

Pengelompokkan Daerah Produksi Kelapa Sawit Menggunakan Algoritma K-Means Clustering

Fina Nasari¹

¹ Universitas Potensi Utama

Jl. K.L Yos Sudarso KM. 6.5 Tanjung Mulia Medan

¹finanasari@gmail.com

Intisari— Riau merupakan salah satu provinsi di Indonesia dengan luas area dan produksi kelapa tertinggi di Indonesia. Penyebaran perkebunan kelapa sawit di provinsi riau menyebar hamper merata diseluruh kabupaten/kota. Produksi kelapa sawit setiap daerah bervariasi, oleh karena itu perlu adanya pengclustering daerah berdasarkan produksi kelapa sawit. Algoritma *k-means clustering* mampu mengelompokkan data yang sama dalam kelompok yang sama. Pengelompokkan data berdasarkan jarak kemasing-masing *centroid*. Tujuan penelitian ini adalah mengcluster daerah sesuai dengan jumlah produksi kelapa sawit yang ada di provinsi riau kedalam 2, 3, dan 5 *cluster*. Hasil dari penelitian ini *algoritma k-means* mampu mengelompokkan daerah berdasarkan produksi kelapa sawit dengan hasil *cluster*, jika data dikelompokkan menjadi 2 *cluster* terdapat 6 daerah masuk pada *cluster* 0 dan 6 daerah masuk pada *cluster* 1. Hasil pengelompokkan 3 *cluster* terdapat 5 daerah pada *cluster* 0, 4 daerah pada *cluster* 1 dan 3 daerah pada *cluster* 2. Hasil pengelompokkan 5 *cluster* terdapat 2 daerah pada *cluster* 0, *cluster*2 dan *cluster* 4 , 3 daerah pada *cluster* 1 dan 3. *Cluster* 2 menjadi *cluster* terbaik berdasarkan nilai *DBI* terkecil dari *cluster* lain dengan nilai -82.338.884.292,014.

Kata kunci— *clustering, K-Means, Nilai DBI, Kelapa Sawit.*

Abstract— Riau is one of the provinces in Indonesia with the highest area and coconut production in Indonesia. The spread of oil palm plantations in Riau Province is spread almost evenly throughout the districts/cities. Oil palm production varies from region to region, therefore it is necessary to cluster regions based on palm oil production. The k-means clustering algorithm is able to group the same data into the same group. Grouping data based on the distance to each centroid. The purpose of this study was to cluster regions according to the amount of oil palm production in Riau province into 2, 3 and 5 clusters. The results of this study the k-means algorithm is able to group regions based on palm oil production with cluster results, if the data is grouped into 2 clusters there are 6 entry areas in cluster 0 and 6 entry areas in cluster 1. The results of grouping 3 clusters have 5 regions in cluster 0 , 4 regions in cluster 1 and 3 regions in cluster 2. The results of grouping 5 clusters are 2 regions in cluster 0, cluster 2 and cluster 4, 3 regions in clusters 1 and 3. Cluster 2 is the best cluster based on the smallest DBI value of another cluster with a value of -82,338,884,292.014.

