

# 1960-5275-1-SM

*by* Wahyu Saputra 2

---

**Submission date:** 25-Jul-2020 04:35AM (UTC+0700)

**Submission ID:** 1361735823

**File name:** 1960-5275-1-SM.docx (140.08K)

**Word count:** 3331

**Character count:** 18685

## ANALISIS ALGORITMA K-MEANS UNTUK KLASTERISASI TINDAK PIDANA KORUPSI DI WILAYAH HUKUM INDONESIA

Wahyu Saputro<sup>1</sup>, Moh Reza Pahlevi<sup>2</sup>, Arief Wibowo<sup>3</sup>

<sup>1,2,3</sup>Magister Ilmu Komputer, Fakultas Ilmu Komputer  
Universitas Budi Luhur

Email: <sup>1</sup>wahyu.saputro6982@gmail.com, <sup>2</sup>onowae848@gmail.com, <sup>3</sup>arief.wibowo@budiluhur.ac.id

(Naskah masuk : ddmmyyy, diterima untuk diterbitkan: ddmmyyy)

### Abstrak

Pasca kemerdekaan, Indonesia masih menghadapi perang yaitu perang melawan korupsi yang sudah menjadi budaya di Indonesia selama ini. Ditahun 1999 pemerintah telah mengeluarkan UU No 31 tentang pemberantasan korupsi dan pada tahun 2001 pemerintah memperbaruinya dengan UU No 20. Menurut UU No. 31 Thn 1999 tentang tindak pidana korupsi sudah dilakukan pembaruan dengan UU No. 20 Thn 2001, definisi korupsi atau rasuah merupakan sebuah perbuatan tidak jujur atau penyelewengan yang dilakukan karena adanya suatu pemberian. Dalam prakteknya, korupsi lebih dikenal sebagai menerima uang yang ada hubungannya dengan jabatan atas imbalan tertentu yang diberikan. Menurut data dari Kejaksaan RI tingkat korupsi di Indonesia makin tahun makin meningkat baik dari tingkat daerah hingga tingkat Pusat. Usaha dalam pemberantasan korupsi di Indonesia sudah dilakukan berbagai cara, namun masih saja terjadi korupsi dengan segala macam cara yang dilakukan oleh berbagai lembaga. Oleh karena itu perlu dilakukan analisis data untuk mengetahui kecenderungan seseorang melakukan korupsi. Tujuan dari penelitian ini mendapatkan analisis data korupsi di Indonesia. Metode yang digunakan K-Means untuk mengklasterisasi pada data korupsi di Indonesia. Hasil penelitian menunjukkan bahwa wilayah dengan kategori rawan tindak pidana korupsi berada pada pulau sumatra ada 5 wilayah, pulau jawa ada 3 wilayah dan pulau sulawesi dan kalimantan ada 1 wilayah.

**Kata kunci:** Korupsi, K-Means, Klustering

## K-MEANS ALGORITHM ANALYSIS FOR CORRUPTION CRIMINAL CLUSTERIZATION IN INDONESIA'S LAW

### Abstract

After independence, Indonesia still faces a war that is a war against corruption which has become a culture in Indonesia so far. In 1999 the government issued Law No. 31 on combating corruption and in 2001 the government updated it with Law No. 20. According to Law No. 31 of 1999 concerning acts of corruption and the renewal of Law No. 20 In 2001, the definition of corruption is a dishonest act or deviation committed because of a gift. In practice, corruption is better known as receiving money that has to do with the position of a certain reward. According to data from the Indonesian Attorney General, the level of corruption in Indonesia has been increasing from year to the central level. Efforts to eradicate corruption in Indonesia have been carried out in various ways, but corruption still occurs in all kinds of ways carried out by various institutions. Therefore it is necessary to do data analysis to determine one's tendency to commit corruption. The purpose of this study is to obtain an analysis of corruption data in Indonesia. The method used by K-Means to classify corruption data in Indonesia. The results showed that the area with a category prone to corruption is in the island of Sumatra there are 5 regions, Java Island there are 3 regions and Sulawesi and Kalimantan islands have 1 region.

