ISSN: 2477-5126
e-ISSN:2548-9356

Pemanfaatan Algoritma K-Means untuk Membuktikan Implementasi Undang-Undang Pelanggaran Hukum Korupsi di Pengadilan Negeri Banjarmasin

Cinantya Paramita[1*)], Fauzi Adi Rafrastara[2], Catur Supriyanto[3]
[123]Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Semarang
[123]Jl. Imam Bonjol No.207, Pendrikan Kidul, Kec. Semarang Tengah, Kota Semarang, Jawa Tengah 50131
email: [1]cinantya.paramita@dsn.dinus.ac.id, [2]fauziadi@dsn.dinus.ac.id, [3]catur.supriyanto@dsn.dinus.ac.id

Abstract* – *This research aims to demonstrate the implementation of the Anti-Corruption Law in the Banjarmasin District Court by utilizing the K-Means algorithm. Corruption, which persists in Indonesia over a prolonged period, has reached a critical level, making it crucial to enforce the law fairly and firmly. In this study, the panel of judges in the Banjarmasin District Court was analyzed using the K-Means Clustering method and silhouette coefficient to decide corruption cases that result in state losses. The research findings indicate that the optimal number of clusters is 3, with a value of 0.686. However, there is also a lowest value among the 4 clusters, which is 0.454. These clusters are then divided into three categories of enforcement, namely cases that have been executed (108 cases), cases that will be executed (26 cases), and cases that have not been executed (2 cases). All clusters have a silhouette score of 0.742, indicating successful enforcement. This research provides concrete evidence that the panel of judges in the Banjarmasin District Court has implemented the Anti-Corruption Law while considering state losses. By utilizing the K-Means algorithm, this study also contributes to a better understanding of enforcement practices in the court. It is expected that the results of this research will support efforts to enhance the implementation of the Anti-Corruption Law in Indonesia, particularly in the Banjarmasin District Court.

***Abstrak* – Penelitian ini bertujuan untuk membuktikan implementasi Undang-Undang Pelanggaran Hukum Korupsi di Pengadilan Negeri Banjarmasin dengan memanfaatkan algoritma K-Means. Korupsi yang terus terulang di Indonesia selama rentang waktu yang lama telah mencapai tingkat darurat, sehingga penting untuk menegakkan hukum dengan adil dan tegas. Dalam penelitian ini, majelis hakim di Pengadilan Negeri Banjarmasin telah dianalisis menggunakan metode K-Means Clustering dan silhouette coefficient untuk memutuskan perkara korupsi yang mengakibatkan kerugian pada negara. Hasil penelitian menunjukkan bahwa jumlah cluster yang optimal adalah 3, dengan nilai sebesar 0.686. Namun, terdapat juga nilai terendah pada jumlah cluster 4, yaitu 0.454. Cluster-cluster tersebut kemudian dibagi menjadi tiga kelompok pelaksanaan pemidanaan, yaitu sudah terlaksana dengan 108 kasus, akan terlaksana dengan 26 kasus, dan belum terlaksana dengan 2 kasus. Seluruh cluster memiliki silhouette score sebesar 0.742, menunjukkan keberhasilan dalam pelaksanaan pemidanaan. Penelitian ini memberikan bukti konkret bahwa majelis hakim di Pengadilan Negeri Banjarmasin telah melaksanakan Undang-Undang kejahatan korupsi dengan mempertimbangkan kerugian negara. Dengan memanfaatkan algoritma K-Means, penelitian ini juga berkontribusi dalam memperoleh pemahaman yang lebih baik tentang pelaksanaan pemidanaan di pengadilan. Diharapkan hasil penelitian ini dapat mendukung upaya meningkatkan implementasi Undang-Undang Pemberantasan Tindak Pidana Korupsi di Indonesia, khususnya di Pengadilan Negeri Banjarmasin.**

***Kata Kunci* –korupsi, data mining, k-means clustering, silhouette coefficient**

*) **penulis korespondensi**: Ciantya Paramita
Email: cinantya.paramita@dsn.dinus.ac.id

I. PENDAHULUAN

Korupsi merupakan tindakan yang merugikan negara dan merusak sendi-sendi kehidupan berbangsa dan bernegara. Di Indonesia, tindak pidana korupsi semakin menjadi perhatian serius karena bahayanya yang semakin tinggi. Untuk mengatasi masalah ini, Pemerintah Indonesia telah mengeluarkan Undang-Undang Nomor 20 Tahun 2001 tentang perubahan atas Undang-Undang Nomor 31 Tahun 1999 tentang Pemberantasan Tindak Pidana Korupsi. Undang-Undang tersebut memfokuskan pada penempatan kerugian keuangan negara. Dalam hal ini, pengadilan khusus korupsi memainkan peran penting dalam menyelesaikan kasus kejahatan korupsi agar harta kekayaan negara yang hilang dapat dikembalikan [1].

