

Penerapan Algoritma K-Means Untuk Menentukan Kelas Besar Tunggakan Pajak Kendaraan Bermotor Pada SAMSAT Toli-Toli

Sukardi, Nur Alinuddin Kaharu, Anwar S. Panyili, Bonitalia

STMIK Adhi Guna

Jl. Undata No. 3, Palu, Sulawesi Tengah

e-mail: sukarvi@gmail.com

Abstrak

Pajak kendaraan bermotor adalah pajak yang dikelola oleh kantor SAMSAT yang berlokasi di setiap provinsi, wilayah dan kota. Pemungutan pajak kendaraan bermotor secara efektif perlu dilakukan untuk meningkatkan pendapatan asli daerah. Kesadaran wajib pajak untuk memenuhi kewajibannya menjadi hal penting dalam permasalahan ini. Jika wajib pajak tidak melakukan penunggakan maka pendapatan daerah akan stabil. Namun demikian, beberapa wajib pajak tidak membayar pajak tepat waktu, sehingga tiap tahun kendaraan yang menunggak semakin meningkat, sehingga pendapatan yang diterima oleh daerah juga mengalami penurunan. Maka perlu adanya implementasi Algoritma K-Means Untuk Menentukan Kelas Besar Tunggakan Pajak Bermotor. Untuk melakukan evaluasi pengelompokan data pada penelitian ini, digunakan metode yang umum dilakukan untuk mengukur performa k-means yaitu silhouette index dan davies bouldin index. Berdasarkan hasil pengujian Silhouette Index memiliki rentang nilai dari -1 hingga 1. Semakin mendekati nilai 1 maka semakin baik performa cluster. Pada eksperimen yang telah dilakukan rata-rata silhouette index sebesar 0,86955 dengan nilai silhouette index terbesar dengan nilai 0,88891 dengan $K=3$.

Kata kunci : Pajak Kendaraan, Algoritma K-Means, Silhouette index, Davies bouldin index

Abstract

Motor vehicle tax is a tax managed by the SAMSAT office, which is located in each province, region, and city. Effective collection of motor vehicle taxes needs to be carried out to increase local revenue. Taxpayer awareness in fulfilling their obligations is an important factor in this issue. If taxpayers do not default on their payments, local revenue will remain stable. However, some taxpayers fail to pay their taxes on time, leading to an increasing number of overdue vehicles each year, which in turn results in a decline in regional revenue. Therefore, the implementation of the K-Means algorithm is necessary to classify the levels of motor vehicle tax arrears. To evaluate the data clustering in this study, commonly used methods for measuring K-Means performance, namely the Silhouette Index and Davies-Bouldin Index, are employed. Based on the test results, the Silhouette Index ranges from -1 to 1, where values closer to 1 indicate better cluster performance. In the experiments conducted, the average Silhouette Index was 0.86955, with the highest Silhouette Index value being 0.88891 at $K=3$.

Keywords: Vehicle Tax, K-Means Algorithm, Silhouette Index, Davies-Bouldin Index

1. Pendahuluan

Membayar pajak merupakan sebuah kewajiban yang harus dilakukan sebagai warga negara Indonesia. Jika tidak bayar pajak, maka dapat memperoleh hukuman, mulai dari denda hingga kurungan penjara. Penerimaan pajak mengalami peningkatan yang cukup signifikan baik dalam jumlah nominal maupun persentase terhadap jumlah keseluruhan pendapatan negara. Namun di sisi lain persentase Wajib Pajak masih sangat kecil jika dibandingkan dengan jumlah seluruh penduduk di Indonesia. Hal ini menunjukkan kesadaran masyarakat Indonesia untuk membayar pajak masih rendah.

Pajak kendaraan bermotor adalah pajak yang dikelola oleh kantor SAMSAT yang berlokasi di setiap provinsi, wilayah dan kota. Pemungutan pajak kendaraan bermotor secara efektif perlu dilakukan untuk meningkatkan pendapatan asli daerah. Di Indonesia, perpajakan merupakan pendapatan tertinggi

untuk meningkatkan perekonomian yang dapat membantu pembangunan suatu daerah dan membiayai kegiatan pemerintahan.

