

Penerapan *Data Mining* Menggunakan Metode *K-Means* Untuk *Clustering* Tindak Pidana Daerah (Studi Kasus: Data BPS 2018, 2019, 2020)

Muhammad Faisal^{*1}, Wiranti Sri Utami², Doni Prastyo³, Janu Ilham Saputro⁴

^{*1,4}Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Raharja

²Program Studi Teknik Informatika, Fakultas Teknik, Universitas Cendekia Abditama

³Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Syekh Yusuf

Email: muhammad.faisal@raharja.info^{*1}, wirantisutami@uca.ac.id², doniprastyo@unis.ac.id³, janu@raharja.info⁴

Abstrak

Tindak pidana merupakan perbuatan melanggar hukum dan pelaku tindak pidana dapat mendapatkan sanksi pidana. Salah satu upaya Pemerintah dalam menekan tindak pidana disetiap daerah adalah dengan membuka Polres di setiap wilayah atau daerah demi memerangi tindak pidana. Setiap tahun terdapat tindak pidana yang beragam walaupun hal tersebut tidak terjadi secara signifikan. Badan Pusat Statistik (BPS) mendata tindak pidana yang terjadi di setiap Wilayah atau Daerah, data tersebut dapat dilakukan proses analisa pengelompokkan agar lebih akurat dan efektif dalam mendata tindak pidana di suatu Daerah. Penelitian yang dilakukan oleh penulis dirasa cukup penting dikarenakan dapat memberikan informasi terhadap tindak pidana pada tiap Provinsi. Dalam melakukan proses analisa tersebut penulis menggunakan teknik data mining serta menerapkan Algoritma Clustering K-Means. Dataset yang diperoleh didapatkan dari BPS periode 2018, 2019, dan 2020 selanjutnya dataset tersebut diolah menggunakan bahasa R untuk dianalisa menggunakan metode K-Means. Hasil dari Algoritma Clustering K-Means menunjukkan 2 cluster tindak pidana dengan total keseluruhan cluster 73,9 %. Total Cluster Means 1 pada tahun 2018 28.842,5 pada tahun 2019 26.439,5 dan tahun 2020 22.508. Kemudian Cluster Means 2 5.963,7 pada tahun 2019 5.425,2 dan tahun 2020 5.239,533.

Kata kunci — *Data Mining, Clustering, K-Means, Tindak Pidana*

Abstract

Criminal acts are acts that violate the law and perpetrators of criminal acts can receive criminal sanctions. One of the Government's efforts to suppress criminal acts in each region is by opening police stations in each region or area to combat criminal acts. Every year there are various criminal acts, although these do not occur significantly. The Central Statistics Agency (BPS) records criminal acts that occur in each region or area. This data can be carried out through a grouping analysis process to make it more accurate and effective in recording criminal acts in a region. The research conducted by the author is considered quite important because it can provide information on criminal acts in each province. In carrying out the analysis process the author uses data mining techniques and applies the K-Means Clustering Algorithm. The dataset obtained was obtained from BPS for the 2018, 2019 and 2020 periods. The dataset was then processed using the R language for analysis using the K-Means method. The results of the K-Means Clustering Algorithm show 2 clusters of criminal acts with a total cluster of 73.9%. Total Cluster Means 1 in 2018 was 28,842.5, in 2019 26,439.5 and in 2020 22,508. Then Cluster Means 2 was 5,963.7 in 2019 5,425.2 and in 2020 5,239,533.

Keywords — *Data Mining, Clustering, K-Means, Crime*

1. PENDAHULUAN

Tindak pidana adalah merupakan istilah pelanggaran hukum dimana pelaku tindak pidana dapat dijatuhi atau diberikan sanksi oleh negara yang diwakili oleh beberapa Hakim. Tindak Pidana merupakan ancaman serius tidak hanya di darat namun juga dapat terjadi di perairan contohnya laut, dalam hal sektor perikanan wilayah laut jelas merupakan ancaman bagi nelayan dikarenakan terdapat ekosistem laut dan sumber daya perikanan bila salah dalam mengelola maka terdapat sanksi tindak pidana [1]. Dalam hal menekan angka Tindak Pidana maka diperlukan penegak hukum, salah satunya aparat kepolisian, aparat polisi merupakan institusi negara yang memiliki peran menciptakan rasa aman, dan ketertiban masyarakat, dan berperan penting dalam penegakkan hukum untuk melindungi, mengayomi, dan melayani masyarakat [2]. Dalam upaya memberantas tindak pidana maka pemerintah mengeluarkan beberapa aturan dalam upaya menekan kasus tindak pidana, Salah satu contoh seseorang yang melanggar hukum maka akan di kenakan sanksi pidana dalam hal ini kejaksaan RI dan Polri memerankan peran penting dalam memberantas tindak pidana [3].

