

Analisis Pengelompokan Bulan Berdasarkan Jenis Surat Tahun 2023 Menggunakan Metode K-Means di UPTD/SAMSAT Kabupaten Bungo

Nurmarita^{1*}, Gusmi Kholijah², Wardi Syafmen³

^{1,2,3} Universitas Jambi, Jambi, Indonesia

ABSTRAK

Pengelolaan data menjadi kunci efisiensi pelayanan publik di era digital, termasuk di UPTD/SAMSAT Kabupaten Bungo yang menangani pajak kendaraan serta pengolahan surat. Penelitian ini bertujuan mengidentifikasi pola aktivitas surat sepanjang tahun 2023 dengan mengelompokkan bulan berdasarkan karakteristik jenis surat menggunakan metode K-Means Clustering, salah satu metode non-hierarkis yang efektif untuk segmentasi data. Empat variabel dianalisis: Surat Masuk, Surat Keluar, Surat Perintah Tugas (SPT), dan Surat Perintah Perjalanan Dinas (SPPD). Jumlah kluster optimal ditentukan dengan metode Elbow, menunjukkan tiga kluster optimal ($k=3$). Hasilnya, kluster 1 (September, November, Desember) menunjukkan aktivitas surat tertinggi; kluster 2 (Januari, Februari, April, Mei, Juni, Agustus, Oktober) memiliki aktivitas normal; dan kluster 3 (Maret, Juli) menonjol dengan jumlah surat masuk yang sangat tinggi. Nilai BCSS sebesar 64,6% mengindikasikan pemisahan kluster yang cukup baik. Temuan ini memberikan wawasan penting bagi instansi dalam merancang strategi pengelolaan surat berbasis waktu, seperti perencanaan beban kerja dan penguatan sistem arsip digital, sehingga pelayanan administrasi dapat meningkat secara terstruktur dan berbasis data.

ABSTRACT

Data management is a key factor in enhancing the efficiency of public services in the digital era, as exemplified by the UPTD/SAMSAT Office of Bungo Regency, which handles vehicle taxes and letter processing. This study aims to identify patterns in letter activity throughout 2023 by clustering months based on letter characteristics using the K-Means Clustering method, a non-hierarchical approach effective for data segmentation. Four variables were analyzed: Incoming Letters, Outgoing Letters, Assignment Letters (SPT), and Official Travel Letters (SPPD). The optimal number of clusters was determined using the Elbow method, which indicated three optimal clusters ($k=3$). Results showed that Cluster 1 (September, November, December) had the highest letter activity; Cluster 2 (January, February, April, May, June, August, October) exhibited normal activity; and Cluster 3 (March, July) was distinguished by a very high number of incoming letters. A BCSS value of 64.6% indicates a fairly good cluster separation. These findings provide important insights for agencies in designing time-based letter management strategies—such as workload planning and digital archive system enhancement—enabling more structured and data-driven administrative service improvements.

Kata Kunci: data management, K-Means clustering, public service, letter activity, digital archive

Email: * nmriita@gmail.com

DOI: <http://dx.doi.org/10.30829/jistech.v10i1.23801>

Diterima 19 Februari 2025; Direvisi 12 Juni 2025; Disetujui 25 Juni 2025



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Pendahuluan

UPTD/SAMSAT Kabupaten Bungo merupakan salah satu Unit Pelaksana Teknis (UPT) yang bertanggung jawab atas pengelolaan dan pelayanan administrasi kendaraan bermotor. Instansi ini didirikan sebagai bagian dari upaya pemerintah untuk meningkatkan pelayanan publik, terutama yang berkaitan dengan pengelolaan perpajakan dan surat-surat kendaraan. SAMSAT (Sistem Administrasi Satu Atap) Kabupaten Bungo adalah unit pelaksana teknis di bawah naungan Dinas Pendapatan Daerah (DISPENDA) Provinsi Jambi. Adapun melalui SAMSAT ini, sehingga dapat menyelenggarakan pelayanan dan pengkajian bahan petunjuk teknis di bidang pajak dan retribusi daerah (Sutanto, 2020)[1].