Keywords— *clustering, K-Means, DBI Value, Palm Oil.*

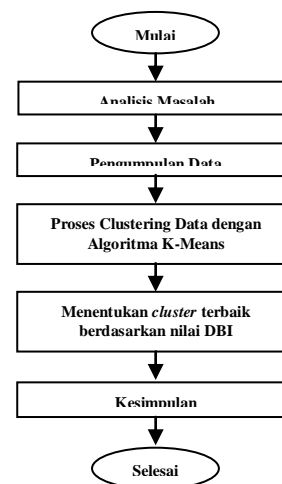
I. PENDAHULUAN

Kelapa sawit merupakan salah satu sumber daya ekspor utama Indonesia. Penyebaran perkebunan kelapa sawit banyak di pulau Sumatera dan Kalimantan. Di pulau Sumatera sendiri kelapa sawit terdapat hampir diseluruh provinsi. Provinsi Riau menjadi salah satu provinsi dengan luas dan jumlah produksi kelapa sawit tertinggi di Sumatera dan Indonesia. Banyaknya produksi kelapa sawit di provinsi Riau salah satunya karena sebaran kebun kelapa sawit terdapat di hampir seluruh kabupaten/kota. Pengelompokkan daerah berdasarkan produksi kelapa sawit bertujuan untuk mengetahui daerah dengan produksi tinggi dan rendah. Algoritma *k-means clustering* merupakan algoritma yang mampu mengelompokkan data yang sama pada kelompok yang sama dan data yang berbeda pada kelompok yang berbeda. Data dikelompokkan berdasarkan jarak setiap kata kemasing-masing *centroid*[1][2][3].

Fokus penelitian ini adalah mengelompokkan daerah berdasarkan produksi kelapa sawit kedalam 2 *cluster*, 3 *cluster*, dan 5 *cluster* serta menemukan bentuk *cluster* dari ke-3 *cluster* yang akan dibuat berdasarkan nilai *DBI*.

II. METODOLOGI PENELITIAN

Tahapan penelitian ini dimulai dari pengumpulan data primer yaitu data produksi kelapa sawit provinsi Riau tahun 2018-2019 yang diperoleh dari data BPS Riau. Tahapan penelitian dapat dilihat pada gambar 1 berikut ini:



Gambar 1. Tahapan Penelitian.

A. Analisis Masalah

Pada tahap ini melakukan analisis masalah terkait pemetaan daerah/kabupaten berdasarkan jumlah produksi kelapa sawit pada provinsi Riau. Dengan adanya pemetaan ini diharapkan mampu mengelompokkan kabupaten dengan produksi tinggi, sedang dan rendah. Selain itu juga melihat jenis pengelompokkan mana yang paling terbaik jika dilihat dari nilai DBI.

B. Pengumpulan Data

Tahap ini melakukan proses pengumpulan data baik data primer maupun data sekunder. Untuk data primer berupa data yang digunakan untuk pengelompokkan daerah/kabupaten berdasarkan produksi kelapa sawit. Data primer menggunakan data yang diperoleh dari BPS riau. Untuk data sekunder berupa literature dari penelitian sebelumnya terkait dengan k-means clustering, nilai DBI hingga pengelompokkan data produksi kelapa sawit.

C. Proses Clustering Data Dengan Algoritma K-Means

Pada Tahap ini melakukan proses clustering algoritma k-means menggunakan tools *rapit miner*. Cluster yang akan diuji dengan 3 jenis yaitu jumlah cluster 2, 3 dan 5. Sehingga nanti akan mendapatkan hasil pengelompokkan daerah/kabupaten dengan jenis 2 kelompok, 3 kelompok dan 5 kelompok.

Clustering merupakan metode yang digunakan untuk membagikan data menjadi beberapa group berdasarkan Kesamaan yang telah ditetapkan sebelumnya[4]. clustering mengelompokkan sejumlah data pada cluster yang sama dan data yang berbeda pada cluster yang berbeda[5][3].

K-Means adalah salah satu metode clustering dengan menggunakan pendekatan partitioning, dimana menggunakan centroid untuk membantu cluster. centroid sendiri merupakan titik tengah suatu cluster[6].

Adapun tahapan algoritma k-means clustering adalah sebagai berikut:

1. Menentukan jumlah cluster
2. Menetapkan Jumlah cluster
3. Menentukan Nilai Centroid
4. Menghitung Jarak setiap data ke masing-masing centroid
5. Mengelompokkan Data berdasarkan jarak terdekat
6. Menghitung nilai centroid
7. Jika ada perubahan pada centroid maka proses kembali ke langkah 3, jika tidak ada perubahan centroid proses selesai[7].