Pajak kendaraan bermotor yang telah habis masa berlakunya dianggap telah menunggak, dan akan diterbitkan surat peringatan berupa Surat Ketetapan Pajak Daerah (SKPD). Semua data tunggakan pajak kendaraan bermotor yang terdaftar akan diatur dan surat tunggakan pajak kendaraan akan disampaikan dalam sistem penilaian resmi (OAS) Dinas Pendapatan Sulawesi Tengah SAMSAT Toli-Toli. Untuk wilayah Toli-Toli, pajak kendaraan bermotor merupakan suatu jenis pajak daerah yang dipungut oleh Dinas Pajak Daerah Provinsi Sulawesi Tengah melalui Kantor Bersama Samsat Toli-Toli sebagai sumber pendapatan yang potensial.

Bertambahnya jumlah kendaraan bermotor maka yang membayar pajak juga semakin banyak, sehingga semakin banyak pula pendapatan daerah yang didapat. Melihat besarnya kontribusi pajak kendaraan bermotor sebagai salah satu pendapatan asli daerah, tentunya pendapatan asli daerah diharapkan untuk meningkat setiap tahunnya, dengan meminimalkan pajak kendaraan yang jatuh tempo serta meningkatkan kesadaran dan kepatuhan wajib pajak dalam melaksanakan kewajibannya. Namun demikian, beberapa wajib pajak tidak membayar pajak tepat waktu, sehingga tiap tahun kendaraan yang menunggak semakin meningkat, sehingga pendapatan yang diterima oleh daerah juga mengalami penurunan. Kesadaran wajib pajak untuk memenuhi kewajibannya menjadi hal penting dalam permasalahan ini. Jika wajib pajak tidak melakukan penunggakan maka pendapatan daerah akan stabil.

2. Metode Penelitian

Penelitian menggunakan Algoritma K-Means mengelompokkan data berdasarkan kedekatan/kemiripan data sehingga data-data yang memiliki karakteristik yang sama akan dimasukkan kedalam cluster yang sama. Kedekatan data tersebut bisa diukur dari seberapa dekat jarak antar data dan jarak data dengan centroid data. Prinsip kerja algoritma K-Means yaitu menentukan jumlah k cluster terlebih dahulu kemudian menentukan titik centroid data secara random/acak. Setelah itu mengalokasikan data ke cluster terdekat dan proses tersebut akan diulang hingga menemukan centroid yang stabil. Keluaran algoritma K-Means sangat bergantung pada penentuan jumlah cluster dan pemilihan centroid awal yang ditentukan secara random/acak. Permasalahan yang ada pada algoritma K-Means adalah menghasilkan centroid akhir yang tidak benar-benar menjadi pusat cluster yang sesungguhnya. Dalam prakteknya algoritma ini harus dijalankan berkali-kali dengan centroid awal yang berbeda-beda untuk mendapatkan centroid akhir yang dianggap paling baik.

Metode evaluasi cluster dapat menyelesaikan masalah tersebut. Metode evaluasi cluster Seperti metode Elbow, Davies Bouldin Index Dan Silhouette Index merupakan metode internal yang dapat membantu untuk mendapatkan klasterisasi ideal pada algoritma K-Means

Pemilihan jumlah k cluster optimal dapat menggunakan metode Elbow, Davies Bouldin Index atau Silhouette Index. Metode Elbow dapat menentukan jumlah k cluster optimal dengan membandingkan persentase antar cluster yang membentuk siku pada titik tertentu. Informasi hasil dari metode Elbow ditunjukkan dengan grafik. Untuk mendapatkan perbandingan tersebut menggunakan nilai Sum of Square Error sebagai berikut:

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_K} \|x_i - C_K\|_2^2 \dots \dots \dots$$