**Keywords:** Keywords: Corruption, K-Means, Clustering

### 1. PENDAHULUAN

Pasca kemerdekaan, Indonesia masih menghadapi perang yaitu perang melawan korupsi yang sudah menjadi budaya di Indonesia selama ini. Ditahun 1999 pemerintah telah mengeluarkan UU No 31 tentang pemberantasan korupsi dan pada tahun 2001 pemerintah memperbaruinya dengan UU No 20, definisi korupsi atau rasuah merupakan sebuah

tindakan perbuatan tidak jujur atau penyelewengan yang dilakukan karena adanya suatu pemberian. Dalam prakteknya, korupsi lebih dikenal sebagai menerima uang yang ada hubungannya dengan jabatan atas imbalan tertentu yang diberikan. Adapun rumusan dari jenis tindak korupsi yang di kelompokkan diantaranya kerugian keuangan negara, penyuapan, penyalahgunaan jabatan, curang, kepentingan dalam pengadaan, gratifikasi.

Dalam upaya memberantas tindak pidana korupsi, pemerintah telah mengeluarkan beberapa regulasi terhadap usaha untuk memberantas korupsi, sebagai contoh peraturan yang mengatur kejahatan korupsi dalam hal ini Kejaksaan RI dan Polri memainkan peranan penting dalam memberantas korupsi. Oleh karena itu perlu adanya terobosan dalam menangani kejahatan korupsi tersebut, yang diantaranya <sup>9</sup> mendorong transparansi penyelenggaraan negara, memberikan rekomendasi kepada kementerian dan lembaga untuk melakukan langkah-langkah perbaikan, memoderinsasi pelayanan publik dengan online atau berbasis IT dan dilakukan pengawasan yang terintegrasi.

Untuk teknik <sup>24</sup> pengolahan data penulis menggunakan teknik data mining, karena data mining menyediakan algoritma yang bisa digunakan untuk mencari inform <sup>23</sup> data tersembunyi dan multidimensi data. Dalam penelitian ini penulis menggunakan salah satu algoritma yaitu algoritma k-means. Algoritma dengan cara mengklaster untuk dianalisis faktor kesamaan (*similarity*) maupun tidak kesamaan (*dissimilarity*) yang terdapat dalam data tersebut.

Dengan memanfaatkan metode K-Means supaya untuk menganalisa korupsi di Indonesia. Pendekatan dibentuk dengan cara digabungkannya teknik klasterisasi pada K-Means diperuntukan untuk pengelompokan data yang sama. Keuntungan atas data-data yang sama sangat signifikan akan timbul jika menggunakan teknik klastering.

Berdasarkan pada latar belakang, terdapat permasalahan yang terjadi yaitu upaya pengelompokan kasus korupsi di wilayah hukum Indonesia. Dalam upaya meminimalisir tindak pidana korupsi diperlukan metode untuk mengelompokan tindak pidana korupsi tersebut sehingga bisa Kejaksaan atau KPK dalam membasmi dan menangkap para pelaku tindak pidana korupsi. Pada penelitian ini, penulis merumuskan permasalahan dengan cara diantaranya Menggunakan kedekatan jarak *euclidean distance* untuk menganalisa K-Means Cluste <sup>22</sup> dan melakukan pengelompokan data ke dalam kategori tindak pidana korupsi paling tinggi, menengah dan cukup menggunakan metode *clustering*.

Dalam penelitian ini peneliti bertujuan untuk bagaimana metode K-means Klasterisasi bisa digunakan dalam mengelompokan data-data tentang tindak pidana korupsi diantaranya dapat membantu KPK, Kepolisian dan Kejaksaan dalam mengelompokan kejahatan korupsi, memberikan gambaran kepada KPK maupun Kejaksaan RI, daerah mana saja yang sering terlibat tindak pidana korupsi, dapat <sup>12</sup> memberikan acuan dalam memetakan daerah rawan tinda <sup>12</sup> pidana korupsi di wilayah Indonesia.

Pada penelitian ini, penulis menggunakan jenis model penelitian eksperimen yaitu mengelompokan tindak pidana korupsi di Indonesia dengan menggunakan Algoritma K-means sebagai fungsi pengelompokan atau klasterisasi yang sangat besar

jumlahnya dan untuk penelitian ini mengambil dataset tindak pidana korupsi pada tahun 2014 dan data diambil dari data dari Kejaksaan RI, variabel yang digunakan adalah tahun SPDP, tahun Penuntutan yang ditangani oleh Kejaksaan, dalam penelitian nanti dilakukan penyamaran nama Kejati menjadi nama pulau atau pulau terdekatnya.

17

## 2. TINJUAN PUSTAKA

### 2.1 Data Mining

*Data mining* dalam artian secara sederhana yaitu suatu penemuan atau penambangan informasi baru dengan cara mencari atau aturan dari sejumlah data yang sangat besar (Davies, 2004). Data mining atau penggalian data juga bisa diartikan suatu proses ekstraksi atau menarik data dalam jumlah yang besar.