Menghitung jarak setiap data ke masing-masing centroid menggunakan rumus sebagai berikut:

$$D = \sqrt{(xi - si)^2 + (yi - ti)^2}$$

Dimana:

- D : Eclidean Distance
- I : Banyaknya Objek
- (x,y) : Koodinat Objek

(s,t) : Koordinat Centroid[8].

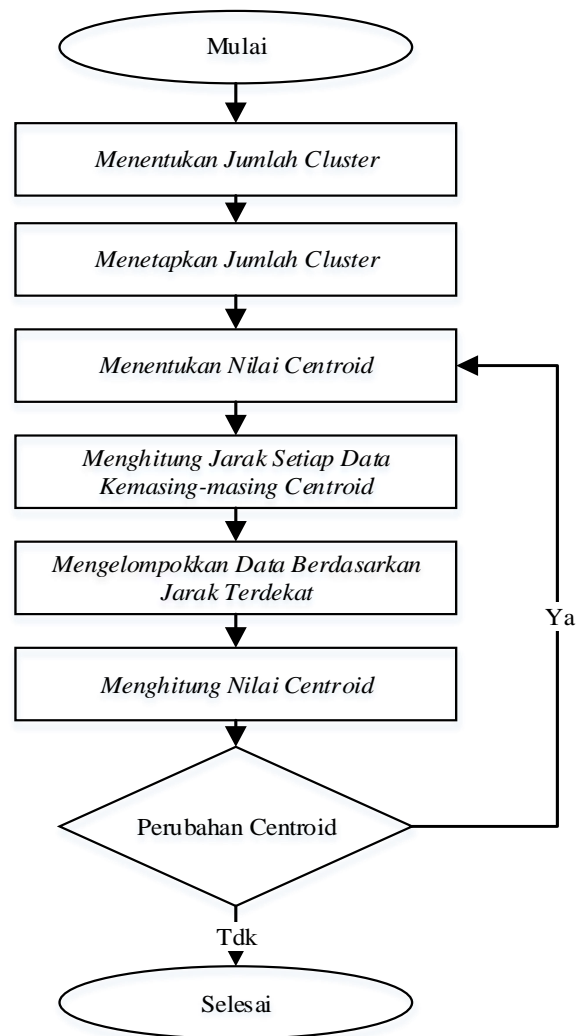
Sedangkan formula yang digunakan untuk menentukan nilai centroid pada tahap Iterasi adalah sebagai berikut:

$$VIJ = \frac{1}{Ni} \sum_{k=0}^{Ni} Xkj$$

Dimana:

- VII: Centorid rata-rata cluster ke-i untuk variabel ke-i
- Ni: Jumlah Cluster ke-i
- i,k: Indeks Cluster
- j: Indeks dari variabel
- Xkj: Nilai dari ke -k variabel k-j dalam cluster[8].

Gambar tahapan algoritma k-means dapat dilihat pada gambar 2.



Gambar. 2 Tahapan Algoritma K-means Clustering[7]

Rapit miner merupakan tools perangkat ilmu data yang dikembangkan oleh perusahaan bernama sama dengan

penyedia lingkungan integrasi untuk persiapan data pembelajaran mesin pembelajaran pengembangan teks dan animasi prediktif[7].

Proses *clustering* yang dilakukan menggunakan *tools rapid miner* dimana data sebelum diproses data simpan dalam *Microsoft excel* dengan format *.xls* [7], selanjutnya data diimport kedalam *rapid miner* dan dibuat pemodelan dengan jumlah *cluster* 2, 3 dan 5.

D. Menentukan Cluster Terbaik Berdasarkan Nilai DBI

Berdasarkan hasil proses *cluster* pada tahap sebelumnya, pada tahap ini dilanjutkan dengan melihat *performance* setiap kelompok yang dibentuk berdasarkan nilai DBI. Dimana kelompok dengan nilai DBI terkecil merupakan kelompok terbaik[9].

