

# MAKALAH PEMODELAN MATEMATIKA

## MENENTUKAN STATUS KEAMANAN DI *MY CITY* MENGUNAKAN *K-MEANS CLUSTERING*



Kelompok 7

Nadya Tjindra	2016710001
Leo Gunawan	2016710003
Geraldry Suryahartanto	2016710013
Yonathan Jeremy Budiman	2016710023
Al-Vinda Tania N.A.P	2016710025
Felix Tandiono	2016710028

20 September 2019

PROGRAM STUDI MATEMATIKA  
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS  
UNIVERSITAS KATOLIK PARAHYANGAN  
2019



## DAFTAR ISI

<b>DAFTAR ISI</b>	<b>iii</b>
<b>DAFTAR GAMBAR</b>	<b>v</b>
<b>DAFTAR TABEL</b>	<b>vii</b>
<b>1 PENDAHULUAN</b>	<b>1</b>
1.1 Latar Belakang . . . . .	1
1.2 Rumusan Masalah . . . . .	1
1.3 Tujuan . . . . .	2
1.4 Manfaat . . . . .	2
1.5 Batasan masalah . . . . .	2
1.6 Metodologi . . . . .	2
1.7 Sistematika Pembahasan . . . . .	3
<b>2 DASAR TEORI</b>	<b>5</b>
2.1 <i>Z-score</i> . . . . .	5
2.2 <i>K-means Clustering</i> . . . . .	5
<b>3 HASIL DAN PEMBAHASAN</b>	<b>7</b>
3.1 Visualisasi Data . . . . .	7
3.1.1 Tabel Frekuensi . . . . .	7
3.1.2 <i>Mosaic Plot</i> . . . . .	8
3.1.3 <i>Heatmap</i> . . . . .	9
3.2 Pengolahan data . . . . .	10
3.3 <i>Clustering</i> . . . . .	10
3.4 Hasil <i>Clustering</i> . . . . .	11
3.5 Validasi . . . . .	15
<b>4 KESIMPULAN DAN SARAN</b>	<b>17</b>
4.1 Kesimpulan . . . . .	17
4.2 Saran . . . . .	17
<b>DAFTAR PUSTAKA</b>	<b>19</b>



## DAFTAR GAMBAR

2.1	Ilustrasi Algoritma <i>Clustering</i> . . . . .	6
3.1	Tabel Frekuensi ARREST dan DOMESTIC . . . . .	7
3.2	Tabel Frekuensi PRIMARY DESCRIPTION dan LOCATION DESCRIPTION . . . . .	8
3.3	Plot <i>mosaic</i> PRIMARY DESCRIPTION terhadap ARREST . . . . .	9
3.4	<i>Heatmap</i> PRIMARY DESCRIPTION terhadap BEAT . . . . .	9
3.5	Tampilan Matriks Kompilasi Data . . . . .	10
3.6	Tampilan Hasil <i>Clustering</i> dari R . . . . .	12
3.7	Visualisasi Hasil <i>Clustering</i> . . . . .	14



## DAFTAR TABEL

3.1	Data <i>My City</i> . . . . .	7
3.2	Hasil <i>Clustering</i> . . . . .	13
3.3	Jenis Kejahatan dengan <i>High Crime Level</i> . . . . .	13
3.4	Jenis Kejahatan dengan <i>Medium Crime Level</i> . . . . .	14
3.5	Jenis Kejahatan dengan <i>Low Crime Level</i> . . . . .	14





# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Keamanan merupakan suatu faktor esensial dalam hidup bermasyarakat untuk menunjang kesejahteraan. Peraturan yang lazim ada pada suatu kota atau daerah semestinya memastikan keamanan dan ketentraman warga yang tinggal disana. Tindak kejahatan dan tindak kriminal merupakan pelanggaran peraturan yang mengancam kesejahteraan yang ada.

Kejahatan dan kriminalitas merupakan kejadian yang merugikan masyarakat dan terjadi di kota-kota setiap harinya. Apabila terjadi kesengajaan atau pelanggaran terhadap hukum dalam masyarakat dan pemerintahan maka harus dikenakan sanksi sesuai dengan ketentuan hukum yang dilanggar tersebut. Tentunya pelaksanaan terhadap hal ini dilakukan oleh suatu lembaga peradilan yang telah ditentukan Undang-Undang. Polisi merupakan salah satu pihak yang memiliki wewenang untuk menindaklanjuti suatu kasus kejahatan dan mencatat setiap kasus kejahatan yang terjadi dan dilaporkan. Ada faktor-faktor penting yang perlu diperhatikan pada suatu kasus kejahatan, faktor tersebut ialah frekuensi kejahatan yang terjadi, penyebaran jenis kejahatan, jenis kejahatan, status penangkapan dan status domisili pelaku.

*My City* merupakan kota pusat perkembangan ekonomi dan teknologi maju yang berisikan penduduk domestik maupun non-domestik. Dari data kasus kejahatan pada kota *My City* tercatat 11162 jumlah kasus dalam rentang waktu 5 Juli 2014 sampai 18 Juli 2014, dengan data kode kasus kejahatan, deskripsi kejahatan (deskripsi primer dan deskripsi sekunder), lokasi kejahatan, domisili pelaku dari kejahatan (domestik atau non-domestik), status penangkapan kejahatan, dan *Beat* (kode daerah) kejahatan.

Model untuk menentukan status keamanan dari kota *My City* berdasarkan data-data yang ada akan disusun untuk menyelesaikan permasalahan yang ada. Data diolah terlebih dahulu untuk menentukan variabel-variabel yang signifikan terhadap model. Variabel-variabel yang telah dipilih selanjutnya dikelompokkan menggunakan algoritma *clustering* untuk membagi jenis kejahatan ke dalam kelompok status keamanan (*low crime level*, *medium crime level*, dan *high crime level*) dengan parameter-parameter tertentu. Dari hasil *clustering*, dapat ditarik suatu kesimpulan untuk status keamanan di kota *My City*.

### 1.2 Rumusan Masalah

Rumusan masalah dari makalah ini adalah:

1. Bagaimana hasil interpretasi data kasus kejahatan di kota *My City*?
2. Bagaimana cara membentuk model untuk menentukan status keamanan di *My City*?
3. Apa status keamanan di kota *My City*?

### 1.3 Tujuan

Tujuan dari makalah ini adalah:

1. Melakukan interpretasi data kasus kejahatan di kota *My City*.
2. Menyusun model untuk menentukan status keamanan kota *My City*.
3. Menentukan apakah status keamanan kota *My City* "AMAN" atau "TIDAK AMAN".

### 1.4 Manfaat

Manfaat dari makalah ini adalah:

1. Mempelajari cara untuk membuat model matematika dan menginterpretasikan hasilnya.
2. Menerapkan teknologi dalam bidang matematika untuk menyelesaikan suatu permasalahan nyata.

### 1.5 Batasan masalah

Batasan masalah dari makalah ini adalah:

1. Data yang digunakan dalam pemodelan adalah data kasus kejahatan di kota *My City*.
2. Deskripsi sekunder dianggap tidak berpengaruh karena deskripsi primer sudah mendeskripsikan jenis kasus kejahatan yang terjadi.
3. Lokasi kasus kejahatan dianggap tidak berpengaruh karena *Beat* (kode daerah) sudah mencakup lokasi yang lebih umum.
4. Tanggal kasus kejahatan tidak diperhatikan karena data diasumsikan sudah menggambarkan seluruh kasus pada tahun 2014.
5. Status kependudukan pelaku kejahatan (domestik atau non-domestik) dianggap tidak berpengaruh, dalam arti seluruh pelaku merupakan warga kota *My City*.