Seiring dengan kemajuan teknologi dan meningkatnya jumlah data administratif, diperlukan pendekatan berbasis data (data-driven) untuk memahami pola-pola distribusi surat yang terjadi dalam satu tahun. Salah satu metode analitik yang relevan untuk tujuan ini adalah *clustering*, yaitu teknik pengelompokan data berdasarkan kemiripan karakteristik. Metode K-Means merupakan salah satu teknik *clustering* non-hierarki yang banyak digunakan karena sederhana, fleksibel, dan efisien untuk dataset berskala menengah hingga besar (Hair et al., 2018). Dalam konteks administrasi publik, metode ini telah berhasil diterapkan untuk mengelompokkan data surat pada beberapa instansi. Hartono (2021) menunjukkan bahwa K-Means efektif dalam memetakan periode aktivitas surat masuk di Dinas Kominfo berdasarkan intensitas bulanan. Pengelompokan ini bertujuan untuk mengidentifikasi pola dalam data surat, mengevaluasi bulan dengan volume surat tertinggi dan terendah, serta memahami perbedaan distribusi jenis surat di setiap bulan [2].

Meski demikian, penerapan metode ini masih jarang ditemukan di lingkungan UPTD/SAMSAT. Oleh karena itu, penelitian ini bertujuan untuk mengelompokkan bulan dalam tahun 2023 berdasarkan empat jenis surat menggunakan metode K-Means *clustering*. Hasilnya diharapkan dapat memberikan dasar analitis untuk peningkatan efisiensi sistem administrasi surat di UPTD/SAMSAT Kabupaten Bungo.

Metodologi Penelitian

Jenis penelitian yang digunakan pada penelitian ini yaitu dengan menggunakan data kuantitatif. Data kuantitatif merupakan data yang berbentuk angka yang dapat diukur secara matematis. Data yang digunakan dalam pengelompokan bulan berdasarkan jenis data surat tahun 2023 di UPTD/SAMSAT Kabupaten Bungo dengan menggunakan metode K-Means *clustering* yaitu terdiri dari 4 jenis data surat. Data tersebut adalah data primer yang diperoleh langsung melalui observasi dan pencatatan jumlah surat yang dikelola oleh kantor SAMSAT dalam periode tertentu. Teknik pengumpulan data dalam penelitian ini dilakukan secara observasi langsung dan pencatatan manual terhadap aktivitas surat menyurat di UPTD/SAMSAT Kabupaten Bungo selama periode Januari hingga Desember tahun 2023 dengan jumlah data sebanyak 12 bulan. Pada laporan penelitian ini, peneliti ingin mengelompokkan bulan berdasarkan data jenis surat di UPTD/SAMSAT Kabupaten Bungo pada tahun 2023. Terdapat 4 jenis surat yang terdiri dari surat masuk, surat keluar, Surat Perintah Tugas (SPT), dan Surat Perintah Perjalanan Dinas (SPPD).

Tabel 1. Variabel Penelitian

Notasi Variabel	Variabel	Definisi Variabel
X_1	Surat Masuk	Jumlah surat yang diterima oleh instansi dalam periode waktu tertentu
X_2	Surat Keluar	Surat resmi yang dikirimkan oleh instansi kepada pihak eksternal
X_3	Surat Perintah Tugas (SPT)	Surat perintah dari pimpinan kepada pegawai untuk melaksanakan tugas tertentu
X_4	Surat Perintah Perjalanan Dinas (SPPD)	Surat resmi untuk penugasan dinas ke lokasi tertentu atas nama instansi

Analisis *cluster* adalah salah satu analisis multivariat yang bertujuan untuk mengelompokkan objek atau subjek berdasarkan kesamaan karakteristik yang dimiliki. Analisis *cluster* memiliki tingkat homogenitas yang tinggi di dalam satu kelompok (*Within Cluster*) dan tingkat heterogenitas yang tinggi antar kelompok (*Between Cluster*). Jika pengelompokan berhasil, individu dalam satu kelompok akan membentuk gerombolan saat dipetakan sementara objek dari kelompok lain akan terpisah dengan jelas. Analisis ini mengasumsikan bahwa variabel-variabel bersifat bebas (tidak berkorelasi) dan minimal menggunakan skala interval [3].

