

## SISTEM CLUSTERING TINDAK KEJAHATAN PENCURIAN DI WILAYAH JAWA BARAT MENGGUNAKAN ALGORITMA K-MEANS

Usep Tatang Suryadi<sup>1</sup>, Yana Supriatna<sup>2</sup>

<sup>1,2</sup>Program Studi Teknik Informatika, STMIK Subang  
Jl. Marsinu No. 5 - Subang, Tlp. 0206-417853 Fax. 0206-411873  
E-mail : [ugie89@gmail.com](mailto:ugie89@gmail.com)<sup>1</sup>, [yanasupriatna126@gmail.com](mailto:yanasupriatna126@gmail.com)<sup>2</sup>.

### Abstrak

Pencurian adalah salah satu tindakan kriminalitas yang banyak kita dapatkan dalam masyarakat. Khususnya di wilayah Jawa Barat, data Polda Jabar pada 2013 mencatat kasus pencurian dengan pemberatan mencapai 3.421 kasus, pencurian dengan kekerasan sebanyak 1.031 kasus serta kasus curanmor sebanyak 7.199 kasus. Kemudian pada tahun 2014, mencatat kasus pencurian dengan pemberatan sebanyak 3.146 kasus, pencurian dengan kekerasan 925 kasus, serta kasus curanmor sebanyak 6.305 kasus. Dari permasalahan tersebut munculah gagasan untuk membuat suatu aplikasi, yang didalamnya dapat melakukan pengelompokan tingkat tindak kejahatan pencurian pada kabupaten/kota di provinsi Jawa Barat. Metodologi yang digunakan adalah algoritma K-Means Clustering. Hasil dari penelitian ini dibentuk 3 cluster, dimana cluster 1 masuk kedalam kategori tingkat pencurian tinggi, cluster 2 masuk kedalam kategori tingkat pencurian sedang dan cluster 3 masuk kedalam tingkat pencurian rendah. Iterasi clustering data pencurian terjadi sebanyak 6 kali iterasi dengan menggunakan WCV dan BCV sebagai acuan iterasi. Terdapat 8 wilayah dengan tingkat pencurian tinggi, 10 wilayah dengan tingkat pencurian sedang dan 9 wilayah dengan tingkat pencurian rendah. Ditemukan pusat cluster dengan Cluster 1 = 297,65 ; 22,375, Cluster 2 = 169,200 ; 11 dan Cluster 3 = 35,222 ; 6,111. Dalam membuat sistem ini menggunakan bahasa pemrograman PHP dan HTML. Datasenya menggunakan MySQL. Pada penelitian ini algoritma K-Means clustering dapat melakukan klusterisasi dengan ideal.

**Kata Kunci :** Kriminalitas, Pencurian, K-Means Clustering, PHP, MySQL.

### I. PENDAHULUAN

Menurut [1] pencurian adalah salah satu tindakan kriminalitas yang banyak kita dapatkan dalam masyarakat. Bentuk kejahatan lainnya yang sering terjadi salah satunya adalah tindak pidana pencurian dengan kekerasan. Khususnya di wilayah Jawa Barat, menurut [2] data Polda Jabar, pada 2013 mencatat kasus pencurian dengan pemberatan mencapai 3.421 kasus, pencurian dengan kekerasan sebanyak 1.031 kasus serta kasus curanmor sebanyak 7.199 kasus. Kemudian pada tahun 2014, mencatat kasus pencurian dengan pemberatan sebanyak 3.146 kasus, pencurian dengan kekerasan 925 kasus, serta kasus curanmor sebanyak 6.305 kasus.

Dengan banyaknya kasus kejahatan pencurian di Jawa Barat, perlu adanya sebuah sistem komputerisasi pengelompokan daerah rawan tindak kejahatan pencurian sebagai salah satu usaha untuk membantu pihak kepolisian dalam mengambil keputusan apakah suatu daerah memerlukan pengawasan ekstra atau tidak, selain itu informasi tersebut dibutuhkan untuk mengetahui intensitas tindak kejahatan pencurian. Pengelompokan tindak kejahatan pencurian dapat dilakukan dengan menggunakan teknik *data mining* berupa analisis kelompok. Menurut [3], salah satu metode analisis kelompok yang dapat digunakan adalah metode non hierarki, yaitu *K-Means Clustering*.

Metode ini merupakan metode yang tepat untuk data yang cukup besar, selain itu proses pengelompokan pada metode non hierarki lebih cepat daripada metode hierarki.

Dalam penelitian yang dilakukan oleh [4], analisis *k-means clustering* dapat menghasilkan pengelompokan daerah rawan kriminalitas di Indonesia. Dan penelitian yang dilakukan oleh [5] *k-means clustering* dapat digunakan sebagai referensi dalam menentukan strategi yang tepat bagi tim promosi kampus STIKOM Dinamika Bangsa Jambi.

Berdasarkan hal tersebut, penulis ingin mengkaji lebih dalam mengenai bagaimana pengelompokan tindak kejahatan pencurian di wilayah Jawa Barat. Sehingga dalam penelitian ini dapat diketahui pengelompokan daerah rawan tindak kejahatan pencurian di Jawa Barat pada tahun 2014 berdasarkan data dari Badan Pusat Statistika (BPS) kabupaten Subang.