Dimana x_i dan C_K adalah:

x_i = nilai atribut dari data ke i

C_K = nilai atribut titik pusat cluster ke i

Davies bouldin index (DBI) menentukan jumlah k cluster optimal berdasarkan nilai kohesi dan separasi data/objek. Kohesi adalah jumlah kedekatan data/objek terhadap centroid nya dalam satu cluster. Separasi adalah jarak antar centroid dari cluster. Untuk mencari matrik kohesi menggunakan Sum of Square Within Cluster (SSW) sebagai berikut[13]:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i)$$

Dimana m_i , c_i , $d(x_j, c_i)$ adalah:

m_i = jumlah data dalam *cluster* ke i

c_i = *centroid cluster* ke i

$d(x_j, c_i)$ = jarak tiap data/objek ke *centroid* yang dihitung dengan metode jarak *Euclidean*

Sedangkan mengetahui separai data/objek menggunakan *Sum of Between Cluster (SSB)* sebagai berikut[13]:

$$SSB_{i,j} = D(C_i, C_j) \dots\dots\dots(4)$$

Kemudian akan dicari nilai rasio untuk mengetahui nilai perbandingan antara *cluster* ke i dan *cluster* ke j , dihitung dengan persamaan[13]:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \dots\dots\dots(5)$$

Nilai perbandingan rasio digunakan untuk mencari nilai DBI, sebagai berikut[13]:

$$DBI = \frac{1}{K} \sum_{k=1}^k \max_{i \neq j} (R_{i,j}) \dots\dots\dots(6)$$

Dimana K adalah jumlah *cluster* yang digunakan.

3. Hasil dan Pembahasan

Penelitian ini dimulai dengan pengumpulan data wajib pajak kendaraan bermotor pada Kabupaten Toli-Toli. Secara umum, ini berisi identitas pemilik kendaraan, identitas kendaraan dan nilai-nilai pendukung pajak kendaraan bermotor. Dataset ini terdiri dari 19 attribut yakni “nama”, “nopol”, “hp”, “Alamat”, “merk”, “tipe”, “tahun buat”, “warna tknb”, “bobot”, “nilai jual”, “dasar pkb”, “skpd”, “lama tunggakan”, “pokok”, “denda”, “total pkb”, “status”. Untuk dapat melakukan proses cluster pada data ini, attribut pada data ini direduksi untuk menghilangkan data yang tidak berpengaruh. Attribut yang dipilih bertipe data angka dan tidak memiliki missing value. Attribut yang digunakan untuk proses cluster adalah “tahun buat”, “bobot”, “nilai jual”, “dasar pkb”, “usia skpd”, “lama tunggakan”, “total pkb”. Dataset dengan attribute awal dapat dilihat pada table 1.

Table 1. Dataset Dengan Attribut Awal

NAMA	NOPOL	HP	...	DENDA	TOTAL PKB	STATUS
RAHMI	DN1201DG		...	421.313	4.196.273	ADA DAN BEROPERASI
ARIFIN	DN1704DE		...	444.938	4.360.388	ADA DAN BEROPERASI
HANAPI	DN2083DO		...	0	184.440	ADA DAN BEROPERASI
NURDIN	DN5111DP		...	0	302.940	ADA DAN BEROPERASI
NURMILASARI	DN4978DT	082223381212	...	0	254.040	ADA DAN BEROPERASI
WAHIDA M.HI.THAYIB	DN3438DW			52.875	501.255	ADA DAN BEROPERASI
MOHAMMAD NUR SA,BAN	DN4270DP			43.125	429.525	ADA DAN BEROPERASI
M.YUNUS S.	DN2669DF			31.500	308.700	ADA DAN

NAMA	NOPOL	HP	...	DENDA	TOTAL PKB	STATUS
						BEROPERASI
ARDI	DN5508DS	082393975078		94.500	918.540	ADA DAN BEROPERASI
MARMIN.L	DN2250DK			31.875	297.075	ADA DAN BEROPERASI
KANTOR DESA TAMPIALA	DN2026T			15.625	144.375	ADA DAN BEROPERASI

Pada proses input data, dataset yang sudah direduksi atributnya secara teknis ditampung ke dalam file excel untuk dapat diproses pada program yang dikembangkan. Hasil dari pengurangan reduksi dapat dilihat pada tabel berikut.