Algoritma K-Means adalah salah satu metode *clustering* non hierarki yang digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik. Metode ini bersifat fleksibel dan dapat diterapkan pada dataset dengan ukuran besar. Metode ini bekerja dengan cara mengelompokkan data ke dalam beberapa *cluster* berdasarkan jarak antar objek. Semakin dekat jarak antara dua objek, maka semakin besar kemungkinan berada dalam satu *cluster*. Data dengan karakteristik akan ditempatkan dalam kelompok lain. Hal ini memastikan setiap kelompok yang terbentuk memiliki tingkat perbedaan yang minim atau sedikit. Algoritma K-Means merupakan metode *clustering* berdasarkan pada jarak yang memisahkan data ke dalam sejumlah *cluster* dan hanya terkait pada atribut numerik [4].

Secara umum, algoritma analisis *cluster* K-Means dimulai dengan langkah berikut [5]:

- Menentukan nilai k sebagai jumlah *cluster* yang akan dibentuk.
Penentuan jumlah *cluster* k biasanya didasarkan pada berbagai pertimbangan, termasuk aspek teoritis dan konseptual yang dapat memberikan panduan dalam menentukan jumlah *cluster* yang optimal.
- Menentukan k titik pusat *cluster* (*centroid*) awal secara acak.
Pemilihan *centroid* awal dilakukan secara acak dari objek-objek yang ada, sebanyak k *cluster*. Selanjutnya, untuk menghitung *centroid cluster* ke- i digunakan rumus sebagai berikut:
- Menentukan jarak antara setiap objek dengan *centroid* masing-masing *cluster*
Jarak antara objek dan *centroid* dihitung dengan menggunakan rumus jarak *Euclidean*
- Alokasikan setiap objek ke *centroid* yang paling dekat
- Lakukan iterasi, lalu tentukan posisi *centroid* yang baru dengan menggunakan rumus berdasarkan persamaan (1)
- Ulangi langkah ke 3 dan 4 hingga tidak ada lagi objek yang berpindah *cluster*.
Pengolahan data dan analisis *clustering* dilakukan menggunakan perangkat lunak RStudio, *software* ini dipilih karena mendukung analisis multivariat dan visualisasi hasil *clustering* secara efisien. Penentuan jumlah *cluster* menggunakan metode *Elbow*, dengan cara memplot nilai *Within-Cluster Sum of Squares* (WSS) terhadap berbagai nilai k . Titik siku yang terlihat pada $k = 3$ menunjukkan bahwa tiga *cluster* sudah optimal karena setelah titik tersebut penurunan WSS melandai. Selain didukung secara visual, pemilihan ini juga relevan secara konseptual karena mencerminkan pola umum: bulan dengan aktivitas tinggi, aktivitas normal, dan lonjakan spesifik (seperti surat masuk tinggi).

Hasil dan Pembahasan

Pada penelitian ini, akan dilakukan pengelompokan sekumpulan objek berdasarkan variabel yang dianalisis yaitu pengelompokan bulan berdasarkan data jenis surat tahun 2023.

a. Standarisasi data

Standarisasi data diperlukan jika data dalam suatu penelitian memiliki satuan yang bervariasi atau berbeda. Namun, jika semua variabel dalam data memiliki satuan yang sama maka standarisasi tidak diperlukan. Adapun dalam penelitian ini, karena semua variabel memiliki satuan yang sama maka pengujian untuk standarisasi data tidak perlu dilakukan [6].

b. Asumsi sampel mewakili populasi

Asumsi sampel mewakili populasi dapat ditentukan melalui uji *Kaiser Mayer Olkin* (KMO). Uji KMO dilakukan untuk mengetahui apakah faktor-faktor dalam penelitian valid atau tidak [7].