## II. TINJAUAN PUSTAKA

### 21. Tindak Pidana Pencurian

Para ahli tidak memberikan definisi tentang pencurian, tetapi unsur unsur dan elemennya saja yang berdasarkan pasal 362 KUHP, diantaranya R. Soesilo dalam [6] mengemukakan bahwa: Barang siapa mengambil barang sesuatu yang seluruhnya atau sebagian kepunyaan orang lain, dengan maksud untuk memiliki secara melawan hukum, diancam karena pencurian dengan, pidana penjara paling lama lima tahun atau pidana denda paling banyak Rp. 900,-.

Menurut Andi Hamzah dalam [6], delik pencurian adalah delik yang paling umum tercantum dalam semua KUHP di dunia, yang disebut delik netral karena terjadi dan diatur oleh semua Negara.

#### 2.4.1 Jenis-jenis Pencurian

Pencurian menurut KUHP terdiri dari lima yaitu:

a) Pencurian biasa

Istilah “pencurian biasa” digunakan oleh pakar hukum pidana untuk Menunjuk pengertian “pencurian dalam arti pokok”. Pencurian biasa ini perumusannya diatur dalam pasal 362 KUHP yang menyatakan: Barangsiaapa mengambil sesuatu barang, yang sama sekali atau sebagian termasuk kepunyaan orang lain, dengan maksud akan memiliki barang itu dengan melawan hak, dihukum, karena pencurian, dengan hukuman penjara selama-lamanya lima tahun atau denda sebanyak-banyaknya Rp.900,-

b) Pencurian ringan

Jenis pencurian ini diatur dalam ketentuan pasal 364 KUHP yang menyatakan: Perbuatan yang diterangkan dalam pasal 362 dan pasal 363 ke-4 begitu juga perbuatan yang diterangkan dalam pasal 365 ke-5, apabila tidak dilakukan dalam sebuah rumah atau pekarangan tertutup yang ada rumahnya, jika harga barang yang dicuri tidak lebih dari dua puluh lima ribu rupiah, dihukum sebagai pencurian ringan, pidana penjara paling lama tiga bulan atau denda paling banyak Sembilan ratus rupiah.

c) Pencurian dengan pemberatan

Suatu perbuatan dapat digolongkan sebagai pencurian berat, apabila memenuhi unsur-unsur Pasal 362 KUHP, juga harus memenuhi unsur lain yang terdapat dalam Pasal 363 KUHP.

d) Pencurian Dengan Kekerasan

Pencurian dengan pemberatan kedua adalah pencurian yang diatur dalam pasal 365 KUHP. Jenis pencurian ini lazim disebut dengan istilah “pencurian dengan kekerasan” atau populer dengan istilah “curas”

e) Pencurian Dalam Kalangan Keluarga

Pencurian sebagaimana diatur dalam ketentuan Pasal 367 KUHP ini merupakan pencurian dikalangan keluarga. Artinya baik pelaku maupun korbannya masih dalam satu keluarga, misalnya yang terjadi, apabila seorang suami atau istri melakukan (sendiri) atau membantu (orang lain) pencurian terhadap harta benda istri atau suaminya.

### 22. Data Mining

Menurut Feelders, Daniels dan Holsheimer dalam [7] data mining adalah proses penggalian informasi dari data set yang besar melalui penggunaan algoritma dan teknik yang diambil dari

bidang statistik, *machine learning* dan database sistem manajemen. Ada 6 langkah penting dalam proses data mining, diantaranya adalah definisi masalah, akuisi pengetahuan, seleksi data, *preprocessing* data, analisis dan interpretasi, serta penggunaan dan pelaporan.

Menurut Larose, dalam [8], data *mining* merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data. Data *mining* merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar.

Dari definisi yang telah disampaikan, hal penting yang terkait dengan data *mining* adalah [8]:

- 1 Data *mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
- 2 Data yang akan diproses berupa data yang sangat besar.
- 3 Tujuan data *mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

### 23. Clustering

Menurut [9] pengelompokan atau Clustering merupakan suatu teknik data mining yang digunakan untuk menganalisis data untuk memecahkan permasalahan dalam pengelompokan data atau lebih tepatnya mempartisi dari dataset ke dalam subset.

Tujuan dari *Clustering* data dapat dibedakan menjadi dua, yakni adalah sebagai berikut

- 1 Pengelompokan untuk pemahaman terbentuk harus menangkap struktur alami data yang bertujuan hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan inti seperti peringkasan atau *summarization* (rata-rata standar deviasi), pemberian label kelas pada setiap kelompok bertujuan untuk digunakan sebagai data latih klasifikasi.
- 2 Pengelompokan untuk penggunaan bertujuan untuk mencari prototipe kelompok yang paling representatif terhadap data, memberikan abstraksi dari setiap objek data dalam kelompok dimana sebuah data terletak di dalamnya..

### 24. Algoritma K-Means

*K-Means* merupakan Algoritma *clustering* yang pertama kali diperkenalkan oleh James B MacQueen pada tahun 1976. Metode ini merupakan suatu metode clustering non-heirarchial yang umum digunakan yang relatif sederhana untuk mengelompokkan data dalam jumlah besar.