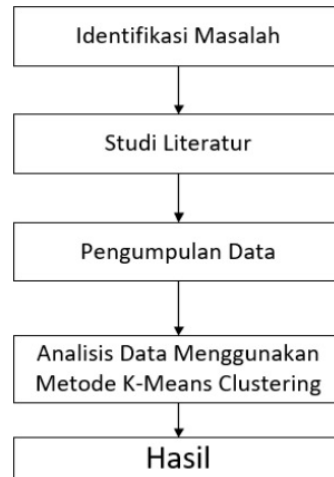
Pengamatan terhadap kasus tindak pidana dapat diamati melalui penelitian menggunakan *Data Mining*. *Data Mining* merupakan konsep dasar yang digunakan dengan tujuan menggali informasi yang belum tersedia dari sebuah *dataset* dan salah satu dalam *Knowledge Discovery in Database (KDD)*, bertujuan menggali informasi dan membantu untuk tugas pengambilan keputusan [4]. *Data Mining* memiliki beberapa teknik atau metode dalam menggali informasi dalam *database*, Salah satunya adalah K-Means. Algoritma K-Means adalah salah satu metode *clustering* data dan dapat mengelompokkan data menjadi beberapa *cluster*, algoritma ini dapat mengelompokkan data kedalam beberapa data yang memiliki karakteristik, dimana karakteristik tersebut menyerupai karakteristik yang sama dan data dengan karakteristik berbeda akan berada pada kelompok yang berbeda [5]. Penelitian mengenai Tindak Pidana kriminalitas juga dapat diproses menggunakan teknik *Data Mining*, Permasalahan mengenai tindak kriminalitas adalah permasalahan yang sangat besar bagi masyarakat Indonesia. Banyak tindakan kriminalitas yang terjadi diberbagai wilayah dan kurun waktu yang berbeda sehingga dibutuhkan informasi geografis tentang tingkat kriminalitas yang tinggi dibutuhkan oleh aparat penegak hukum [6]. Algoritma *Cluster K-Means* yang dipilih pada penelitian ini karena algoritma K-Means merupakan algoritma klusterisasi untuk pengelompokkan data berdasarkan *centroid* dan Nilai rata-rata. Berbeda dengan Algoritma K-Medoids, algoritma K-Medoids tidak menetapkan nilai rata-rata obyek dalam suatu *cluster* melainkan menggunakan median, yaitu menggunakan obyek yang berada pada tengah *cluster* [7].

Algoritma *Cluster K-Means* juga banyak digunakan dalam beberapa penelitian, salah satunya digunakan untuk memprediksi waktu kelulusan mahasiswa, penelitian ini membahas tentang kebutuhan analisis yang berkaitan tentang prediksi kelulusan Mahasiswa data yang diperoleh diantaranya adalah NIM, Nama, Jenis kelamin, Jurusan, IPK, dan kehadiran sebagai data Akademik atau data Mahasiswa [8]. Algoritma *Clustering K-Means* juga dapat digunakan sebagai penelitian dalam pengelompokkan potensi Padi, Penerapan algoritma *Clustering K-Means* dapat digunakan sebagai pendukung keputusan oleh dinas Pertanian Kota Pagar Alam dalam menentukan kelurahan mana yang memiliki potensi Padi rendah, sedang, dan Tinggi [9].

Teknik *Data Mining* dengan penerapan Algoritma *Clustering K-Means* bertujuan menggali informasi pengelompokkan atau *Clustering* tindak pidana berdasarkan dataset yang diperoleh dari BPS. Dengan dilakukannya penelitian ini penulis berharap dapat memberikan informasi terkait tindak Pidana pada setiap daerah.