Secara umum statistik uji KMO yaitu sebagai berikut:

$$r_{x_i, x_j} = \frac{\sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x}_i) (x_j - \bar{x}_j)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2} \sqrt{\sum_{j=1}^n (x_j - \bar{x}_j)^2}}$$

Dimana:

r_{x_i, x_j} : Koefisien korelasi antara variabel i dan j

x_i, x_j : Nilai objek pada variabel ke- i dan ke- j

\bar{x}_i, \bar{x}_j : Nilai rata-rata variabel i dan j

Berikut akan ditampilkan hasil perhitungan uji korelasi dan kecukupan sampel menggunakan *software* RStudio yang disajikan pada gambar di bawah ini:

```
> bart_spher(faktor)

Bartlett's Test of Sphericity

Call: bart_spher(x = faktor)

      x2 = 7.357
      df = 6
p-value = .28906

> KMOS(faktor)

Kaiser-Meyer-Olkin Statistics

Call: KMOS(x = faktor)

Measures of Sampling Adequacy (MSA):
      x1      x2      x3      x4
0.5226035 0.6347531 0.7251013 0.5658144

KMO-Criterion: 0.6206507
```

Gambar 1. KMO and Bartlett's Test

Berdasarkan gambar 1 terlihat bahwa taraf signifikan sebesar 0,620. Nilai tersebut <0,5 yang artinya H_0 ditolak sehingga sampel yang digunakan layak untuk dianalisis secara lanjut.

c. Asumsi Non Multikolinearitas

Uji asumsi non multikolinearitas digunakan untuk mengidentifikasi hubungan antar variabel yang digunakan. Pengujian ini dapat dilakukan dengan melihat nilai *Variance Inflation Factor* (VIF) [8]. Berikut merupakan hasil perhitungan uji multikolinearitas yang dicari menggunakan *software* RStudio:

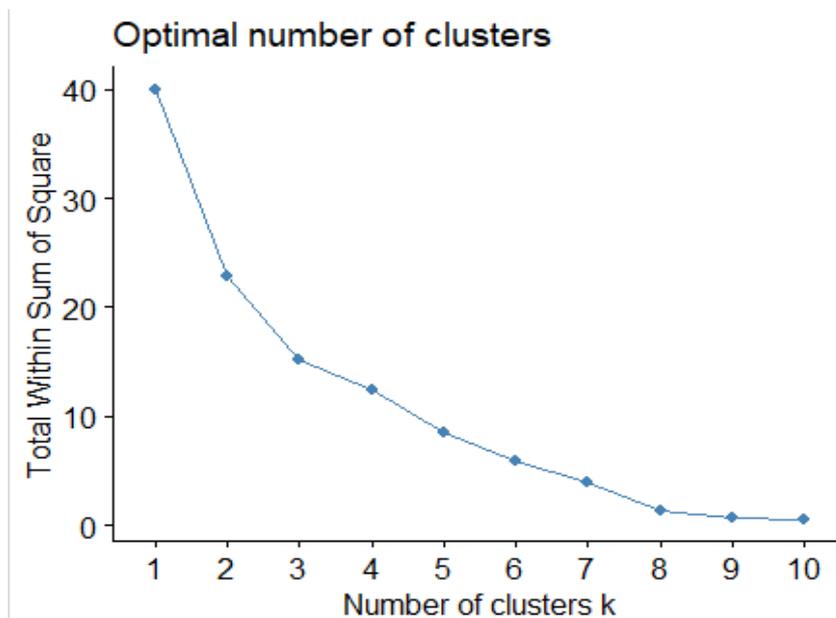
```
> #Uji Multikolinearitas
> library(car)
> vif(rit)
      x1      x2      x3      x4
1.262840 1.698616 1.438505 1.498048
```