*K-Means* merupakan metode klasterisasi yang sering digunakan diberbagai bidang karena penggunaannya sederhana, mudah untuk di implementasikan, mampu untuk mengklaster data yang besar. Algoritma *K-Means* merupakan metode berbasis jarak yang membagi data kedalam sejumlah cluster dan dalam setiap tahapan tertentu setiap objek harus masuk dalam kelompok, pada tahap selanjutnya objek dapat berpindah ke kelompok lain. Dalam algoritma ini pada dasarnya melakukan proses clustering tetapi tergantung dari data yang didapat dan konklusi yang dicapai. Maka dari itu algoritma *K-Means* mempunyai aturan yaitu ada jumlah cluster yang akan diinputkan dan hanya dapat memiliki atribut yang bertipe numerik.

Pada awalnya dalam algoritma *K-Means* melakukan pengambilan sebagian dari banyaknya populasi untuk dijadikan cluster awal. Ada banyak cara dalam memberi nilai awal misalnya dengan pengambilan data sampel awal dari objek. Pusat cluster dipilih secara acak yang berada dari beberapa populasi data. Setelah mendapatkan pusat *cluster* awal, algoritma *K-Means* melakukan pengujian masing-masing komponen ke salah satu pusat *cluster* yang telah didefinisikan jarak minimumnya antar komponen dengan tiap-tiap pusat cluster. Posisi pusat cluster akan melakukan perhitungan kembali sampai semua komponen data dapat digolongkan kedalam setiap *cluster* dan akan membentuk posisi *cluster* baru.

Algoritma *K-Means* 3 komponen didalamnya yaitu,

#### 1. Jumlah Cluster K

Dalam metode ini jumlah k harus ditentukan dulu, setelah jumlah k didapatkan dengan melalui pendekatan metode hirarki dapat melakukan pengambilan cluster awal. Aturan khusus dalam menentukan jumlah *cluster* k bahkan tidak ada, namun ada juga jumlah *cluster* yang diinginkan sesuai dengan kebutuhan subjektif seseorang.

## 2. Cluster Awal

Ada banyak pendapat saat melakukan pengambilan *cluster* awal untuk metode *K-Means* misalnya pemilihan terhadap interval dari jumlah setiap observasi, melalui pendekatan salah satu metode hirarki dan ada juga dengan melalui pemilihan *cluster* secara acak dari sekumpulan observasi. Maka dengan adanya beberapa cara pengambilan *cluster* awal tersebut dapat memungkinkan solusi terbaik yang dihasilkan.

## 3. Ukuran Jarak

Pada tahapan ini ukuran jarak juga penting dalam menempatkan observasi ke dalam *cluster* berdasarkan nilai *centroid* terdekat. *Euclidian Distance* adalah jarak yang digunakan untuk mengukur jarak dalam metode *K-Means*.

Berikut ini adalah rumus *Euclidian Distance*:

$$Dik = \sqrt{\sum_{j=1}^m (Xij - Ckj)^2} \dots\dots\dots (2.1)$$

Keterangan :

*Dik* = titik dokumen/jarak data ke-i

*M* = jumlah variable

*Xij* = data yang akan dilakukan pengclusteran

*Ckj* = pusat dari cluster

Jarak yang terpendek antara *centroid* dengan dokumen menentukan posisi *cluster* suatu dokumen.

Berikut merupakan rumus penentuan *centroid* baru :

$$Vij = \frac{\sum_{k=0}^{Ni} Xkj}{Ni} \dots\dots\dots (2.2)$$

Keterangan :

*Vij* = centroid/ rata-rata cluster ke-i untuk variable ke-j

*Ni* = jumlah data yang menjadi anggota cluster ke-i

*i.k* = indeks dari cluster

*j* = indeks dari variable

*Xkj* = nilai data ke-k yang ada di dalam cluster tersebut untuk variable ke-j

Algoritma *K-Means* bersifat *partitional clustering* yaitu dengan cara membagi himpunan objek data ke dalam sub himpunan (*cluster*) yang tidak overlap, sehingga hasil dari setiap objek data berada tepat dalam satu *cluster*.

### 2.4.1 Tahapan Algoritma *K-Means*

Menurut [10] menyatakan:

1. Memilih objek k secara acak, setelah mendapatkan objek k tersebut data akan diproses sebagai *mean* pada *cluster*.
2. Setiap objek akan dimasukkan kedalam *cluster* yang mempunyai kemiripan terhadap *cluster*. Tingkat kemiripan dapat ditentukan dengan mencari jarak objek terhadap *mean* atau *centroid* cluster tersebut.
3. Lakukan perhitungan nilai *centroid* yang baru pada setiap cluster.
4. Proses perhitungan nilai *centroid* tersebut dilakukan berulang-ulang hingga didapati anggota pada kelompok *cluster* tersebut tidak berubah.

Setelah didapat hasil dari proses *clustering* dengan algoritma *K-Means* maka perlu diadakan suatu evaluasi pengujian. Pengujian tersebut dengan menggunakan metode *sum squared error (SSE)*. Selain menggunakan *SSE* bisa juga dengan cara melakukan perhitungan perbandingan antara *Between-Class Variation (BCV)* dan *Within-Class Variation (WCV)* pada iterasi terakhir yang sering disebut dengan rasio. Apabila hasil perhitungan pengujian yang diperoleh besar, maka semakin bagus juga tingkat kualitas *clustering* tersebut [11].

