

ANALISIS *CLUSTER* DENGAN METODE *K-MEANS CLUSTER* PADA JENIS DATA SURAT DI BPPRD SUMATERA UTARA

Rima Aprilia^{1*}, Khaila Afsari², Radita Rahma³, Nurjannah Nasution⁴,
Sauqi Ouri Dwi Putri⁵

^{1,2,3,4,5}Prodi Matematika FST, Universitas Islam Negeri Sumatera Utara

*Korespondensi: rima_aprillia@uinsu.ac.id

Abstrak

Pengabdian masyarakat adalah penerapan bidang pendidikan dan ekonomi untuk menyelesaikan masalah di masyarakat baik warga kota atau daerah maupun masyarakat industri dan pemerintah. Pengabdian masyarakat ini dilakukan dalam rangka menghasilkan karya monumental bagi masyarakat sekitar. Metode yang digunakan untuk menganalisis kasus pengelompokan *K-Means Cluster* terhadap Data Jenis Surat Pada Setiap Bulan di Badan Pengelolaan Pajak dan retribusi Daerah Sumatera Utara (BPPRDSU). *K-Means Cluster* adalah salah satu metode dan *clustering* non - hierarki yang berusaha mengelompokkan data ke dalam suatu *cluster* sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster* yang sama. Tujuan pengelompokan adalah untuk meminimalkan *objective function* yang di set dalam proses *clustering*, yang pada dasarnya berusaha untuk meminimalkan variasi dalam satu *cluster* dan memaksimalkan variasi antar *cluster*. Pengelompokan sekumpulan objek terhadap variabel yang di analisis yaitu pengelompokan setiap bulan di dalam satu tahun terhadap jenis data surat ke dalam *cluster-cluster* kecil namun berbeda karakteristiknya dari *cluster-cluster* yang lainnya. Hasil analisis pengelompokan pada jenis data surat tahun 2021 pada BPPRD-SU untuk SPT, SPPD dan surat keluar nota dinas menunjukkan adanya perbedaan di setiap bulannya pada ketiga *cluster* yang terbentuk dengan nilai $F = 32,171$ dan $\text{sig} = 0$ serta $F = 7,365$ dan $\text{sig} = 0,13$.

Kata Kunci : Data Surat, Variabel, *K-Means Cluster*

Abstract

Community service is the application of education and economics to solve problems in society, both urban or regional residents as well as industrial communities and the government. This community service is carried out in order to produce monumental works for the surrounding community. Every Month at the North Sumatra Regional Tax and Retribution Management Agency (BPPRDSU) *K-Means Cluster* is a method and non-hierarchical clustering that attempts to group data into a cluster so that data that has the same characteristics are grouped into the same cluster. The purpose of grouping is to minimize the objective function set in the clustering process, which basically seeks to minimize variation within one cluster and maximize variation between clusters. Grouping a set of objects against the variables being analyzed is grouping every month in one year for the type of letter data into clusters. Clusters are small but have different characteristics from other clusters. The results of the grouping analysis on the type of letter data in 2021 at BPPRD SU for SPT, SPPD and outgoing official notes show that there are differences in each month on the three clusters formed with a value of $F = 32,171$ and $\text{sig} = 0$ and $F = 7,365$ and $\text{sig} = 0,13$.

Keywords: Letter Data, Variable, *K-Means Cluster*

1. PENDAHULUAN

Pengabdian masyarakat adalah penerapan bidang pendidikan dan ekonomi untuk menyelesaikan masalah di masyarakat baik warga kota atau daerah maupun masyarakat industri dan pemerintah. Pengabdian masyarakat ini dilakukan dalam rangka menghasilkan karya monumental bagi masyarakat sekitar.

Dalam pengabdian ini dilaksanakan pada Badan Pengelolaan Pajak dan Retribusi Daerah (BPPRD-SU), dengan menerapkan metode *K-Means Cluster* terhadap data jenis surat pada tahun 2021. BPPRD Sumatera Utara merupakan instansi yang menangani pengelolaan tentang perpajakan kendaraan dan retribusi daerah. Sistem pendataan surat masuk dan surat keluar pada instansi

BPPRD Sumatera Utara saat ini masih berjalan secara semi komputerisasi. Seiring dengan perkembangan teknologi dan informasi, dengan adanya sistem pendataan surat masuk dan surat keluar yang sudah terkomputerisasi akan dapat lebih memudahkan pegawai dalam melakukan kegiatan pengarsipan surat. Sistem pendataan surat masuk dan surat keluar merupakan salah satu upaya untuk membantu dan mempermudah pegawai dalam melakukan pengarsipan, sehingga dapat lebih efektif dan efisien.

