ESTIMASI: Journal of Statistics and Its Application
Vol. 6, No. 1, Januari, 2025, Hal. 61-69
DOI: 10.20956/ejsa.v6i1.27200

Pengelompokan Daerah Produksi Tanaman Biofarmaka Menurut Jenis Tanaman dengan Metode *K – Means Clustering*

Fitri Nur Kholifah[1*], Saiful Bahri[2]
[12]Prodi Matematika, Fakultas Sains dan Teknologi UIN Sunan Ampel Surabaya
*** Corresponding author, email: fitrinurkholifah27@gmail.com**

Abstract

Biopharmaceutical plants or commonly called medicinal plants are plants in which one, several or all parts of these plants contain substances or active ingredients that are useful for body health, disease healing or cosmetic ingredients. Based on data from the Central Bureau of Statistics (BPS), the production of biopharmaceutical plants varies in each region. So that in an effort to equalize the results of biopharmaceutical production in Indonesia, it is necessary to group areas that have the potential to produce biopharmaceutical plants to find out which areas produce in large or small quantities. In this research, a method is needed to facilitate the grouping of biopharmaceutical producing regions, one of which is the clustering analysis method. . One of the methods that can be used in cluster analysis is the k-means clustering algorithm. The results of this study indicate that there are 5 provinces that are included in cluster 1 (C1) with the category of low biopharmaceutical production in 2021. Meanwhile, 29 other provinces are included in the cluster. 2 (C2) with the category of high biopharmaceutical production areas in 2021.

***Keywords**: Biopharmaceuticals, BPS, Clustering, C1, C2, K-Means*

Abstrak

Tanaman biofarmaka atau biasa disebut tanaman obat adalah tumbuhan yang salah satu, beberapa atau seluruh bagian tumbuhan tersebut mengandung zat atau bahan aktif yang berguna untuk kesehatan tubuh, penyembuhan penyakit maupun bahan kosmetik. Berdasarkan data Badan Pusat Statistika (BPS) produksi tanaman biofarmaka bervariasi jumlahnya disetiap daerah. Sehingga dalam uapaya pemerataan hasil produksi tanaman biofarmaka di Indonesia diperlukan pengelompokkan daerah yang berpotensi menghasilkan tanaman biofarmaka untuk mengetahui daerah mana saja yang memproduksi dalam jumlah banyak ataupun sedikit. Pada penelitian ini dibutuhkan metode untuk memudahkan pengelompokan daerah penghasil tanaman biofarmaka salah satunya dengan metode analisis clustering. Salah satu metode yang dapat digunakan dalam analisis *cluster* yaitu algoritma *k-means clustering* Hasil dari penelitian ini menunjukkan bahwa terdapat 5 provinsi yang masuk kedalam *cluster* 1 (C1) dengan kategori produksi tanaman biofarmaka yang rendah pada tahun 2021. Sedangkan 29 provinsi lainnya masuk kedalam *cluster* 2 (C2) dengan kategori daerah produksi tananman biofarmaka tinggi pada tahun 2021.

Kata Kunci: Biofarmaka, BPS, *Clustering*, C1, C2, *K-Means*.

1. Pendahuluan

Negara Indonesia merupakan salah satu negara yang memproduksi banyak tanaman biofarmaka sehingga memiliki potensi untuk menemukan obat baru. Tanaman biofarmaka atau biasa disebut tanaman obat adalah tumbuhan yang salah satu, beberapa atau seluruh

bagian tumbuhan tersebut mengandung zat atau bahan aktif yang berguna untuk kesehatan tubuh, penyembuhan penyakit maupun bahan kosmetik [1].

Keberadaan tanaman obat atau biofarmaka sebagai penawar penyakit telah digunakan sejak zaman nenek moyang hingga saat ini masuk ke dunia farmasi dan diakui secara ilmiah [2]. Mayoritas masyarakat di Indonesia memilih budidaya tanaman obat karena selain memiliki manfaat untuk kesehatan juga dapat digunakan sebagai bumbu masakan ataupun tanaman hias [3]. Produksi tanaman obat atau biofarmaka dalam negeri diharapkan dapat memenuhi kebutuhan masyarakat di Indonesia sehingga tidak perlu mengimpor dari negara lain. Untuk itu pemerintah perlu melakukan upaya agar dapat mengoptimalisasi dan dapat melakukan pemerataan produksi tanaman biofarmaka di Indonesia [4].

