JOISIE Journal Of Information System And Informatics Engineering
Vol. 6, No.1, Juni 2022, Hlm 39-48

p- ISSN: 2503-5304
e- ISSN: 2527-3116

IMPLEMENTASI ALGORITMA K-MEANS UNTUK MENGKLASTER KELOMPOK SEKTOR PERKEBUNAN DI INDONESIA

Hesti Pratiwi[1)], Ari Purno Wahyu Wibowo[2)]
[1,2] Program Studi S1 Teknik Informatika, Fakultas Teknik, Universitas Widyatama
Jl. Cikutra No.204A, Sukapada, Cibeunying Kidul, Kota Bandung, Jawa Barat 40125
email: [1]hesti.1402@widyatama.ac.id, [2]ari.purno@widyatama.ac.id

Abstract

Known to be the agricultural country, Indonesia offers plenty of potential in terms of agricultural productivity, mainly soil-related production. In addition to that, of all sectors, a plantation sector also helps to sustain national economy. Therefore, the development from the agricultural sector, in this case refers to plantation, needs to be monitored by the associated institution in an attempt for optimizing the output or production of the aforementioned sector. This study is aimed at providing the classification of information on the agricultural sector area based on its productivity and the crop production. A grouping of these plantation sectors was developed by data mining method that implemented K-Means algorithm, which then visualized on the Business Intelligence Tableau application. From that classification, the study obtained three clusters from the plantation sectors, such as "great" that meant above the target, "good" that referred to inability to reach the target yet still above average, and "underperformed" which meant the plantation sectors in these clusters needed to be properly looked at because of its underperforming productivity. The cluster results obtained from this study can be used as supporting material for decision-making that will be made by relevant agencies for future policies so that the distribution of plantations can be controlled. The accuracy rate of this research reached 79.41%.

Keywords: *Clustering, Kmeans, Tableau, Indonesian Plantation*

Abstrak

Dikenal dengan julukan negara agraris, Indonesia memiliki banyak potensi hasil bumi dari pertanahan, tak terkecuali dari sektor perkebunan yang memegang peran penting dari keseluruhan perekonomian nasional. Oleh karena itu, perkembangan produktivitas hasil bumi dari sektor pertanian, dalam hal ini perkebunan perlu dipantau supaya bisa menjadi perhatian instansi terkait dalam mengoptimalkan daerah penghasil tersebut. Penelitian ini bertujuan untuk menyediakan informasi klasifikasi terhadap daerah sektor pertanian berdasarkan produktivitas serta produksi hasil bumi. Pengelompokan sektor perkebunan ini dikembangkan dengan metode data mining dengan mengimplementasikan algoritma K-Means yang divisualisasikan pada aplikasi Business Intelligence Tableau. Dari hasil klasifikasi tersebut akan di dapat tiga cluster sektor perkebunan yang diantaranya adalah "great" yang berarti di atas target, "good" yang berarti belum mencapai target namun masih di atas rata-rata, dan "underperformed" yang berarti sektor perkebunan dalam cluster ini perlu perhatian khusus karena produktivitasnya di bawah performa. Hasil cluster yang didapatkan dari penelitian ini dapat dijadikan sebagai bahan pendukung pengambilan keputusan yang akan dibuat oleh instansi terkait untuk kebijakan di masa depan sehingga distribusi perkebunan dapat terkontrol. Tingkat akurasi dari penlitian ini mencapai 79.41%.

Kata Kunci: *Clustering, Kmeans, Tableau, Perkebunan Indonesia*

1. PENDAHULUAN

Pandemic Covid-19 telah memberikan dampak bagi semua sektor dalam kehidupan, tak terkecuali adalah sektor perkebunan yang merupakan penunjang pada sektor ekonomi negara. Banyak negara yang mengalami hal serupa, maka sebagai negara agraris, Indonesia diharapkan dapat memanfaatkan peluang dalam sektor perkebunan dengan meningkatkan angka ekspor ke berbagai negara di dunia. Status negara agraris yang disandang oleh Indonesia menunjukkan sektor pertanian mempunyai potensi dan dapat

menjadi andalan nasional untuk menjalankan dan memperkuat ketahanan pangan nasional (Darmawan, Didit; Genua, Veronika; Kristianto, Sonny; Murdaningsih; Hutubessy, 2019).