Pada tahun 2020 telah diberlakukan Peraturan Mahkamah Agung Republik Indonesia Nomor 1 Tahun 2020 tentang Pedoman Pemidanaan Pasal 2 dan 3 Undang-Undang Pemberantasan Korupsi [2][3] yang dipergunakan majelis hakim Pengadilan Negeri Banjarmasin dalam memutuskan perkara tindak pidana korupsi.

Dalam rangka membuktikan implementasi Undang-Undang Pemberantasan Tindak Pidana Korupsi, penelitian ini memfokuskan pada pemanfaatan algoritma K-Means untuk mengelompokkan data pemidanaan di Pengadilan Negeri Banjarmasin. Salah satu keunggulan dari algoritma K-Means dalam membuktikan implementasi Undang-Undang Pelanggaran Hukum Korupsi di Pengadilan Negeri Banjarmasin adalah kemampuannya untuk mengelompokkan data pemidanaan secara efisien dan mudah dipahami. Algoritma K-Means dapat mengelompokkan data berdasarkan kesamaan karakteristik, sehingga memungkinkan identifikasi pola dan hubungan antar data pemidanaan korupsi. Dengan menggunakan algoritma ini, pengadilan dapat mengklasifikasikan kasus-kasus korupsi ke dalam kelompok yang relevan, memberikan pemahaman yang lebih baik tentang berbagai aspek pelanggaran hukum korupsi yang diadili di Pengadilan Negeri Banjarmasin. Selain itu, algoritma K-

Means juga dapat memberikan hasil pengelompokan secara cepat dan efisien, sehingga mempercepat proses analisis dan pemahaman terhadap implementasi Undang-Undang Pelanggaran Hukum Korupsi. Penelitian ini diharapkan dapat memberikan kontribusi pada peningkatan penerapan Undang-Undang Pemberantasan Tindak Pidana Korupsi di Banjarmasin.

II. PENELITIAN YANG TERKAIT

Sistem Informasi Penelusuran Perkara (SIPP) merupakan aplikasi [4] penyediaan layanan informasi dan administrasi perkara untuk pihak internal dan eksternal pengadilan. Data yang di proses dalam penelitian ini diambil dari sistem tersebut. Penelitian ini memanfaatkan algoritma K-Means dalam membuktikan implementasi Undang-Undang Tindak Pidana Korupsi di Pengadilan Negeri Banjarmasin, dengan menggunakan data putusan pengadilan korupsi di Banjarmasin dalam beberapa tahun terakhir. Data tersebut akan diproses menggunakan algoritma K-Means untuk mengklasifikasikan putusan pengadilan menjadi beberapa *cluster*.

Penelitian [5] membahas implementasi algoritma *K-Means* dengan data mengenai kasus imunisasi campak pada anak balita berdasarkan wilayah provinsi, menghasilkan 3 tiga cluster yang terbentuk yaitu cluster tinggi yakni C1, cluster sedang C2 dan cluster rendah C3, dimana terdapat 21 provinsi pada cluster tinggi, 12 provinsi ada cluster sedang dan 1 provinsi pada cluster rendah. hasil klasterisasi dijadikan bahan referensi untuk cluster tertinggi untuk menarik perhatian yang lebih dan meningkatkan upaya sosialisasi imunisasi campak pada anak balita.

Selain itu [6], mengimplementasikan K-Mean untuk mengetahui tingkat kejahatan dimana pada penelitian tersebut menghasilkan *cluster* 1 dengan centroid (3, 1.33, 1.33) rata-rata usia 31 sampai 37 tahun telah melanggar pasal 112 dan pasal 114 KUHP dengan jenis kejahatan narkotika. Pada cluster 2 dengan centroid (3, 7, 7) rata-rata usia 31 sampai 37 telah melanggar pasal 362-363 KUHP dengan jenis kejahatan pencurian. Pada cluster 3 dengan centroid (2.5, 7, 7) rata-rata usia 24 sampai 30 tahun telah melanggar pasal 362-363 KUHP. Hasil penelitian ini dapat menjadikan sebuah informasi pengelompokan kejahatan berdasarkan jenis kejahatan, usia, dan implementasi UU sesuai dengan tindak kejahatan yang dilakukan. Telah berbagai penelitian yang menerapakan silhouette coefficient untuk mengevaluasi K-Means Clustering [7].

Tujuan dari penelitian ini dapat memberikan rujukan kepada para pengambil kebijakan dalam mengevaluasi efektivitas implementasi Undang-Undang Tindak Pidana Korupsi di Pengadilan Negeri Banjarmasin.

III. METODE PENELITIAN

A. Sumber Data

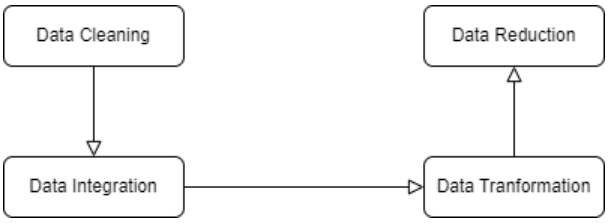
Penelitian ini menggunakan data primer yang diperoleh dari [4] (open-source). Data bersifat kualitatif dan kuantitatif. Adapun data yang diambil adalah data umum dan data putusan majelis hakim. Data umum terdiri dari: nomor perkara, tahun, pkerjaan, dan pasal. Sedangkan data putusan majelis yakni: kerugian negara, lama penjara (bulan), dan denda.