### 1.6 Metodologi

Metode penelitian yang digunakan dalam makalah ini adalah algoritma *K-means*. Pertama, data kasus kejahatan dari *My City* dilakukan interpretasi dan pengolahan awal. Selanjutnya, dipilih variabel yang signifikan terhadap model. Data kemudian dilakukan standardisasi dan dijadikan input dari algoritma untuk mengelompokkan 27 jenis kejahatan menjadi kelompok *high crime level*, *medium crime level*, dan *low crime level*. Hasil *clustering* dijadikan acuan penilaian status keamanan di *My City*.

## 1.7 Sistematika Pembahasan

Sistematika pembahasan pada laporan ini terdiri dari empat bab, yaitu:

### **Bab 1: Pendahuluan**

Bab ini berisi tentang latar belakang, rumusan masalah, tujuan, batasan masalah, metodologi, dan sistematika pembahasan.

### **Bab 2: Dasar Teori**

Bab ini membahas tentang dasar teori yang digunakan dalam pemodelan.

### **Bab 3: Hasil dan Pembahasan**

Bab ini berisi hasil dan pembahasan dari pemodelan.

### **Bab 4: Kesimpulan dan Saran**

Bab ini berisi kesimpulan dari isi makalah dan saran untuk pemodelan selanjutnya.



## BAB 2

### DASAR TEORI

Bab ini akan membahas tentang materi-materi yang digunakan dalam skripsi ini. Materi-materi yang akan dibahas adalah *Z-score* dan *K-means Clustering*.

#### 2.1 *Z-score*

*Z-score*, atau yang disebut sebagai standardisasi, merupakan prosedur dimana nilai data diskalakan agar memiliki sifat-sifat distribusi normal, yaitu  $\mu = 0$  dan  $\sigma = 1$ . *Z-score* didapatkan dengan cara sebagai berikut:

$$Z = \frac{x_j - \mu_j}{\sigma_j}$$

dengan  $j = 1, 2, \dots, n$  [4].

#### 2.2 *K-means Clustering*

*K-Means Clustering* (KMC) adalah salah satu algoritma *machine learning* yang digunakan dalam melakukan pengelompokan data. Misalkan  $a$  adalah *dataset* dengan  $n$  jumlah data sehingga  $a = (a_1, a_2, \dots, a_n)$ , dimana setiap  $a_i$  memiliki  $d$  banyak *feature*. Pada *machine learning*, *feature* atau yang disebut juga sebagai atribut adalah sifat-sifat atau karakteristik terukur dari data. Pada statistika, *feature* dikenal sebagai variabel dari data. Tujuan dari algoritma ini adalah untuk mengklasifikasi  $n$  jumlah data menjadi  $k$  kelompok. Masing-masing kelompok  $P_1, \dots, P_k$  mempunyai *centroid* (titik pusat)  $c_1, \dots, c_k$ . Setiap *centroid*  $c_k$  dioptimalkan dengan cara meminimalkan total jarak *centroid* terhadap setiap titik pada masing-masing kelompok. Jarak total di setiap kelompok  $P_j$  adalah:

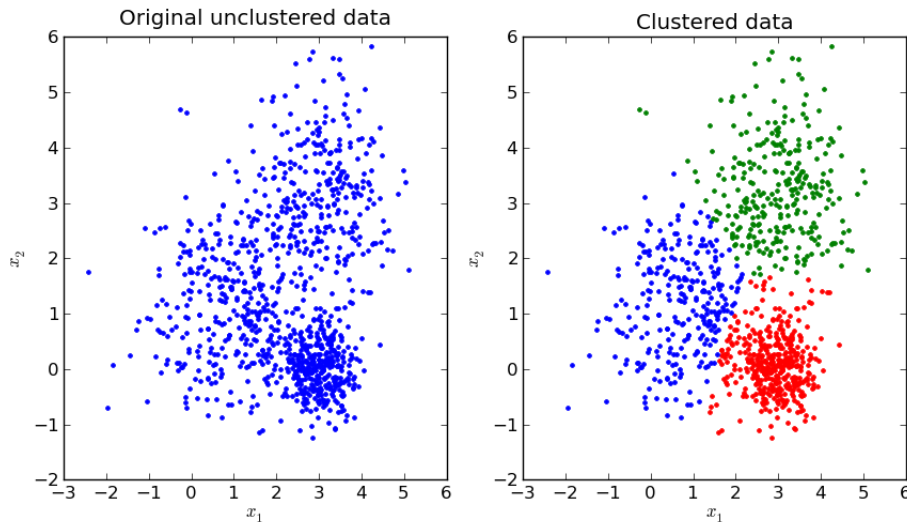
$$D_j = ||c_j - a||^2.$$

Dengan kata lain, tujuan algoritma adalah untuk meminimalkan  $D = D_1 + \dots + D_k$  untuk seluruh  $a_i$  di  $P_j$  [3].

Algoritma *K-means* mempartisi data ke dalam kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu kelompok yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [1]. Langkah-langkah algoritma dasar dari *K-Means* adalah sebagai berikut:

1. Menentukan  $k$  jumlah *cluster*  $P_1, \dots, P_k$ .
2. Memilih  $k$  jumlah *centroid* pertama secara acak, yaitu  $c_1, \dots, c_k$ .
3. Mengalokasikan setiap titik pada data ke dalam suatu *cluster* berdasarkan *centroid* yang terdekat dengan titik tersebut.

4. Menghitung posisi rata-rata titik di dalam *cluster*  $P_1, \dots, P_k$  untuk menjadi *centroid* yang baru  $c_1, \dots, c_k$ .
5. Apabila masih ada data yang berpindah *cluster* atau apabila ada perubahan titik *centroid*, lakukan pengulangan langkah 3 dan 4 sampai konvergen.



Gambar 2.1: Ilustrasi Algoritma *Clustering*

Sumber: [mubaris.com/posts/kmeans-clustering/](https://mubaris.com/posts/kmeans-clustering/)

Gambar 2.1 menunjukkan ilustrasi dasar dari *clustering*. Di sebelah kiri adalah plot dari suatu *dataset* dengan 2 *feature* yaitu  $x_1$  dan  $x_2$ . Seluruh data diberi warna biru sebagai tanda bahwa data belum dikelompokkan. Di sebelah kanan adalah plot hasil *clustering* dimana data dikelompokkan menjadi 3 kelompok, yang berarti  $k = 3$ . Data diberi warna biru, hijau, dan merah tergantung dari kelompok mereka masing-masing.

## BAB 3

### HASIL DAN PEMBAHASAN

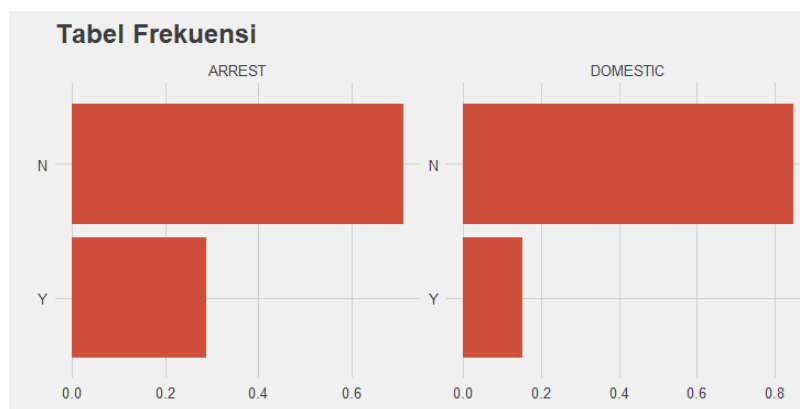
Populasi *My City* saat ini sebesar 2,8 juta orang yang terkena dampak oleh area metropolitan dengan tambahan sekitar 6 juta orang. Dari data yang telah diterima, terdapat 11162 kasus kejahatan yang terjadi selama 14 hari di *My City* dengan beberapa variabel lainnya yang dijelaskan dengan tabel berikut:

Nama Kolom	Deskripsi	Tipe Data	Deskripsi Data
CASE	Kode kasus kejahatan	<i>numeric</i>	11162
DATE.OF.OCCURENCE	Tanggal kejadian kasus	<i>date</i>	14 hari
PRIMARY.DESCRPTION	Jenis kejahatan primer	<i>factor</i>	27 jenis
SECONDARY.DESCRPTION	Jenis kejahatan sekunder	<i>factor</i>	213 jenis
LOCATION.DESCRPTION	Lokasi kasus kejahatan	<i>factor</i>	93 jenis
ARREST	Status penangkapan pelaku kejahatan	<i>factor</i>	Y / N
DOMESTIC	Status domestik pelaku kejahatan	<i>factor</i>	Y / N
BEAT	Kode daerah	<i>character</i>	274 jenis

Table 3.1: Data *My City*

## 3.1 Visualisasi Data

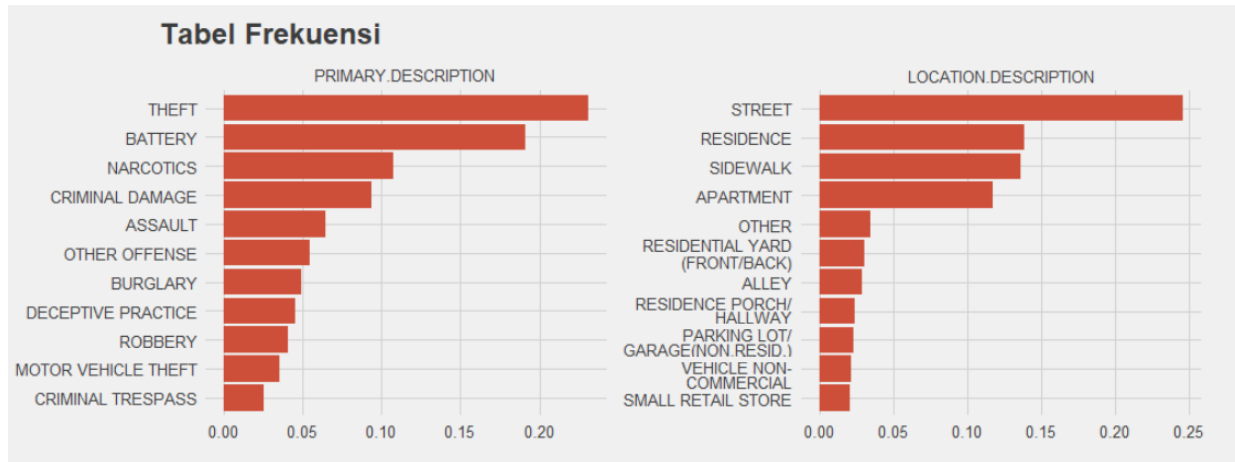
### 3.1.1 Tabel Frekuensi



Gambar 3.1: Tabel Frekuensi ARREST dan DOMESTIC

Tabel frekuensi ARREST menunjukkan seberapa banyak terjadinya penangkapan dari kejadian kriminal yang terjadi. Bisa terlihat dari grafik, bahwa mayoritas sebanyak 71,1% dari kejadian kejahatan bisa diatasi dengan menangkap tersangka dan sebanyak 28,9% tersangka belum tertangkap.

Tabel frekuensi DOMESTIC menunjukkan perbandingan antara penduduk lokal dan non-lokal yang melakukan kejahatan di *My City*. Dari grafik bisa terlihat bahwa sebagian besar penduduk yang melakukan kejahatan adalah penduduk domestik yaitu sebesar 84,8% dan penduduk non lokal sebanyak 15,2%.



Gambar 3.2: Tabel Frekuensi PRIMARY DESCRIPTION dan LOCATION DESCRIPTION

Deskripsi kejahatan dibagi atas 2 macam yaitu deskripsi primer dan sekunder. Deskripsi primer mencakup jenis kejahatan primer menurut polisi, dan deskripsi sekunder mencakup keterangan lebih lanjut mengenai kejahatan tersebut. Jenis kejahatan menurut deskripsi primer terdiri dari 27 jenis kejahatan, dan deskripsi kejahatan sekunder terdiri atas 213 jenis kejahatan.

Dari 27 jenis kejahatan yang ada pada deskripsi primer, didapatkan 11 kejahatan yang paling banyak terjadi di *My City*. Kejahatan yang paling banyak dilakukan di *My City* adalah pencurian (THEFT) yaitu sebanyak 23%.

Dari data yang telah diterima, dari 274 daerah di *My City* lokasi kejahatan yang terjadi ada di 93 titik lokasi. Dari 93 titik lokasi kejahatan, didapatkan 11 lokasi yang paling banyak terjadi tindak kejahatan. Lokasi yang paling sering terjadi tindak kejahatan adalah jalanan yaitu sebanyak 24%.

### 3.1.2 Mosaic Plot

Plot *mosaic* menunjukkan banyaknya kasus kejahatan yang terjadi di kota *My City* berdasarkan jenis kejahatan primer beserta proporsi penangkapannya. Warna merah pada plot menunjukkan kasus kejahatan yang pelakunya belum ditangkap, sementara warna biru merepresentasikan kasus yang pelakunya sudah ditangkap. Sebagai contoh, semakin panjang ukuran daerah yang berwarna merah, maka semakin kecil frekuensi penangkapan kejahatan tersebut. Sebaliknya, semakin pendek daerah tersebut, maka semakin tinggi frekuensi penangkapan kejahatan tersebut.

Frekuensi kejahatan dapat dilihat dari tinggi daerah tersebut yang ditentukan berdasarkan jumlah kejahatan yang terjadi. Semakin tinggi daerah tersebut maka semakin besar frekuensi kejahatan yang terjadi. Di sisi lain, semakin pendek (vertikal) daerah tersebut maka semakin kecil frekuensinya. Sebagai ilustrasi, semakin luas daerah yang berwarna merah maka kejahatan tersebut dianggap sering terjadi dan tingkat penangkapannya rendah.





Gambar 3.3: Plot *mosaic* PRIMARY DESCRIPTION terhadap ARREST

### 3.1.3 Heatmap

*Heatmap* menunjukkan penyebaran jenis kejahatan di setiap *Beat*. Sumbu-*x* menunjukkan *Beat* atau kode daerah dari kota *My City* sedangkan sumbu-*y* menunjukkan jenis kejahatan yang dapat terjadi. Semakin muda warna dari *heatmap* tersebut, maka semakin sering frekuensi kejahatan tersebut terjadi di setiap *Beat*. Namun semakin tua warna *heatmap* tersebut maka semakin kecil frekuensi kejahatan yang terjadi di setiap *Beat*.



Gambar 3.4: *Heatmap* PRIMARY DESCRIPTION terhadap BEAT

### 3.2 Pengolahan data

Sebelum memulai pemodelan, langkah awal adalah memilih variabel yang akan dimasukkan ke dalam model. Variabel yang tidak terpakai dalam model ini adalah kode kasus, tanggal kejadian kasus, status kependudukan pelaku, dan deskripsi sekunder. Pengeluran variabel tersebut akan memudahkan pemodelan. Variabel yang kelompok kami pilih untuk model adalah jenis kejahatan primer, daerah *Beat* terjadinya kejahatan, dan status penangkapan.

Langkah berikutnya adalah menghitung jumlah setiap jenis kejahatan primer yang terjadi di setiap *Beat* beserta jumlah penangkapan yang tidak terjadi. Sebagai contoh, salah satu jenis kejahatan primer adalah ARSON. Dilakukan perhitungan kejadian ARSON yang terjadi pada setiap 274 *Beat* dan total penangkapan yang tidak terjadi di seluruh *Beat*. Berikut adalah potongan dari tabel data yang sudah dikompilasi.