K-Means Cluster adalah salah satu metode dan *clustering* non-hierarki yang berusaha mengelompokkan data ke dalam suatu cluster sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu *cluster* yang sama. Ukuran kesamaan yang digunakan adalah ukuran jarak antar objek. Kedua objek yang memiliki jarak paling dekat akan bergabung menjadi satu *cluster*. Kedekatan jarak yang dimiliki menunjukkan bahwa kedua objek tersebut memiliki tingkat kesamaan karakteristik.

2. METODE PELAKSANAAN

Pada pelaksanaan pengabdian masyarakat di BPPRD-SU merupakan penerapan bidang untuk meningkatkan sistem pendataan dalam melakukan pengarsipan dengan cara mengelompokkan data. Metode dokumentasi yang digunakan untuk mendapatkan informasi tersebut adalah melalui data yang telah di input ke dalam Ms. Excel pada komputer di sub bagian tata usaha BPPRDSU. Observasi yang dilakukan yaitu mengamati dan mengetahui secara langsung sistem yang sedang berjalan dan proses kerja serta melihat laporan dalam kantor tersebut yang digunakan saat ini. Dimana, metode penyelesaian dilakukan dengan menggunakan bantuan software SPSS. Adapun variabel yang digunakan, yaitu

X_1 (surat masuk), X_2 (surat keluar), X_3 (surat keluar nota dinas), X_4 (undangan), dan X_5 (surat perintah tugas (SPT) dan surat permohonan perjalanan dinas (SPPD)).

3. HASIL DAN PEMBAHASAN

Dalam kegiatan pengabdian ini dilakukan pengelompokkan sekumpulan objek terhadap variabel yang di analisis yaitu pengelompokkan setiap bulan di tahun 2021 terhadap jenis data surat ke dalam *cluster-cluster* kecil.

➤ Standardisasi / Transformasi

Dari data tersebut dapat kita ketahui bahwa data yang terkumpul mempunyai variabilitas satuan, dikarenakan variabel-variabel tersebut memiliki satuan yang serupa, maka tidak perlu dilakukan langkah standardisasi atau transformasi terhadap variabel-variabel yang relevan ke bentuk *Z-score*.

➤ Sampel Mewakili Populasi

Tabel 1. Uji KMO dan Bartlett's Test

<i>KMO and Bartlett's Test</i>		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,517
Bartlett's Test of Sphericity	Approx. Chi-Square	19,612
	df	10
	Sig.	,033

Dapat dilihat pada Tabel 1 diketahui bahwa taraf signifikansi sebesar $0,033 < \alpha = 0,05$, H_0 ditolak sehingga sampel yang digunakan sudah memadai atau layak untuk dianalisis.

➤ Non Multikolinearitas

Berdasarkan Tabel 2 semua nilai VIF setiap variabel kurang dari 10, sehingga dapat dikatakan bahwa tidak terjadi multikolinearitas pada semua variabel. Maka asumsi dalam *K-Means Cluster* telah terpenuhi.

Tabel 2. Nilai VIF Setiap Variabel

VARIABEL INDIKATOR	R_k^2	VIF
--------------------	---------	-----

SURAT MASUK	0,590	1,176
SURAT KELUAR	0,260	2,121
SURAT KELUAR NOTA DINAS	0,600	2,098
UNDANGAN	0,724	1,517
SPT & SPPD	0,671	1,739

➤ Analisis Cluster

Analisis cluster dengan metode non hierarki (*K-Means Cluster*) digunakan untuk pengelompokan objek dimana banyaknya *cluster* yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur penggerombolan, kemudian metode ini dapat diterapkan pada data yang lebih besar.