Dalam pernyataan *data mining* bahwa *data mining* masih melibatkan campur tangan manusia dalam setiap fase atau prosesnya. Salah satu kelebihan *data mining* adanya perangkat lunak yang mampu menganalisis kemungkinan terjadinya kasalahan yang fatal, dalam pemahamannya dapat dijabarkan yang seperti gambar 1.



Gambar 1. Bidang Ilmu *Data Mining*

### 2.2 Clustering

*Clustering* atau kata lain klastering merupakan salah satu proses umum yang terdapat dalam data mining untuk menganalisa data, memecahkan permasalahan atau mempartisi dataset menjadi <sup>11</sup> bset. Teknik *cluster* atau klasterisasi mempunyai dua metode yaitu *hierarchical clustering* dan *non-hierarchical clustering*. <sup>7</sup>

*Clustering* atau klasterisasi adalah suatu metode suatu metode untuk pengelompokan dokumen yang mana dokumen dikelompokan dengan konten untuk mengurangi pencarian ruang dengan merespon query data.

*Hierarchical clustering* adalah suatu model *clustering* yang berfungsi untuk mengelompokan data berdasarkan urutan partisi, metode ini cara mengelompokan dengan 2 cara yaitu *agglomerative* dan *devisiv*. *Clustering* bermanfaat untuk menganalisa segmentasi pasar, marketing dan pemetaan zonasi wilayah, *clustering* juga bermanfaat untuk mengidentifikasi obyek dalam berbagai bidang.

Konsep untuk menghasilkan tingkat kesamaan hanya mengenal 4 tipe data atau variable diantaranya *variable* berskala besar, *biner*, normal, *ordinal*, *rasio* dan *variable* tipe lainnya.

Menurut pendapat peneliti Han dan Kamber, 2012 *clustering* ada beberapa syarat yang harus dipenuhi diantaranya skalabilitas, bisa menemukan bentuk data yang tidak terduga, memiliki kemampuan untuk menangani noise

11  
2.3 **Algoritma K-Means**

K-Means salah satu algoritma yang dimiliki oleh data mining yang bisa mengelompokkan proses *clustering*/pengelompokan.

K-Means juga salah satu *clustering* yang bertujuan membagi diantara data untuk menjadi beberapa kelompok, yang mempunyai karakteristik utama antara karakteristik satu dengan karakteristik yang lainnya dan mempunyai karakteristik yang berbeda dengan data yang sudah ada di kelompok lainnya.

Algoritma K-Means untuk melakukan *clustering* diantaranya dengan cara tentukan K untuk menentukan *centroid* secara random, kelompokkan data sampai terbentuk K atau *cluster* dengan titik *centroid* sudah ditentukan, perbarui titik *centroid*, ulangi sampai nilai *centroid* nya tidak berubah.

K-Means juga memiliki beberapa kelebihan diantaranya mudah diimplementasikan atau dijalankan, waktu proses pembelajaran lebih cepat, mudah untuk diadaptasi, banyak yang menggunakan.

3. METODE PENELITIAN

Penelitian ini akan menggunakan data yang diolah merupakan dataset tindak pidana korupsi dari tahun 2014. Dataset tindak pidana korupsi terdiri dari atribut Jumlah SPDP, Jumlah Penuntutan, Belum Selesai, berdasarkan v21 yah. Data yang akan diuji terdiri dari 1102 data tindak pidana korupsi seluruh Indonesia

Pada penelitian ini pengelompokan tindak pidana korupsi dari 5 data yang diperoleh menggunakan algoritma K-Means. Karena K-means merupakan cara *clustering* yang paling simpel. Karena K-means 5 memiliki kemampuan untuk menggabungkan data yang cukup besar jumlah dalam waktu yang relatif cepat dan efektif. Metode K-Means adalah suatu pola yang menggunakan algoritma klasterisasi data yang berbasis jarak untuk dibagi ke dalam *cluster* dan untuk atributnya hanya berbentuk numerik.