2. METODOLOGI PENELITIAN

Metode penelitian ini akan dirumuskan kedalam kerangka penelitian. Kerangka penelitian adalah urutan atau langkah-langkah penelitian yang akan dilakukan pada penelitian ini, kerangka penelitian tersebut dapat dilihat pada gambar dibawah ini.



Gambar 1. Kerangka penelitian

Proses penelitian ini penulis telah menyusun beberapa metode penelitian yang diperlukan pada penelitian ini diantaranya sebagai berikut:

2.1 Data Mining

Data Mining adalah teknik proses penggalian informasi yang bermanfaat dari suatu *database* yang sangat besar [10]. Proses ini dilakukan untuk menemukan suatu aturan dalam sebuah penelitian. *Data Mining* teknik dari *Artificial Intelligence* dan *Machine Learning* yang melalui proses ekstrak dan menghasilkan suatu pengetahuan dan informasi yang terdapat didalam *database* [11].

2.2 Clustering

Teknik *clustering* adalah metode yang memiliki tujuan untuk analisa data digunakan sebagai penyelesaian tentang pengelompokkan data [12]. *Clustering* adalah teknik yang digunakan untuk membedakan suatu kumpulan data yang diperoleh dari *database*, hal tersebut bertujuan untuk melihat banyaknya kelompok yang memiliki kemiripan dari kumpulan data [13].

2.3 K-Means

K-Means merupakan sebuah Algoritma yang sederhana karena mudah digunakan, efisien, dan sederhana. Berikut ini beberapa tahapan yang digunakan dalam penerapan algoritma K-Means, diantaranya sebagai berikut [14].

1. Mengimport *dataset* yang akan dikelompokkan
2. Menentukan nilai K
3. Menentukan *centroid* secara acak
4. Mengelompokkan data yang memiliki kemiripan
5. Input setiap data
6. Tentukan nilai *centroid* yang baru
7. Bila semua data sudah dikelompokkan dan nilai *centroid* tidak berubah maka proses selesai
8. Menampilkan bentuk data dalam tampilan visualisasi agar mudah dipahami

K-Means dapat digunakan untuk pengelompokkan atau *cluster* data. Setiap *cluster* yang terbentuk akan meningkatkan kriteria Partisi, misalkan fungsi dan perbedaan berdasarkan jarak. Hasil dari proses tersebut objek-objek didalam *cluster* menjadi sama, dan banyak obyek pada *cluster* berbeda dan tidak memiliki kesamaan dalam hal *attribute dataset*. Untuk mengukur jarak pada tiap *dataset* rumus yang digunakan adalah *Euclidean Distance* sebagai ukuran jarak yang digunakan pada metode K-Means. Rumus *Euclidean Distance* dapat dilihat dibawah ini.

$$d = (b_i, a_t) = \sqrt{\sum_j^l (b_{ij} - a_{tj})^2}$$

Dimana:

d = Jarak dari nilai data dan nilai *centoroid cluster*

b_i = nilai dari data, $i = 1,2,3, \dots, n, n =$ jumlah data

a_t = nilai pusat *cluster*, $t = 1,2,3, \dots, k, k =$ jumlah *cluster*

l = jumlah atribut

3. HASIL DAN PEMBAHASAN

Software yang digunakan dalam penelitian ini adalah *R Studio* yang sekarang telah berganti nama menjadi *Posit Cloud*. *Software* ini digunakan untuk mengelompokkan metode K-Means dengan menggunakan *dataset* yang diperoleh dari BPS. Data tindak pidana akan dianalisa sehingga didapatkan informasi mengenai Tindak Pidana dengan jumlah tinggi di suatu Provinsi.