*BCV* merupakan rata-rata dari *centroid*, sedangkan *WCV* adalah *SSE* itu sendiri. Rumus perhitungannya adalah sebagai berikut :

$$BCV = \frac{1}{nk} \sum_{i=1}^k d(mi, mi) \dots\dots\dots (2.3)$$

Keterangan :

k = jumlah cluster

i = nama yang mewakili cluster yang dibentuk

mi = jumlah anggota dari cluster ke-i

$$WCV(SSE) = \sum_j^N \sum_{p \in C_i} d(p, mi)^2 \dots\dots\dots (2.4)$$

Keterangan :

p ∈ Ci = jumlah semua data

k = jumlah cluster

p = cluster jarak terdekat

mi = jumlah anggota dari cluster ke-i

$$\text{Rasio} = \frac{BCV}{WCV} \dots\dots\dots (2.5)$$

### III. METODE

Pada penelitian ini penulis menggunakan metode deskriptif, yaitu metode yang menggambarkan suatu keadaan atau permasalahan yang sedang terjadi berdasarkan fakta dan data – data yang diperoleh dan dikumpulkan pada waktu melaksanakan penelitian.

#### 3.1 Metode Pengumpulan Data

Metode Pengumpulan data yang dilakukan adalah :

##### a. Field Research / Penelitian Langsung

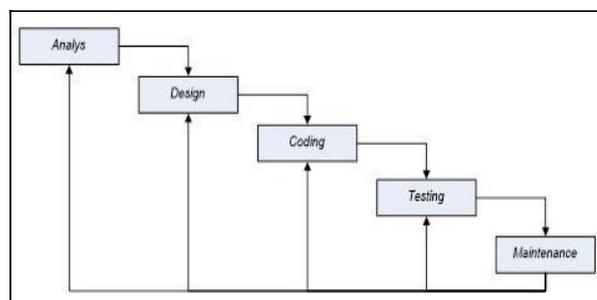
Melakukan Penelitian langsung ke lapangan yaitu penelitian dilakukan langsung kepada pihak Badan Pusat Statistika (BPS) kabupaten Subang untuk mendapatkan data dimana penulis melakukan pengumpulan data dengan cara mencatat data tindak kejahatan yang telah terjadi di provinsi Jawa Barat pada tahun 2014.

##### b. Library Search(Study Pustaka)

Mengadakan penelitian dengan mencari pendukung sumber data tersebut guna mengumpulkan data yang bersifat teoritis yang berkaitan dengan penelitian dan penyusunan tugas akhir. Adapun metode dalam aplikasi ini menggunakan Metode *K-means Clustering*.

#### 3.2 Metode Perancangan Aplikasi

Pengembangan aplikasi ini menggunakan metode skensila linier (*Waterfall*). Metode *waterfall* merupakan metode pengembangan perangkat lunak sistematis dan skuensial yang mulai pada tingkat dan kemajuan sistem sampai analisis, desain, kode, test, dan pemeliharaan. Berikut adalah tahapan dari model *waterfall* [12].



Gambar 1. 1 Model waterfall (Pressman, 2001)

Berdasarkan model metode pengembangan sistem yang telah digambarkan diatas, maka dapat diuraikan pembahasan masing – masing tahap dalam penelitian sebagai berikut :

##### a. Analisis

Proses menganalisis dan pengumpulan kebutuhan sistem yang sesuai dengan domain informasi tingkah laku, untuk kerja, melakukan perhitungan *K-Means Clustering*.

*b. Design*

Dalam tahap ini penulis akan merancang desain dan model aplikasi yang akan dikembangkan berdasarkan hasil analisa pada tahap sebelumnya, diantaranya adalah model data (ERD), diagram konteks, DFD, struktur tabel, rancangan antar muka, dan struktur menu.

*c. Coding*

Pengkodean (coding) merupakan proses penerjemahan desain ke dalam suatu bahasa yang bisa dimengerti oleh komputer. Dalam proses ini penulis menggunakan PHP sebagai bahasa pemrograman dan DBMS MySQL.

*d. Testing*

Proses pengujian berfokus pada logika *internal software*, memastikan bahwa semua pernyataan sudah diuji, dan pada eksternal fungsional, yaitu mengarahkan pengujian untuk menemukan kesalahan-kesalahan dan memastikan bahwa input yang dibatasi akan memberikan hasil *actual* yang sesuai dengan hasil yang dibutuhkan. Pada penelitian ini penulis menggunakan teknik *black box* untuk menguji fitur-fitur sistem yang telah dibangun.

*e. Maintenance*

Menjalankan dan melakukan pemeliharaan sistem yang telah dibuat. Pemeliharaan ini termasuk memperbaiki kesalahan yang tidak ditemukan pada tahap sebelumnya.