B. Metode Pengumpulan Data

Proses pengumpulan data dari Sistem Informasi SIPP [4] Pengadilan Negeri Banjarmasin menggunakan teknik studi dokumen. Data yang diambil dari Sistem Informasi Penelusuran Perkara (SIPP) terhitung sejak Januari 2017 sampai November 2022 dan dimasukkan secara manual pada microsoft excel kemudian dikonversi ke file CSV.

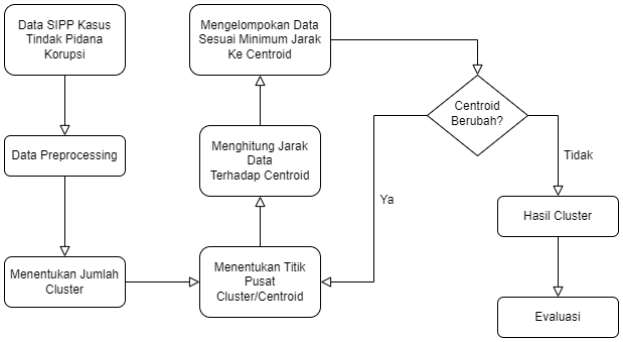
C. Preprocessing

Preprocessing berguna agar data memenuhi persyaratan untuk input K-Means Clustering [8][9]. Gbr. 1 menunjukkan tahapan dari preprocessing data. K-Mean memiliki standarisasi seperti variable data numerik, K-Means juga sensitive terhadap data *outlier*, dimana penyebaran variable yang simetris, dan variabel pada skala yang sama memiliki mean dan varian yang sama [10] dengan rentang -1.0 hingga 1.0 (data standar) atau 0.0 hingga 1.0 (data yang dinormalisasi).



Gbr. 1 Alur Processing Data

Untuk dapat mengetahui penyebaran data dapat menggunakan formula standar deviasi dimana semakin tinggi standar deviasi, semakin besar penyebaran data, sedangkan semakin rendah standar deviasi, semakin dekat data dengan nilai rata-rata.



Gbr. 2 Flowchart Metode

$$x' = \frac{x - \bar{x}}{\sigma} \quad (1)$$

Keterangan: x' = Nilai standarisasi data, x = Nilai data normal, $\bar{x}$=Rata-rata, dan σ = Standar deviasi.

D. K-Means Clustering

Perhitungan manual dengan menggunakan microsoft excel untuk memberikan gambaran proses klasterisasi dengan menggunakan algoritma K-Means. Tahap awal yakni dengan menentukan jumlah cluster atau nilai K [11], dengan menetapkan centroid atau titik pusat yang dipilih secara acak, tahap kedua mengkalkulasikan jarak antara data dengan pusat klaster dengan mengimplementasikan formula *euclidean distance*:

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x_i - y_i)^2} \quad (2)$$

Keterangan: d = Jarak antara x dan y, x = Data, y = Centroid, i = Iterasi, dan m = Jumlah data.

Tahap ketiga mengkategorikan data pada setiap cluster dengan jarak minimum, tahap keempat menghitung centroid dengan menggunakan formula:

$$C_k = \left(\frac{1}{n_k}\right)\sum d_i \quad (3)$$

Keterangan: Ck = Centroid, k = Cluster, n = Jumlah data, d = Data, i = Iterasi.

E. Silhouette Coefficient

Silhouette Coefficient [12] digunakan untuk mengevaluasi kualitas pengelompokan data tindak pidana. Nilai Silhouette Coefficient berkisar antara -1 hingga 1, dengan nilai yang lebih tinggi menunjukkan pengelompokan yang lebih baik. Tahap pertama dalam melakukan *Silhouette Coefficient* yakni rata-rata jarak data ke-i pada semua data di cluster:

$$a(i) = \frac{1}{|A|-1}\sum_{j \in A, j \neq 1} d(i,j) \quad (5)$$

Keterangan: a(i) = Perbedaan rata-rata objek (i) ke semua objek lain pada A, A = Jumlah data pada cluster, dan d(i, j) = Jarak antara data i ke j.

Tahap kedua menghitung nilai minimum dari rata-rata jarak dari data ke-i terhadap semua data di klister.

$$d(i,C) = \frac{1}{|A|}\sum_{j \in C} d(i,j) \quad (6)$$

$$b(i) = \min_{c \neq A} d(i,C) \quad (7)$$

Keterangan: d(i, C) merupakan perbedaan rata-rata antara objek i dan semua objek lain di dalam cluster C, dengan C sebagai cluster selain cluster A, dan b(i) sebagai rata-rata jarak antara objek i dan semua objek lain di cluster-cluster lainnya

Pada persamaan 6 dan 7 berfungsi untuk mencari cluster C yang berbeda dengan cluster A di mana data ke-i berada, dengan tujuan memilih cluster dengan jarak minimum sebagai cluster yang memiliki rata-rata jarak terkecil dari data ke-i. Maka tahap akhir menghitung silhouette coefficient score dengan persamaan:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (8)$$

Keterangan:

s(i) = Silhouette Coefficient.

