14 Hari	60
2	Ramadhan I'Tikaf (Jt)	56
10	Maulid Promo	53
17	Promo Jan-Feb 12 Hari 16 Januari	45

Tabel 4. 6 Dataset Paket Terendah

Insalisasi	Paket	Jumlah Data
16	Promo Jan-Feb 19 Hari 24 Januari	44
18	Maulid 13 Hari	41
15	Maulid 10 Hari	37
4	Promo Jan-Feb 10H 02 FEB	36
1	Ramadhan Reguler Akhir	33
6	Rajab 10 Hari	25
13	Maulid 10 Hari Reguler	22
8	Rajab Promo Reguler 13 Hari	17
7	Rajab Reguler 13 Hari	14

```
plt.hist(data['PAKET'])  
  
(array([ 89., 96., 85., 31., 129., 114., 112., 119., 89., 109.]),  
 array([ 1. , 2.8, 4.6, 6.4, 8.2, 10. , 11.8, 13.6, 15.4, 17.2, 19. ]),  
 <a list of 10 Patch objects>)
```

Segment Program 4. 6 Plot data paket



Gambar 4. 9 Grafik Dataset Paket

Tampilan histogram pada data paket tingkatan histogram pada paket sangatlah normal pada histogram 7.5 mengalami penurunan dan peningkatan pada data 9 total data set jama mencapai 129 Jama'ah.

4.10 Dataset Jenis Kelamin

Pada tahap perhitungan data keseluruhan untuk jenis kelamin yang di hasilkan Wanita 556 Jamah'ah Umroh yang berangkat pada tahun 2019.

```
data['L/P'].value_counts()
```

Segment Program 4. 7 count Jenis Kelamin

```
2    556  
1    417  
Name: L/P, dtype: int64
```

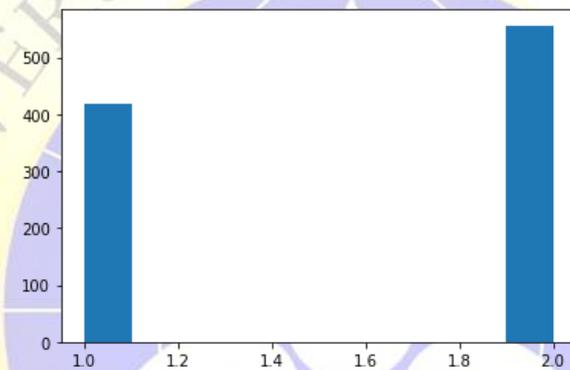
Gambar 4. 10 Perhitungan Data Rata-Rata Pada Jenis kelamin

Tabel 4. 7 Data Jenis Kelamin

Insalisasi	Keterangan Data	Hasil Data
1	Pria	417 Data
2	Wanita	556 Data

```
plt.hist(data['L/P'])  
(array([417., 0., 0., 0., 0., 0., 0., 0., 0., 556.]),  
 array([1. , 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2. ]),  
 <a list of 10 Patch objects>)
```

Segment Program 4. 8 Plot data Jenis Kelamin



Gambar 4. 11 Grafik Dataset jenis Kelamin

Pada data Jenis kelamin cukuplah singkat hanya pria dan wanita dataset terbanyak adalah dataset pada wanita dan yang kedua pria, pada dataset Jenis kelamin kita bisa mengetahui jumlah jama'ah haji yang terdaftar lebih banyak wanita.

4.11 Dataset Pekerjaan

Pengelompokan dataset pada pekerjaan jama'ah umroh meperdiski pekerjaan apa saja yang rata-rata berangkat Umroh disini kita bisa mengetahui pekerjaan A jumlah jama'ah terbanyak dan ada juga data jama'ah yg tidak terdaftar sebagai jama'ah umrah

```
data['PEKERJAAN'].value_counts()
```

Segment Program 4. 9 count dataset pekerjaan

```
data['PEKERJAAN'].value_counts()
4      236
2      181
5      125
7       83
1       74
8       56
6       49
14      40
90      25
3       24
9       23
11      9
12      7
15      7
10      5
16      4
17      4
20      4
25      3
13      2
19      2
21      2
23      2
18      2
27      1
22      1
24      1
26      1
Name: PEKERJAAN, dtype: int64
```

Gambar 4. 12 Dataset Pekerjaan.