Berdasarkan data Badan Pusat Statistika Indonesia produksi tanaman biofarmaka bervariasi jumlahnya disetiap daerah. Sehingga diperlukan pengelompokkan daerah yang berpotensi menghasilkan tanaman biofarmaka untuk mengetahui daerah mana saja yang memproduksi dalam jumlah banyak ataupun sedikit. Sehingga hasil dari penelitian ini dapat menjadi sedikit informasi untuk pihak yang berkepentingan dalam mengambil kebijakan terhadap daerah yang memiliki produksi tanaman biofarmaka sedikit. Oleh karena itu dibutuhkan metode untuk memudahkan pengelompokan daerah penghasil tanaman biofarmaka salah satunya dengan metode analisis *clustering* [5].

Analisis *cluster* merupakan metode yang digunakan untuk membagikan data menjadi beberapa grup berdasarkan kesamaan yang telah ditetapkan sebelumnya. *Clustering* mengelompokkan sejumlah data pada *cluster* yang sama dan data yang berbeda pada *cluster* yang berbeda. Salah satu metode yang dapat digunakan dalam analisis *cluster* yaitu algoritma *K-Means clustering* [6]. K-Means *clustering* merupakan salah satu teknik dari algoritma Data Mining. Sebagaimana diketahui bahwa Algoritma Data Mining telah banyak digunakan untuk pemecahan masalah komputasi seperti yang berhubungan dengan klasifikasi [7].

Pada penelitian sebelumnya algoritma *K-Means* digunakan dalam pengelompokan daerah resapan air dan hasilnya menyimpulkan bahwa algoritma *K-Means* mampu mengelompokkan daerah dengan tingkat resapan air yang dimiliki. Sehingga memperoleh hasil terdapat 16 provinsi dengan tingkat resapan air tinggi dan 18 provinsi lainnya dengan tingkat resapan air rendah [8]. Pada penelitian ini akan dilakukan pengelompokkan daerah produksi tanaman biofarmaka berdasarkan produksi jenis tanaman di Indonesia menggunakan algoritma *K-Means clustering*. Tujuan dari penelitian ini yaitu untuk mengetahui daerah yang berpotensi mengasilkan tanaman dalam jumlah tinggi maupun rendah sehingga dapat dilakukan pemerataan disetiap daerah sebagai upaya pemerintah

2. Material dan Metode

2.1. *Clustering*

Clustering adalah teknik data mining yang digunakan untuk menganalisis dan mengkaji data untuk menyelesaikan permasalahan dalam pengelompokkan data membagi dari suatu dataset ke dalam subset [9]. Teknik *cluster* mempunyai dua cara dalam pengelompokkan yaitu hierarchical *clustering* dan non-hierarchical *clustering*. Hierarchical *clustering* digunakan untuk mengelompokkan data yang memiliki kemiripan paling dekat. Sedangkan non hierarchical *clustering* digunakan untuk mengelompokkan data yang berukuran besar dan harus menentukan jumlah klaster yang akan dibentuk [10]. *K-Means* yaitu salah satu dari metode pengelompokan data nonhierarki (sekatan) yang dapat mempartisi data ke dalam bentuk dua kelompok ataupun lebih [11]**.**

2.2. Algoritma *K-Means Clustering*

K-Means *clustering* adalah salah satu metode pengelompokkan bersifat partitional. Secara prinsip, metode *K-Means* bekerja dengan memasukkan *K* sebagai konstanta jumlah *cluster* yang diinginkan. Sedangkan, Means dalam hal ini berarti nilai satu rata-rata dari suatu grup data yang dalam hal ini didefinisikan sebagai *cluster* [12]. Algoritma ini didasarkan pada penentuan jarak antara centroid dan data latih. Kemudian, jumlah *cluster* centroid berdasarkan jumlah yang diinginkan. Sementara itu, centroid inisialisasi yang dihasilkan secara acak dengan mempertimbangkan pelatihan data. Dengan kata lain, *centroid* harus berada dalam ruang data pelatihan [13]. Berikut Langkah-langkah *clustering* data menggunakan metode *K-Means* [14]:

1. Menentukan banyaknya *k* sebagai jumlah *cluster* yang akan dibentuk
2. Menghitung pusat *cluster* dengan menggunakan rumus :

$$\frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

dengan:
xi : data ke-*i*
n : jumlah data

3. Menghitung jarak tiap objek ke centroid menggunakan *Euclidian distance* dengan persamaan berikut :

$$d(x,y) = \|x - y\| = \sqrt{\sum_{i=0}^{n} (x_i - y_i)^2} \tag{2}$$

dengan:

$d(x, y)$: Jarak antar data x dan data y
x_i : Data x pada pusat *cluster* i
y_i : Data y pada pusat data i
n : Jumlah data

4. Kelompokkan data berdasarkan jarak pada centroid
5. Menentukan pusat *cluster* baru dengan menghitung rata – rata data di pusat yang sama.
6. Mengulangi langkah 3 hingga 5 sampai anggota yang ada pada tiap *cluster* tidak berubah
7. Jika anggota *cluster* sudah tidak berubah, maka centroid pada perulangan terakhir akan digunakan sebagai parameter untuk pengelompokkan data. Kemudian menghitung *Silhouette Coefficient* dengan k dan anggota-anggota *cluster* yang didapat.

2.3. Koefesien *Silhouette*

Silhouette coefficient merupakan salah satu metode yang digunakan untuk menguji kualitas dan kekuatan dari sebuah *cluster*. Nilai silhouette *coefficient* berkisar antara -1 dan 1. Hasil *cluster* dikatakan baik jika nilai silhouette *coefficient* adalah 1 Tahap perhitungan silhouette coefficient adalah [15] :

1. Menghitung rata-rata jarak tiap data ke-i dengan semua data yang berada dalam satu *cluster*.
2. Menghitung rata-rata jarak tiap data ke-i dengan semua data di *cluster* lain llau mengambil nilai terkecil dari semua jarak rata – rata.
3. Menghitung nilai silhouette coefficient dengan menggunakan persamaan:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \tag{3}$$

dengan:
a(i) : rata – rata data ke-i dengan semua data pada *cluster* yang sama
b(i) : rata – rata data ke-I dengan semua data pada *cluster* yang berbeda
S(i) : nilai silhouette coefficient

2.4. Normalisasi Data

Normalisasi adalah proses penskalaan nilai atribut dari data sehingga bisa terletak pada rentang tertentu. Normalisasi Z-score atau Z-score Normalization merupakan metode normalisasi berdasarkan mean (nilai rata-rata) dan standard deviation (deviasi standar) dari data. Metode ini sangat berguna jika tidak diketahui nilai aktual minimum dan maksimum dari data. Normalisasi data dapat dihitung menggunakan persamaan berikut [16] :

$$Z_i = \frac{X_i - \bar{X}}{S} \tag{4}$$

dengan:
$\bar{X}$: nilai rata – rata
X_i : data mentah
Z_i : data sekunder
S : simpangan baku

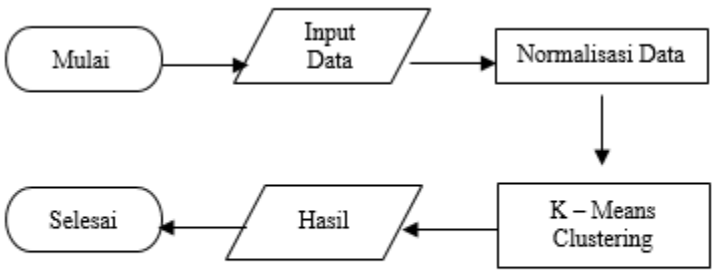
2.5. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data sekunder yang didapatkan dari website resmi Badan Pusat Statistika (BPS) dengan alamat https://bps.go.id. Data yang digunakan adalah data jumlah produksi tanaman biofarmaka menurut jenis tanamannya pada tahun 2021 di Indonesia. Data tersebut akan dianalisis dengan menggunakan metode algoritma *K-Means Clustering* dan bantuan *softwere* Rstudio. Berikut merupakan sampel data yang digunakan.

Tabel 1. Sampel Data Hasil Produksi Tanaman Biofarmaka Tahun 2021.