**IV. HASIL DAN PEMBAHASAN**

Dalam penelitian ini terdapat 27 data kabupaten/ kota di provinsi Jawa Barat, yang akan di proses kedalam perhitungan *k-means clustering* dengan berdasarkan dua atribut tindak kejahatan yaitu pencurian dan kekerasan. Tahap awal dalam menentukan pengelompokan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, yaitu:

1. Tentukan jumlah cluster yang diinginkan. Dalam penelitian ini data-data yang ada akan dikelompokkan menjadi tiga *cluster*.
2. Tentukan titik pusat awal dari setiap cluster. Dalam penelitian ini titik pusat awal ditentukan secara random dan didapat titik pusat dari setiap cluster, dapat dilihat pada Tabel 4.1 berikut:

Tabel 4.1 Titik pusat awal setiap *cluster*

Titik Pusat Awal	Kabupaten/Kota	Pencurian	Pencurian dengan Kekerasan
C1	Sukabumi	344	14
C2	Indramayu	223	31
C3	Kota Banjar	18	1

3. Tempatkan setiap data pada cluster. Dalam penelitian ini digunakan metode *hard k-means* untuk mengalokasikan setiap data ke dalam suatu *cluster*, sehingga data akan dimasukkan dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat dari setiap *cluster*.
4. Untuk mengetahui *cluster* mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap *cluster*.

Untuk perhitungan pertama, akan dihitung jarak dari data kabupaten/kota pertama ke pusat *cluster* pertama dengan menggunakan rumus 2.1.

Data: (391, 36) , centroid C1: (344, 14), centroid C2: (223, 31), centroid C3: (18, 1).

Jarak data kabupaten/kota pertama ke pusat *centroid* 1, 2 dan 3 yaitu:

$$\text{Data C1} = \sqrt{(391 - 344)^2 + (36 - 14)^2} = 51.89412298$$

$$\text{Data, C2} = \sqrt{(391 - 223)^2 + (36 - 31)^2} = 168.0743883$$

$$\text{Data, C3} = \sqrt{(391 - 18)^2 + (36 - 1)^2} = 374.6384924$$

Dengan perhitungan jarak seluruh data terhadap 3 *centroid* yang sudah ditetapkan, maka dapat dilihat jarak semua kabupaten/kota ke *centroid* 1, 2 dan 3 terdapat pada tabel 4.2 berikut:

Tabel 4.2 Hasil perhitungan iterasi ke- 1

NO	Kabupaten/Kota	Pencurian	Pencurian dengan Kekerasan	Jarak Ke Centroid			Jarak Terdekat
				Cluster 1	Cluster 2	Cluster 3	
1	Bogor	391	36	51.8941	168.0744	374.6385	1
2	Sukabumi	344	14	0	122.1884	326.2591	1
3	Cianjur	268	16	76.0263	47.4342	250.4496	2
4	Bandung	255	19	89.1403	34.1760	237.6826	2
5	Garut	310	13	34.0147	88.8426	292.2465	1
6	Tasikmalaya	219	2	125.5747	29.2746	201.0025	2
7	Ciamis	161	2	183.3930	68.4471	143.0035	2
8	Kuningan	176	1	168.5022	55.7584	158.0000	2
9	Cirebon	278	32	68.4105	55.0091	261.8416	2
10	Majalengka	270	10	74.1080	51.4782	252.1607	2
11	Sumedang	175	8	169.1065	53.2259	157.1560	2
12	Indramayu	223	31	122.1884	0.0000	207.1835	2
13	Subang	176	8	168.1071	52.3259	158.1550	2
14	Purwakarta	141	3	203.2978	86.6487	123.0163	2
15	Karawang	265	39	82.8613	42.7551	249.9060	2
16	Bekasi	165	17	179.0251	59.6657	147.8682	2
17	Bandung Barat	123	7	221.1108	102.8397	105.1713	2
18	Pangandaran	27	0	317.3090	198.4364	9.0554	3
19	Kota Bogor	54	7	290.0845	170.6956	36.4966	3
20	Kota Sukabumi	22	3	322.1878	202.9409	4.4721	3
21	Kota Bandung	133	31	211.6837	90	118.8486	2
22	Kota Cirebon	15	5	329.1231	209.6187	5	3
23	Kota Bekasi	44	13	300.0017	179.9028	28.6356	3
24	Kota Depok	58	20	286.0629	165.3663	44.2832	3
25	Kota Cimahi	15	0	329.2977	210.2974	3.1623	3
26	Kota Tasikmalaya	64	6	280.1143	160.9534	46.2709	3
27	Kota Banjar	18	1	326.2591	207.1835	0	3

Pada langkah selanjutnya hitung pula rasio antara besaran BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*).

Tabel 4.3 Centroid awal

Titik Pusat Awal	Kabupaten/ Kota	Pencurian	Pencurian dengan Kekerasan
C1	Sukabumi	344	14
C2	Indramayu	223	31
C3	Kota Banjar	18	1

Untuk menghitung BCV nya adalah sebagai berikut:

$$BCV = \frac{1}{nk} \sum_{i=1}^k d(m_i, m_i)$$

$$M1 = (344, 14), M2 = (223, 31), M3 = (18, 1)$$

$$d(m1, m2) = \sqrt{(344 - 223)^2 + (14 - 31)^2} = 122.1883792$$

$$d(m1, m3) = \sqrt{(344 - 18)^2 + (14 - 1)^2} = 326.2590995$$

$$d(m2, m3) = \sqrt{(223 - 18)^2 + (31 - 1)^2} = 207.1834936$$

$$BCV = d(m1, m2) + d(m1, m3) + d(m2, m3) = 122.1883792 + 326.2590995 + 207.1834936 = 655.630972$$