Tabel 4. 8 Data Pekerjaan Tertinggi

No	Insalisasi	Pekerjaan	Jumlah Data
1	4	MRT	236
2	2	Petani/Berkebun	181
3	5	Wiraswasta	125
4	7	Pedagang	83
5	1	PNS	74
6	8	Karyawan Swasta	56
7	6	Mahasiswa/Pelajar	49

8	14	Tidak Bekerja	40
9	90	Tidak ada data	25
10	3	Guru	24
11	9	Pensiunan	23
12	11	Dosen	9
13	12	Nelayan	7
14	15	Karyawan BUMN	7

Tabel 4. 9 Data Pekerjaan Terendah

No	Insalisasi	Pekerjaan	Jumlah Data
1	10	Ustad/Mubaligh	5
2	16	Buruh Harian Lepas	4
3	17	Perangkat Desa	4
4	20	Polisi	4
5	25	Supir/Transportasi	3
6	13	TNI	2
7	19	Industri	2
8	21	Karyawan Honorer	2
9	23	Kontruksi	2
10	18	Kepala Desa	2
11	27	Tukang Batu	1
12	22	Dokter	1
13	24	MRS	1
14	26	Mekanik	1

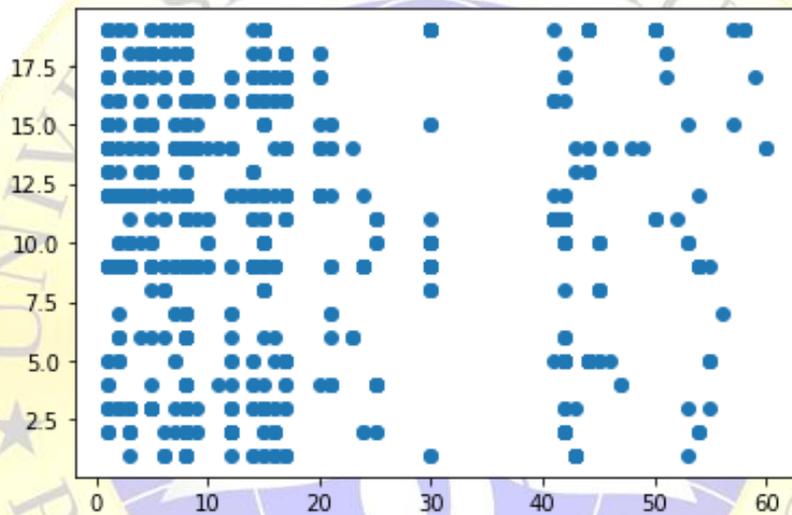
4.12 Proses Clustering

Pada tahap proses data kita melakukan data set objek dengan menggunakan clustering dan proses implementasi pada objek yang telah ditentukan.

```
[45] plt.scatter(data['ALAMAT'],data['PAKET'])
```

```
<matplotlib.collections.PathCollection at 0x7fe8554d9a90>
```

Segment Program 4. 10 Scatter objek awal



Gambar 4. 13 Objek Awal Pada Clustering

Pada Segmentasi Program di atas data plt scatter dat Alamat dan paket tapi pada tahap 1 data masih belum di cluster hanya menampilkan titik data pada alamat dan paket, tahap selanjutnya proses pembagian kepada 3 cluster.

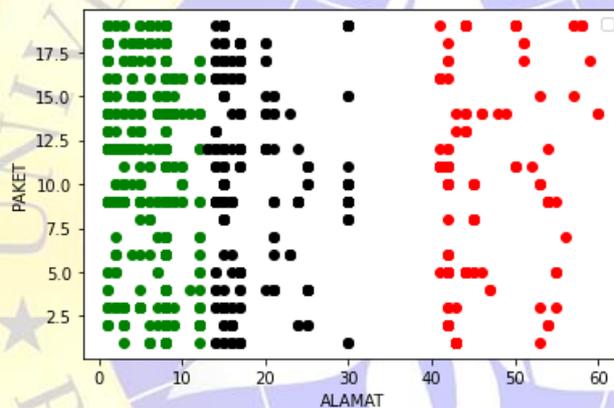
```
▶ data1 = data[data.cluster==0]
data2 = data[data.cluster==1]
data3 = data[data.cluster==2]

plt.scatter(data1.ALAMAT,data1['PAKET'],color='green')
plt.scatter(data2.ALAMAT,data2['PAKET'],color='red')
plt.scatter(data3.ALAMAT,data3['PAKET'],color='black')

plt.xlabel('ALAMAT')
plt.ylabel('PAKET')
plt.legend()

⌘ No handles with labels found to put in legend.
<matplotlib.legend.Legend at 0x7fe85e1eaad0>
```

Segment Program 4. 11 Scatter cluster 1,2,3



Gambar 4.13 Prosesing Menggunakan K-Means Label Clustering

Tahap proses membagi data menjadi 3 cluster atau 3 kelompok Label X Alamat dan L Paket jarak antara 1 dan 2 cluster sangat berdekatan pada cluster 3 data berjarak jauh dan cluster 1 data yang terlihat lebih banyak dari pada data cluster 2 dan 3

```

k_rng = range(1,10)
sse = []
for k in k_rng:
    km = KMeans(n_clusters=k)
    km.fit(data[['ALAMAT', 'PAKET']])
    sse.append(km.inertia_)

```

Segment Program 4. 12 Proses Range 1-10

Pada tahap selanjutnya, mencoba menggunakan teknik elbow yang digunakan untuk membuat data dalam menentukan jumlah kelompok terbaik dengan melihat akibat dari korelasi antara jumlah tandan yang membentuk titik siku pada suatu titik.