PRIMARY.DESCRPTION	BEAT1012	BEAT1013	BEAT1014	BEAT1021	BEAT1022	BEAT1023
1 ARSON	0	0	0	1	0	0
2 ASSAULT	0	1	2	2	1	1
3 BATTERY	1	2	2	2	2	2
4 BURGLARY	1	1	1	1	1	1
5 CONCEALED CARRY LICENSE VIOLATION	0	0	0	0	0	0
6 CRIM SEXUAL ASSAULT	0	0	0	0	1	1
7 CRIMINAL DAMAGE	2	1	1	1	1	2
8 CRIMINAL TRESPASS	2	0	1	1	1	2

Showing 1 to 10 of 27 entries. 276 total columns

Gambar 3.5: Tampilan Matriks Kompilasi Data

Hasil kompilasi dari data menghasilkan matriks berukuran  $27 \times 276$ . Setiap baris dari matriks merepresentasikan 27 jenis kejahatan primer, sementara setiap kolom merepresentasikan jumlah kejadian yang terjadi di 274 *Beat*, dengan tambahan 2 kolom yaitu kolom nama jenis kejahatan primer dan kolom jumlah penangkapan yang tidak terjadi.

Berikutnya, standarisasi dilakukan terhadap matriks data yang ada agar data memiliki rata-rata 0 dan standar deviasi 1. Proses ini dilakukan agar data dipusatkan dan memiliki satuan yang sama untuk mempermudah tahapan *K-means* yang dilakukan selanjutnya.

*K-means* merupakan algoritma yang menghitung jarak antar titik, sehingga eksistensi dari data pencilan (*outlier*) pada *dataset* akan sangat mempengaruhi hasil dari pengelompokkan. Oleh karena itu, standarisasi merupakan tahapan yang penting untuk memastikan bahwa data *outlier* ada pada daerah yang sama dengan data lainnya.

### 3.3 Clustering

Data yang telah diolah dan distandarisasi kemudian menjadi input untuk algoritma *K-means* menggunakan software R. *K-means clustering* menerima input 27 titik data dengan masing-masing 275 *feature*. Algoritma *K-means* membagi 27 titik data tersebut menjadi 3 kelompok *cluster*, yaitu *high crime level*, *medium crime level*, dan *low crime level*. Oleh karena itu, dari jumlah kelompok tersebut ditentukan bahwa  $k = 3$ .

Langkah pertama dari algoritma *K-means* adalah memilih 3 *centroid* pertama secara acak dari 27 jumlah data untuk iterasi 1. Setiap data mempunyai peluang yang sama untuk dipilih, yaitu  $\frac{1}{27}$ , dan pengambilan dilakukan secara bersamaan.

Langkah kedua, setiap titik kecuali 3 titik *centroid* dihitung jaraknya terhadap masing-masing *centroid*. Setiap titik dikelompokkan sesuai dengan *centroid* yang jaraknya paling dekat dengan mereka. Setelah semua data dikelompokkan, langkah ketiga adalah menghi-

tung rata-rata posisi masing-masing kelompok. Hasil rata-rata di setiap kelompok menjadi *centroid* yang baru.

Kemudian, langkah kedua dan ketiga, yang merupakan iterasi dari *K-means*, diulang terus menerus sampai kelompok *centroid*, titik centroid, dan jarak antar *centroid* dan titik di setiap kelompok konvergen. Iterasi dianggap konvergen ketika setiap titik tidak mengalami perubahan *cluster* dan titik *centroid* tidak berubah.