Langkah – langkah penelitian untuk pengambilan sampel yang diantaranya sebagai berikut:

- Dilakukan *cleansing* data dan merubah identitas kejadi menjadi wilayah berdasarkan pulau
- Tentukan jumlah klaster yang akan dibentuk dan k sebagai nilainya

- Hitung jarak setiap inputan terhadap *centroid – centroid* menggunakan persamaan *Euclidean*, dengan rumus i

$$D(x_i, \mu_k) = \sqrt{\sum_k^m (c_{ij} - c_{kj})^2} \quad (i)$$

- Data tersebut akan dikelompokkan kedalam *cluster* dengan cara mencari jarak terpendek dengan persamaan

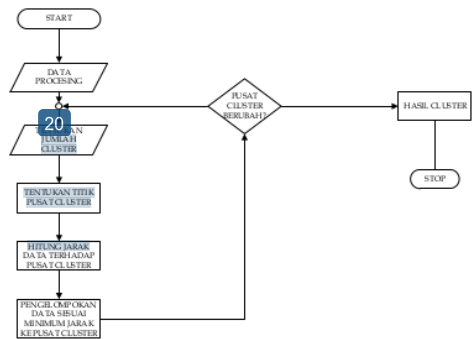
$$M_m \sum_k^k dik = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (ii)$$

- Penghitungan pusat *cluster* yang baru dapat menggunakan persamaan

$$C_{ij} = \frac{\sum_{i=1}^k x_{ij}}{p} \quad (3)$$

Dimana:

$$X_{ij} \in \text{cluster } ke = k$$



Gambar 1. Flowchar Algoritma K-Means Klasterisasi

4. ANALISA DAN PEMBAHASAN

Data Tindak pidana korupsi di Indonesia ditangani oleh 33 wilayah di tahun 2014 dan semua data tindak pidana Korupsi masing-masing wilayah akan dijadikan contoh pemrosesan dengan cara *Data Mining*. Format data tersebut berupa data Microsoft Excel dan dilakukan proses pengujian menggunakan tool aplikasi *rapidminer* yang bisa mengolah data file tipe .xls. Dari hasil pengolahan data penelitian ini menggunakan data tindak pidana korupsi tahun 2014 yang dapat dilihat dalam tabel berikut.

Tabel ke-1 Data Master Tindak Pidana Korupsi Tahun 2014

No	Wilayah	SPDP	Putusan	Belum Selesai
1	Sumatra1	140	91	49
2	Bali1	6	2	4
3	Sumatra2	4	2	2
4	Jawa1	8	8	0
5	Sumatra3	50	37	13
6	Jawa2	18	12	6
7	Jawa3	41	24	17
8	Sulawesi1	7	6	1

No	Wilayah	SPDP	Putusan	Belum Selesai
9	Sumatra4	97	46	51
10	Jawa4	26	20	6
11	Jawa5	55	33	22
12	Jawa6	41	25	16
13	Kalimantan1	28	14	14
14	Kalimantan2	29	24	5
15	Kalimantan3	11	9	2
16	Kalimantan4	13	4	9
17	Sumatra5	4	4	0
18	Sumatra6	25	17	8
19	Papua1	31	11	20
20	Papua2	17	8	9
21	Bali2	8	5	3
22	Bali3	37	24	13
23	Papua3	4	0	4
24	Papua4	1	1	0
25	Sumatra7	78	55	23
26	Sulawesi2	3	2	1
27	Sulawesi3	12	9	3
28	Sulawesi4	8	5	3
29	Sulawesi5	10	3	7
30	Sulawesi6	15	8	7
31	Sulawesi7	95	56	39
32	Sumatra8	107	69	38
33	Sumatra9	73	49	24

Data yang akan digunakan penelitian ini adalah Jumlah Kasus, Kasus Selesai, untuk menentukan berapa jumlah cluster dan awal titik *cluster* pusat secara acak.

Dari hasil tersebut, bisa disimpulkan dari tiap-tiap titik awal *cluster* yang berdasarkan urutan. Untuk menentukan wilayah yang terjadi tindak pidana korupsi tidak rawan pidana korupsi sedang dan rawan tindak pidana korupsi sebagai berikut:

- *Cluster* pusat I: Data ke-1 Wilayah Sumater 1 dengan SPDP = 140 dan Putusan 91
- *Cluster* pusat II: Data ke-14 Wilayah Kalimantan2 dengan SPDP = 29 dan Putusan 24
- *Cluster* pusat III: Data ke-29 Wilayah Sulawesi5 dengan SPDP = 10 dan Putusan 3.