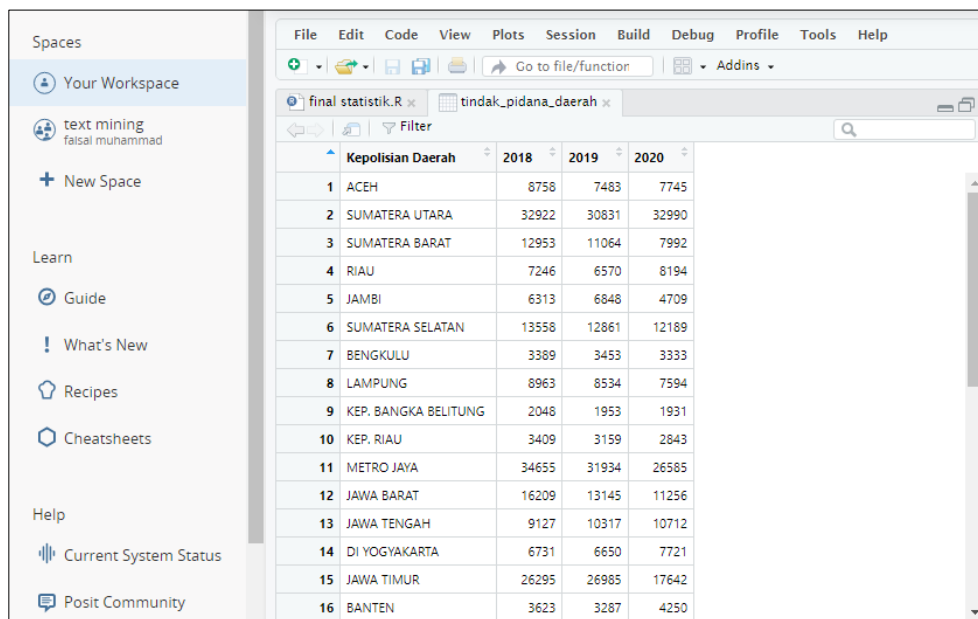
Tabel dibawah ini merupakan *dataset* dari tindak pidana dengan kurun waktu tahun 2018, 2019, dan 2020 yang diperoleh data BPS.

Tabel 1. Tindak Pidana

Kepolisian Daerah	2018	2019	2020
Aceh	8758	7483	7745
Sumatera Utara	32922	30831	32990
Sumatera Barat	12953	11064	7992
Riau	7246	6570	8194
Jambi	6313	6848	4709
Sumatera Selatan	13558	12861	12189
Bengkulu	3389	3453	3333
Lampung	8963	8534	7594
Kep. Bangka Belitung	2048	1953	1931
Kep. Riau	3409	3159	2843
Metro Jaya	34655	31934	26585
Jawa Barat	16209	13145	11256
Jawa Tengah	9127	10317	10712
DI Yogyakarta	6731	6650	7721
Jawa Timur	26295	26985	17642
Banten	3623	3287	4250
Bali	3212	3047	2597
Nusa Tenggara Barat	6451	8185	8591
Nusa Tenggara Timur	6257	5865	4790
Kalimantan Barat	5814	4721	3858

Kalimantan Tengah	2667	2444	2629
Kalimantan Selatan	5699	5375	5206
Kalimantan Timur	6287	4417	3609
Kalimantan Utara	396	876	1015
Sulawesi Utara	10247	7425	6274
Sulawesi Tengah	9379	6265	5454
Sulawesi Selatan	21498	16008	12815
Sulawesi Tenggara	1263	1213	2148
Gorontalo	2836	2367	2518
Sulawesi Barat	1817	1863	1704
Maluku	2751	3495	5350
Maluku Utara	722	718	850
Papua Barat	3475	2972	3162
Papua	7311	6994	6962

Gambar dibawah ini merupakan *dataset* Tindak Pidana yang sudah berbentuk Excel dan di *import* Posit Cloud.



Gambar 2. *Dataset* Tindak Pidana

Gambar dibawah ini merupakan hasil rata-rata Tindak Pidana yang sudah di proses pada *Software* Posit Cloud dengan menggunakan *R Programming* dibawah ini.

```

9 data_statis <- dataset_t_pidana [2:4]
10 summary(data_statis)

```

Gambar 3. *R Programming* Untuk Mencari Nilai Rata-Rata

Maka hasil dari *R Programming* menggunakan Posit Cloud *output* dari nilai rata-rata Tindak Pidana terdapat pada gambar dibawah ini

	2018	2019	2020
1	8758	7483	7745
2	32922	30831	32990
3	12953	11064	7992
4	7246	6570	8194
5	6313	6848	4709
6	13558	12861	12189
7	3389	3453	3333
8	8963	8534	7594
9	2048	1953	1931
10	3409	3159	2843
11	34655	31934	26585
12	16209	13145	11256
13	9127	10317	10712
14	6731	6650	7721
15	26295	26985	17642
16	3623	3287	4250