Perlunya memberikan pengukuran atau penilaian terhadap kualitas cluster berdasarkan rentang Silhouette Coefficient (SC) yakni:

TABEL I
KRITERIA PENILAIAN KUALITAS CLUSTER [12]

Silhouette Index	Kualitas
≤ 0.25	Struktur Buruk
0.26 – 0.50	Struktur Lemah
0.51 – 0.70	Struktur Layak
0.71 – 1.00	Struktur Bagus

IV. HASIL DAN PEMBAHASAN

Proses data mining pada penelitian ini adalah pengumpulan data, diamana data diinputkan secara manual pada microsoft excel kemudian dikonversi ke file CSV (comma delimited). Data dengan format CSV tersebut akan diolah menggunakan bahasa pemrograman python dan Jupyter Notebook [13]. Pada penelitian ini klasterisasi data berupa 3 dimensi antara kerugian negara, penjara (bulan), dan denda. Dataframe ditampilkan menggunakan library open-source pada python pandas.

Pada Gbr. 3 terlihat bahwa dari 136 kasus korupsi yang mengakibatkan kerugian keuangan negara di Pengadilan Negeri Banjarmasin dimana jumlah tertinggi kasus tersebut pada tahun 2021 dengan jumlah 31 kasus. Berdasarkan Gbr. 2 dan UU Pemberantasan Tindak Pidana Korupsi terdapat 110 kasus terjerat pasal 3 dan 26 kasus terjerat kasus pasal 2 ayat 1, dan Jumlah kasus tertinggi terjerat pasal 3.

	No Perkara	Tahun	Pekerjaan	Pasal	Kerugian Negara	Penjara (Bulan)	Denda
0	27/Pid.Sus-TPK/2022/PN Bjm	2022	ASN	2 ayat 1	1356851255	72	30000000
1	26/Pid.Sus-TPK/2022/PN Bjm	2022	Kepala Desa	2 ayat 1	579620700	54	20000000
2	25/Pid.Sus-TPK/2022/PN Bjm	2022	Swasta	2 ayat 1	2258335620	84	30000000
3	23/Pid.Sus-TPK/2022/PN Bjm	2022	Perangkat Desa	3	319170146	24	5000000
4	22/Pid.Sus-TPK/2022/PN Bjm	2022	ASN	2 ayat 1	2796725900	60	20000000
5	21/Pid.Sus-TPK/2022/PN Bjm	2022	ASN	2 ayat 1	2796725900	84	20000000
6	16/Pid.Sus-TPK/2022/PN Bjm	2022	Kepala Desa	3	191813407	24	5000000
7	15/Pid.Sus-TPK/2022/PN Bjm	2022	Swasta	3	550926727	18	5000000
8	14/Pid.Sus-TPK/2022/PN Bjm	2022	Swasta	3	550926727	30	5000000
9	12/Pid.Sus-TPK/2022/PN Bjm	2022	Kepala Desa	3	192178208	36	10000000

Gbr. 3 Dataframe Kasus

Dari Gbr. 2 dapat kita dapat melakukan standarisasi data menggunakan standar deviasi dengan standardscaler yaitu salah satu class dari library sklearn [14] untuk menormalisasi data agar data tidak memiliki penyimpangan yang besar. Sklearn sendiri merupakan salah satu modul dari bahasa pemrograman python yang dibangun berdasarkan *matplotlib*, *scipy*, dan *numpy*. Standarisasi data no perkara 14/Pid.Sus-TPK/2022/PN Bjm:

Kerugian (X):

$$\bar{x} = 920614977$$
$$\sigma = 957811848{,}2$$
$$x' = \frac{x - \bar{x}}{\sigma} = \frac{550926727 - 920614977}{957811848{,}2} = -0{,}385972$$

Penjara (Bulan) (Y):

$$\bar{y} = 42$$
$$\sigma = 21{,}13$$
$$y' = \frac{y - \bar{y}}{\sigma} = \frac{30 - 42}{21{,}13} = -0{,}567962$$

Sehingga menghasilkan Tabel II.

TABEL II
STANDARISASI DATASET

No Perkara	Kerugian Negara (X)	Penjara (Bulan) (Y)
15/Pid.Sus-TPK/2022/PN Bjm	-0,385972	-1,135924
14/Pid.Sus-TPK/2022/PN Bjm	-0,385972	-0,567962
12/Pid.Sus-TPK/2022/PN Bjm	-0,760522	-0,283981

TABEL III PENENTUAN CENTROID SECARA ACAK

No Perkara	Centroid	Kerugian Negara (X)	Penjara (Bulan)(Y)
21/Pid.SusTPK/2022/PN Bjm	1	1,958747	1,987866

15/Pid.SusTPK/2022/PN Bjm	2	-0,385972	-1,135924
11/Pid.SusTPK/2022/PN Bjm	3	-0,604570	0,283981