Untuk hasil tersebut penulis ilustrasikan sebagai berikut:<sup>2</sup>

a. Iterasi I

Dt-1 ke Cluster Pusat I

$$= \sqrt{(140 - 140)^2 + (91 - 91)^2}$$

= 0

Dt-2 ke Cluster Pusat I

$$= \sqrt{(6 - 140)^2 + (2 - 91)^2}$$

= 160,863296

Dt-3 ke Cluster Pusat I

$$= \sqrt{(4 - 140)^2 + (2 - 91)^2}$$

= 162,5330736

Dt-4 ke Cluster Pusat I

$$= \sqrt{(8 - 140)^2 + (8 - 91)^2}$$

= 155,9262646

Dt-5 ke Cluster Pusat I

$$= \sqrt{(50 - 140)^2 + (37 - 91)^2}$$

= 104,9571341

Dt-6 ke Cluster Pusat I

$$= \sqrt{(18 - 140)^2 + (12 - 91)^2}$$

= 145,3444185

Dt-7 ke Cluster Pusat I

$$= \sqrt{(41 - 140)^2 + (24 - 91)^2}$$

= 119,540788

Dt-8 ke Cluster Pusat I

$$= \sqrt{(7 - 140)^2 + (6 - 91)^2}$$

= 157,8416928

Dt-9 ke Cluster Pusat I

$$= \sqrt{(97 - 140)^2 + (46 - 91)^2}$$

= 62,24146528

Dt-10 ke Cluster Pusat I

$$= \sqrt{(26 - 140)^2 + (20 - 91)^2}$$

= 134,3018987

Dt-11 ke Cluster Pusat I

$$= \sqrt{(55 - 140)^2 + (33 - 91)^2}$$

= 102,9028668

Dt-12 ke Cluster Pusat I

$$= \sqrt{(41 - 140)^2 + (25 - 91)^2}$$

= 118,9831921

Dt-13 ke Cluster Pusat I

$$= \sqrt{(28 - 140)^2 + (14 - 91)^2}$$

= 135,915414

Dt-14 ke Cluster Pusat I

$$= \sqrt{(29 - 140)^2 + (24 - 91)^2}$$

= 129,6533841

Dt-15 ke Cluster Pusat I

$$= \sqrt{(11 - 140)^2 + (9 - 91)^2}$$

= 152,8561415

Dt-16 ke Cluster Pusat I

$$= \sqrt{(13 - 140)^2 + (4 - 91)^2}$$

= 153,9415473

Dt-17 ke Cluster Pusat I

$$= \sqrt{(4 - 140)^2 + (4 - 91)^2}$$

= 161,4465856

Dt-18 ke Cluster Pusat I

$$= \sqrt{(25 - 140)^2 + (17 - 91)^2}$$

= 136,7515996

Dt-19 ke Cluster Pusat I

$$= \sqrt{(31 - 140)^2 + (11 - 91)^2}$$

= 135,2072483

Dt-20 ke Cluster Pusat I

$$= \sqrt{(17 - 140)^2 + (8 - 91)^2}$$

= 148,3846353

Dt-21 ke Cluster Pusat I

$$= \sqrt{(8 - 140)^2 + (5 - 91)^2}$$

= 157,5436447

Dt-22 ke Cluster Pusat I

$$= \sqrt{(37 - 140)^2 + (24 - 91)^2}$$

$$= 122,8739191$$

Dt-23 ke Cluster Pusat I

$$= \sqrt{(4 - 140)^2 + (0 - 91)^2}$$

$$= 163,6367929$$

Dt-24 ke Cluster Pusat I

$$= \sqrt{(1 - 140)^2 + (1 - 91)^2}$$

$$= 165,5928742$$

Dt-25 ke Cluster Pusat I

$$= \sqrt{(78 - 140)^2 + (55 - 91)^2}$$

$$= 71,69379332$$

Dt-26 ke Cluster Pusat I

$$= \sqrt{(3 - 140)^2 + (2 - 91)^2}$$

$$= 163,370744$$

Dt-27 ke Cluster Pusat I

$$= \sqrt{(12 - 140)^2 + (9 - 91)^2}$$

$$= 152,0131573$$

Dt-28 ke Cluster Pusat I

$$= \sqrt{(8 - 140)^2 + (5 - 91)^2}$$

$$= 157,5436447$$

Dt-29 ke Cluster Pusat I

$$= \sqrt{(10 - 140)^2 + (3 - 91)^2}$$

$$= 156,9840756$$

Dt-30 ke Cluster Pusat I

$$= \sqrt{(15 - 140)^2 + (8 - 91)^2}$$

$$= 150,0466594$$

Dt-31 ke Cluster Pusat I

$$= \sqrt{(95 - 140)^2 + (8 - 91)^2}$$

$$= 57,00877125$$

Dt-32 ke Cluster Pusat I

$$= \sqrt{(107 - 140)^2 + (69 - 91)^2}$$

$$= 39,66106403$$

Dt-33 ke Cluster Pusat I

$$= \sqrt{(73 - 140)^2 + (49 - 91)^2}$$

$$= 79,07591289$$

Dari perhitungan tersebut diatas yang menggunakan persamaan *Euclidian Distance* terhadap iterasi I tersebut diatas dapat diamati di tabel ini:

Tabel ke-2 Iterasi I

No	Wilayah	C1	C2	C3
1	Sumatra1	0,00	129,65	156,98
2	Bali1	160,86	31,83	4,12
3	Sumatra2	162,53	33,30	6,08
4	Jawa1	155,93	26,40	5,39
5	Sumatra3	104,96	24,70	52,50
6	Jawa2	145,34	16,28	12,04
7	Jawa3	119,54	12,00	37,44
8	Sulawesi1	157,84	28,43	4,24
9	Sumatra4	62,24	71,47	97,05
10	Jawa4	134,30	5,00	23,35
11	Jawa5	102,90	27,51	54,08
12	Jawa6	118,98	12,04	38,01
13	Kalimantan1	135,92	10,05	21,10
14	Kalimantan2	129,65	0,00	28,32
15	Kalimantan3	152,86	23,43	6,08
16	Kalimantan4	153,94	25,61	3,16
17	Sumatra5	161,45	32,02	6,08

No	Wilayah	C1	C2	C3
18	Sumatra6	136,75	8,06	20,52
19	Papua1	135,21	13,15	22,47
20	Papua2	148,38	20,00	8,60
21	Bali2	157,54	28,32	2,83
22	Bali3	122,87	8,00	34,21
23	Papua3	163,64	34,66	6,71
24	Papua4	165,59	36,24	9,22
25	Sumatra7	71,69	57,98	85,60
26	Sulawesi2	163,37	34,06	7,07
27	Sulawesi3	152,01	22,67	6,32
28	Sulawesi4	157,54	28,32	2,83
29	Sulawesi5	156,98	28,32	0,00
30	Sulawesi6	150,05	21,26	7,07
31	Sulawesi7	57,01	73,35	100,17
32	Sumatra8	39,66	90,05	117,32
33	Sumatra9	79,08	50,61	78,01

Dan untuk tahap selanjutnya peneliti menentukan *cluster* pusat yang jaraknya terpendek terhadap awal *cluster* yang pemilihannya dilakukan secara acak, dimana untuk jaraknya terpendek dengan *cluster* ditentukan dengan rumus  $MIN(C_{ij} : C_{kj} : C_{lj})$  yang mana untuk menentukan hasil jarak terpendek tersebut terhadap *cluster* juga bisa dilihat sesuai tabel dibawah

Tabel ke-3 Biner jarak terdekat data terhadap Cluster

Data Ke	Wilayah	C1	C2	C3	Cluster
1	Sumatra1	1	0	0	C1
2	Bali1	0	0	1	C3
3	Sumatra2	0	0	1	C3
4	Jawa1	0	0	1	C3
6	Sumatra3	0	0	1	C3
7	Jawa2	0	1	0	C2
8	Jawa3	0	0	1	C3
9	Sulawesi1	1	8	0	C1
10	Sumatra4	0	1	0	C2
11	Jawa4	0	1	0	C2
12	Jawa5	0	1	0	C2
13	Jawa6	0	1	0	C2
14	Kalimantan1	0	1	0	C2
15	Kalimantan2	0	0	1	C3
16	Kalimantan3	0	0	1	C3
17	Kalimantan4	0	0	1	C3
18	Sumatra5	0	1	0	C2
19	Sumatra6	0	1	0	C2
20	Papua1	0	0	1	C3
21	Papua2	0	0	1	C3
22	Bali2	0	1	0	C2
23	Bali3	0	0	1	C3
24	Papua3	0	0	1	C3
25	Papua4	0	1	0	C2
26	Sumatra7	0	0	1	C3
27	Sulawesi2	0	0	1	C3
28	Sulawesi3	0	0	1	C3
29	Sulawesi4	0	0	1	C3
30	Sulawesi5	0	0	1	C3
31	Sulawesi6	1	0	0	C1

32	Sulawesi7	1	0	0	C1
33	Sumatra8	0	1	0	C2

Berdasarkan hasil minimum nilai yang didapat pada penggunaan rumus dan hasilnya juga dapat dilihat maka dapat dikelompokkan seperti berikut:

**Tabel ke-4 Hasil iterasi / pengelompokan**

Cluster Group	Data Cluster	Jumlah Data
1	Wilayah 1,9,31,32	4
2	Wilayah 5,7,10,11,12,13,14,18,19,22,25,33	12
3	Wilayah 2,3,4,6,8,15,16,17,20,21,23,24,26,27,28,30	17

Selanjutnya untuk iterasi ke-2 s/d iterasi terakhir bisa digambarkan sesuai tabel berikut:

**Tabel 5 Iterasi biner ke 4**

Data Ke	Wilayah	Jarak Centroid		
		C1	C2	C3
1	Sumatra1	0	0	1
2	Bali1	0	1	0
3	Sumatra2	0	0	1
4	Jawa1	0	1	0
5	Sumatra3	0	0	1
6	Jawa2	0	1	0
7	Jawa3	0	0	1
8	Sulawesi1	0	1	0
9	Sumatra4	1	0	0
10	Jawa4	0	1	0
11	Jawa5	0	0	1
12	Jawa6	0	0	1
13	Kalimantan1	0	0	1
14	Kalimantan2	0	1	0
15	Kalimantan3	0	1	0
16	Kalimantan4	0	1	0
17	Sumatra5	0	0	1
18	Sumatra6	0	1	0
19	Papua1	0	1	0
20	Papua2	0	1	0
21	Bali2	0	1	0
22	Bali3	0	1	0
23	Papua3	0	1	0
24	Papua4	0	0	1
25	Sumatra7	0	0	1
26	Sulawesi2	0	0	1
27	Sulawesi3	0	1	0
28	Sulawesi4	0	1	0
29	Sulawesi5	0	1	0
30	Sulawesi6	0	1	0
31	Sulawesi7	1	0	0
32	Sumatra8	1	0	0
33	Sumatra9	0	0	1

Hasil diatas tersebut bila dibandingkan antara hasil iterasi ke-3 yang sebelumnya maka dapat menghasilkan hasil yang sama. maka untuk pencarian iterasi selanjutnya bisa dihentikan sudah ditentukan hasilnya. Untuk membuktikan hasil dari

pengelompokan bisa dilihat di tabel yang jarak / C3 adalah wilayah rawan terjadinya korupsi berada di wilayah dengan kategori rawan tindak pidana korupsi berada pada pulau sumatra ada 5 wilayah, pulau jawa ada 3 wilayah dan pulau sulawesi dan kalimantan ada 1 wilayah.

## 5. KESIMPULAN

Pada penelitan yang sudah dilakukan penulis, penulis bisa mengambil kesimpulan yang diantaranya sebagai berikut:

1. Pengelompokan daerah rawan tindak pidana korupsi di wilayah indonesia belum bisa dijadikan, dikarenakan belum ada inovasi untuk memetakan wilayah yang dianggap sebagai kelompok rawan dan saat ini pengelompokan belum secara modern.
2. Pembagian kategori yang didapat dari data-data SPDP, penuntutan kemudian kategory tersebut dirubah menjadi wilayah tidak rawan tindak pidana korupsi, cukup dan rawan sebagai nilai yang bisa dihasilkan pada masing-masing cluster.

## DAFTAR PUSTAKA

- [1] R. Hablum, A. Khairan, and R. Rosihan, "Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 2, no. 1, pp. 26–33, 2019, doi: 10.33387/jiko.v2i1.1053.
- [2] M. Metode and S. M. Som, "Clustering Kelompok Swadaya Masyarakat (KSM) dalam Menentukan Kebijakan Bantuan Badan Pemberdayaan Masyarakat di Kota Surabaya dengan," no. September, 2012.
- [3] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *J. Sist. Komput. dan Inform.*, vol. 1, no. 2, p. 151, 2020, doi: 10.30865/json.v1i2.1955.
- [4] T. Panorama, N. Cerah, O. D. Nurhayati, R. R. Isnanto, O. D. Nurhayati, and R. R. Isnanto, "Perbandingan Metode Segmentasi K-Means Clustering dan Segmentasi Region Growing untuk Pengukuran Luas Wilayah Hutan Mangrove Comparison of K-Means Clustering and Growing Region Segmentation Methods for Area Measurement of Mangrove Forests," vol. 7, no. March, pp. 31–37, 2019, doi: 10.14710/jtsiskom.7.1.2019.31-37.
- [5] U. T. Suryadi and Y. Supriatna, "Sistem Clustering Tindak Kejahatan Pencurian Di Wilayah Jawa Barat Menggunakan Algoritma K-Means," *J. Teknol. Inf. dan Komun.*, no. April, pp. 1–13, 2019.