Nilai *DBI (Davies Bouldin Index)* merupakan salah satu metode untuk mengukur validitas jumlah cluster paling optimal dimana kohesi didefinisikan sebagai jumlah pendekatan data terhadap titik pusat cluster dari cluster yang diikuti[9][10].

E. Kesimpulan

Pada tahap ini menarik kesimpulan dari penelitian yang dilakukan. Mulai dari hasil pengelompokan hingga kelompok terbaik yang dihasilkan berdasarkan nilai DBI.

III. HASIL DAN PEMBAHASAN

A. Data Penelitian

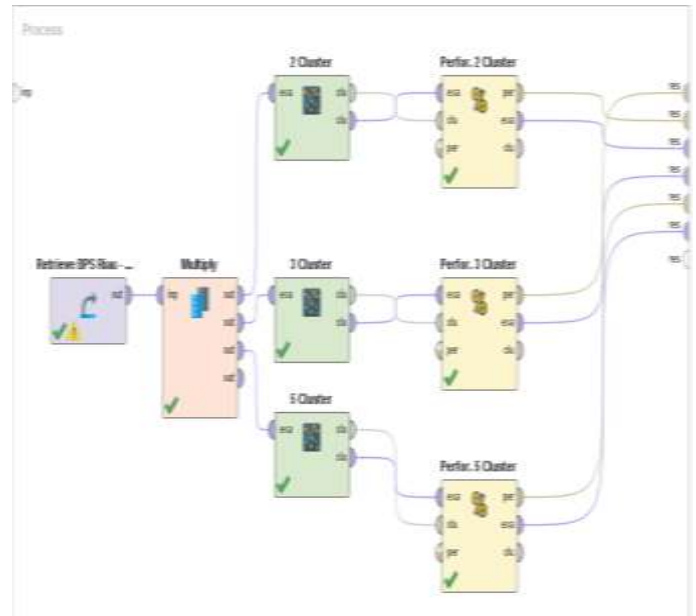
Data yang digunakan pada penelitian ini adalah data BPS provinsi riau Terkait data Produksi Kelapa Sawit Tahun 2018-2019 terlihat pada Tabel 1:

TABEL I
DATA PRODUKSI KELAPA SAWIT TAHUN 2018-2019

Kabupaten	2018	2019
Kuantan Singingi	452218	450804
Indragiri Hulu	286243	469273
Indragiri Hilir	733009	731009
Pelalawan	1339609	1339609
Siak	1193290	1098665
Kampar	1222465	955735
Rokan Hulu	1195460	1195460
Bengkalis	334066	334066
Rokan Hilir	813834	813832
Kepulauan Meranti	0	0
Pekanbaru	31219	36612
Dumai	82122	41195

B. Pemodelan Algoritma K-Means Pada Rapid Miner

Pemodelan *algoritma K-Means* yang akan diuji adalah jumlah *cluster* 2, 3 dan 5. Bentuk pemodelan terlihat pada gambar 3 berikut:



Gambar. 3 Pemodelan pengujian k-means dengan jumlah cluster 2, 3 dan 5

Berdasarkan gambar 1 diatas pemodelan yang akan digunakan untuk menguji algoritma k-means yaitu pemodelan 2 cluster, 3 cluster dan 5 cluster. Data set dihubungkan dengan *multiply* agar bisa melakukan pengujian bersamaan untuk 3 model cluster yang mau dibuat, selanjutnya dihubungkan dengan *performance distance* untuk melihat nilai *performance DBI* pada setiap model.