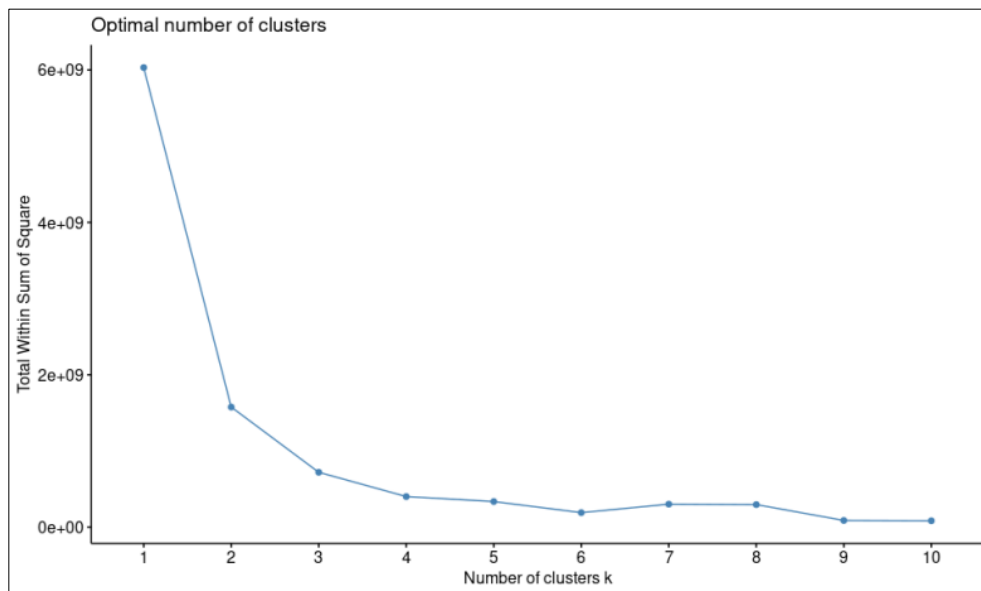
Gambar 4. Nilai Rata-rata Tindak Pidana

Setelah menentukan nilai rata-rata pada Tindak Pidana selanjutnya menentukan adalah menentukan total *cluster* dengan memanfaatkan metode *Elbow* dan *Silhouette*. Pada Tampilan visualisasi dibawah ini menunjukkan total *cluster* dengan menerapkan metode *Elbow* dengan R Programming.

```

11
12 #menentukan jumlah cluster dengan elbow
13 fviz_nbclust(data_statis,kmeans,method = "wss")
    
```

Gambar 5. R Programming Untuk Mencari Jumlah Cluster Dengan Metode Elbow



Gambar 6. Hasil Cluster Dengan Metode Silhouette

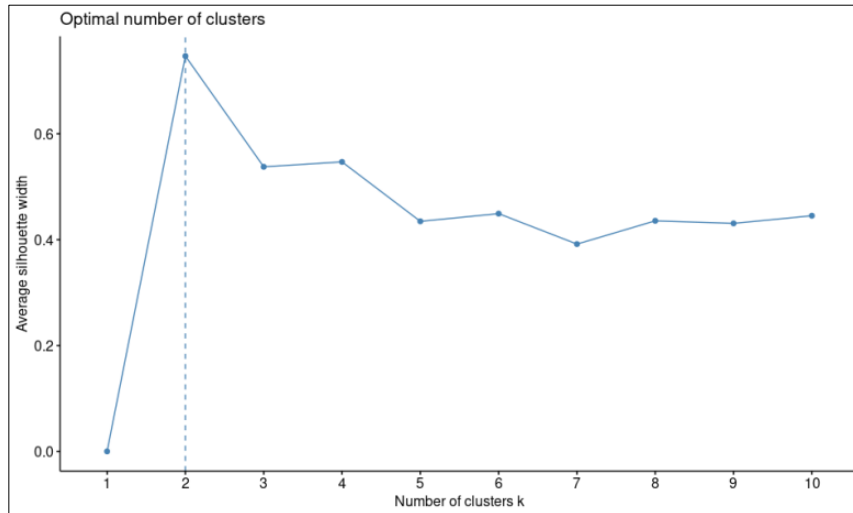
Dari hasil gambar diatas menunjukkan diagram hasil garis dari metode *elbow*. Pada gambar diatas menunjukkan titik tumpul berada pada titik tiga dan empat setelahnya tidak ada lagi penurunan yang terlalu signifikan namun dititik tujuh dan delapan terdapat kenaikan setelahnya turun kembali, dan dapat diambil kesimpulan jumlah *cluster* dari metode *Elbow* adalah 2 Cluster.