Provinsi	Jahe (kg)	Lengkuas (kg)	kunyit (kg)	kencur (kg)	lempuyang (kg)	Temulawak (kg)	Temuireng (kg)
Aceh	2327504	612026	271305	1246598	154	17418	0.00
Sumatera utara	52245300	1099804	1345654	14114104	40779	180468	4412
Sumatera barat	13772833	3589529	1195254	4354848	2957	132735	3345
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Papua barat	203323	74862	64764	807076	1298	36728	470
Papua	113016	84069	35413	62437	9319	13280	8876

Adapun langkah – langkah dari penelitian ini dalam bentuk flowchart yaitu sebagai berikut:



Gambar 1. Flowchart Penelitian

3. Hasil dan Diskusi

3.1. Normalisasi data

Analisis *clustering* menggunakan algritma *K-Means* menggunakan data yang sudah dinormalisasikan terlebih dahulu supaya mendapatkan hasil yang optimal. Berikut merupakan sampel normalisasi data penelitian yang telah dilakukan normalisasi data.

Tabel 2. Sampel Data Normalisasi

Provinsi	Jahe (kg)	Lengkuas (kg)	kunyit (kg)	kencur (kg)	lempuyang (kg)	Temulawak (kg)	Temuireng (kg)
Aceh	-0.41650	-0.35910	-0.41035	-0.28538	-0.27576	-0.23117	-0.25973
Sumatera utara	2.68246	-0.25263	-0.07862	0.59114	-0.23054	-0.19073	-0.25376
Sumatera barat	0.29404	0.29081	-0.12506	-0.07365	-0.27264	-0.20257	-0.25520
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Papua barat	-0.54838	-0.47635	-0.47413	-0.31532	-0.27449	-0.22638	-0.25910
Papua	-0.55398	-0.47434	-0.48319	-0.36605	-0.26556	-0.23220	-0.24771

3.2. Menemtukan Jumlah *Cluster*

Pada analisis *clustering* menggunakan algoritma *K-Means* dengan bantuan software Rstudio, langkah pertama yang dilakukan adalah menentukan jumlah *cluster*. Dalam penelitian ini digunakan metode uji akurasi data menggunakan koefisisen silhouette untuk mengidentifikasi jumlah *cluster* terbaik. Hasil *clustering* akan semakin baik jika nilai koefisien silhouette yang dihasilkan mendekati 1 seperti terlihat pada tabel berikut.

Tabel 3. Nilai Validasi Koefisien Silhouette

Jumlah *Cluster*	Koefisien Silhouette
2 *cluster*	0.8627
3 *cluster*	0.8118
4 *cluster*	0.7888

Pada tabel 3 terlihat bahwa nilai koefisien silhouette yang paling besar dan mendekati angka 1 yaitu pada 2 *cluster* dengan nilai koefisien silhouette sebesar 0.8627. Sehingga pada penelitian ini akan digunakan 2 *cluster* sebagai pengelompokkan data.

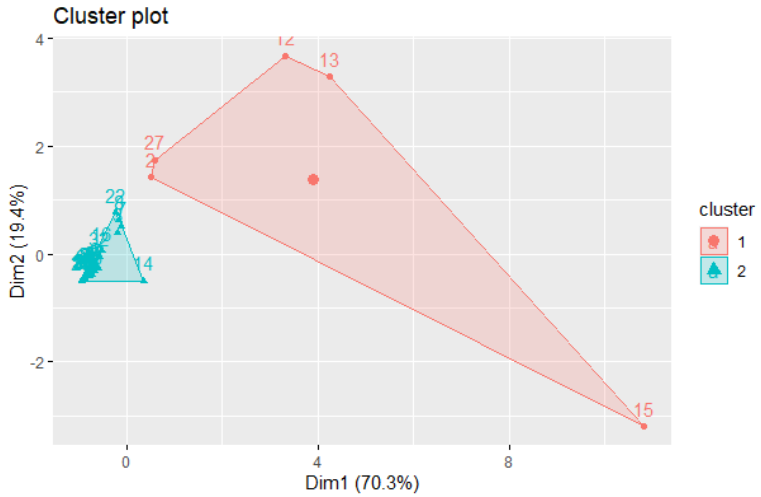
3.3. Menentukan Pusat *Cluster (Centroid)*

Langkah selanjutnya dalam algoritma *K-Means* adalah mementukan nilai centroid atau pusat *cluster* dari data yang didapatkan dengan jumlah *cluster* yang telah ditentukan yaitu 2 *cluster*. Hasil perhitungan nilai centroid dapat dilihat pada tabel 4 berikut.