Tabel 3. Proses Clustering Sebelum Iterasi

	Cluster		
	1	2	3
SURAT MASUK	242	209	257
SURAT KELUAR	194	262	138
SURAT KELUAR NOTA DINAS	41	19	24
UNDANGAN	64	12	51
SPT & SPPD	361	53	231

Pada Tabel 3 menunjukkan hasil proses *clustering* sebelum dilakukan iterasi yaitu hasil proses sementara pengelompokan data yang dilakukan, yang setelah itu akan dilakukan proses selanjutnya.

Tabel 4. Hasil Proses Iterasi
Iteration History^a

Iteration	Change in Cluster Centers		
	1	2	3
1	,000	96,905	44,182
2	,000	,000	,000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 2. The minimum distance between initial centers is 143,941.

Berdasarkan Tabel 4 diatas diketahui bahwa tahapan *clustering*

cukup dilakukan dengan 2 iterasi dengan jarak minimum sebesar 143,941.

Tabel 5. Keanggotaan Cluster
Cluster Membership

Case Number	Bulan	Cluster	Distance
1	Januari	2	53,086
2	Februari	2	96,905
3	Maret	2	74,675
4	April	2	51,066
5	Mei	2	57,304
6	Juni	3	41,442
7	Juli	3	54,852
8	Agustus	3	56,884
9	September	3	25,595
10	Oktober	3	44,182
11	November	1	,000
12	Desember	3	65,616

Pada Tabel 5 dapat dilihat bahwa:

- Pada *Cluster* 1 hanya berisikan 1 bulan saja yaitu bulan November dengan jarak terhadap pusat *cluster* 1 sebesar 0.
- Pada *cluster* 2 berisikan 5 bulan yaitu bulan Januari, Februari, Maret, April, dan Mei dengan masing-masing jarak terhadap pusat *cluster* 2 sebesar 53,086; 96,905; 74,675; 51,066; dan 57,304.
- Pada *cluster* 3 berisikan 6 bulan yaitu bulan Juni, Juli Agustus, September, Oktober dan Desember dengan masing-masing jarak terhadap pusat *cluster* 3 sebesar 41,442; 54,852; 56,884; 25,595; 44,182, dan 65,616.

Pada Tabel 6 menunjukkan hasil analisis *cluster* untuk masing-masing variabel dan *cluster* yang dibentuk dengan ketentuan sebagai berikut :

- Nilai negatif (-) berarti data berada di bawah rata-rata total.

- Nilai positif (+) berarti data berada di atas rata-rata total.

Sehingga interpretasi yang didapat ialah pada setiap variabel yaitu surat masuk, surat keluar, surat keluar nota dinas, undangan, SPT & SPPD memiliki nilai di atas rata-rata total di tiap *cluster*, karena nilai yang dihasilkan di setiap *cluster* untuk masing-masing variabel bernilai positif semua.

Tabel 6. Hasil Analisis Cluster

	Cluster		
	1	2	3
SURAT MASUK	242	238	247
SURAT KELUAR	194	172	131
SURAT KELUAR NOTA DINAS	41	19	22
UNDANGAN	64	12	35
SPT & SPPD	361	53	192

Adapun untuk mengetahui nilai rata-rata dari 5 variabel yang berada pada 3 *cluster* tersebut dapat menggunakan rumus umum sebagai berikut :

$$X = \mu + z \cdot \sigma \dots\dots\dots(1)$$

Tabel 7. Anova

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
SURAT MASUK	116,442	2	1270,226	9	,092	,91
SURAT KELUAR	3224,108	2	1359,189	9	2,372	,14
SURAT KELUAR NOTA DINAS	232,442	2	31,559	9	7,365	,01
UNDANGAN	488,392	2	164,904	9	2,962	,10
SPT & SPPD	46890,458	2	1457,556	9	32,171	,00

The F tests should be used only for descriptive purposes because the clusters have been chosen to maximize the differences among cases in different clusters. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the cluster means are equal.