Maka Nilai *Between Cluster Variation* (BCV) nya adalah 655,630972

Selanjutnya adalah menghitung WCV yaitu dengan memilih jarak terkecil antara data dengan centroid pada masing-masing cluster:

Tabel 4.4 Jarak terdekat dan WCV

NO	Kabupaten/ Kota	Jarak Terdekat	Proses WCV
1	Bogor	51.89412298	2693
2	Sukabumi	0	0
3	Cianjur	47.4341649	2250
4	Bandung	34.17601498	1168
5	Garut	34.0147027	1157
6	Tasikmalaya	29.27456234	857
7	Ciamis	68.44705983	4685
8	Kuningan	55.75840744	3109
9	Cirebon	55.00909016	3026
10	Majalengka	51.4781507	2650
11	Sumedang	53.22593353	2833
12	Indramayu	0	0
13	Subang	52.32590181	2738
14	Purwakarta	86.64871609	7508
15	Karawang	42.75511665	1828
16	Bekasi	59.66573556	3560
17	Bandung Barat	102.8396811	10576
18	Pangandaran	9.055385138	82
19	Kota Bogor	36.49657518	1332
20	Kota Sukabumi	4.472135955	20
21	Kota Bandung	90	8100
22	Kota Cirebon	5	25
23	Kota Bekasi	28.63564213	820
24	Kota Depok	44.28317965	1961
25	Kota Cimahi	3.16227766	10
26	Kota Tasikmalaya	46.27094121	2141
27	Kota Banjar	0	0

**Nilai WCV 65129**

Jadi Nilai *Within Cluster Variation* nya adalah **65129**

Pada tahapan akhir dilakukan perhitungan rasio dengan cara menghitung perbandingan BCV dan WCV atau SSE pada akhir iterasi.

Jika hasil perhitungan pengujian rasio yang diperoleh besar, maka semakin bagus juga tingkat kualitas *clustering*.

$$\begin{aligned} \text{Rasio} &= \frac{BCV}{WCV} \\ &= \frac{655,630972}{65129} = 0,010066652 \end{aligned}$$

Karena langkah ini merupakan proses awal maka dilanjutkan pada tahap selanjutnya dengan mengulang langkah yang sama dengan kata lain iterasi berlanjut.

Untuk proses iterasi selanjutnya perlu dilakukan penentuan centroid baru dengan menghitung rata-rata nilai pada masing-masing cluster. Seperti pada Tabel 4.5 di bawah ini.

Tabel 4.4 Pembaruan centroid iterasi ke 1

Cluster 1		
Kabupaten/ Kota	Pencurian	Pencurian dengan Kekerasan
Bogor	391	36
Sukabumi	344	14
Garut	310	13
<b>Mean</b>	<b>348.3333333</b>	<b>21</b>

Cluster 2		
Kabupaten/ Kota	Pencurian	Pencurian dengan Kekerasan
Cianjur	268	16
Bandung	255	19
Tasikmalaya	219	2
Ciamis	161	2
Kuningan	176	1
Cirebon	278	32
Majalengka	270	10
Sumedang	175	8
Indramayu	223	31
Subang	176	8
Purwakarta	141	3
Karawang	265	39
Bekasi	165	17
Bandung Barat	123	7
Kota Bandung	133	31
<b>Mean</b>	<b>201.8666667</b>	<b>15.06666667</b>

Cluster 3		
Kabupaten/ Kota	Pencurian	Pencurian dengan Kekerasan

Cluster 3		
Kabupaten/ Kota	Pencurian	Pencurian dengan Kekerasan
Pangandaran	27	0
Kota Bogor	54	7
Kota Sukabumi	22	3
Kota Cirebon	15	5
Kota Bekasi	44	13
Kota Depok	58	20
Kota Cimahi	15	0
Kota Tasikmalaya	64	6
Kota Banjar	18	1
<b>Mean</b>	<b>35.11111111</b>	<b>6.11111111</b>

Dari hasil perhitungan di atas di dapatkan centroid baru yaitu:

$c1 = (348.3333333, 21)$

$c2 = (201.8666667, 15.06666667)$ ,

$c3 = (35.11111111, 6.11111111)$ .

Dengan nilai *centroid* baru tersebut lakukan perhitungan iterasi selanjutnya hingga terlihat bahwa tidak ada data yang berpindah cluster serta nilai ratio sama dengan nilai ratio pada iterasi sebelumnya. Ketika nilai ratio sama dan tidak ada data yang berpindah cluster maka perhitungan iterasi dihentikan.

Dalam penelitian ini perhitungan iterasi berhenti pada iterasi ke-6 karena pada iterasi ke-6 tidak ada data yang berpindah cluster dan nilai ratio sama dengan iterasi sebelumnya.