### 3.4 Hasil *Clustering*

Berikut adalah tampilan yang dikeluarkan R setelah *clustering*

```
> print(crime.km)
K-means clustering with 3 clusters of sizes 3, 16, 8

Cluster means:
  BEAT1012  BEAT1013  BEAT1014  BEAT1021  BEAT1022  BEAT1023  BEAT1024  BEAT1031
1  0.2304973  1.7863538  1.7863538  1.3985216  1.5479055  1.1639873  1.0675210  1.7903953
2 -0.5150173 -0.5150173 -0.7094994 -0.5962447 -0.6738726 -0.5621530 -0.4036564 -0.6266384
3  0.9435982  0.3601520  0.7491161  0.6680438  0.7672807  0.6878107  0.4069924  0.5818785
  BEAT1032  BEAT1033  BEAT1034  BEAT111  BEAT1111  BEAT1112  BEAT1113  BEAT1114
1  1.6587112  1.1639873  2.22864569  2.1499354  1.1182425  0.9085135  1.0232679  1.7047242
2 -0.5805489 -0.3835867 -0.45766831 -0.5374838 -0.6032624 -0.3974747 -0.5520261 -0.5450200
3  0.5390811  0.3306782  0.07959449  0.2687419  0.7871839  0.4542568  0.7203267  0.4507684
  BEAT1115  BEAT1112  BEAT1121  BEAT1122  BEAT1123  BEAT1124  BEAT1125  BEAT113
1  1.1143228  2.14993540  1.1645969  1.4719601  0.6886385  1.6886655  1.3877773  2.0167270
2 -0.4377697 -0.43670563 -0.5386261 -0.4293217 -0.6394500 -0.4403632 -0.5963106 -0.5564445
3  0.4576683  0.06718548  0.6405283  0.3066584  1.0206606  0.2474768  0.6722046  0.3566164
  BEAT1131  BEAT1132  BEAT1133  BEAT1134  BEAT1135  BEAT114  BEAT121  BEAT1211
1  1.9081938  1.274755  1.5918898  2.22864569  1.1639873  1.3877773  1.8027756  1.4557461
2 -0.4987325 -0.610820 -0.5273135 -0.36812451 -0.6514361 -0.5963106 -0.4882517 -0.6405283
3  0.2818923  0.743607  0.4576683 -0.09949311  0.8663769  0.6722046  0.3004626  0.7351518
  BEAT1212  BEAT1213  BEAT1214  BEAT1215  BEAT122  BEAT1221  BEAT1222  BEAT1223
1  1.6124515  1.2018504  1.4557461  0.6009252  0.9813068  0.8673608  1.5918898  1.8027756
2 -0.5374838 -0.6009252 -0.6405283 -0.6009252 -0.4906534 -0.5963106 -0.7064011 -0.4882517
3  0.4702984  0.7511565  0.7351518  0.9765035  0.6133167  0.8673608  0.8158435  0.3004626
  BEAT1224  BEAT1225  BEAT123  BEAT1231  BEAT1232  BEAT1233  BEAT1234  BEAT1235
1  1.1146274  1.9798147  1.1639873  1.2735118  2.0599204  1.6401640  1.2677350  2.2951702
2 -0.5450200 -0.6405283 -0.5621530 -0.6367559 -0.5885487 -0.5621530 -0.5150173 -0.5871366
3  0.6720547  0.5386261  0.6878107  0.7959449  0.4046272  0.5092445  0.5546341  0.3135843
  BEAT124  BEAT131  BEAT132  BEAT133  BEAT1411  BEAT1412  BEAT1413  BEAT1414
1  1.4557461  1.5079737  1.3253593  0.7407192  0.7491161  1.4691855  1.6124515  1.7047242
2 -0.6405283 -0.6429085 -0.5546341 -0.5092445 -0.7094994 -0.7226157 -0.5374838 -0.6556632
3  0.7351518  0.7203267  0.6122584  0.7407192  1.1380803  0.8942868  0.4702984  0.6720547
  BEAT1421  BEAT1422  BEAT1423  BEAT1424  BEAT1431  BEAT1432  BEAT1433  BEAT1434
1  1.9798147  1.3877773  1.6401640  1.3253593  1.6124515  2.1692491  1.2018504  1.9081938
2 -0.5422654 -0.4987325 -0.5621530 -0.5546341 -0.5374838 -0.6878107 -0.6009252 -0.6938887
3  0.3421003  0.4770485  0.5092445  0.6122584  0.4702984  0.5621530  0.7511565  0.6722046

  BEAT1511  BEAT1512  BEAT1513  BEAT1522  BEAT1523  BEAT1524  BEAT1531  BEAT1532
1  1.5918898  1.2018504  1.5742343  1.4424590  1.4793756  1.8147857  1.4053078  1.5248424
2 -0.6168573 -0.6009252 -0.7027832 -0.6228800 -0.5732580 -0.4970645 -0.5709063 -0.6886385
3  0.6367559  0.7511565  0.8152285  0.7048379  0.5917502  0.3135843  0.6148222  0.8504611
  BEAT1533  BEAT1611  BEAT1612  BEAT1613  BEAT1614  BEAT1621  BEAT1622  BEAT1623
1  1.5265551  1.1146274  0.7959449  0.9122911  1.9143847  1.4090744  1.2794675  0.8673608
2 -0.5811318 -0.5450200 -0.2338088 -0.5805489 -0.5469671 -0.4931760 -0.7526279 -0.6938887
3  0.5898054  0.6720547  0.1691383  0.8189886  0.3760399  0.4579492  1.0254556  1.0625170
  BEAT1624  BEAT1631  BEAT1632  BEAT1633  BEAT1634  BEAT1651  BEAT1652  BEAT1653
1  1.3877773  2.36789559  1.6124515  1.8439782  1.8027756  0.5115415  1.1369894 -0.2775555
2 -0.5637845 -0.43619129 -0.5374838 -0.5546341 -0.6009252 -0.4092332 -0.3248541 -0.2775555
3  0.6071526 -0.01557826  0.4702984  0.4177763  0.5258096  0.6266384  0.2233372  0.6591942
  BEAT1654  BEAT1655  BEAT1711  BEAT1712  BEAT1713  BEAT1722  BEAT1723  BEAT1724
1  1.2990468 -0.2775555  1.3877773  1.6930725  2.0599204  1.9081938  1.1639873  1.3877773
2 -0.5469671 -0.2775555 -0.5963106 -0.5092445 -0.4892311 -0.5963106 -0.3835867 -0.6938887
3  0.6067916  0.6591942  0.6722046  0.3835867  0.2059920  0.4770485  0.3306782  0.8673608
  BEAT1731  BEAT1732  BEAT1733  BEAT1811  BEAT1812  BEAT1813  BEAT1814  BEAT1821
1  1.0971343  1.7863538  1.9798147  1.0005328  0.7959449  1.5122953  1.0490611  2.1692491
2 -0.7542798 -0.7094994 -0.4440026 -0.4892311 -0.6367559 -0.6367559 -0.5245305 -0.5985276
3  1.0971343  0.7491161  0.1455746  0.6032624  0.9750325  0.7064011  0.6556632  0.3835867
  BEAT1822  BEAT1823  BEAT1824  BEAT1831  BEAT1832  BEAT1833  BEAT1834  BEAT1911
1  1.6124515  1.4557461  1.6676442  1.8849672  1.7040257  1.6851582  1.7347217  1.9798147
2 -0.5374838 -0.5422654 -0.7021660 -0.4476797 -0.4868645 -0.5451982 -0.4770485 -0.6405283
3  0.4702984  0.5386261  0.7789654  0.1884967  0.3347193  0.4584622  0.3035763  0.5386261
  BEAT1912  BEAT1913  BEAT1914  BEAT1915  BEAT1921  BEAT1922  BEAT1923  BEAT1924
1  1.5122953  1.8439782  1.4214934  1.9081938  0.4076089  1.9081938  1.3877773  1.3253593
2 -0.6367559 -0.5546341 -0.6774305 -0.5963106 -0.5422654 -0.5963106 -0.5963106 -0.5546341
3  0.7064011  0.4177763  0.8218009  0.4770485  0.9316775  0.4770485  0.6722046  0.6122584
  BEAT1925  BEAT1931  BEAT1932  BEAT1933  BEAT1934  BEAT1935  BEAT2011  BEAT2012
1  0.8498366  2.1312295  1.1146274  0.9316775  1.3344013  1.6176363  1.6401640  1.5396007
2 -0.6108200 -0.5451982 -0.6556632 -0.4440026 -0.5871366 -0.7117600 -0.6514361 -0.5532940
3  0.9029514  0.2911854  0.8933411  0.5386261  0.6738726  0.8169063  0.6878107  0.5292377
  BEAT2013  BEAT2022  BEAT2023  BEAT2024  BEAT2031  BEAT2032  BEAT2033  BEAT211
1  2.0599204  1.4557461  1.9798147  0.8888573  1.3344013  1.0490611  1.8027756  0.2645426
2 -0.5885487 -0.6405283 -0.6405283 -0.7163488 -0.4970645 -0.3770063 -0.6009252 -0.6878107
3  0.4046272  0.7351518  0.5386261  1.0993762  0.4937285  0.3606147  0.5258096  1.2764179
```

	BEAT212	BEAT213	BEAT214	BEAT215	BEAT221	BEAT2211	BEAT2212	BEAT2213
1	0.6176363	1.0352872	1.3253593	1.6176363	1.3253593	1.2168958	1.6996732	2.0599204
2	-0.7117600	-0.6025695	-0.5546341	-0.6025695	-0.5546341	-0.5092445	-0.5311479	-0.5885487
3	0.8169063	0.8169063	0.6122584	0.5985254	0.6122584	0.5621530	0.4249183	0.4046272
	BEAT222	BEAT2221	BEAT2222	BEAT2223	BEAT223	BEAT2232	BEAT2233	BEAT2234
1	0.8067404	1.2168958	1.4373408	1.3253593	1.8439782	1.6719243	1.5079737	1.3300079
2	-0.7491161	-0.5092445	-0.5421005	-0.5546341	-0.6518751	-0.5851735	-0.5520261	-0.5403157
3	1.1957046	0.5621530	0.5451982	0.6122584	0.6122584	0.5433754	0.5385621	0.5818785
	BEAT224	BEAT225	BEAT231	BEAT232	BEAT233	BEAT234	BEAT235	BEAT2411
1	0.8942868	1.9081938	1.5396007	1.8834121	1.1146274	1.4557461	1.6587112	1.3877773
2	-0.6148222	-0.5963106	-0.5532940	-0.6257389	-0.6556632	-0.6405283	-0.5805489	-0.5637845
3	0.8942868	0.4770485	0.5292377	0.5451982	0.8933411	0.7351518	0.5390811	0.6071526
	BEAT2412	BEAT2413	BEAT2422	BEAT2423	BEAT2424	BEAT2431	BEAT2432	BEAT2433
1	0.3849002	1.0675210	2.1163406	1.2168958	0.4076089	1.7047242	1.8834121	0.6878107
2	-0.6615472	-0.4937285	-0.7407192	-0.6878107	-0.5422654	-0.6556632	-0.7093773	-0.2050205
3	1.1787568	0.5871366	0.6878107	0.9192854	0.9316775	0.6720547	0.7124750	0.1521120
	BEAT2511	BEAT2512	BEAT2513	BEAT2514	BEAT2515	BEAT2521	BEAT2522	BEAT2523
1	1.8027756	1.4557461	2.0599204	1.2168958	1.0675210	1.8439782	1.1143228	1.1639873
2	-0.4882517	-0.6405283	-0.4892311	-0.5985276	-0.5838006	-0.5546341	-0.6168573	-0.5621530
3	0.3004626	0.7351518	0.2059920	0.7407192	0.7672807	0.4177763	0.8158435	0.6878107
	BEAT2524	BEAT2525	BEAT2531	BEAT2532	BEAT2533	BEAT2534	BEAT2535	BEAT311
1	1.1645969	1.2018504	1.3877773	1.6401640	1.6719243	1.5079737	1.7347217	1.7510788
2	-0.6368889	-0.6009252	-0.4987325	-0.6514361	-0.5851735	-0.4611438	-0.5746266	-0.6367559
3	0.8370540	0.7511565	0.4770485	0.6878107	0.5433754	0.3567974	0.4987325	0.6168573
	BEAT312	BEAT313	BEAT314	BEAT321	BEAT322	BEAT323	BEAT324	BEAT331
1	1.2677350	1.7047242	1.8439782	1.3985216	1.4373408	1.6719243	1.5479055	1.035563
2	-0.5180173	-0.5450200	-0.5546341	-0.6805306	-0.7093773	-0.6657841	-0.6738726	-0.573258
3	0.5546341	0.4507684	0.4177763	0.8366156	0.8797517	0.7045967	0.7672807	0.758180
	BEAT332	BEAT333	BEAT334	BEAT411	BEAT412	BEAT413	BEAT414	BEAT421
1	1.362770	1.8834121	1.8834121	1.1143228	1.5479055	1.7347217	1.7309327	1.7309327
2	-0.567821	-0.6257389	-0.6257389	-0.5273135	-0.5838006	-0.5746266	-0.7163488	-0.6374043
3	0.624603	0.5451982	0.5451982	0.6367559	0.5871366	0.4987325	0.7835979	0.6257088
	BEAT422	BEAT423	BEAT424	BEAT431	BEAT432	BEAT433	BEAT434	BEAT511
1	1.2115822	1.5918898	0.8696206	1.5479055	1.2143052	1.5918898	1.1639873	0.7754126
2	-0.4785750	-0.4377697	-0.5403157	-0.6738726	-0.5746266	-0.6168573	-0.5621530	-0.5603568
3	0.5028066	0.2785807	0.7545237	0.7672807	0.6938887	0.6367559	0.6878107	0.8299338