Table 2. Dataset Hasil Reduski Atribut

THN. BUAT	BOBOT	NILAI JUAL	DASAR PKB	USIA SKPD	LAMA TUNGGAKAN	TOTAL PKB
2012	1,05	113.000.000	118.650.000	721	1	4.360.388
2012	1,05	113.000.000	118.650.000	498	0	1.957.725
2012	1,05	111.000.000	116.550.000	821	1	4.143.353
2012	1	13.100.000	13.100.000	388	0	231.870
2012	1	11.100.000	11.100.000	705	1	407.925
2012	1	12.600.000	12.600.000	699	1	463.050
2012	1	11.000.000	11.000.000	818	1	391.050
2012	1,085	62.000.000	67.270.000	390	0	1.190.679
2012	1	11.200.000	11.200.000	764	1	404.880
...
2021	1,085	93.000.000	100.905.000	543	0	1.634.661
2021	1	14.800.000	14.800.000	364	0	261.960
2021	1	18.000.000	18.000.000	565	0	286.200
2021	1	11.900.000	11.900.000	411	0	207.060
2021	1	13.000.000	13.000.000	439	0	222.300
2021	1	12.400.000	12.400.000	442	0	212.040
2021	1	14.300.000	14.300.000	411	0	248.820
2021	1,05	134.000.000	140.700.000	451	0	2.405.970
2021	1,05	117.000.000	122.850.000	504	0	2.027.025
2021	1	12.400.000	12.400.000	521	0	200.880

Tabel yang ditunjukkan sebelumnya berisi 10 data pertama dan 10 data terakhir dari 241 data yang dikumpulkan.

Dari hasil perancangan program k-means cluster menggunakan python pada bagian sebelumnya, dilakukan evaluasi terhadap performa algoritma terhadap data yang diproses. Seperti yang telah dijelaskan sebelumnya, evaluasi performa dilakukan dengan menggunakan silhouette index dan davies bouldin index. Evaluasi dilakukan dengan menjalankan program dan memproses cluster dengan jumlah kelas yang berbeda. Pada penelitian ini, proses evaluasi dilakukan sebanyak 3 kali dengan jumlah kelas 2, 3 dan 4.

Eksperimen pertama dilakukan dengan menentukan jumlah kelas sebanyak 2. Eksperimen ini menghasilkan nilai silhouette index sebesar 0,86061 dan davies bouldin index sebesar 0,35156.