Berikutnya menentukan jumlah *cluster* K-Means dengan metode *Silhouette*, gambar dibawah ini menunjukkan garis dari hasil metode *Silhouette*. Pada gambar dibawah ini menunjukkan titik tertinggi berada pada *cluster* 2, *cluster* lainnya mengalami naik turun. Sehingga jumlah *cluster* menurut metode *Silhouette* berjumlah 2 *cluster*.

```

14
15 #menentukan jumlah cluster
16 fviz_nbclust(data_statis,kmeans)
    
```

Gambar 7. R Programming Untuk Mencari Jumlah Cluster Dengan Metode Silhouette



Gambar 8. Hasil Cluster Silhoutte

Dari hasil pencarian jumlah *cluster* menggunakan metode *Elbow* dan *Silhouette* hasilnya sama sebanyak dua *cluster*. Sehingga jumlah *cluster* yang digunakan untuk penelitian ini adalah dua *cluster*.

3.1 Clustering dengan K-Means

Total *cluster* yang telah ditentukan oleh penulis tahap selanjutnya menentukan *clustering* dengan Algoritma K-Means. Dengan memasukkan beberapa perintah menggunakan R Programming pada Posit Cloud, untuk menampilkan hasil *cluster*. Gambar dibawah ini merupakan hasil *cluster* dari jumlah Tindak Pidana yang didapatkan Penulis melalui dataset BPS.

```

> final=kmeans(data_statis,2)
> print(final)
K-means clustering with 2 clusters of sizes 4, 30

Cluster means:
      2018      2019      2020
1 28842.5 26439.5 22508.000
2  5963.7  5452.2  5239.533

Clustering vector:
[1] 2 1 2 2 2 2 2 2 2 1 2 2 2 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 1 2 2 2 2 2 2 2

Within cluster sum of squares by cluster:
[1] 513573936 1062966799
   (between_SS / total_SS = 73.9 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"
[5] "tot.withinss" "betweenss"    "size"         "iter"
[9] "ifault"
>
    
```

Gambar 9. Hasil Cluster

Gambar diatas merupakan hasil *cluster* data dari Tindak Pidana *dataset* BPS yang telah melalui proses R *Programming* pada Posit Cloud dan kode pemogramman yang telah di *insert* pada gambar diatas.

Dari temuan hasil *Cluster* pada gambar diatas kemudian akan ditampilkan dalam bentuk tabel, Tabel dibawah ini secara keseluruhan merupakan hasil akhir *cluster* Tindak Pidana menggunakan Metode K-Means