Tabel 4. Nilai Centroid

Cluster	Jahe	Lengkuas	Kencur	Kunyit	Lempuyang	Temulawak	Temuireng
C1	-0.03491	-0.9290	-0.0184	-0.1609	-0.1636	-0.16657	-0.16302
C2	1.1521	3.06581	0.6095	5.2829	5.39722	5.49665	5.37971

Pada tabel 4 diatas berdasarkan nilai centroid masing – masing variabel diketahui bahwa *cluster* 1 (C1) nilai centroid setiap variabelnya bernilai negatif sehingga dikategorikan sebagai tingkat produksi rendah. Sedangkan pada *cluster* 2 (C2) nilai centroid tiap variabelnya bernilai positif sehingga dapat dikategorikan sebagai tingkat produksi tinggi. Hasil analisis *K-Means clustering* dapat dilihat pada gambar berikut:



Gambar 2. Plot Hasil *K-Means Clustering*

Gambar 2 menunjukkan visualisasi hasil analisis menggunakan *K-Means clustering*. Berdasarkan hasil tersebut dapat diketahui hasil pengelompokkan pada tabel berikut:

Tabel 5. Hasil *K-Means Clustering*

Cluster	Provinsi
C1	Sumatera Utara, Jawa Barat, Jawa Tengah, Jawa Timur, dan Sulawesi tenggara.
C2	Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua.

Tabel 5 menunjukkan hasil pengelompokkan daerah menggunakan algoritma *K-Means*. Berdasarkan hasil tersebut terdapat 5 provinsi yang masuk dalam C1 yaitu dengan

kategori produksi rendah tanaman biofarmaka pada tahun 2021, wilayah tersebut yaitu Sumatera Utara, Jawa Barat, Jawa Tengah, Jawa Timur, dan Sulawesi tenggara. Sedangkan 29 provinsi lainnya masuk kedalam C2 dimana daerah tersebut masuk kedalam kategori produksi tinggi tanaman biofarmaka pada tahun 2021, wilayah tersebut diantaranya yaitu Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kep. Bangka Belitung, Kep. Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Timur, Kalimantan Utara, Sulawesi Utara, Sulawesi Tengah, Sulawesi Selatan, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua.

4. Kesimpulan

Berdasarkan penelitian yang telah dilakukan menggunakan metode algoritma *K-Means clustering* dengan bantuan *software* Rstudio maka dapat disimpulkan bahwa terdapat 5 provinsi yang masuk kedalam C1 dengan kategori produksi tanaman biofarmaka yang rendah pada tahun 2021. Sedangkan 29 provinsi lainnya masuk kedalam C2 dengan kategori daerah produksi tananman biofarmaka tinggi pada tahun 2021.Saran untuk penelitian selanjutnya dapat ditambahkan variabel yang mempengaruhi dengan metode pengelompokkan yang berbeda. Maka tidak menutup kemungkinan hasil yang didapatkan juga akan berbeda dari penelitian ini.

Daftar Pustaka

[1] Ariyanto, S. E., & Indaryani, M. Potensi Tanaman Obat (Biofarmaka) di Kabupaten Jepara. *Semin. Nas.*, 2017, pp. 277–285, 2022.

[2] Listyana, N. H., Darsono, D., & Sutrisno, J. Potensi Pengembangan Tanaman Obat Di Wilayah Aglomerasi Solo Raya. *J. Tumbuh. Obat Indones.*, 15(1), pp. 27–30, 2022, doi: 10.22435/jtoi.v15i1.5846.

[3] Parawansah, A. E., & Saida. Sosialisasi Pemanfaatan Tanaman Obat Keluarga Sebagai Upaya Untuk Meningkatkan Imunitas Tubuh Ditengah Pandemi di Kota Kendari. *J. Community Engagem. Heal.*, 3(2), pp. 325–328, 2020.