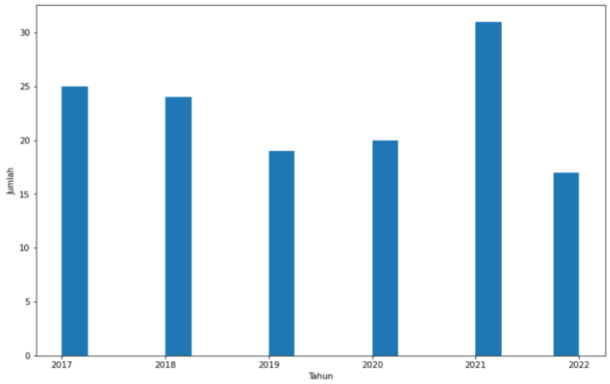
Berdasarkan hasil dari Tabel II, dapat diterapkannya algoritma K-Means dengan penentuan centroid secara acak yang dapat dilihat pada Tabel III, selanjutnya menetapkan jarak terdekat sebagai contoh diambilnya 1 data dengan no perkara 14/Pid.Sus-TPK/2022/PN Bjm.

$$C1 = \sqrt{(-0{,}385972 - 1{,}958747)^2 + (-0{,}567962 - 1{,}987866)^2} = 3{,}468424$$

$$C2 = \sqrt{(-0{,}385972 + 0{,}385972)^2 + (-0{,}567962 + 1{,}135924)^2} = 0{,}567962$$

$$C3 = \sqrt{(-0{,}385972 + 0{,}385972)^2 + (-0{,}567962 + 1{,}135924)^2} = 0{,}879541$$

Dari hasil perhitungan, jarak minimum untuk data dengan no perkara 14/Pid.Sus-TPK/2022/PN Bjm adalah 0,567962 dan terletak pada cluster 2.



Gbr. 4 Histogram Jumlah Kasus (2017-2022)

No Perkara	Kerugian Negara (X)	Kurungan Penjara (Y)	C1	C2	C3	Minimum	Cluster
22/Pid.Sus-TPK/2022/PN Bjm	1,958747	0,851943	1,13592	3,073974	2,625486	1	1
21/Pid.Sus-TPK/2022/PN Bjm	1,958747	1,987866	0	3,905863	3,077957	0	1
16/Pid.Sus-TPK/2022/PN Bjm	-0,760903	-0,851943	3,932049	0,470339	1,146631	0	2
15/Pid.Sus-TPK/2022/PN Bjm	-0,385972	-1,135924	3,905863	0	1,436633	0	2
14/Pid.Sus-TPK/2022/PN Bjm	-0,385972	-0,567962	3,468424	0,567962	0,879541	0,567962	2
12/Pid.Sus-TPK/2022/PN Bjm	-0,760522	-0,283981	3,543404	0,930642	0,588984	0,588984	3
11/Pid.Sus-TPK/2022/PN Bjm	-0,604570	0,283981	3,077957	1,436633	0	0	3
9/Pid.Sus-TPK/2022/PN Bjm	-0,575443	0,283981	3,053743	1,432490	0,029127	0	3
8/Pid.Sus-TPK/2022/PN Bjm	-0,417055	-1,419905	4,154195	0,285677	1,714173	0,285677	2
5/Pid.Sus-TPK/2022/PN Bjm	-0,027057	0,851943	2,287737	2,020008	0,810001	0,810001	3

Gbr. 5 Perhitungan Jarak Minimum Centroid Acak

Gbr.5 menghasilkan pengelompokkan cluster berdasarkan perhitungan jarak minimum antara data dan centroid acak. Nilai minimum diperoleh melalui penetapan jumlah centroid yang telah ditentukan sebesar 3 dalam Gbr.5.

Pada tahap awal, centroid awal dipilih secara acak, dan titik-titik data juga dipilih secara acak sebagai centroid awal. Selanjutnya, jarak antara setiap titik data dengan centroid dihitung menggunakan rumus Euclidean. Tahap berikutnya melibatkan penetapan centroid baru yang terlihat di Tabel IV, untuk setiap cluster dengan memilih titik data yang memiliki jarak minimum ke centroid. Proses ini diulangi dengan menghitung kembali jarak antara setiap titik data dengan centroid yang baru ditetapkan. Langkah ini diulangi hingga tidak ada perubahan dalam penempatan centroid atau hingga mencapai konvergensi, di mana tidak ada lagi perubahan dalam jarak antara titik data. Pada titik ini, centroid yang ditetapkan merupakan nilai minimum dalam perhitungan jarak minimum centroid acak.

Dalam algoritma K-means, pemilihan titik-titik data secara acak sebagai centroid awal dapat berdampak pada hasil akhir. Maka dari itu, sangat penting untuk menjalankan algoritma ini beberapa kali dengan inisialisasi yang berbeda untuk memastikan stabilitas data yang dihasilkan

Untuk hasil berikut dilakukannya iterasi yang dipergunakan untuk centroid baru, sebagai berikut dimana:

$$X = \frac{-0{,}760903 + (-0{,}385972) + (-0{,}38597) + (-0{,}575443)}{4} = -0{,}487475$$

$$Y = \frac{0{,}851943 + (-1{,}135924) + (0{,}567962) + (0{,}283981)}{4} = -0{,}993933$$

Dengan tujuan menetapkan label atau kategori cluster yang sesuai untuk setiap objek yang belum diklasifikasikan, maka dihasilkan Tabel V.