Tabel 2. Hasil akhir *Cluster* K-Means

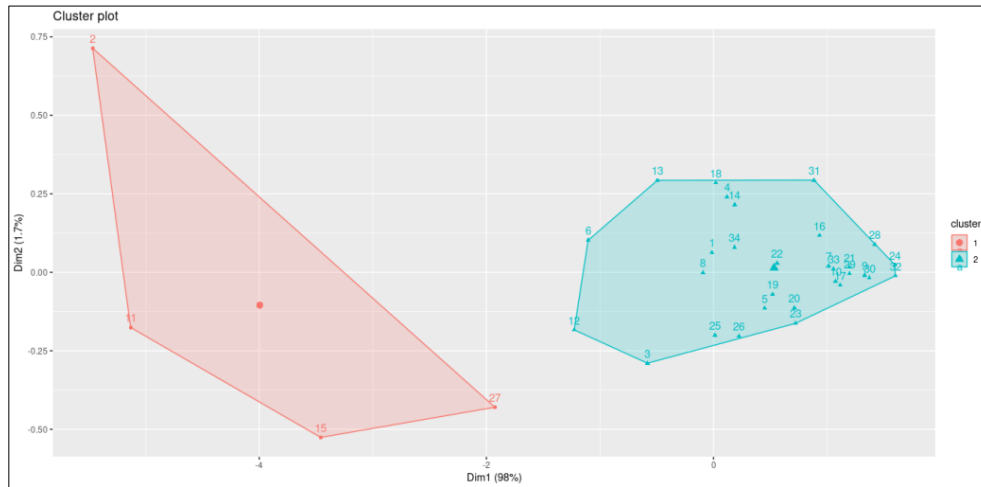
No.	Kepolisian Daerah	Final. Cluster
1	Aceh	2
2	Sumatera utara	1
3	Sumatera barat	2
4	Riau	2
5	Jambi	2
6	Sumatera selatan	2
7	Bengkulu	2
8	Lampung	2
9	Kep. Bangka belitung	2
10	Kep. Riau	2
11	Metro jaya	1
12	Jawa barat	2
13	Jawa tengah	2
14	DI yogyakarta	2
15	Jawa timur	1
16	Banten	2
17	Bali	2
18	Nusa tenggara barat	2
19	Nusa tenggara timur	2
20	Kalimantan barat	2
21	Kalimantan tengah	2
22	Kalimantan selatan	2
23	Kalimantan timur	2
24	Kalimantan utara	2
25	Sulawesi utara	2
26	Sulawesi tengah	2
27	Sulawesi selatan	1
28	Sulawesi tenggara	2
29	Gorontalo	2
30	Sulawesi barat	2
31	Maluku	2
32	Maluku utara	2
33	Papua barat	2
34	Papua	2

Hasil *cluster* plot menggunakan K-Means yang di masukkan kedalam R *Programming* pada Posit Cloud dapat dilihat pada gambar dibawah ini. Terdapat dua dimensi Dim1 sebesar 98% dan Dim 2 sebesar 1,7%.

```

20
21 fviz_cluster(final,data = data_statis)
22
    
```

Gambar 10. R Programming Cluster Plot



Gambar 11. Diagram Cluster Plot

Hasil akhir dari *cluster* Tindak Pidana ini diharapkan dapat memberikan informasi mengenai tingkat Tindak Pidana pada setiap Daerah. Hasil *cluster* tinggi diharapkan dapat perhatian khusus dari pihak terkait mengenai tindak pidana, sementara hasil yang rendah pula diharapkan dapat ditekan kembali agar tindak pidana tersebut dapat teratasi.

Penulis memahami masih banyaknya keterbatasan dalam penelitian ini termasuk data yang penulis gunakan, dimana data tersebut didapatkan dari Badan Pusat Statistik (BPS) bukan dengan cara mendatangi lembaga terkait. Penulis berharap akan ada penelitian terkait yang membahas Tindak Pidana dengan menerapkan letak Geografis Tindak Pidana menggunakan *Data Mining*.

4. KESIMPULAN

Metode K-Means pada penelitian Tindak Pidana ini data *cluster* Tindak Pidana baik yang tinggi maupun rendah diharapkan dapat perhatian khusus dari aparat penegak Hukum. Dalam menentukan *Centeroid* berperan besar pada penelitian menggunakan metode K-Means dan memiliki pengaruh besar pada penemuan *cluster*. Berdasarkan Penelitian ini didapatkan dua *cluster* yaitu *cluster* 1 dengan kategori cukup tinggi dan *cluster* 2 dengan kategori rendah. *Cluster* dengan kategori cukup tinggi atau *cluster* 1 terdapat 4 *cluster* dan *cluster* rendah atau *cluster* 2 terdapat 30 *cluster*.

5. SARAN

Setelah melakukan penelitian ini penulis berharap dapat dimanfaatkan untuk mendapatkan informasi mengenai kasus tindak pidana di setiap daerah dan cara penanggulangannya dalam menekan tingginya kasus tindak pidana. Harapan penulis akan ada penelitian lebih lanjut dan metode lainnya mengenai tindak pidana yang membentuk suatu *cluster* wilayah dengan memanfaatkan letak geografis suatu wilayah.