- [6] W. Astuti, A. Widodo, J. T. Elektro, F. Teknik, and U. N. Semarang, "Pemetaan Tindak Kejahatan Jalanan di Kota Semarang Menggunakan Algoritma K-Means Clustering," vol. 8, no. 1, pp. 5–7, 2016.
- [7] J. J. Purnama, "Analisa Algoritma K-Means Clustering Pemetaan Jumlah Tindak Pidana," *Klik - Kumpul. J. Ilmu Komput.*, vol. 6, no. 2, p. 128, 2019, doi: [10.20527/klik.v6i2.208](https://doi.org/10.20527/klik.v6i2.208).
- [8] I. T. Utami, R. Rais, and W. Seftiani, "Analisis Klaster Pautan Lengkap Untuk Mengelompokkan Kabupaten/Kota Di Provinsi Sulawesi Tenggara Berdasarkan Indikator Kriminalitas," *Ilm. Mat. Dan Terap.*, vol. 16, no. 1, pp. 79–88, 2019, doi: [10.22487/2540766x.2019.v16.i1.12757](https://doi.org/10.22487/2540766x.2019.v16.i1.12757).
- [9] M. S. Fajar, "DI KOTA SEMARANG MENGGUNAKAN METODE K-MEANS CLUSTERING," 2015.
- [10] I. Method, K. C. Based, S. Value, W. Interface, C. Study, and I. U. M. M. Magelang, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," vol. 18, no. 1, pp. 76–82, 2015..

ORIGINALITY REPORT

---

10%

SIMILARITY INDEX

4%

INTERNET SOURCES

3%

PUBLICATIONS

7%

STUDENT PAPERS

---

PRIMARY SOURCES

---

1	Submitted to Universitas Brawijaya Student Paper	1%
2	prosiding.seminar-id.com Internet Source	1%
3	mukhsonrofi.wordpress.com Internet Source	1%
4	Submitted to Universitas 17 Agustus 1945 Surabaya Student Paper	1%
5	Submitted to Syiah Kuala University Student Paper	1%
6	Imam Kurniawan, Ajib Susanto. "Implementasi Metode K-Means dan Naïve Bayes Classifier untuk Analisis Sentimen Pemilihan Presiden (Pilpres) 2019", Eksplora Informatika, 2019 Publication	1%
7	Submitted to Universitas Dian Nuswantoro Student Paper	1%

---

8

Morikawa, Yoji Nishimura et al. "Identification of chromosomal aberrations of metastatic potential in colorectal carcinoma", Genes, Chromosomes and Cancer, 2010

Publication

<1%

9

Submitted to Sekolah Ciiputra High School

Student Paper

<1%

10

Submitted to KYUNG HEE UNIVERSITY

Student Paper

<1%

11

Submitted to Universitas Atma Jaya Yogyakarta

Student Paper

<1%

12

Submitted to Universitas Islam Indonesia

Student Paper

<1%

13

[www.scilit.net](http://www.scilit.net)

Internet Source

<1%

14

[worldwidescience.org](http://worldwidescience.org)

Internet Source

<1%

15

Yulia Retno Sari, Lidya Rizki Ananda, Maha Rani. "The Analysis of Fuzzy Logic Method and Multiple Linear Regression in Determining National Rice Production to Improve Food Self-Sufficiency in Indonesia", IOP Conference Series: Materials Science and Engineering, 2020

Publication

<1%

16

[journal2.uad.ac.id](http://journal2.uad.ac.id)

Internet Source

<1%

---

17

[media.neliti.com](http://media.neliti.com)

Internet Source

<1%

---

18

Submitted to Universitas Sam Ratulangi

Student Paper

<1%

---

19

Submitted to Universitas Putera Batam

Student Paper

<1%

---

20

Submitted to State Islamic University of  
Alauddin Makassar

Student Paper

<1%

---

21

Ulang Mangun Sosiawan. "Peran Komisi  
Pemberantasan Korupsi (KPK) Dalam  
Pencegahan dan Pemberantasan Korupsi",  
Jurnal Penelitian Hukum De Jure, 2019

Publication

<1%

---

22

[klik.ulm.ac.id](http://klik.ulm.ac.id)

Internet Source

<1%

---

23

Submitted to UIN Syarif Hidayatullah Jakarta

Student Paper

<1%

---

24

Submitted to Surabaya University

Student Paper

<1%

---

25

Submitted to Lambung Mangkurat University

Student Paper

<1%

---

---

Exclude quotes      On

Exclude matches      Off

Exclude bibliography      Off