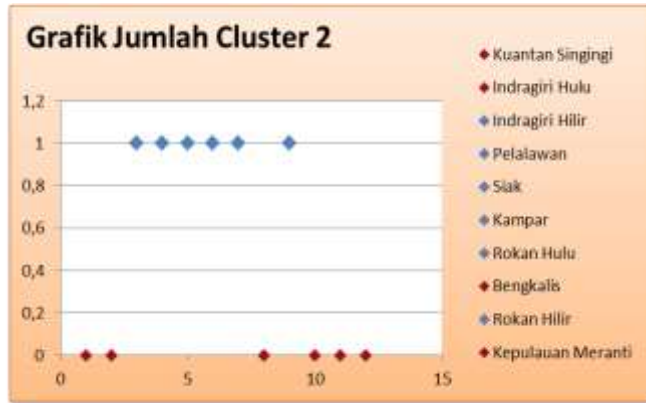
C. Hasil Pengujian K-Means Dengan Jumlah Cluster 2

Hasil pengujian *k-means* untuk jumlah cluster 2 terlihat pada Tabel 2 dan gambar 4 berikut:

TABEL II
HASIL PENGUJIAN K-MEANS 2 CLUSTER

Kabupaten/Kota	2018	2019	Kluster
Kuantan Singingi	452218	450804	cluster_0
Indragiri Hulu	286243	469273	cluster_0
Indragiri Hilir	733009	731009	cluster_1
Pelalawan	1339609	1339609	cluster_1
Siak	1193290	1098665	cluster_1
Kampar	1222465	955735	cluster_1
Rokan Hulu	1195460	1195460	cluster_1
Bengkalis	334066	334066	cluster_0

Kabupaten/Kota	2018	2019	Kluster
Rokan Hilir	813834	813832	cluster_1
Kepulauan Meranti	0	0	cluster_0
Pekanbaru	31219	36612	cluster_0
Dumai	82122	41195	cluster_0



Gambar. 4 Grafik Jumlah Cluster 2

Berdasarkan hasil pengujian terlihat pada tabel 2 diperoleh hasil, untuk setiap cluster masing-masing memiliki 6 data. 6 data pada cluster 0 dan 6 data pada cluster 1. Dimana kabupaten Kuantan Singingi, Indragiri Hulu, Bengkalis, Kepulauan Meranti, Pekanbaru dan Dumai masuk dalam satu kelompok. Indragiri Hilir, Pelalawan, Siak, Kampar, Rokan Hulu Dan Rokan Hilir masuk dalam kelompok kedua.

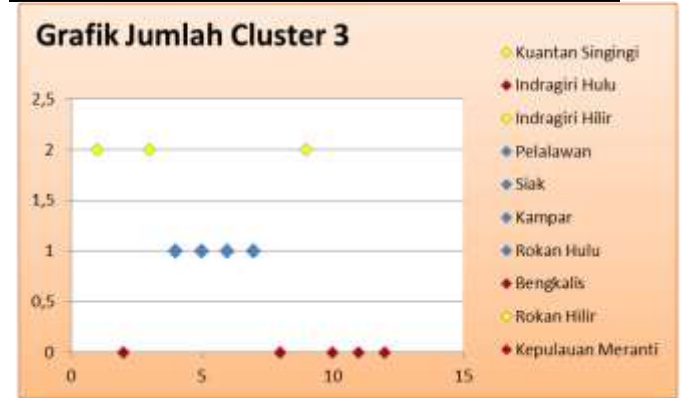
D. Hasil Pengujian K-Means Dengan Jumlah Cluster 3

Hasil pengujian k-means untuk jumlah cluster 3 terlihat pada Tabel 3 dan gambar 5 berikut:

TABEL III
HASIL PENGUJIAN K-MEANS 3 CLUSTER

Kabupaten/Kota	2018	2019	Kluster
Kuantan Singingi	452218	450804	cluster_2
Indragiri Hulu	286243	469273	cluster_0
Indragiri Hilir	733009	731009	cluster_2
Pelalawan	1339609	1339609	cluster_1
Siak	1193290	1098665	cluster_1
Kampar	1222465	955735	cluster_1
Rokan Hulu	1195460	1195460	cluster_1
Bengkalis	334066	334066	cluster_0
Rokan Hilir	813834	813832	cluster_2
Kepulauan Meranti	0	0	cluster_0