Gambar 2. Nilai VIF setiap variabel

Berdasarkan gambar 2, nilai VIF untuk setiap variabel <10 sehingga dapat disimpulkan bahwa tidak terdapat multikolinearitas pada semua variabel. Hal ini menunjukkan bahwa asumsi dalam analisis *cluster* K-Means telah terpenuhi.

d. Analisis *cluster* K-Means

Pada tahap awal yang dilakukan yaitu menentukan jumlah *cluster* yang optimal dan *centroid* awal. Untuk analisis *centroid* awal yang akan digunakan dalam perhitungan dengan metode K-Means di tentukan jumlah *cluster* optimal yaitu $K = 3$ yang dapat dilihat dari grafik metode *Elbow* berikut ini:



Gambar 3. Grafik Metode *Elbow*.

Berdasarkan grafik metode *Elbow* yang diperoleh dari *output* RStudio, menunjukkan hubungan antara jumlah *cluster* k dan nilai *Within Sum of Square* (TWSS) untuk setiap nilai K . Dalam konteks untuk analisis *cluster*, jumlah *cluster* optimal ditentukan ketika terjadi penurunan yang signifikan pada nilai eigen, diikuti oleh penurunan yang lebih landai atau stabil. Hal ini menandakan titik "siku" (*Elbow*) di mana nilai eigen tidak lagi turun secara signifikan [9]. Pada grafik di atas titik "*Elbow*" tampak di sekitar komponen kedua.

Berikut adalah hasil K-Means *clustering* yang diperoleh menggunakan bantuan *software* RStudio:

```

> #Mencari hasil K-Means Clustering
> Clustering=kmeans(datanumerik,centers=3,nstart=25)
> Clustering
k-means clustering with 3 clusters of sizes 3, 7, 2

Cluster means:
  Surat Masuk Surat Keluar      SPT      SPPD
1  25.33333    36.00000 25.00000 23.33333
2  22.28571    20.42857 16.57143 13.00000
3  46.50000    15.00000 10.00000 13.50000

Clustering vector:
[1] 2 2 3 2 2 3 2 1 2 1 1

within cluster sum of squares by cluster:
[1] 489.3333 534.8571 143.0000
 (between_ss / total_ss = 64.6 %)

Available components:
[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"
[7] "size"         "iter"         "ifault"

```

Gambar 4. Hasil Analisis K-Means Clustering

Pada proses hasil K-Means clustering di atas dapat diinterpretasikan yaitu sebagai berikut:

1. Jumlah *cluster* yang optimal menghasilkan 3 *cluster*. Adapun untuk algoritma K-Means didefinisikan dengan $nstart=25$ yang artinya algoritma dijalankan sebanyak 25 kali dengan inialisasi acak untuk memastikan hasil yang terbaik.
 - a. *Cluster 1*(September, November, Desember), memiliki rata-rata tertinggi untuk hampir semua jenis surat, menunjukkan bulan-bulan ini merupakan periode aktivitas surat menyurat yang paling padat.
 - b. *Cluster 2*(Januari, Februari, April, Mei, Juni, Agustus, Oktober), menunjukkan aktivitas surat menyurat yang cenderung stabil dan tidak ekstrem, baik tinggi maupun rendah.
 - c. *Cluster 3*(Maret dan Juli), menonjol karena memiliki jumlah surat masuk yang jauh lebih tinggi dibanding bulan lainnya, sementara variabel surat lainnya cenderung lebih rendah. Perhitungan untuk menentukan nilai rata-rata dari 4 variabel yang terdapat dalam 3 *cluster* tersebut bisa digunakan rumus umum sebagai berikut:

$$X = \mu + z \cdot \alpha$$

2. Nilai *Within Cluster Sum of Squares*(WCSS) digunakan untuk mengukur seberapa dekat data dalam *cluster* terhadap pusatnya(*centroid*). WCSS yang kecil menunjukkan bahwa data dalam *cluster* tersebut homogen dan saling berdekatan. Dalam hasil ini:
 - a. *Cluster 3* memiliki nilai WCSS terkecil (143,000), menunjukkan penyebaran data dalam *cluster* ini sangat rapat, atau hampir seragam.
 - b. *Cluster 2* memiliki WCSS terbesar (534,8571), yang mengindikasikan variasi data cukup besar dalam *cluster* tersebut.
 - c. *Cluster 1* berada ditengah (489,3333), menunjukkan tingkat variasi sedang. Semakin kecil nilai WCSS, semakin baik kualitas *cluster* karena data dalam satu kelompok menunjukkan kemiripan yang tinggi. Hal ini memperkuat bahwa *cluster 3*, meskipun anggotanya hanya dua bulan, merupakan kelompok dengan karakteristik paling konsisten.
3. Persentase *Between-Cluster Sum of Squares* (BCSS) terhadap *Total Sum of Squares* (TSS) sebesar 64.6%, menunjukkan seberapa baik *cluster* terpisah satu sama lain. Semakin tinggi nilainya, semakin baik pemisahan antar *cluster*. Nilai 64.6% menunjukkan pemisahan yang cukup baik, meskipun masih bisa ditingkatkan untuk mencapai *cluster* yang lebih terdefinisi.

Hasil K-Means clustering menghasilkan tiga *cluster* dengan karakteristik dan ukuran berbeda. *Cluster 3* paling rapat (WCSS terkecil), sedangkan *cluster 2* paling tersebar (WCSS terbesar). Pemisahan antar *cluster* cukup baik (BCSS 64,6%) dan pola masing-masing *cluster* dapat dimanfaatkan untuk analisis atau pengambilan keputusan lebih lanjut.

Selanjutnya untuk mengetahui bulan mana saja di tahun 2023 berdasarkan jumlah surat masuk, surat keluar SPT dan SPPD yang berada pada *cluster 1*, *cluster 2* dan *cluster 3*, disajikan pada tabel berikut:

Tabel 2. Keanggotaan Cluster

No	Bulan	Surat Masuk	Surat Keluar	SPT	SPPD	Cluster
1	Januari	19	19	22	15	2
2	Februari	20	21	15	12	2
3	Maret	51	15	7	20	3
4	April	19	17	12	15	2
5	Mei	20	19	9	9	2
6	Juni	33	20	16	12	2
7	Juli	42	15	13	7	3
8	Agustus	6	28	24	10	2
9	September	25	47	17	21	1
10	Oktober	29	19	18	18	2
11	November	33	28	32	19	1
12	Desember	18	33	26	30	1

Berdasarkan Tabel 2, pembagian bulan ke dalam tiga cluster menunjukkan pola yang jelas. Cluster 1 terdiri dari September, November, dan Desember, yang memiliki aktivitas surat menyurat tertinggi di hampir semua jenis surat. Cluster 2, yang berisi tujuh bulan Januari, Februari, April, Mei, Juni, Agustus, dan Oktober, menunjukkan volume surat yang stabil atau sedang. Cluster 3, yaitu Maret dan Juli, menonjol karena jumlah surat masuk yang sangat tinggi dibandingkan bulan lain, terutama Maret yang mencapai 51 surat masuk.

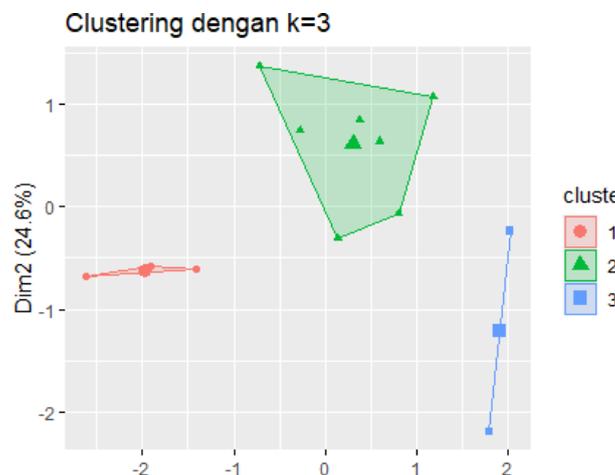
e. Tabel ringkasan statistik hasil cluster

Selanjutnya, untuk bisa dipahami lebih lanjut karakteristik masing-masing cluster, berikut ditampilkan ringkasan statistik yang mencakup rata-rata jumlah surat per jenis, jumlah anggota, serta variabel dominan dalam tiap cluster.