	BEAT512	BEAT513	BEAT522	BEAT523	BEAT524	BEAT531	BEAT532	BEAT533
1	2.0282899	1.8480464	1.2115822	1.3985216	1.0232679	1.4793756	1.2677350	1.9231883
2	-0.6738726	-0.5119588	-0.4785750	-0.6805306	-0.5520261	-0.4068283	-0.7094994	-0.5732580
3	0.5871366	0.3309002	0.5028066	0.8366156	0.7203267	0.2588907	0.9435982	0.4253205
	BEAT611	BEAT612	BEAT613	BEAT614	BEAT621	BEAT622	BEAT623	BEAT624
1	0.9489968	1.7309327	0.9085135	1.9926796	1.4116965	1.3253593	1.0490611	1.0148433
2	-0.2591011	-0.7163488	-0.6529941	-0.5520261	-0.4623306	-0.5546341	-0.6966421	-0.5835349
3	0.1623284	0.7835979	0.9652956	0.3567974	0.3952750	0.6122584	0.9998863	0.7865036
	BEAT631	BEAT632	BEAT633	BEAT634	BEAT711	BEAT712	BEAT713	BEAT714
1	1.2018504	1.1102219	1.6930725	2.0282899	1.4373408	1.7347217	1.2115822	1.7347217
2	-0.6510023	-0.6591942	-0.5985276	-0.4937285	-0.4584622	-0.4770485	-0.6421385	-0.6722046
3	0.8513107	0.9020553	0.5621530	0.2268482	0.3779215	0.3035763	0.8299338	0.6938887
	BEAT715	BEAT722	BEAT723	BEAT724	BEAT725	BEAT726	BEAT731	BEAT732
1	0.7491161	1.9798147	1.8480464	1.3471506	1.4719601	1.6719243	1.5079737	1.5479055
2	-0.4177763	-0.6405283	-0.6805306	-0.6374909	-0.6133167	-0.5045629	-0.5520261	-0.4036564
3	0.5546341	0.5386261	0.6680438	0.7698004	0.6746484	0.3821541	0.5385621	0.2268482
	BEAT733	BEAT734	BEAT735	BEAT811	BEAT812	BEAT813	BEAT814	BEAT815
1	1.6719243	1.0675210	1.2735118	1.9081938	1.6176363	1.6930725	1.2794675	1.5396007
2	-0.5851735	-0.5838006	-0.4875162	-0.5963106	-0.6025695	-0.5985276	-0.6256220	-0.5532940
3	0.5433754	0.7672807	0.4974656	0.4770485	0.5985254	0.5621530	0.7714436	0.5292377
	BEAT821	BEAT822	BEAT823	BEAT824	BEAT825	BEAT831	BEAT832	BEAT833
1	1.2677350	1.5918898	1.5742343	1.3475204	1.7347217	1.9231883	1.4373408	1.4053078
2	-0.6122584	-0.5273135	-0.5509820	-0.5866361	-0.4770485	-0.5732580	-0.5421005	-0.6786998
3	0.7491161	0.4576683	0.5116261	0.6679519	0.3035763	0.4253205	0.5451982	0.8304092
	BEAT834	BEAT835	BEAT911	BEAT912	BEAT913	BEAT914	BEAT915	BEAT921
1	0.7491161	1.274755	0.9912695	1.2735118	1.5918898	1.5396007	1.6930725	1.5079737
2	-0.5150173	-0.610820	-0.6257389	-0.6367559	-0.6168573	-0.6615472	-0.5985276	-0.6429085
3	0.7491161	0.743607	0.8797517	0.7959449	0.6367559	0.7457441	0.5621530	0.7203267
	BEAT922	BEAT923	BEAT924	BEAT925	BEAT931	BEAT932	BEAT933	BEAT934
1	1.7863538	1.5079737	1.7047242	1.5396007	1.6676442	1.2677350	1.0971343	1.1389006
2	-0.7094994	-0.4611438	-0.4343768	-0.5532940	-0.6281094	-0.7094994	-0.5074246	-0.6036173
3	0.7491161	0.3567974	0.2294821	0.5292377	0.6308522	0.9435982	0.6034239	0.7801469
	BEAT935	ARREST_N						
1	1.4557461	2.1126285						
2	-0.5422654	-0.5550560						
3	0.5386261	0.3178763						

```
Clustering vector:
[1] 2 1 1 3 2 2 3 3 3 2 2 2 2 2 3 3 2 2 3 2 2 3 2 2 1 2
```

```
within cluster sum of squares by cluster:
[1] 471.9529 954.6197 1463.4746
(between_SS / total_SS = 59.4 %)
```

```
Available components:
```

```
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"
[7] "size"         "iter"         "ifault"
```

Gambar 3.6: Tampilan Hasil *Clustering* dari R

Tampilan memberikan informasi bahwa data telah dikelompokkan menjadi 3 *cluster* yang sementara bernama *cluster* 1, *cluster* 2, dan *cluster* 3, yang masing-masing berukuran 3, 16, dan 8. Setiap titik pusat dari masing-masing *cluster* memiliki 275 sumbu koordinat.

Standarisasi telah dilakukan terhadap data. Oleh karena itu, ketika titik memiliki besaran yang mendekati 0, maka titik tersebut dianggap merepresentasikan data yang secara numerik rendah. Sebaliknya, jika data memiliki besaran yang menjauh dari 0, baik positif maupun negatif, titik tersebut dianggap merepresentasikan data yang secara numerik tinggi. Tanda positif atau negatif dari data hanya menunjukkan arah penyebaran.

Matriks dari data yang sudah dikompilasi sebelumnya menghitung jumlah kasus yang terjadi pada setiap *Beat* dan jumlah penangkapan yang tidak terjadi. Oleh karena itu, semakin tinggi nilai titik pada data, semakin banyak kasus yang terjadi dan jumlah penangkapan yang tidak terjadi. Artinya, nilai yang tinggi pada matriks kompilasi menunjukkan indikasi tidak aman. Jika data sudah distandarisasi, semakin jauh nilai tersebut dari 0, semakin menunjukkan indikasi tidak aman.