Kabupaten/Kota	2018	2019	Kluster
Pekanbaru	31219	36612	cluster_0
Dumai	82122	41195	cluster_0



Gambar. 5 Grafik Jumlah Cluster 3

Berdasarkan tabel 3 diperoleh hasil 5 data masuk pada cluster 0, 4 data pada cluster 1 dan 3 data pada cluster 2. Indragiri hulu, Bengkalis, Kepulauan Meranti, Pekanbaru dan Dumai masuk kedalam cluster 0. Pelalawan, Siak, Kampar dan Rokan Hulu masuk pada cluster 1. Kuantan Singingi, Indragiri Hilir dan Rokan Hilir masuk pada cluster 3.

E. Hasil Pengujian K-Means Dengan Jumlah Cluster 5

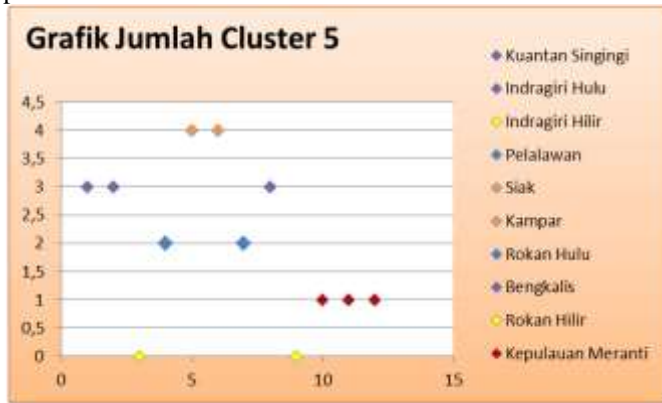
Hasil pengujian k-means untuk jumlah cluster 3 terlihat pada Tabel 4 dan gambar 6 berikut:

TABEL IV
HASIL PENGUJIAN K-MEANS 5 CLUSTER

Kabupaten/Kota	2018	2019	Kluster
Kuantan Singingi	452218	450804	cluster_3
Indragiri Hulu	286243	469273	cluster_3
Indragiri Hilir	733009	731009	cluster_0
Pelalawan	1339609	1339609	cluster_2
Siak	1193290	1098665	cluster_4
Kampar	1222465	955735	cluster_4
Rokan Hulu	1195460	1195460	cluster_2
Bengkalis	334066	334066	cluster_3
Rokan Hilir	813834	813832	cluster_0
Kepulauan Meranti	0	0	cluster_1
Pekanbaru	31219	36612	cluster_1
Dumai	82122	41195	cluster_1

Berdasarkan tabel 4 diperoleh hasil 2 data pada cluster 0, 3 data pada cluster 1, 2 data pada cluster 2, 3 data pada cluster 3 dan 2 data pada cluster 4. Dimana Indragiri Hilir dan Rokan Hilir masuk pada cluster 0. Kepulauan Meranti, Pekanbaru

dan Dumai masuk pada *cluster* 1. Pelalawan dan Rokan Hulu masuk pada *cluster* 2. Kuantan Singingi, Indragiri Hulu dan Bengkalis masuk pada *cluster* 3. Siak dan Kampar masuk pada *cluster* 4.



Gambar. 6 Grafik Jumlah Cluster 3

F. Hasil Pengukuran Performance Nilai DBI

Hasil pengukuran *performance* setiap model yang diuji berdasarkan *Nilai DBI* terlihat pada tabel 5 berikut:

TABEL V
HASIL PERFORMANCE NILAI DBI

Pengujian	Nilai performa
K-Means 2	-82.338.884.292
K-Means 3	-42.529.203.084
K-Means 5	-5.659.950.362

Berdasarkan nilai *dbi* bentuk *cluster* terbaik adalah *cluster* 2 dengan nilai *dbi* -82.338.884.292, selanjutnya *cluster* 3 dengan nilai *dbi* -42.529.203.084 dan *cluster* 5 pada peringkat terakhir dengan nilai *dbi* -5.659.950.362. Grafik Hasil pengukuran nilai *dbi* terlihat pada Gambar 2 berikut:



Gambar. 4 Pemodelan pengujian k-means dengan jumlah *cluster* 2, 3 dan 5

IV. KESIMPULAN

Berdasarkan hasil pengujian algoritma *k-means* untuk model pertama jumlah *cluster* 2 diperoleh hasil 6 data pada

cluster 0 dan 6 data pada *cluster* 1. Untuk model kedua jumlah *cluster* 3 diperoleh hasil 5 data pada *cluster* 0, 4 data pada *cluster* 1 dan 3 data pada *cluster* 2. Untuk model terakhir jumlah *cluster* 5 diperoleh hasil 2 data pada *cluster* 0, 3 data pada *cluster* 1, 2 data pada *cluster* 2, 3 data pada *cluster* 3 dan 2 data pada *cluster* 4.

Berdasarkan hasil pengukuran nilai *dbi* model terbaik ada pada bentuk *cluster* 2 dengan nilai *dbi* -82.338.884.292. Data daerah berdasarkan model *cluster* 2 yaitu kabupaten Kuantan Singingi, Indragiri Hulu, Bengkalis, Kepulauan Meranti, Pekanbaru dan Dumai masuk dalam satu kelompok atau *cluster* 0 atau Jumlah Produksi Rendah. Indragiri Hilir, Pelalawan, Siak, Kampar, Rokan Hulu Dan Rokan Hilir masuk dalam kelompok kedua atau *cluster* 1 atau jumlah produksi tinggi.

REFERENSI

- [1] F. Nasari, C. Jhony, and M. Sianturi, "Penerapan Algoritma K-Means Clustering..."
- [2] F. Nasari and S. Darma, "Seminar Nasional Teknologi Informasi dan Multimedia 2015 PENERAPAN K-MEANS CLUSTERING PADA DATA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS POTENSI UTAMA)"
- [3] A. Nofiar, S. Defit, and Sumijan, "Penentuan Mutu Kelapa Sawit Menggunakan Metode K-Means Clustering," *J. KomtekInfo*, vol. 5, no. 3, pp. 1-9, 2019, doi: 10.35134/komtekinfo.v5i3.26.
- [4] I. M. Pulungan, S. Saifullah, M. Fauzan, and A. P. Windarto, "Implementasi Algoritma K-Means Clustering dalam Menentukan Blok Tanaman Sawit Paling Produktif," *Pros. Semin. Nas. Rts. Inf. Sci.*, vol. 1, no. September, p. 338, 2019, doi: 10.30645/senaris.v1i0.39.
- [5] F. Nasari, "ALGORITMA K-MEANS CLUSTERINGDALAM PENYEBARAN PENYAKIT DIARE".
- [6] S. Hajar, A. A. Novany, A. P. Windarto, A. Wanto, and E. Irawan, "Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Penerapan K-Means Clustering Pada Ekspor Minyak Kelapa Sawit Menurut Negara Tujuan," *Semin. Nas. Teknol. Komput. Sains*, pp. 314-318, 2020.
- [7] D. F. Pasaribu, I. S. Damanik, E. Irawan, Suhada, and H. S. Tambunan, "Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 2, no. 1, pp. 11-20, 2021, doi: 10.37148/bios.v2i1.17.
- [8] D. Haryadi, "Penerapan Algoritma K-Means Clustering Pada Produksi Perkebunan Kelapa Sawit Menurut Provinsi," *J. Informatics Commun. Technol.*, vol. 3, no. 1, pp. 50-64, 2021, doi: 10.52661/j_ict.v3i1.71.
- [9] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95-100, 2021, [Online]. Available: www.bps.go.id
- [10] A. Bates and J. Kalita, "Counting Clusters in Twitter Posts," in *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*, 2016, doi: 10.1145/2905055.2905295.