Indonesia termasuk ke dalam tiga besar negara di dunia dengan penghasil hasil bumi dari perkebunan terbesar. Untuk itu, informasi mengenai klasifikasi dari sektor perkebunan di Indonesia dibutuhkan sebagai alat pantau instansi terkait dalam mengoptimalkan produksi hasil buminya.

Informasi klasifikasi yang dimaksud adalah mengelompokkan perkebunan di Indonesia berdasarkan produktivitas pada masing-masing daerahnya dengan tujuan untuk mengetahui bagaimana keadaan produktivitas hasil bumi pada masing-masih daerah di Indonesia, sehingga instansi terkait dapat melakukan pemantauan dan memberikan strategi khusus pada daerah-daerah yang produktivitasnya di bawah performa.

Clustering adalah proses pengelompokan data dari sekumpulan data yang tidak teratur dengan memperhatikan kemiripan antar satu data dengan yang lainnya. *Clustering* telah menjadi instrumen yang valid untuk memecahkan masalah kompleks ilmu komputer dan statistic (Amelio & Tagarelli, 2019). Proses *clustering* ini mencoba mengumpulkan data serupa hanya ke dalam satu grup dan data tersebut tidak diizinkan untuk muncul di grup lain mana pun (Al., 2019).

Penelitian ini menggunakan metode k-means di mana algoritma ini merupakan salah satu algoritma yang paling banyak diminati dalam metode clustering dan termasuk ke dalam algortima yang mudah diimplementasikan. Terlepas dari hal itu, algoritma k-means ini sendiri tentu memiliki kelemahan diantaranya memiliki sifat yang sensitif dalam menentukan jumlah cluster (k) awal dan solusi akhir menyatu pada local minima.

Analisa menggunakan metode *cluster* ini dulunya hanya familiar di sektor kesehatan saja. Namun, dengan semakin berkembangnya zaman, metode analisis ini juga mulai dilirik dan diimplementasikan pada sektor-sektor lain, salah satunya adalah sektor pendidikan (Kusuma & Aryati, 2019). Bahkan yang menarik lagi, Budi Setiyono (Dwiarni & Setiyono, 2020) menggunakan metode ini untuk menganalisa kebiasaan orang-orang dalam menggunakan sosial media. Hal ini membuktikan bahwa penelitian menggunakan ini dapat membantu di banyak sektor, tak terkecuali di sektor perkebunan. Manfaat dari penelitian ini adalah untuk membantu mengelola dan menentukan kebijakan apa yang harus dipakai untuk memaksimalkan daerah yang memiliki produktivitas perkebunan yang bagus sehingga dapat menyeimbangkan antara daerah penyedia dengan daerah yang mengkonsumsi.

Metode *clustering* menggunakan algoritma k-means sudah sangat banyak digunakan. Salah satunya adalah penggunaan metode *k-means* dalam membantu mengklasifikasikan kelas unggulan di salah satu sekolah (Satria & Anggrawan, 2021). Selain itu juga Poerwanto (Poerwanto & Fa'rifah, 2019) mengelompokkan kecamatan di Tana Luwu berdasarkan produktivitas hasil pertaniannya dengan menggunakan bantuan algoritma ini.

Penerapan algoritma *k-means* dan algoritma *k-medoids* juga digunakan oleh Ninda Nurul Rhamadani, dkk untuk memudahkan peserta didik dalam mendapatkan informasi kategori sekolah (Ninda et al., 2020). Penelitian yang dilakukan oleh Dini Marlina, dkk yang melakukan pengelompokan pada data sebaran anak cacat yang ada pada Provinsi Riau menjadi tiga *cluster* menggunakan algortima yang sama (Marlina et al., 2018). Castaka Agus Sugianto, dkk melakukan penelitian untuk mengklasifikasi data kemiskinan di sebuah provinsi di Indonesia dengan metode *k-means* (Sugianto Castaka et al., 2021).