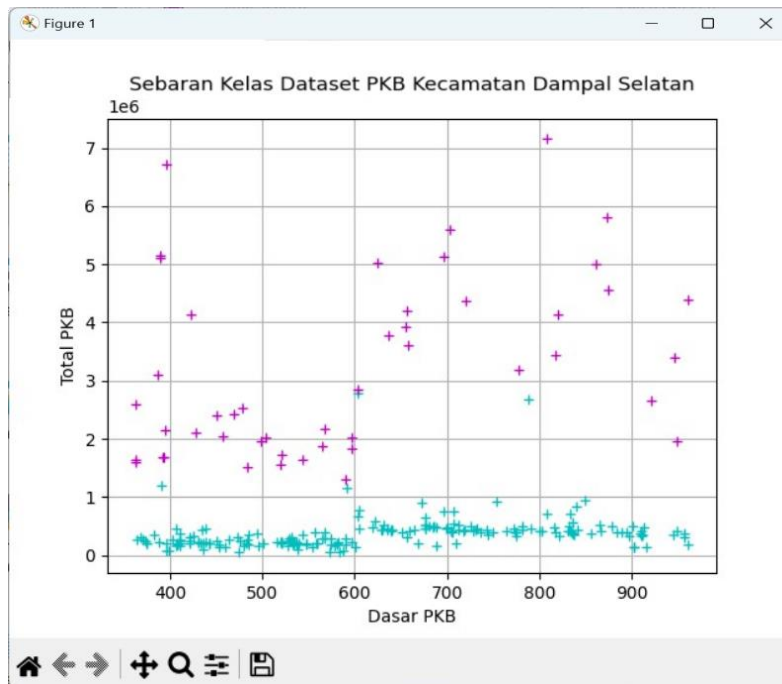
NOPOL	Kelas	BOBOT	NILAI JUAL
DN1201DG	1	1.05	10700000
DN1704DE	1	1.05	11300000
DN2083DO	0	1.0	10600000
DN5111DP	0	1.0	18700000
DN4978DT	0	1.0	14600000
DN3915D	0	1.0	11200000
DN1854DG	1	1.05	14000000
DN1406DG	1	1.05	13700000
DN1579DY	1	1.05	11300000
DN1428DE	1	1.05	86000000
DN1336DG	1	1.05	73000000
DN1564DG	1	1.05	111000000
DN1114D	1	1.05	163000000
DN8216DD	1	1.085	85000000
DN8330DD	1	1.085	89000000
DN3618DW	0	1.0	11800000
DN318DM	0	1.0	13100000
DN6481DQ	0	1.0	16700000
DN2596DP	0	1.0	11000000

Cluster	silhouette score	Davies Bouldin score
Cluster K-2	0.8606113727909047	
Cluster K-3		0.3515572575217268
Cluster K-4		

Gambar sebelumnya menunjukkan hasil pembagian kelas atau cluster dengan jumlah kelas ditentukan sebanyak 2 kelas menggunakan program yang dikembangkan. Sebagai perbandingan, peneliti juga melakukan cluster menggunakan rapidminer yang dapat dilihat pada gambar berikut.

Row No.	NOPOL	cluster	BOBOT	NILAI JUAL	DASAR PKB	USIA SKPD	LAMA TUNGG.	TOTAL PKB
1	DN1201DG	cluster_1	1.050	10700000	112350000	657	1	4195273
2	DN1704DE	cluster_1	1.050	11300000	118650000	721	1	4360388
3	DN2083DO	cluster_0	1	10600000	10600000	408	0	184440
4	DN5111DP	cluster_0	1	18700000	18700000	532	0	302940
5	DN4978DT	cluster_0	1	14600000	14600000	421	0	254040
6	DN3915D	cluster_0	1	11200000	11200000	584	0	58240
7	DN1854DG	cluster_1	1.050	140000000	147000000	362	0	2601900
8	DN1406DG	cluster_1	1.050	137000000	143850000	469	0	2416680
9	DN1579DY	cluster_1	1.050	113000000	118650000	498	0	1957725
10	DN1428DE	cluster_1	1.050	86000000	90300000	363	0	1598310
11	DN1336DG	cluster_1	1.050	73000000	76650000	921	1	2655923
12	DN1564DG	cluster_1	1.050	111000000	116550000	821	1	4143353
13	DN1114D	cluster_1	1.050	163000000	171150000	950	1	1959668
14	DN8216DD	cluster_1	1.085	85000000	92225000	484	0	1521713
15	DN8330DD	cluster_1	1.085	89000000	96565000	393	0	1680231
16	DN3618DW	cluster_0	1	11800000	11800000	443	0	201780

Gambar sebelumnya membuktikan cluster yang dihasilkan menggunakan program yang dikembangkan dengan cluster yang dihasilkan menggunakan rapidminer bernilai sama. Grafik sebaran kelas yang dihasilkan menggunakan program yang dikembangkan dapat dilihat pada gambar berikut.