TABEL IV
PENENTUAN CENTROID BARU PERTAMA

Cluster baru	X	Y
Centroid Baru 1	1,958747	2,839809
Centroid Baru 2	-0,487475	-0,993933
Centroid Baru 3	-0,491898	0,283981

No Perkara	Kerugian Negara (X)	Kurungan Penjara (Y)	C1	C2	C3	Minimum	Cluster	Keterangan
22/Pid.Sus-TPK/2022/PN Bjm	1,958747	0,851943	1,987866	3,064516	2,515600	1,98787	1	Aman
21/Pid.Sus-TPK/2022/PN Bjm	1,958747	1,987866	0,851943	3,856829	2,984776	0,85194	1	Aman
16/Pid.Sus-TPK/2022/PN Bjm	-0,760903	-0,851943	4,585360	0,308097	1,167341	0,308097	2	Aman
15/Pid.Sus-TPK/2022/PN Bjm	-0,385972	-1,135924	4,615643	0,174540	1,423850	0,174540	2	Aman
14/Pid.Sus-TPK/2022/PN Bjm	-0,385972	-0,567962	4,136497	0,437898	0,858503	0,437898	2	Aman
12/Pid.Sus-TPK/2022/PN Bjm	-0,760522	-0,283981	4,141556	0,760649	0,628283	0,628283	3	Aman
11/Pid.Sus-TPK/2022/PN Bjm	-0,604570	0,283981	3,619786	1,283268	0,112672	0	3	Aman
9/Pid.Sus-TPK/2022/PN Bjm	-0,575443	0,283981	3,599219	1,280938	0,083545	0	3	Aman
8/Pid.Sus-TPK/2022/PN Bjm	-0,417055	-1,419905	4,877458	0,431753	1,705528	0	2	Aman
5/Pid.Sus-TPK/2022/PN Bjm	-0,027057	0,851943	2,809810	1,902431	0,733933	0,733933	3	Aman

Gbr. 6 Komputasi Jarak Minimum Centroid Baru Pertama

ISSN: 2477-5126
e-ISSN:2548-9356

Setelah menentukan *centroid* baru berdasarkan cluster pertama setiap data dihitung jarak minimumnya kembali menggunakan rumus *euclidean distance* untuk memastikan tidak ada data yang berpindah cluster. Misalkan untuk menghitung jarak minimum no perkara 14/Pid.Sus-TPK/2022/PN Bjm pada centroid baru iterasi 1:

$$C1 = \sqrt{(-0{,}385972 - 1{,}958747)^2 + (-0{,}567962 - 2{,}839809)^2} = 4{,}136497$$

$$C2 = \sqrt{(-0{,}385972 - 0{,}487475)^2 + (-0{,}567962 - 0{,}993933)^2} = 0{,}437898$$

$$C3 = \sqrt{(-0{,}385972 - 0{,}491898)^2 + (-0{,}567962 - 0{,}283981)^2} = 0{,}858503$$

Dari hasil perhitungan, jarak minimum untuk data dengan no perkara 14/Pid.SusTPK/2022/PN Bjm, pada centroid baru iterasi 1 adalah 0,437898 dan terletak pada cluster 2, maka dapat dikatakan tidak berpindah cluster informasi tersebut terdokumentasi pada Gbr. 6. Untuk mengetahui kualitas struktur suatu cluster yang terbentuk dapat diketahui dari SC pada Gbr. 7.

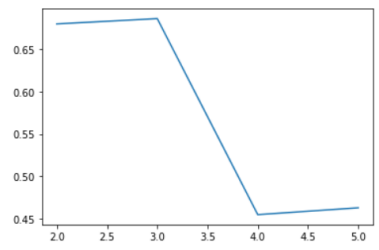
No Perkara	Kerugian Negara (X)	Kurungan Penjara (Y)	Cluster	a(i)	d(i,1)	d(i,2)	d(i,3)	b(i)	s(i)
22/Pid.Sus-TPK/2022/PN Bjm	1,958747	0,851943	1	1,135924	-	3,077909	1,583655	1,583655	0,28272
21/Pid.Sus-TPK/2022/PN Bjm	1,958747	1,987866	1	1,135924	-	3,865133	1,848117	1,848117	0,385362
16/Pid.Sus-TPK/2022/PN Bjm	-0,760903	-0,851943	2	0,470388	3,570683	-	0,661661	0,661661	0,28908
15/Pid.Sus-TPK/2022/PN Bjm	-0,385972	-1,135924	2	0,416749	3,489919	-	0,918325	0,918325	0,546186

Gbr. 7 Perhitungan Silhouette Coefficient

K-Means Clustering merupakan metode data mining unsupervised yang digunakan untuk pengelompokan data berdasarkan feature. Penentuan nilai hasil cluster berdasarkan jumlah centroid dan jarak antar objek data. Algoritma K-Means memiliki kelemahan pada menentukan jumlah K terbaik dalam mengklasterisasi suatu dataset.