DAFTAR PUSTAKA

- [1] H. D. Alhadi, N. and I. Kurniawan, "Tinjauan Kriminologis Terhadap Faktor-Faktor Penyebab Terjadinya Tindak Pidana Dibidang Perikanan Dan Penanggulangannya Di Laut Kabupaten Bintan," *Delicti*, vol. 1, no. 2, pp. 26-35, 2023.
- [2] K. Lie, A. Zurnetti and E. Elda, "Penegakan Hukum Tindak Pidana Pemerasan Dengan Ancaman Dalam Pembangunan Properti Di Kota Padang," *Delicti*, vol. 1, no. 2, pp. 54-62, 2023.
- [3] W. Saputro, M. R. Pahlevi and A. , "ANALISIS ALGORITMA K-MEANS UNTUK KLASTERISASI TINDAK PIDANA KORUPSI DI WILAYAH HUKUM INDONESIA," *JIKO (Jurnal Informatika dan Komputer)*, vol. 3, no. 3, pp. 137-142, 2020.
- [4] P. M. S. Tarigan, J. T. Hardinata, H. Qurniawan, M. Safii and R. Winanjaya, "IMPLEMENTASI DATA MINING MENGGUNAKAN ALGORITMA APRIORI DALAM MENENTUKAN PERSEDIAAN BARANG (STUDI KASUS : TOKO SINAR HARAHAP)," *JUST IT (Jurnal Sistem Informasi, Teknologi Informasi dan Komputer)*, vol. 12, no. 2, pp. 51-61, 2022.
- [5] M. Faisal, S. and D. Prastyo, "Implementasi Data Mining Menggunakan Metode K-Means Dalam Pengelompokan Bayi Penerima ASI Eksklusif," *ICIT (Innovative Creative and Information Technology)*, vol. 9, no. 2, pp. 198-207, 2023.
- [6] F. Rahmadayanti and R. Rahayu, "PENERAPAN METODE DATA MINING PADA KASUS KRIMINALITAS INDONESIA," *Jurnal Teknologi Informasi Mura*, vol. 15, no. 1, pp. 52-61, 2023.
- [7] Y. Diana and F. Hadi, "ANALISA PENJUALAN MENGGUNAKAN ALGORITMA K-MEDOIDS UNTUK MENGOPTIMALKAN PENJUALAN BARANG," *JOISIE*, vol. 7, no. 1, pp. 97-103, 2023.
- [8] H. Priyatman, F. Sajid and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 5, no. 1, p. 62, 2014.
- [9] D. t. Cahaya, D. Puspita and R. Syahri, "PENERAPAN METODE K-MEANS CLUSTERING UNTUK PENGELOMPOKAN POTENSI PADI DI KOTA PAGAR ALAM," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 2187-2193, 2024.
- [10] A. "Data Mining," in *DATA MINING DAN KNOWLEDGE DISCOVERY PROCESS*, Tangah Padang Sumatera Barat , PT GLOBAL EKSEKUTIF TEKNOLOGI, 2023, pp. 1-16.
- [11] J. Nasir, "PENERAPAN DATA MINING CLUSTERING DALAM MENGELOMPOKAN BUKU DENGAN METODE K-MEANS," *SIMETRIS*, vol. 11, no. 2, pp. 690-703, 2020.
- [12] N. A. Maori, "METODE ELBOW DALAM OPTIMASI JUMLAH CLUSTER PADA K-MEANS CLUSTERING," *SIMETRIS*, vol. 14, no. 2, pp. 277-287, 2023.
- [13] E. A. Saputra and Y. Nataliani, "Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means," *ISI (Information Systems and Informatics)*, vol. 3, no. 3, pp. 424-439, 2021.
- [14] N. Febrian and N. , "Perbandingan Manhattan dan Euclidean Distance Untuk Pengelompokan Penyakit Jantung Menggunakan Algoritma K-Means," *ICI (Innovative Creative and Information Technology)*, vol. 10, no. 1, pp. 61-70, 2024.
- [15] A. W. S, I. G. I. Sudipa, A. A. E. Putra, A. J. Wahidin, W. A. Syukrilla, A. K. Wadhani, N. Heryana, T. Indriyani and L. W. Santoso, *Data Mining*, Padang Sumatera Barat: PT Global Eksekutif Teknologi, 2023.