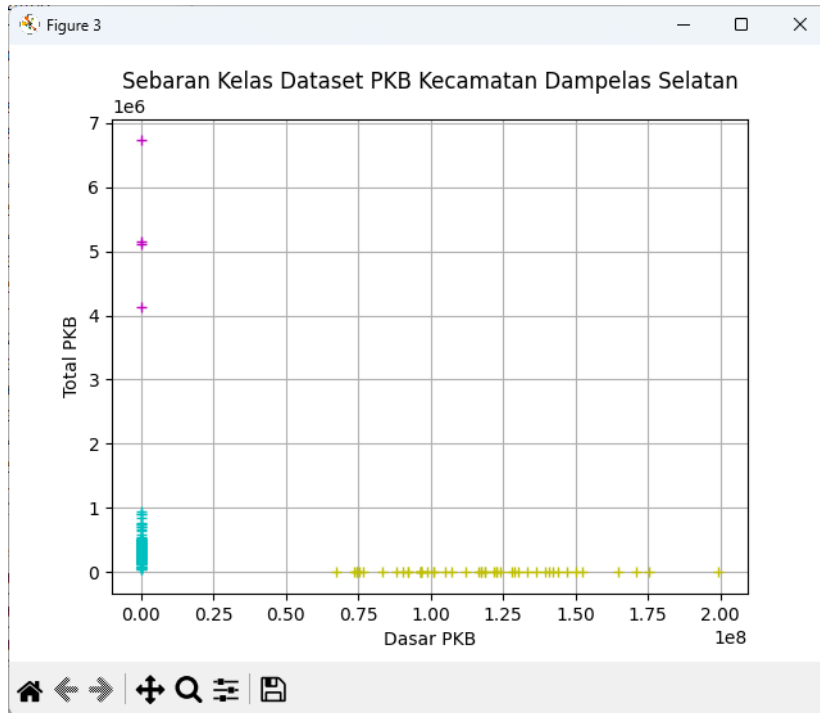
Hasil evaluasi dari ketiga eksperimen menggunakan kedua metode yang telah disebutkan sebelumnya dapat dilihat pada tabel berikut.

Eksperimen kedua dilakukan dengan menentukan kelas sebanyak 3 kelas. Eksperimen ini menghasilkan nilai silhouette index sebesar 0,8849 dan davies bouldin index sebesar 0,3838.

K-Means Cluster				
Hasil Cluster				
NOPOL	Kelas	BOBOT	NILAI JUAL	
DN1201DG	2	1.05	10700000	
DN1704DE	2	1.05	11300000	
DN2083DO	0	1.0	10600000	
DN5111DP	0	1.0	18700000	
DN4978DT	0	1.0	14600000	
DN3915D	0	1.0	11200000	
DN1854DG	2	1.05	14000000	
DN1406DG	2	1.05	13700000	
DN1579DY	2	1.05	11300000	
DN1428DE	2	1.05	8600000	
DN1336DG	2	1.05	7300000	
DN1564DG	2	1.05	11100000	
DN1114D	2	1.05	16300000	
DN8216DD	2	1.085	8500000	
DN8330DD	2	1.085	8900000	
DN3618DW	0	1.0	11800000	
DN3318DM	0	1.0	13100000	
DN6481DQ	0	1.0	16700000	
DN2596DP	0	1.0	11000000	
DN1540DO	0	1.0	13200000	

Dataset	
Cluster K-2	silhouette score: 0.8849157294170119
Cluster K-3	Davies Bouldin score: 0.38377907155337904
Cluster K-4	

Gambar sebelumnya menunjukkan hasil pembagian kelas atau cluster dengan jumlah kelas ditentukan sebanyak 3 kelas menggunakan program yang dikembangkan. Grafik sebaran kelas yang dihasilkan menggunakan program yang dikembangkan dapat dilihat pada gambar berikut.