Hasil *clustering* secara keseluruhan menunjukkan bahwa *cluster* 1 memiliki nilai titik yang paling jauh dengan 0, sementara *cluster* 3 memiliki nilai titik yang paling dekat dengan 0. *Cluster* 1 dinamakan sebagai kelompok *high crime level* karena menunjukkan indikasi tidak aman yang paling tinggi. Demikian, *cluster* 2 dan 3 juga dinamakan dengan metode yang sama, sehingga didapatkan tabel sebagai berikut:

Nama Lama <i>Cluster</i>	Nama Baru <i>Cluster</i>	Besar <i>Cluster</i>	Presentase <i>Cluster</i>
<i>Cluster</i> 1	<i>High</i>	3	0,111
<i>Cluster</i> 2	<i>Medium</i>	16	0,592
<i>Cluster</i> 3	<i>Low</i>	8	0,296

Table 3.2: Hasil *Clustering*

Dari tabel diatas dapat diinterpretasikan bahwa terdapat 3 kejahatan primer di *high crime level*, artinya ada 3 kejahatan primer yang tingkat frekuensinya sangat tinggi, penyebarannya tersebar di setiap daerah, dan tingkat penangkapan kejahatan yang sangat rendah. Hal ini berarti 3 jenis kejahatan tersebut mempunyai kemungkinan yang tinggi untuk terjadi kembali, karena para pelaku masih berkeliaran. Dari keseluruhan *cluster*, *high crime level* mempunyai persentase paling kecil yaitu sekitar 11%.

*Medium crime level* masih dikategorikan berbahaya, meskipun tingkatannya tidak mencapai *high crime level*. Presentase *cluster* ini adalah yang paling tinggi, yaitu sekitar 59% yang berarti lebih dari setengah kejahatan di *My City* masuk kedalamnya.

*Low crime level* dikategorikan tidak terlalu berbahaya dikarenakan beberapa faktor yang mempengaruhi, dari tingkat penangkapan yang tinggi, frekuensi terjadinya kasus sedikit, serta tingkat penyebarannya rendah. Presentase *cluster* adalah sebesar 29%, maka dapat diartikan bahwa sekitar 29% persen kasus di *My City* sudah diatasi dengan baik.

<b>HIGH CRIME LEVEL</b>
ASSAULT
BATTERY
THEFT

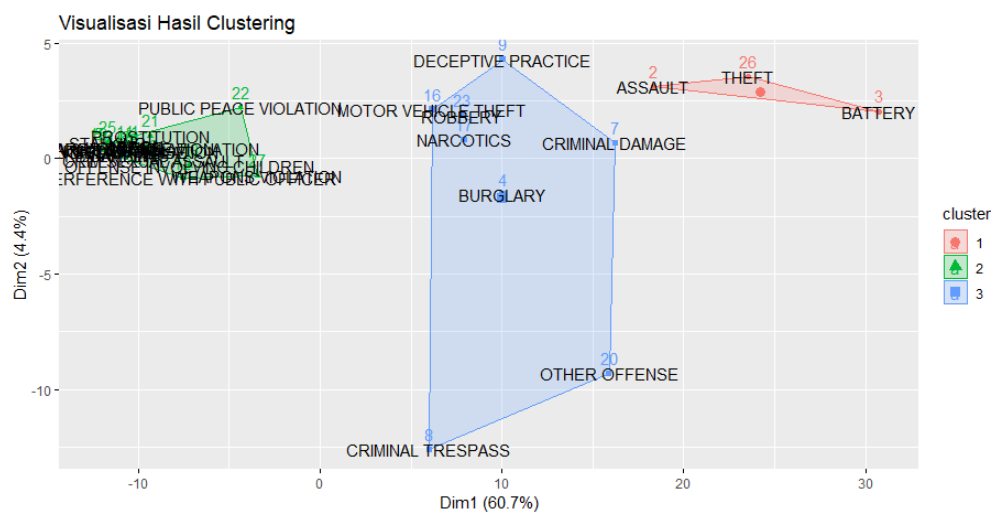
Table 3.3: Jenis Kejahatan dengan *High Crime Level*

MEDIUM CRIME LEVEL
ARSON
CONCEALED CARRY LICENSE VIOLATION
CRIME SEXUAL ASSAULT
GAMBLING
HOMICIDE
INTERFERENCE WITH PUBLIC OFFICER
INTIMIDATION
KIDNAPPING
LIQUOR LAW VIOLATION
OFFENSE INVOLVING CHILDREN
OTHER NARCOTIC VIOLATION
PROSTITUTION
PUBLIC PEACE VIOLATION
SEX OFFENSE
STALKING
WEAPONS VIOLATION

Table 3.4: Jenis Kejahatan dengan *Medium Crime Level*

LOW CRIME LEVEL
BURGLARY
CRIMINAL DAMAGE
CRIMINAL TRESPASS
DECEPTIVE PRACTICE
MOTOR VEHICLE THEFT
NARCOTICS
OTHER OFFENSE
ROBBERY

Table 3.5: Jenis Kejahatan dengan *Low Crime Level*



Gambar 3.7: Visualisasi Hasil *Clustering*



Data yang dikelompokkan memiliki 275 atribut atau *feature*, sehingga secara eksplisit tidak dapat divisualisasi oleh plot. Gambar 3.7 merupakan visualisasi dari hasil *clustering* dimana sumbu-*x* merepresentasikan 60,7% dari *feature* dan sumbu-*y* merepresentasikan 4,4% dari *feature*.

Hasil dan pembahasan dari *clustering* menghasilkan kesimpulan bahwa *My City* berstatus “TIDAK AMAN”. Hampir 70% jenis kejahatan masuk kedalam kategori *high crime level* dan *medium crime level*, yang berarti 70% dari jenis kejahatan memiliki frekuensi, penyebaran, dan tingkat tidak tertangkap yang cukup tinggi. Secara keseluruhan, sebagian besar kejahatan yang terjadi di *My City* memiliki intensitas yang tinggi sehingga dapat mengancam ketentraman bagi warga yang tinggal di kota tersebut.

### 3.5 Validasi

Dilakukan validasi terhadap model untuk memastikan model yang telah dibentuk sesuai kondisi asli dan data yang ada. Validasi dilakukan dengan 3 aspek yaitu, frekuensi jenis kejahatan primer, penyebaran kasus pada *Beat*, dan status penangkapan. Validasi untuk aspek frekuensi jenis kejahatan primer dan status penangkapan pada suatu kasus dapat dilihat melalui *mosaic* plot.

Berdasarkan gambar 3.3, dapat dilihat bahwa THEFT, BATTERY dan ASSAULT, yang merupakan anggota *high crime level*, memiliki frekuensi yang paling tinggi dengan tingkat penangkapan yang rendah. NARCOTICS yang dikelompokkan menjadi *low crime level*, memiliki frekuensi yang tinggi, namun tingkat penangkapannya pun tinggi yaitu sekitar 99%, sehingga pengelompokkannya dikatakan sesuai.

Gambar 3.4 memberikan visualisasi frekuensi jenis kejahatan pada setiap *Beat* di kota *My City*. Semakin terang suatu kotak berarti semakin tinggi frekuensi suatu jenis kejahatan pada *Beat* tersebut. Dapat dilihat THEFT, BATTERY, dan ASSAULT memiliki warna yang cenderung terang hampir di seluruh *Beat*, yang berarti mereka memiliki frekuensi yang cenderung tinggi dan tersebar luas hampir di seluruh daerah.

Tidak menutup kemungkinan bahwa terjadi kesalahan atau ketidakakuratan pada model yang telah dibuat. Pada kasus ini, CRIMINAL DAMAGE memiliki area yang luas dan tingkat penangkapan yang rendah pada plot *mosaic* serta memiliki penyebaran yang luas di *heatmap*, tetapi, ia dikategorikan sebagai *low crime level*.

Berdasarkan validasi tersebut, dapat disimpulkan bahwa hasil *clustering* dapat dikatakan valid atau cukup akurat, karena sebagian besar dari jenis kejahatan dapat dikatakan masuk ke dalam *cluster* yang sesuai. Ketidakakuratan dalam pemodelan terjadi relatif rendah. Oleh karena itu, metode yang digunakan dalam pemodelan ini dikatakan cukup sesuai dan akurat untuk menarik sebuah kesimpulan.