Agustina Heryati, dkk meneliti pengelompokan pada data mahasiswa baru (Heryati & Herdiansyah, 2020) menggunakan metode *k-means*. Penelitian yang dilakukan oleh Nabila Amalia Khairani, dkk meenerapkan metode *clustering k-means* untuk klasifikasi daerah rawan hotspot (Khairani & Sutoyo, 2020). Henrique Jos de Paula Alves, dkk melakukan penelitian terhadap klasifikasi efektivitas dari biaya intervensi pada kesehatan masyarakat (Alves et al., 2020). Sama halnya dengan Nayuni Dwitri, dkk yang melakukan penelitian untuk menetapkan klasifikasi penyebaran Covid-19 (Sianipar et al., 2020). Ali Mahmudan melakukan penelitian terhadap klasifikasi kabupaten/kota di sebuah provinsi di Indonesia

berdasarkan jumlah kasus covid-19 yang terjadi (Mahmudan, 2020).

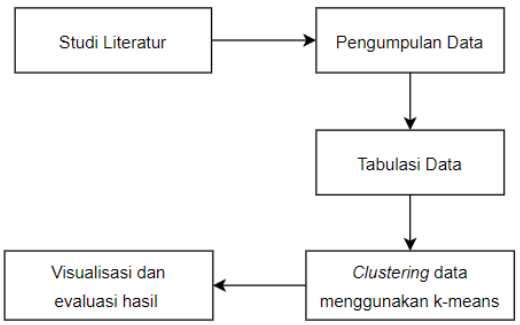
Mayoritas perbedaan dari penelitian-penelitian sebelumnya dengan penelitian ini adalah terletak pada objek yang dikelompokkan dan jumlah *cluster* yang dibuat.

Penelitian ini terbagi menjadi beberapa tahapan proses dalam melakukan klasifikasi kelompok perkebunan pada masing-masing daerah dan proses visualisasi. Pada proses visualisasi juga ada perhitungan growth year on year dari produktivitas untuk masing-masing daerah yaitu perbandingan pertumbuhan tahun berjalan dengan tahun sebelumnya.

2. METODE PENELITIAN

2.1 TAHAPAN PENELITIAN

Sebuah penelitian pasti melalui satu demi satu proses atau tahapan supaya penelitian tersebut dapat berjalan dengan baik, tak terkecuali pada penelitian ini. Langkah-langkah proses yang dilakukan pada penelitian ini dapat digambarkan sebagai berikut:



Gambar 1. Tahapan Penetitian

Dari gambar di atas dapat dijelaskan tahapan dari penelitian, diantaranya.

a. Studi literatur merupakan tahapan pertama dalam memulai penelitian, di mana pada tahap ini peneliti mengumpulkan, mencari, serta menganalisa teori, data dan metode yang relevan dari berbagai sumber yang akan digunakan pada penelitian.
b. Pengumpulan data, teknik pengumpulan data yang dipakai pada penelitian ini adalah data sekunder. Di mana peneliti mendapatkan sumber data dari *website* instansi Kementerian Pertanian Republik Indonesia. Data tersebut berupa catatan report tahunan produktivitas sektor perkebunan rakyat, perkebunan negara, dan perkebunan swasta dari seluruh provinsi di Indonesia.
c. Tabulasi data, pada tahap ini dilakukan proses penginputan data ke dalam bentuk tabel dalam hal ini menginputkan data ke dalam *database* setelah semua data sudah terkumpul. Jika data masih dalam format *non-numeric* maka perlu dilakukan proses inisiasi ke dalam bentuk *numeric*. Namun jika data sudah dalam bentuk *numeric* maka tidak diperlukan inisiasi (Andrean et al., 2019).
d. *Clustering* data menggunakan metode *k-means* merupakan proses pengolahan data yang sebelumnya tak beraturan dan tidak memiliki *pattern* menjadi beberapa kelompok data yang memiliki arti dan informasi. Kelompok tersebut diantaranya adalak kelompok *great, good,* dan *underperformed.* Di mana ketiga kelompok tersebut me-representasikan kondisi produktivitas pada masing-masing sektor perkebunan dan daerah di Indonesia. Detail dari proses ini bisa dilihat di point **D. Algoritma K-Means.**
e. Visualisasi dan evaluasi hasil, tahap ini adalah menampilkan hasil dari proses-proses sebelumnya. Pada tahap ini juga peneliti melakukan analisa dan evaluasi terhadap hasil *clustering* apakah sudah sesuai atau tidak.