Penentuan jumlah K [15] merupakan proses penting pada K-Means Clustering. Pada penelitian ini dalam menentukan jumlah K terbaik [16] menggunakan metode silhouette coefficient dengan library sklearn [14]. Pengujian dilakukan dengan menguji jumlah klaster mulai dari K = 2 hingga K = 5.

```
2 [0.6802143657192251]
3 [0.6802143657192251, 0.6864697647968059]
4 [0.6802143657192251, 0.6864697647968059, 0.4544744596833943]
5 [0.6802143657192251, 0.6864697647968059, 0.4544744596833943, 0.4625847169214095]
```



Gbr. 8 Silhouette Score

Pada Gbr. 4 diatas dapat diketahui jumlah cluster dan *silhouette score* terbaik adalah K = 3 dengan *silhouette score* = 0.686. *Silhouette score* semakin turun ketika K = 4 dengan silhouette score = 0.454.

TABEL V
SILHOUETTE SCORE

No	*Cluster*	*Silhouette Score*
1	K=2	0.680
2	K=3	0.686
3	K=4	0.454
4	K=5	0.462

Label cluster untuk K = 3 (0, 1 ,2) berfungsi untuk mengetahui letak cluster masing-masing data. Untuk menampilkan array label cluster data dapat menggunakan library sklearn.

	Kerugian Negara	Penjara (Bulan)	Denda	Cluster
0	0.364540	2.405598	2.784734	2
1	-0.095613	1.386933	1.459366	2
2	0.898256	3.084709	2.784734	2
3	-0.249810	-0.310843	-0.528685	0
4	1.217005	1.726488	1.459366	2
...	...	...	...	...
131	0.840423	-0.310843	-0.528685	0
132	-0.096534	1.047378	1.459366	2
133	-0.379568	0.368268	0.796682	0
134	-0.423971	-0.989953	-0.528685	0
135	-0.122170	-0.989953	-0.528685	0

136 rows × 4 columns

Gbr. 9 Label Cluster Pada Dataframe Transformation

Untuk dapat dengan mudah dalam mengidentifikasi dan evaluasi kualitas cluster, sebaran data, dan hubungan antar cluster maka perlunya mencari centroid cluster dari label cluster, dimana diperoleh Tabel X, sebagai berikut:

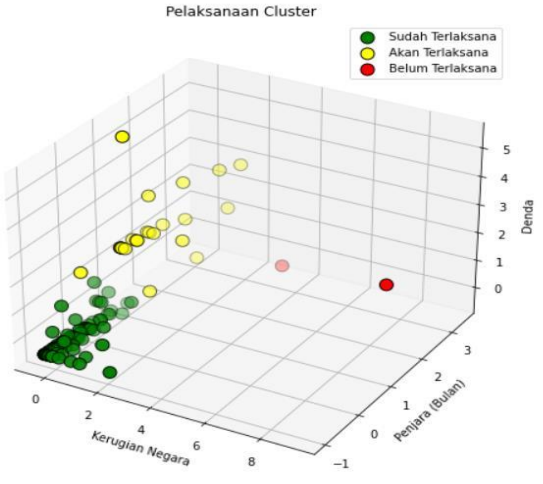
TABEL VI
NILAI PUSAT CLUSTER

Centroid	X	Y	Z
1	-0.17	-0.42	-0.45
2	7.11	1.56	-1.79
3	0.16	1.64	1.71

Pada visualisasi cluster [17] diatas dapat diketahui terdapat 3 cluster pelaksanaan yaitu sudah terlaksana (hijau), akan terlaksana (kuning), dan belum terlaksana (merah). Pada cluster sudah terlaksana terdapat 108 kasus, akan terlaksana terdapat 26 kasus, dan belum terlaksana terdapat 2 kasus. Terlihat pada Gbr. 5, dimana hasil *silhouette score cluster* pelaksanaan untuk jumlah cluster 3 adalah 0.742 dimana pada Tabel I. Kriteria Penilaian Kualitas Cluster menunjukkan bahwa *Silhouette Index* dengan kriteria nilai 0.71 – 1.00 diartikan kuat secara struktur cluster.

V. KESIMPULAN

Dengan menggunakan metode K-Means Clustering dapat diketahui ada tidaknya disparitas pemidanaan dalam penjatuhan sanksi pidana yang diterapkan untuk tindak pidana korupsi yang menghasilkan kerugian bagi negara yang terjadi di Pengadilan Negeri Banjarmasin. Disparitas dalam pemidanaan tindak pidana korupsi terjadi ketika terdapat perbedaan yang signifikan dalam hukuman yang diberikan dalam kasus-kasus serupa, yang mengakibatkan ketidakadilan dan ketidakpercayaan masyarakat.