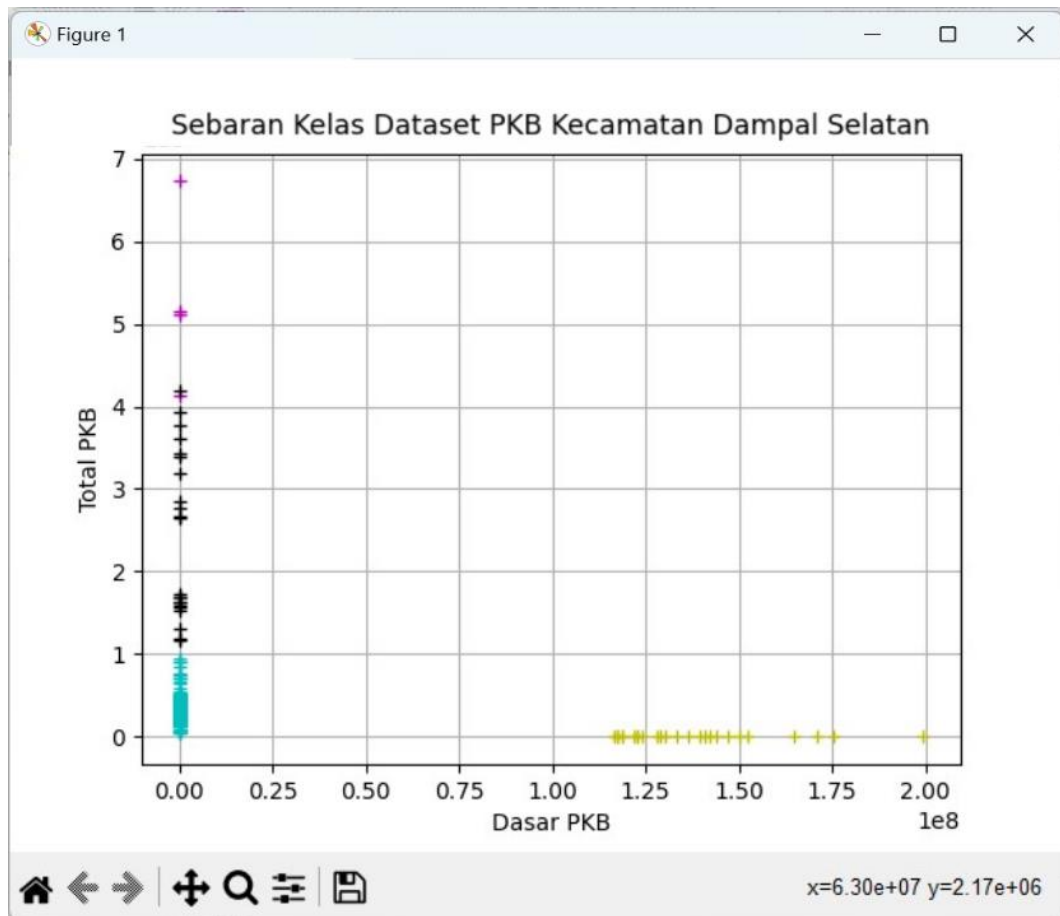
Ekspirimen kedua dilakukan dengan menentukan kelas sebanyak 3 kelas. Ekspirimen ini menghasilkan nilai silhouette index sebesar 0,8631 dan davies bouldin index sebesar 0,4383.

K-Means Cluster

NOPOL	Kelas	BOBOT	NILAI JUAL
DN1201DG	3	1.05	107000000
DN1704DE	2	1.05	113000000
DN2083DO	0	1.0	106000000
DN5111DP	0	1.0	187000000
DN4978DT	0	1.0	146000000
DN3915D	0	1.0	112000000
DN1854DG	2	1.05	140000000
DN1406DG	2	1.05	137000000
DN1579DY	2	1.05	113000000
DN1428DE	3	1.05	860000000
DN1336DG	3	1.05	730000000
DN1564DG	2	1.05	111000000
DN1114D	2	1.05	163000000
DN8216DD	3	1.085	850000000
DN8330DD	3	1.085	890000000
DN3618DW	0	1.0	118000000
DN3318DM	0	1.0	131000000
DN6481DQ	0	1.0	167000000
DN2596DP	0	1.0	110000000
DN1540DD	0	1.0	112000000