Berikut adalah perbandingan WCV dan RATIO pada proses iterasi 1, 2, 3, 4, 5 dan 6 seperti pada tabel 4.5 dibawah ini :

Tabel 4.5 Perbandingan BCV, WCV dan Ratio iterasi ke 1, 2, 3, 4, 5 dan 6

Nilai	Iterasi 1	Iterasi 2	Iterasi 3	Iterasi 4	Iterasi 5	Iterasi 6
BCV	655.630972	627.1585412	592.1312937	520.5593861	525.9010396	525.9010396
WCV	65129	50498.64398	45341.24753	34104.6049	31355.79444	31355.79444
RASIO	0.010066652	0.012419314	0.01305944	0.01526361	0.016772053	0.016772053

Setelah proses perhitungan iterasi selesai atau terhenti maka selanjutnya akan ditampilkan hasil dari masing-masing data kabupaten/kota dengan jarak terdekat dengan masing-masing cluster.

Tabel 4.6 Data wilayah cluster 1

NO	Kabupaten/Kota	Pencurian	Pencurian dengan Kekerasan	Jarak Terdekat Ke		
				C1	C2	C3
1	Bogor	391	36	OK		
2	Sukabumi	344	14	OK		
3	Cianjur	268	16	OK		
4	Bandung	255	19	OK		
5	Garut	310	13	OK		
6	Cirebon	278	32	OK		
7	Majalengka	270	10	OK		
8	Karawang	265	39	OK		

Tabel 4.7 Data wilayah cluster 2

NO	Kabupaten/Kota	Pencurian	Pencurian dengan Kekerasan	Jarak Terdekat Ke		
				C1	C2	C3
1	Tasikmalaya	219	2		OK	
2	Ciamis	161	2		OK	
3	Kuningan	176	1		OK	
4	Sumedang	175	8		OK	
5	Indramayu	223	31		OK	
6	Subang	176	8		OK	
7	Purwakarta	141	3		OK	
8	Bekasi	165	17		OK	
9	Bandung Barat	123	7		OK	
10	Kota Bandung	133	31		OK	

Tabel 4.8 Data wilayah cluster 3

NO	Kabupaten/Kota	Pencurian	Pencurian dengan Kekerasan	Jarak Terdekat Ke		
				C1	C2	C3
1	Pangandaran	27	0			OK
2	Kota Bogor	54	7			OK
3	Kota Sukabumi	22	3			OK
4	Kota Cirebon	15	5			OK
5	Kota Bekasi	44	13			OK
6	Kota Depok	58	20			OK
7	Kota Cimahi	15	0			OK
8	Kota Tasikmalaya	64	6			OK
9	Kota Banjar	18	1			OK

Dari hasil cluster 1, terlihat bahwa kabupaten/kota yang mempunyai tingkat kejahatan pencurian tinggi terdapat 8 wilayah yaitu: Bogor, Sukabumi, Cianjur, Bandung, Garut, Cirebon, Majalengka dan Karawang.

Kemudian, dari hasil cluster 2 dapat dilihat bahwa kabupaten/kota yang mempunyai tingkat kejahatan pencurian sedang terdapat 10 wilayah yaitu: Tasikmalaya, Ciamis, Kuningan, Sumedang, Indramayu, Subang, Purwakarta, Bekasi, Bandung Barat dan Kota Bandung.

Sedangkan dari hasil cluster 3, dapat dilihat bahwa kabupaten/kota yang mempunyai tingkat kejahatan pencurian rendah terdapat 9 wilayah yaitu: Pangandaran, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya dan Kota Banjar.

## V. PENUTUP

### Kesimpulan

Setelah menyelesaikan pembuatan laporan dan sistem *clustering* tindak kejahatan pencurian di wilayah Jawa Barat menggunakan metode *k-means*, maka penulis mengambil kesimpulan sebagai berikut:

1. Algoritma *k-means clustering* dapat digunakan untuk mengelompokan tingkat tindak kejahatan pencurian di wilayah Jawa Barat.
2. Pengujian yang dilakukan dalam penelitian ini, iterasi *clustering* data pencurian di wilayah Jawa Barat terjadi sebanyak 6 kali iterasi.
3. Terdapat 8 wilayah dengan tingkat pencurian tinggi, yaitu Bogor, Sukabumi, Cianjur, Bandung, Garut, Cirebon, Majalengka dan Karawang. Terdapat 10 wilayah dengan tingkat tindak kejahatan pencurian sedang, yaitu Tasikmalaya, Ciamis, Kuningan, Sumedang, Indramayu, Subang, Purwakarta, Bekasi, Bandung Barat dan Kota Bandung. Sedangkan wilayah dengan tingkat pencurian rendah terdapat 9 wilayah, yaitu Pangandaran, Kota Bogor, Kota Sukabumi, Kota Cirebon, Kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya dan Kota Banjar.
4. Pengujian yang dilakukan ditemukan pusat *cluster* dengan *Cluster 1* = 297,65 ; 22,375, *Cluster 2* = 169,200 ; 11 dan *Cluster 3* = 35,222 ; 6,111.