Gbr. 10 Visualisasi Cluster Pelaksanaan

Melalui visualisasi cluster disparitas 136 data pada Sistem Informasi Penelusuran Perkara (SIPP) [4] Pengadilan Negeri Banjarmasin yang dihasilkan K-Means Clustering menggunakan silhouette coefficient untuk 3 cluster dapat diketahui terdapat 108 kasus sudah terlaksana, 26 kasus akan terlaksana, dan 2 kasus belum terlaksana dengan silhouette score 0.742.

DAFTAR PUSTAKA

[1] H. M. Sitohang, "Analisis Hukum Terhadap Tindak Pidana Korupsi Dengan Penyalagunaan Jabatan Dalam Bentuk Penyuapan Aktif," *PATIK J. Huk.*, vol. 07, no. 2086–4434, pp. 75–88, 2018.

[2] K. R. Indonesia, "Peraturan Mahkamah Agung No.1 Th 2020, Pedoman pasal 2 dan 3 UU Pemberantasan Tindak Pidana Korupsi." .

[3] W. Saputro, M. Reza Pahlevi, and A. Wibowo, "Analisis Algoritma K-Means Untuk Klasterisasi Tindak Pidana Korupsi Di Wilayah Hukum Indonesia," *JIKO (Jurnal Inform. dan Komputer)*, vol. 3, no. 3, pp. 137–142, 2020, doi: 10.33387/jiko.v3i3.1960.

[4] M. A. R. Indonesia, "Sistem Informasi Penelusuran Perkara - Pengadilan Negeri Banjarmasin," 2023. https://sipp.pn-banjarmasin.go.id/.

[5] R. W. Sari, A. Wanto, and A. P. Windarto, "Implementasi Rapidminer Dengan Metode K-Means (Study Kasus: Imunisasi Campak Pada Balita Berdasarkan Provinsi)," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 224–230, 2018, doi: 10.30865/komik.v2i1.930.

[6] Y. Mayona, R. Buaton, and M. Simanjutak, "Data Mining Clustering Tingkat Kejahatan Dengan Metode Algoritma K-Means (Studi Kasus: Kejaksaan Negeri Binjai)," *J. Inform. Kaputama*, vol. 6, no. 3, pp. 2548–9739, 2022.

[7] P. W. Cahyo and L. Sudarmana, "A Comparison of K-Means and Agglomerative Clustering for Users Segmentation based on Question Answerer Reputation in Brainly Platform," *Elinvo (Electronics, Informatics, Vocat. Educ.*, vol. 6, no. 2, pp. 166–173, 2021, [Online]. Available: https://journal.uny.ac.id/index.php/elinvo/article/view/44486.

[8] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.

[9] Z. Mustakim and R. Kamal, "K-Means Clustering for Classifying the Quality Management of Secondary Education in Indonesia," *Cakrawala Pendidik.*, vol. 40, no. 3, pp. 725–737, 2021, doi: 10.21831/cp.v40i3.40150.

[10] F. N. Musid, "Implementasi Algoritma K-Means Clustering Dalam Pengelompokkan Data Jumlah Kerusakan Rumah Berdasarkan Kondisi Di Jawa Barat," vol. 1, no. 3, pp. 101–114, 2023.

[11] D. O. Dacwanda and Y. Nataliani, "Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan," *Aiti*, vol. 18, no. 2, pp. 125–138, 2021, doi: 10.24246/aiti.v18i2.125-138.

[12] R. T. Vulandari, W. L. Y. Saptomo, and D. W. Aditama, "Application of K-Means Clustering in Mapping of Central Java Crime Area," *Indones. J. Appl. Stat.*, vol. 3, no. 1, p. 38, 2020, doi: 10.13057/ijas.v3i1.40984.

[13] B. M. Randles, I. V. Pasquetto, M. S. Golshan, and C. L. Borgman, "Using the Jupyter Notebook as a Tool for Open Science: An Empirical Study," *Proc. ACM/IEEE Jt. Conf. Digit. Libr.*, pp. 17–18, 2017, doi: 10.1109/JCDL.2017.7991618.

[14] A. L. Ramdani and H. B. Firmansyah, "Clustering Application for UKT Determination Using Pillar K-Means Clustering Algorithm and Flask Web Framework," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 2, p. 53, 2018, doi: 10.24014/ijaidm.v1i2.5126.

[15] M. T. Islam and M. A. Yousuf, "Development of a Corruption Detection Algorithm using K-means Clustering," *2018 Int. Conf. Adv. Electr. Electron. Eng. ICAEEE 2018*, no. November 2018, 2019, doi: 10.1109/ICAEEE.2018.8642985.

[16] C. Yuan and H. Yang, "Research on K-Value Selection Method of K-Means Clustering Algorithm," *J*, vol. 2, no. 2, pp. 226–235, 2019, doi: 10.3390/j2020016.

[17] G. Andrienko, N. Andrienko, I. Kopanakis, A. Ligtenberg, and S. Wrobel, "Visual analytics methods for movement data," *Mobility, Data Min. Priv. Geogr. Knowl. Discov.*, pp. 375–410, 2008, doi: 10.1007/978-3-540-75177-9_14.