Tabel 3. Ringkasan Statistik Tiap Cluster Berdasarkan Jenis Surat

Cluster	Jumlah Bulan	Variabel Dominan	Rata-rata Surat Masuk	Rata-rata Surat Keluar	Rata-rata SPT	Rata-rata SPPD
1	3(Sep, Nov, Des)	Aktivitas Surat Tinggi	25,3	36,0	25,0	23,3
2	7(Jan, Feb, Apr, Mei, Jun, Agt, Okt)	Aktivitas Normal	21,0	20,4	16,6	13,0
3	2(Mar, Jul)	Surat Masuk Sangat Tinggi	46,5	15,0	10,0	13,5

Adapun untuk visualisasi pada cluster yang telah terbentuk dapat dilihat pada plot cluster berikut ini:



Gambar 5. Plot Cluster

Berdasarkan gambar plot cluster diatas, dapat diinterpretasikan sebagai berikut:

1. Pada *cluster* 1 yang berwarna merah lingkaran memiliki titik-titik yang berdekatan. menunjukkan bahwa bulan-bulan dalam *cluster* ini memiliki kesamaan yang kuat.
2. Pada *cluster* 2 yang berwarna hijau segitiga terlihat bahwa titik tersebut berada di bagian tengah dan variasi titik dalam *cluster* ini relatif lebih besar dibandingkan *cluster* lain, yang dapat menunjukkan keragaman karakteristik dalam *cluster*.
3. Pada *cluster* 3 yang berwarna biru persegi yang terletak terpisah di sisi kanan bawah, menunjukkan bahwa bulan Maret dan Juli memiliki pola yang sangat berbeda dan unik dibandingkan *cluster* lainnya.

Dapat disimpulkan bahwa *cluster* 1 memiliki kesamaan yang kuat, sedangkan pada *cluster* 2 menunjukkan keragaman yang lebih besar dibandingkan *cluster* lainnya. Kemudian pada *cluster* 3 sangat terisolasi dan berbeda secara signifikan dari dua *cluster* lainnya, hal ini menunjukkan kedua bulan tersebut memiliki sifat yang unik. Hasil ini sejalan dengan penelitian oleh (Hartono, 2021)[2] di Dinas Kominfo yang juga menggunakan K-Means untuk pengelompokan surat berdasarkan bulan. Hartono melaporkan 3 *cluster* dengan karakteristik serupa: bulan sibuk di akhir tahun, bulan normal tersebar di tengah tahun, dan bulan ekstrem yang menonjol di satu jenis surat. Namun, pada penelitian ini, *cluster* 3 lebih menonjol dalam surat masuk, sedangkan dalam studi Hartono, variasi ekstrem terjadi pada surat keluar. Selain itu, penelitian oleh (Aprilia et al., 2022)[12] di BPPRD Sumatera Utara menemukan bahwa bulan Desember cenderung masuk *cluster* dengan volume surat tinggi, yang konsisten dengan temuan dalam penelitian ini. Hal ini menunjukkan adanya tren musiman administratif, khususnya di kuartal akhir tahun. Oleh karena itu, hasil penelitian ini tidak hanya konsisten secara internal tetapi juga memperkuat temuan dari beberapa studi sejenis mengenai pola administrasi musiman di lembaga pemerintahan.