## Saran

Dengan sistem yang masih terbatas, maka penulis tugas akhir ini masih memiliki kekurangan-kekurangan yang perlu diperhatikan supaya lebih baik untuk kedepannya. Untuk itu saran untuk pengembang selanjutnya adalah sebagai berikut:

1. Untuk pengembangan penelitian selanjutnya dapat menggunakan algoritma *Clustering* lainnya.
2. Pada penelitian selanjutnya dapat dilakukan komparasi dengan metode lain, kemudian dipilih metode yang terbaik untuk digunakan di sistem.
3. Tambahkan kriteria atau variable-variabel tindak kejahatan lain selain tindak kejahatan pencurian seperti penipuan, penganiayaan, pembunuhan dll.
4. Sistem ini dapat dikembangkan dengan menggunakan *Geographic Information System* (GIS).
5. Sistem ini dapat dikembangkan dengan memberikan tambahan fitur *print out* data hasil klasterisasi sebagai keluaran cetak data, sehingga sistem ini bisa menyediakan rekapitulasi data.

## VI. DAFTAR PUSTAKA

- [1] C. Q. Akyunin, "Kajian Viktimologi Tentang Perlindungan Hukum Bagi Korban Pengendara Kendaraan Bermotor Yang Mengalami Tindak Pidana Pencurian Dengan Kekerasan," *Jurnal Ilmiah*, p. Dapat diakses di: <http://hukum.studentjournal.ub.ac.id/index.php/hukum/article/view/121/116>. Diakses 10 Februari 2018, 2013.
- [2] Basri., dkk, "Kajian Yuridis Unsur Tindak Pidana Pencurian Dengan Kekerasan," *The 2nd University Research Coloquium 2015 ISSN 2407-918*, pp. Fakultas Hukum Universitas Muhammadiyah Magelang. Dapat diakses di: <https://media.neliti.com/media/publications/176168-ID-kajian-yuridis-unsur-tindak-pidana-pencu.pdf>. Diakses 10 Februari 2018, 2015.
- [3] J. Supranto, Analisis Multivariat: Arti dan Interpretasi, Jakarta: PT. Rineka Cipta, 2004.
- [4] Hapsari dan Widodo, "Pengelompokan Daerah Rawan Kriminalitas di Indonesia Menggunakan Analisis K-Means Clustering," *Prosiding SI MaNIs (Seminar Nasional Integrasi Matematika dan Nilai Islami) Vol.1, No.1, Juli 2017*, pp. Hal. 147-153, Program Studi Statistika Fakultas MIPA Universitas Islam Indonesia. Dapat diakses di: <http://conferences.uin-malang.ac.id/index.php/SIMANIS/article/view/52>. Diakses 10 Ferbruari 2018, 2017.
- [5] Jasmir, "Analisis Profil Akademik Alumni Dengan Menggunakan Metode Klasterisasi K-Means Untuk Menentukan Strategi Promosi STIKOM Dinamika Bangsa Jambi," *Vol.11, No.1, April 2017*, pp. Teknik Informatika, STIKOM Dinamika Bangsa Jambi. Dapat diakses di: <http://ejournal.stikom-db.ac.id/index.php/mediasisfo/article/download/213/200/>. Diakses 10 Februari 2018, 2017.
- [6] A. R. Yusuf, "Tinjauan Kriminologis Terhadap Kejahatan Pencurian Disertai Kekerasan Yang Dilakukan Oleh Anak Dijalanan (Studi Kasus di Kota Makassar Tahun 2014-2016)," *Departemen Hukum Pidana Fakultas Hukum Universitas Hasanuddin Makassar*, pp. Dapat diakses di: <http://repository.unhas.ac.id/bitstream/handle/123456789/23549/SKRIPSI%20LENGKAP-PIDANA-ALFISYAHHRIN%20R.%20YUSUF.pdf?sequence=1>. Diakses 30 Maret 2018, 2017.
- [7] J. Ranjan, "Application of Data Mining Technique in Pharmaceutical Industry," *Journal of Theoretical and Applied Information Technology*, pp. Page 61 – 67. Dapat diakses di: <http://www.jatit.org/volumes/research-papers/Vol3No4/7vol3no4.pdf>. Diakses 10 Februari 2018, 2009.
- [8] Kusri dan Emha, Algoritma Data Mining, Yogyakarta: Andi Offset, 2009.

- [9] A. K. Wardhani, "Implementasi Algoritma K-Means Untuk Pengelompokan Penyakit Pasien Pada Puskesmas Kajen Pekalongan," *JURNAL TRANSFORMATIKA*, pp. Volume 14, Nomor 1, Juli 2016. Dapat diakses di: <http://journals.usm.ac.id/index.php/transformatika/article/viewFile/387/248>. Diakses 15 Mei 2018, 2016.
- [10] J. Han dan M. Kamber, *Data Mining Concepts and Techniques*, Morgan Kaufmann, 2006.
- [11] Alfina, T., dkk, "Analisa Perbandingan Metode Hierarchical Clustering, K-means dan Gabungan Keduanya dalam Cluster Data (Studi kasus : Problem Kerja Praktek Jurusan Teknik Industri ITS)," *JURNAL TEKNIK ITS Vol. 1, (Sept, 2012) ISSN: 2301-9271*, pp. A-521 - A-525. Dapat diakses di: <http://digilib.its.ac.id/public/ITS-paper-25497-2508100055-Paper.pdf>. Diakses 15 Mei 2018, 2012.
- [12] R. S. Pressman, *Software Engineering : A Practitioner's Approach*, Fifth Ed., New York, McGraw-Hill Book Company, 2001.