2.2 SEKTOR PERKEBUNAN DI INDONESIA

Perkebunan adalah salah satu sektor andalan dari pendapatan nasional Indonesia dan penyumbang devisa terbesar negara. Secara status pengusahaannya, perkebunan di Indonesia dibagi ke dalam tiga:

1) Perkebunan Rakyat, yaitu usaha perkebunan yang dikelola oleh perorangan dan tidak berbadan hokum.
2) Perkebunan Besar Negara, yaitu usaha perkebunan di bawah pengelolaan negara dalam hal ini pemerintah.
3) Perkebunan Besar Swasta, yaitu usaha perkebunan di bawah pengelolaan perusahaan di luar pemerintah dan berbadan hukum.

2.3 DATA MINING

Data Mining merupakan suatu proses yang menguraikan pencarian atau penemuan informasi menarik di dalam suatu *database* atau *raw data* dengan memakai metode atau teknik tertentu. Proses yang terjadi di dalam *data mining* ini sendiri menggunakan beberapa metode diantaranya teknik statistik, konsep matematika, kecerdasan buatan, dan *machine learning* dalam menemukan dan menguraikan informasi yang berguna untuk pengetahuan terkait dengan berbagai *database* berkapasitas besar.

Output dari proses *data mining* ini nantinya mampu dijadikan sebagai bekal untuk mengidentifikasi dan menentukan langkah di masa depan. Beberapa teknik yang populer digunakan pada proses *data mining* diantaranya adalah asosiasi, estimasi, prediksi, *clustering* serta klasifikasi. Pelaksanaan *data mining* dalam penanganan dan pengolahan data pada sebuah kasus, akan sangat membantu dalam meningkatkan daya tarik digital serta dapat mengasilkan informasi yang lebih bermakna (Yunita, 2018).

2.4 ALGORITMA K-MEANS

K-means adalah algoritma *clustering* yang sangat sederhana yang membagi kelompok data menjadi beberapa *cluster k*. *K-means* merupakan salah satu algoritma pelatihan *unsupervise*, awal mula algoritma ini dipublikasikan oleh Stuart Lloyd di tahun 1984 yang hingga saat ini merupakan alogiritma yang banyak digunakan. Alasan algoritma ini banyak digunakan adalah karena *k-means* cukup mudah untuk diimplementasikan, relatif cepat, dan mudah dimodifikasi. Algoritma *Clustering K-Means* dapat dengan mudah diimplementasikan dengan Algoritma Genetika (Hussein, 2018).