## BAB 4

### KESIMPULAN DAN SARAN

#### 4.1 Kesimpulan

Kesimpulan dari makalah ini adalah:

1. Penentuan status keamanan di *My City* dilakukan dengan mengolah 11163 data dengan 8 *feature* menjadi 27 data dengan 275 *feature*. Selanjutnya, data diolah menggunakan metode *K-Means* dengan membagi daerah menjadi 3 *cluster* atau  $k = 3$ .
2. Hasil *clustering* menunjukkan dari 27 jenis kejahatan primer, 11,1% jenis kejahatan primer masuk ke dalam kategori *high crime level*, 59,2% jenis kejahatan primer masuk dalam kategori *medium crime level*, dan 29,6% jenis kejahatan primer masuk dalam kategori *low crime level*.
3. Status keamanan dari *My City* adalah “TIDAK AMAN” karena sekitar 70% dari jenis kejahatan primer masuk ke dalam kategori *high crime level* dan *medium crime level* atau dapat diartikan kecenderungan tingkat frekuensi tinggi, penyebarannya luas, dan status penangkapannya rendah.

#### 4.2 Saran

Saran dari makalah ini adalah:

1. Menggunakan lebih banyak variabel dari data agar hasil pemodelan lebih sesuai dengan kondisi nyata.
2. Mencari status keamanan di setiap beat agar mengetahui daerah mana yang lebih rawan terjadi kejahatan.
3. Membandingkan status keamanan dengan kota-kota lain agar dapat melakukan validasi hasil pemodelan.



## DAFTAR PUSTAKA

- [1] Kamath, R.S., Kamat, R. K. 2016. *K-MEANS CLUSTERING FOR ANALYZING PRODUCTIVITY IN LIGHT OF R D SPILLOVER*. International Journal of Information Technology, Modeling and Computing (IJITMC).Volume 4(2):55-63.
- [2] MacQueen, J. B. 1967. *Some Methods for classification and Analysis of Multivariate Observations, Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley: University of California Press.Volume 1: 281-297.
- [3] Strang, G. 2007. *Computational Science and engineering*. USA: Wellesley-Cabridge Press.
- [4] Wendenhall III, W., Beaver, R. J., dan Beaver, B. M. 2012. *Introduction to Probability and Statistics 14-th Edition*. Boston: Brooks/Cole.



## LAMPIRAN

### *Coding R*

#### *Load library*

```
# load library
# untuk visualisasi
library(tidyverse)
library(ggmosaic)
library(ggplot2)

# untuk mengolah data
library(dplyr)
library(reshape)
library(tidyr)
library(lubridate)

# untuk clustering
library(cluster)
library(factoextra)
library(fpc)
```

#### *Load data*

```
# load working directory
getwd()
setwd("C:/Users/user/Downloads")
crime <- read.csv("crime.csv")
```

#### *Pengolahan data*

```
# membuang kode kasus dan secondary description
crime <- crime[,2:8]
crime <- crime[,!names(crime) %in% c("SECONDARY.DESRIPTION")]
str(crime)

# pre-processing data
crime$DATE..OF.OCCURRENCE <- as.Date(crime$DATE..OF.OCCURRENCE, "%d/%m/%Y")
crime$BEAT <- as.character(crime$BEAT)

# membuang data yang kosong
crime <- na.omit(crime)

# membuang level yang tidak terpakai
crime <- droplevels(crime)
```

#### *Plot tabel frekuensi*

```
freq_val <-
  crime %>%
  select_if(is.factor) %>%
  select_if(function(x) !is.ordered(x)) %>%
  gather("var", "value") %>%
  group_by(var) %>%
  count(var, value) %>%
  mutate(prop = prop.table(n)) %>%
  filter(prop > .02)

freq_plot <-
```

```

ggplot(data = cat_value_freq,
       aes(x = reorder(stringr::str_wrap(value, 20), prop),
           y = prop)) + ggtitle("Tabel Frekuensi")+
geom_bar(stat = "identity", fill = "tomato3") +
coord_flip() +
facet_wrap(~var, ncol = 3, scales = "free") +
ggthemes::theme_fivethirtyeight()

freq_plot

```

### Plot *mosaic*

```

mosaic_plot <- ggplot(data = crime) +
  geom_mosaic(aes(x =product(PRIMARY.DESRIPTION), fill=ARREST), na.rm=
    TRUE) +
  labs(x="Primary Description ", y="Arrest",
       title='Plot Mosaic Primary Description terhadap Arrest') +
  coord_flip()

mosaic_plot

```

### Plot *Heatmap*

```

heatmap_data <- data.frame(heatmap_data)
heatmap_plot <- ggplot(heatmap_data, aes(BEAT, PRIMARY.DESRIPTION)) +
  geom_tile(aes(fill = Freq)) +
  scale_fill_gradient(low = "red", high = "yellow")

print(heatmap_plot + labs(
  title= "Heatmap Primary Description terhadap Beat",
  y="Primary Description", x = "Beat (tanpa digit terakhir)")
)

```

### Algoritma *Clustering*

```

# mengelompokkan berdasarkan primary description per beat and per arrest
crime.second <- crime %>%
  group_by(PRIMARY.DESRIPTION) %>%
  count(BEAT, ARREST)

# membuat model matrix dari primary description
crime.second.matrix <- cbind(as.character(crime.second$PRIMARY.DESRIPTION)
,
model.matrix(PRIMARY.DESRIPTION~., crime.second)[,-1]) %>% as.data.frame()

# memperbaiki tipe data
crime.second.matrix[,2:276] <- lapply(crime.second.matrix[,2:276], as.
  character)
crime.second.matrix[,2:276] <- lapply(crime.second.matrix[,2:276], as.
  numeric)

# menjumlahkan kolom berdasarkan primary description
crime.second.matrix.a <- crime.second.matrix %>%
  group_by(V1) %>%
  summarise_all(funs(sum))

# menamakan ulang kolom
names(crime.second.matrix.a)[1]<-"PRIMARY.DESRIPTION"

# membuat kolom baru untuk ARREST N
crime.second.matrix.a$ARREST_N <- crime.second.matrix.a$n - crime.second.
  matrix.a$ARRESTY

```

```

# membuang kolom tidak terpakai
crime.primary.matrix <- crime.second.matrix.a[-275:-276]

# standarisasi data
crime.z <- crime.primary.matrix %>%
  mutate_if(is.numeric, scale)

# set seed untuk model
set.seed(101)

# k-means dengan 3 clusters
crime.km <- kmeans(crime.z[, -1], 3)

# print hasil clustering
print(crime.km)

# memasukkan cluster ke data asli
crime.second.matrix.a$cluster <- as.factor(crime.km$cluster)

# menamakan ulang level
levels(crime.second.matrix.a$cluster)[levels(crime.second.matrix.a$cluster)
  == "1"] <- "High"
levels(crime.second.matrix.a$cluster)[levels(crime.second.matrix.a$cluster)
  == "2"] <- "Medium"
levels(crime.second.matrix.a$cluster)[levels(crime.second.matrix.a$cluster)
  == "3"] <- "Low"

# menyimpan hasil clustering di csv
crime_cluster_result <- cbind(crime.second.matrix.a[1], crime.second.matrix.
  a[278])
crime_cluster_result <- data.frame(crime_cluster_result)
write.csv(crime_cluster_result, 'crime_cluster_result.csv')

# visualisasi clustering
vis_cluster <- fviz_cluster(crime.km, crime.primary.matrix[, -1])
vis_cluster + geom_text(label=crime.primary.matrix$PRIMARY.DESCRPTION) +
  labs(title = "Visualisasi Hasil Clustering")

```