[4] Silalahi, C. J., Situmorang, A., & Naibaho, J. F. Implementasi Metode K-Means Clustering Untuk Memetakan Daerah Potensial Penghasil Padi di Provinsi Sumatera Utara. *Methotika J. Ilm. Tek. Inform.*, 2(2), pp. 49–57, 2022, [Online]. Available: http://ojs.fikom-methodist.net/index.php/methotika

[5] Herviany, M., Delima, S. P., & Nurhidayah, T. Comparison of K-Means and K-Medoids Algorithms for Grouping Landslide Prone Areas in West Java Province. *MALCOM Indones. J. Mach. Learn. Comput. Sci. J.*, 1(4), pp. 34–40, 2021.

[6] Nasari, F. Pengelompokkan Daerah Produksi Kelapa Sawit Menggunakan

Algoritma K-Means Clustering. *J. Sains dan Ilmu Terap.*, 4(2), pp. 34–38, 2021.

[7] Syaifuddin, S., Ramlah, R., Hakim, I., Berliana, Y., & Nurhayati, N. Pemetaan Produksi Tanaman Tomat di Indonesia Berdasarkan Provinsi Menggunakan Algoritma K-Means Clustering. *J. Comput. Syst. Informatics*, 3(4), pp. 222–228, 2022, doi: 10.47065/josyc.v3i4.2206.

[8] Batubara, D. N., Windarto, A. P., Hartama, D., & Satria, H. Analisis Metode K-MEANS Pada Pengelompokan Keberadaan Area Resapan Air Menurut Provinsi. *Semin. Nas. Sains Teknol. Inf.*, pp. 345–349, 2019, [Online]. Available: https://prosiding.seminar-id.com/index.php/sensasi/article/view/324 (05 Juni 2020)

[9] Alkhairi P., & Windarto, A. P. Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara. *Semin. Nas. Teknol. Komput. Sains*, pp. 762–767, 2019.

[10] Khomsatun, Ikhsan, D., Ali, M., & Kusrini. Sistem Pengambilan Keputusan Pemilihan Lahan Tanam di Kabupaten Wonosobo dengan K-Means Clustering dan Topsis. *J. Nas. Pendidik. Tek. Inform. JANAPATI*, 9(1), pp. 55–63, 2020.

[11] Aditya, A., Jovian, I., & Sari, B. N. Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018 / 2019. *J. MEDIA Inform. BUDIDARMA,* 4(1), pp. 51–58, 2020, doi: 10.30865/mib.v4i1.1784.

[12] Pakpahan, H. S., Widians, J. A., & Firmanda, H. D. A. Implementasi Metode K-Means Untuk Pengelompokan Potensi Produksi Komoditas Perkebunan. *Adopsi Teknol. dan Sist. Inf.*, 1(1), pp. 52–60, 2022, doi: 10.30872/atasi.v1i1.49.

[13] Purnaansyah, Haviluddin, Fanany, O. G. A., & Tahyudin, I. Comparison Between K-Means and Fuzzy C-Means Clustering in Network Traffic. *Netw. Traffic Act*, 2, pp. 301–310, 2018, doi: 10.1007/978-3-319-59280-0.

[14] Akbar, F., Dekapriyo, E. L., Moefad, A. M., & Haris, A. M. Grouping of the Second Wave of Covid-19 Infection Areas in East Java Province Using K-Means Clustering, *J. Mat. MANTIK*, 8(1), pp. 53–62, 2022.

[15] Adhe, D., Rachman, C., Goejantoro, R., & Tisna, D. Implementation ff Text Mining For Grouping Thesis Documents Using K-Means Clustering. *J. EKSPONENSIAL*, 11(2), pp. 167–174, 2020.

[16] Abdurrahman, D. D., Agus, F., & Putra, G. M. Implementasi Algoritma Partitioning Around Medoids (PAM) untuk Mengelompokkan Hasil Produksi Komoditi Perkebunan (Studi Kasus : Dinas Perkebunan Provinsi Kalimantan Timur). *J. Ilm. Ilmu Komput.*, 16(2), pp. 130–138, 2021, doi: http://dx.doi.org/10.30872/jim.v16i2.6520.