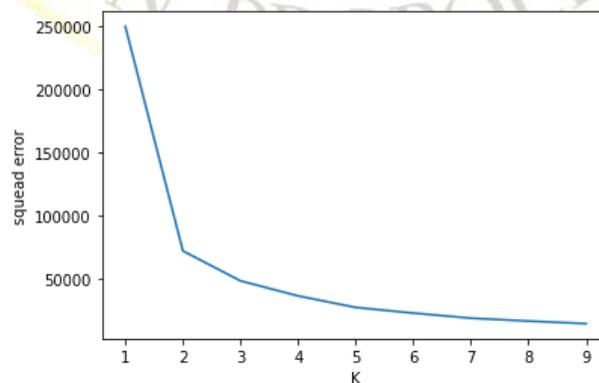
```

[15] sse
[249599.35662898252,
 71753.08840569595,
 48062.66753280694,
 36155.398552895174,
 27020.12564128213,
 22570.218656612436,
 18524.326733359783,
 16247.785814932618,
 14213.723811490087]

```

Gambar 4. 14 Hasil SSE (Sum of Square Error)

Untuk mendapatkan perbandingannya adalah dengan menghitung *SSE* (*Sum of Square Error*) dari masing-masing nilai cluster. Karena semakin besar jumlah cluster K maka nilai *SSE* akan semakin kecil. Rumus *SSE* pada *K-Means*



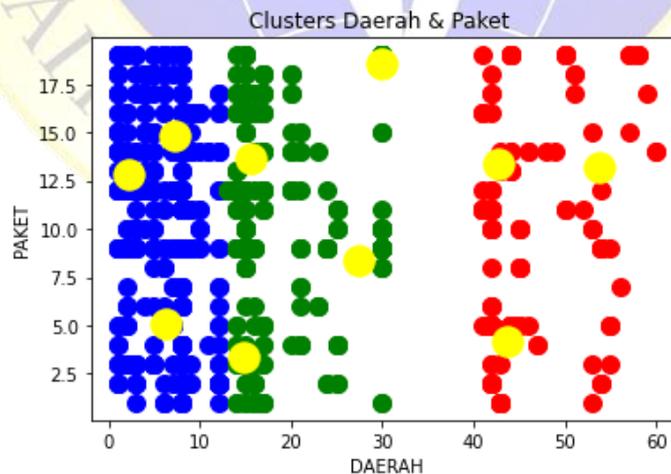
Gambar 4. 15 Grafik SSE

Hasil presentase yang berbeda dari setiap nilai tandan dapat ditampilkan dengan menggunakan grafik sebagai sumber data. Setelah melihat bahwa akan ada beberapa kualitas K yang mengalami penurunan terbaik dan kemudian efek samping dari nilai K secara bertahap akan berkurang sampai konsekuensi dari nilai K stabil. Misal nilai grup K=1 menjadi K=2, maka pada titik tersebut dari K=3 terdapat pengurangan yang tidak biasa karena titik pada pemandu K=1 menuju K=2 maka nilai ideal grup k adalah K=3.

```
[ ] plt.scatter(X[y_kmeans == 0, 0], X[y_kmeans == 0, 1], s = 100, c = 'red', label = 'Cluster 1')
plt.scatter(X[y_kmeans == 1, 0], X[y_kmeans == 1, 1], s = 100, c = 'blue', label = 'Cluster 2')
plt.scatter(X[y_kmeans == 2, 0], X[y_kmeans == 2, 1], s = 100, c = 'green', label = 'Cluster 3')
plt.scatter(kmeans.cluster_centers[:, 0], kmeans.cluster_centers[:, 1], s = 300, c = 'yellow', label = 'Centroids')
plt.title('Clusters Daerah & Paket ')
plt.xlabel('DAERAH')
plt.ylabel('PAKET')
# plt.legend()
plt.show()
```

Segment Program 4. 13 Clustering & Titik Centroid

Clustering termasuk dalam klasifikasi tanpa pengawasan (*unsupervised classification*). Pengertian clustering adalah cara paling umum untuk mengumpulkan atau mengkarakterisasi objek berdasarkan data yang diperoleh dari informasi yang menggambarkan hubungan antar objek dengan standar meningkatkan kemiripan antar individu dari satu kelas dan membatasi kemiripan antar kelas/kelompok.



Gambar 4. 16 Clustering & Titik Centroid