Tahapan proses pada algoritma *K-means* dapat dijelaskan sebagai berikut (Larose, 2005; Prasetyo, 2012; Khotimah, 2014; Parlina et al., 2018; Primatha, 2018):

1. Tahap pertama adalah menentukan jumlah *cluster* (k) atau kelompok pada suatu *dataset*
2. Selanjutnya mentukan nilai pusat (*centroid*). Penentuan nilai awal *centroid* dilakukan secara acak. Setelah itu ketika tahap perulangan, rumus perhitungan yang dipakai seperti persamaan di bawah ini:

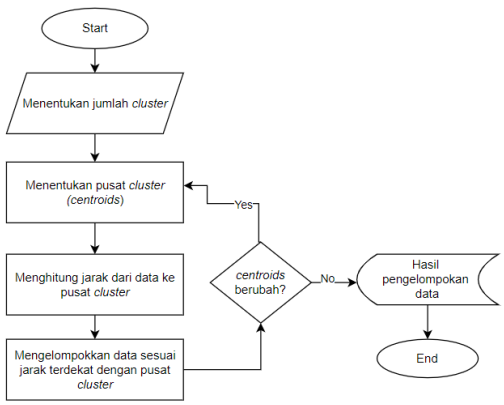
$$V_{ij} = \frac{1}{Ni} + \sum_{k=0}^{Ni} Xkj$$

Keterangan:

V_{ij}	:	*Centroid* / nilai rata-rata *cluster* indeks ke-i pada *variable* indeks ke-j
N_i	:	Jumlah data *cluster* indeks ke-i
i, k	:	Indeks atau penanda pada *cluster*
j	:	Indeks atau penanda pada *variable*
Xkj	:	Nilai data indeks k *variable* indeks ke-j pada *cluster* tersebut

3. Untuk setiap *row data*, hitung jarak minimum data tersebut dengan *centroid*. Terdapat beberapa metode perhitungan yang dapat digunakan dalam mengukur jarak minimum data ke *centroid*, diantaranya adalah metode *Euclidean*, *Manhattan / City Block*, serta *Minkowsky*.
4. Mengelompokkan objek data berdasarkan hitungan jarak ke *centroid* dengan jarak hitungan minimum terdekat.
5. Kembali ke tahapan nomor 2 hingga tahapan nomor 4, lakukan perulangan sampai didapat centroid yang bernilai optimal.

Dari langkah-langkah di atas, dapat divisualisasikan dalam bentuk *flow* proses algoritma *k-means* ini bekerja adalah sebagai berikut:



Gambar 2. Flowchart Algoritma K-Means

Gambar di bawah ini merupakan ilustrasi dari proses yang dilakukan dan didapatkan oleh algoritma *k-means*:

Gambar 3. Ilustrasi Proses *K-Means Clustering*

Dari gambar tersebut, dapat disimpulkan bahwa tujuan pada penelitian ini adalah untuk mengklasifikan *raw data* menjadi kelompok-kelompok data yang bisa dianggap sebagai informasi yang lebih dapat diolah menjadi sesuatu yang lebih bermanfaat.

2.5 TABLEAU

Tableau adalah aplikasi *business intelligence* yang digunakan dalam memvisualisasikan data dan mengolah data menjadi sebuah informasi yang lebih interaktif. Beberapa kelebihan dari fitur tableau adalah sebagai berikut:

1. Memiliki kecepatan analisis. Pengguna tidak diwajibkan untuk mahir dalam Bahasa pemrograman tertentu atau ahli *coding*.
2. Tidak bergantung pada aplikasi lain. Memudahkan pengguna supaya tidak perlu untuk mengolah data di banyak aplikasi.
3. Menyajikan dalam bentuk visualisasi yang interaktif. Tableau menyediakan berbagai macam fitur yang bisa menghasilkan visualisasi yang menarik dan interaktif.
4. Dapat berkolaborasi secara *real-time.* Artinya aplikasi ini memungkinkan pengguna untuk menggunakan berbagai sumber data dalam satu waktu yang sama.
5. Mengimpor berbagai ukuran dan *range data*, semua ukuran dan *range data* bisa diolah di Tableau.