Dataset	silhouette score:	Davies Bouldin score:
Cluster K-2	0.8631348748926856	
Cluster K-3		0.43839974899049544
Cluster K-4		

Gambar sebelumnya menunjukkan hasil pembagian kelas atau cluster dengan jumlah kelas ditentukan sebanyak 3 kelas menggunakan program yang dikembangkan. Grafik sebaran kelas yang dihasilkan menggunakan program yang dikembangkan dapat dilihat pada gambar berikut.



Hasil evaluasi dari ketiga eksperimen di atas menggunakan kedua metode yang telah disebutkan sebelumnya dapat dilihat pada tabel berikut.

Jumlah kelas	Silhouette Score	Davies Bouldin score
K = 2	0.86061	0,35156
K = 3	0.88491	0,3838
K = 4	0.86313	0,4383
Rata-rata	0,86955	0,39122

4. Kesimpulan

Penilaian performa dilakukan untuk mengetahui seberapa dekat suatu data dengan data yang lain dalam kelas yang sama dan seberapa jauh satu data dengan data yang lain pada kelas yang berbeda. Eksperimen yang dilakukan pada bab sebelumnya menghasilkan dua jenis score yaitu silhouette index dan davies bouldin index. Kedua metode penilaian ini umum digunakan untuk menilai performa k-means clustering. Silhouette Index memiliki rentang nilai dari -1 hingga 1. Semakin mendekati nilai 1 maka semakin baik performa cluster. Pada eksperimen yang telah dilakukan rata-rata silhouette index sebesar 0,86955 dengan nilai silhouette index terbesar dengan nilai 0,88891 dengan K=3.

Daftar Pustaka

- [1] Auliasari.Karina & Mariza Kertaningtyas,2019.Penerapan Algoritma K-Means untuk Segmentasi Konsumen Menggunakan R.Jurnal Teknologi & Manajemen Informatika.Vol.5,No.
- [2] Dudih Gustian, Kudin Rustaman, Aditya NurfitriYunita, Gusti Nurani Neng, Resti Noviansyah, Anggy Pradiftha Junfithrana (2019), Pemantauan Pola Pengunjung Perpustakaan dengan Metode FP-Growth Guna Meningkatkan Fungsi Perusahaan, Jurnal Informatics For Educators And Professionals.

-
- [3] Erene Gernaria Sihombing, 2017, Klasifikasi Data Mining Pada Rumah Tangga Menurut Provinsi dan Status Kepemilikan Rumah Kontrak/Sewa Menggunakan K-MEANS Clustering Method. STMIK Nusa Mandiri, Vol. 2, No. 2 Tahun 2017, p-ISSN :2502-7131 e-ISSN :2502-714x.
- [4] Normah,Siti Nurajizah, Arinda Salbinda,2021.Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko fashion Hijab Banten.Jurnal Teknik Komputer AMIK BSI.Vol.7,No.2.E-ISSN:2550-0120
- [5] Mohamad Fauzy Dkk (2016), Penerapan Metode Association Rule Menggunakan Algoritma Apriori Pada Simulasi Prediksi Hujan Wilayah Kota Bandung. Jurnal Ilmiah Teknologi Informasi Terapan. [4] Prasetyo, E., “Data Mining: Konsep dan Aplikasi Menggunakan MATLAB”, Yogyakarta: Andi,2012.
- [6] Wijayanto,Sena & M Yoka Fathoni,2021.Pengelompokkan Produktivitas Tanaman Padi di Jawa Tengah Menggunakan Metode Clustering K-Means.Jurnal JUPITER.Vol.13,No.2,Hal 212-219
- [7] Dayat Suryana dalam Controls Visual Basic Jilid 1.2019, pengertian dataset adalah representasi data yang disimpan di memori dalam kondisi tidak terhubung (disconnected).