Pada Tabel 7 dapat dilihat bahwa terdapat perbedaan nilai F dan nilai probabilitas (*Sig*) dari masing-masing variabel pada *cluster* yang terbentuk.

$$F = \frac{MS \text{ Between}}{MS \text{ Within}} \dots\dots\dots(2)$$

Dapat diinterpretasikan bahwa untuk SPT & SPPD dan surat keluar nota dinas adalah variabel yang paling menunjukkan adanya perbedaan di setiap bulannya pada ketiga *cluster* yang terbentuk. Hal ini ditunjukkan dengan

masing-masing untuk nilai $F = 32,171$ dan $sig = 0$ serta $F = 7,365$ dan $sig = 0,13$.

Tabel 8. Jumlah Obyek Pada Tiap Cluster

<i>Number of Cases in each Cluster</i>	
Cluster 1	1,000
2	5,000
3	6,000
Valid	12,000
Missing	,000

Pada Tabel 8 memberikan gambaran jumlah objek yang masuk ke dalam masing-masing *cluster*. Sehingga dapat kita lihat bahwa pada *cluster* 1 hanya ada 1 bulan saja, pada *cluster* 2 ada 5 bulan, dan untuk *cluster* 3 ada 6 bulan.

4. KESIMPULAN

Kegiatan pengabdian masyarakat di BPPRD-SU sangat bermanfaat, dengan menerapkan metode *K-Means Clustering* untuk SPT & SPPD dan surat keluar nota dinas, masing masing surat tersebut merupakan variabel yang paling menunjukkan adanya perbedaan di setiap bulannya pada ketiga *cluster* yang terbentuk. Hal ini ditunjukkan dengan masing-masing untuk nilai $F = 32,171$ dan $sig = 0$ serta $F = 7,365$ dan $sig = 0,13$. Metode ini juga mudah diterapkan, sehingga dapat membantu pegawai BPPRD-SU dalam melakukan *cluster* pada surat-surat yang ada di BPPDR-SU.

REFERENSI

Alwi, Imran. 2018. *Analisis Cluster dengan Menggunakan SPSS*. Diunduh 16 November 2021. <https://swanstatistics.com/analisis-cluster-dengan-menggunakan-an-spss/amp/>

- Azizah, Novia Farah dan Agung Kuswantoro. 2021. *Pengelolaan Surat Masuk dan Surat Keluar*. (EEAJ) Economic Education Analysis Journal. 10 (1). Hal 65-81
- Bimo, Suseno. 2021. Cara Analisis K-Mean Cluster. Diunduh 19 November 2022. <http://www.statistikolahdata.com/2021/01/cara-analisis-k-mean-cluster.html?m=1>
- Darlianto, Andi dan Inggih Permana. 2016. *Sistem Informasi Pencatatan Surat Masuk (Studi Kasus: Kantor Camat Kampar Kiri Kabupaten Kampar Provinsi Riau*. Jurnal Rekayasa dan Manajemen Sistem Informasi. Vol.2 No.1. Hal 37-43
- Hidayat, Anwar. 2018. Analisis Cluster Non Hirarki dengan SPSS. Diunduh 19 November 2022. <https://www.statistikian.com/2014/03/analisis-cluster-non-hirarki-dengan-spss.html?amp>
- Hidayat, Anwar. 2018. Interpretasi Analisis Cluster Non Hirarki dengan SPSS. Diunduh 19 November 2022. <https://www.statistikian.com/2014/03/interpretasi-analisis-cluster-non-hirarki-dengan-spss.html?amp>
- Mulyati, Nasril Sary dan Muhammad Kurniawan. 2020. *Sistem Informasi Surat Masuk pada Pengelolaan Rantai Suplai Satuan Kerja Khusus Migrasi*. TMJ (Technomedia Journal). Vol.5 No.1. Hal 27-38
- Sari, Desy Noor Permata. 2020. *Analisis Cluster Dengan Metode K-Means Cluster pada Persebaran Kasus Covid-19 Berdasarkan Provinsi di Indonesia*. Tugas Akhir. Semarang: Universitas Negeri Semarang
- Sholihah, Siti Azizatus. 2021. *Analisis Cluster untuk Pemetaan Data Kasus Covid-19 di Indonesia dengan K-Means*. Skripsi. Surabaya: UIN Sunan Ampel
- Sirojuddin, Achmad. 2016. *Analisis Cluster Pada Kabupaten/Kota Di Provinsi Jawa Timur*. Skripsi. Malang: UIN Maulana Malik Ibrahim.