Kesimpulan

Penelitian ini mengelompokkan bulan dalam tahun 2023 menjadi tiga *cluster* menggunakan metode K-Means. Hasilnya, *cluster* 1 (aktivitas surat tinggi: September, November, Desember), *cluster* 2 (aktivitas normal: Januari, Februari, April, Mei, Juni, Agustus, Oktober), dan *cluster* 3 (surat masuk sangat tinggi: Maret, Juli). Pemisahan *cluster* cukup baik dengan BCSS sebesar 64,6%, dan *cluster* 3 menunjukkan sebaran data paling rapat (WCSS terkecil). Hasil penelitian ini dapat mendukung efisiensi pengelolaan surat melalui penerapan sistem digital, otomatisasi arsip, dan analisis *e* untuk pengelompokan yang lebih terstruktur. Sebagai saran, hasil pengelompokan ini dapat dimanfaatkan untuk peningkatan sistem administrasi berbasis data, misalnya dengan menerapkan otomatisasi arsip dan pemrosesan surat pada bulan-bulan dengan beban tinggi serta perencanaan sumber daya yang lebih efisien berdasarkan pola musiman yang teridentifikasi dari data.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dan kontribusi dalam penyelesaian penelitian ini. Berkat bantuan, bimbingan dan dukungan, baik secara langsung maupun tidak langsung, sehingga penelitian ini dapat diselesaikan dengan baik. Dukungan, kerja sama, dan kontribusi yang diberikan sangat berarti bagi keberhasilan penelitian ini, dan penulis sangat menghargai setiap bentuk perhatian yang telah diberikan.

Daftar Pustaka

- [1] B. Sutanto, "Manajemen Pelayanan Publik dan Administrasi Pemerintahan," Yogyakarta: Andi Publisher, 2020.
- [2] R. Hartono, "Implementasi Algoritma K-Means dalam Pengelompokan Data Surat Masuk di Dinas Komunikasi dan Informatika," *Sist. Inf.*, vol. 17, no. 3, hal. 89–97, 2021.
- [3] Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate Data Analysis* (8th ed.). Cengage Learning.
- [4] D. Putra, *Clustering Data dengan Algoritma K-Means*. Bandung: Informatika, 2023.
- [5] E. Rahmawati, *Pemanfaatan Algoritma K-Means dalam Pengelompokan Data*. Surabaya: Techno Data Press, 2024.
- [6] R. Fadilah, *Standarisasi Data dalam Analisis Statistik dan Machine Learning*. Jakarta: Data Science Press, 2024.
- [7] R. Hidayat, "Analisis Faktor dengan Uji KMO dan Bartlett dalam Penelitian Statistik," *Stat. Terap.*, 2020.
- [8] R. Susanto, *Analisis Statistik dengan Uji Multikolinearitas*. Jakarta: Statistika Press, 2024.
- [9] P. J. Kaufman, L., & Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: Wiley.
- [10] S. M. Bungo, "Laporan Tahunan UPTD/Samsat Muara Bungo," Jambi, 2023.
- [11] F. A., "Perancangan Sistem Informasi Surat Perintah Perjalanan Dinas Berbasis Web Pada Kantor Imigrasi Kelas II TPI Belawan," Sumatera Utara: Universitas Muhammadiyah, 2023.

- [12] S. O. D. . Aprilia, R., Afsari, K., Rahma, R., Nasution, N., & Putri, “Analisis Cluster Dengan Metode K-Means Pada Jenis Data Surat Di BPPRD Sumatera Utara Amaliah,” *Pengabd. Kpd. Masy.*, vol. 6, no. 2, hal. 369–373, 2022.
- [13] Wu, J. (2016). *Advances in K-Means Clustering: A Data Mining Thinking*. Berlin: Springer.
- [14] I. Purnama, “Akuntabilitas Penggunaan Surat Perintah Perjalanan Dinas (SPPD) Pegawai di Sekretariat DPRD Kota Makassar,” Makassar: Universitas Muhammadiyah Makassar, 2017.
- [15] Laporan Tahunan UPTD/SAMSAT Muara Bungo. (2023). Kabupaten Bungo: SAMSAT.