2.6 ETL (Extract Transform Load)

ETL merupakan akronim dari kata *Extract, Transform,* dan *Load*. ETL merupakan proses integrasi data yang memadukan data dari berbagai sumber ke dalam satu repositori yang konsisten dan dimasukkan ke dalam *database* atau sistem lain. Pada awal munculnya ETL di tahun 1970, ETL diperkenalkan dengan tujuan untuk mengumpulkan atau mengintegrasikan proses pemuatan data ke dalam superkomputer yang selanjutnya akan dianalisis lebih dalam lagi. Sejak akhir 1980-an hingga pertengahan tahun 2000, ETL merangkak menjadi proses primer yang diperuntukkan untuk membangun gudang data sehingga bisa mendukung aplikasi *Business Intelligence* (BI).

Tiga tahapan proses yang digambarkan pada ETL adalah sebagai berikut:

1) *Extract*

Ini merupakan tahapan pertama dari proses yang melibatkan ekstraksi data dari sumber data yang sesuai. Sumber data yang digunakan bisa berbentuk *flat file* dengan ekstensi seperti csv, xls, dan atau txt.

2) *Transform*

Pada tahapan ini dilakukan proses cleansing sesuai dengan target skema yang sudah disediakan. Proses-proses tersebut bisa berupa normalisasi data, pengecekan duplikasi, filtering data, pengurutan serta pengelompokan data.

3) *Load.*

Dan terakhir adalah tahapan load data. Pada proses ini, data yang sudah melalui proses transformasi sebelumnya akan disimpan ke dalam wadah seperti table yang sudah disiapkan pada data *warehouse*.

3. HASIL DAN PEMBAHASAN

Setelah rangkaian pengumpulan dan tabulasi data sebelumnya, berikut ini dijelaskan data perkebunan apa saja yang dipakai pada penelitian ini.

Table 1. Daftar Data Perkebunan

No	Perkebunan
1	Teh
2	Tembakau
3	Tebu
4	Kakao
5	Cengkeh
6	Lada
7	Kopi
8	Karet

9	Kelapa sawit
10	Kelapa
11	Jambu mete
12	Kapas
13	Nilam
14	Pala
15.	Sagu

Proses pengklasifikasian sektor perkebunan di daerah-daerah Indonesia pada penelitian ini dikelompokkan berdasarkan *variable* produktivitas pembentuk kelompok. *Variable* tersebut tersedia datanya dalam basis *yearly* atau tahunan. Pengelompokkan ini juga dibagi menjadi dua tahap, yang pertama adalah pengelompokkan daerah untuk per sektor perkebunan, dan yang kedua adalah pengelompokkan daerah untuk semua sektor perkebunan.

Tahapan proses pengujian algoritma *K-Means* adalah seperti berikut:

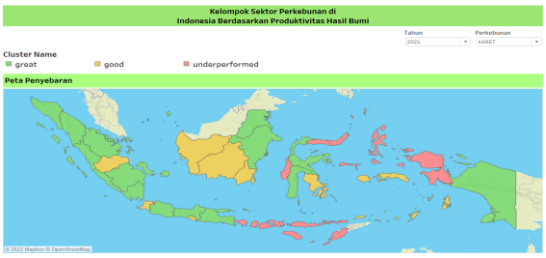
1) Membuat *cluster* untuk pengelompokkan data, diantaranya:
   a. *Cluster great*, untuk kelompok daerah yang memiliki produktivitas bagus.
   b. *Cluster good*, untuk kelompok daerah yang memiliki produktivitas cukup baik.
   c. *Cluster underperformed*, untuk kelompok daerah yang memiliki produktivitas di bawah target.
2) Menetapkan *centroids* (nilai pusat) bagi masing-masing *cluster*. Penentuan nilai ini dilakukan acak bedasarkan tabel data yang ada (Muliono & Sembiring, 2019).

Table 2. Data Centorid awal yang dipilih secara acak

Centroids	Nilai
C. *Great*	13,000
C. *Good*	5,000
C. *Underperformed*	3,900

3) Proses pengelompokkan data ke dalam tiga *cluster* menggunakan algoritma k-means.
4) Visualisasi menggunakan *tools* Tableau untuk menampilkan hasil dari pengelompokkan sektor perkebunan pada daerah-daerah di Indonesia.

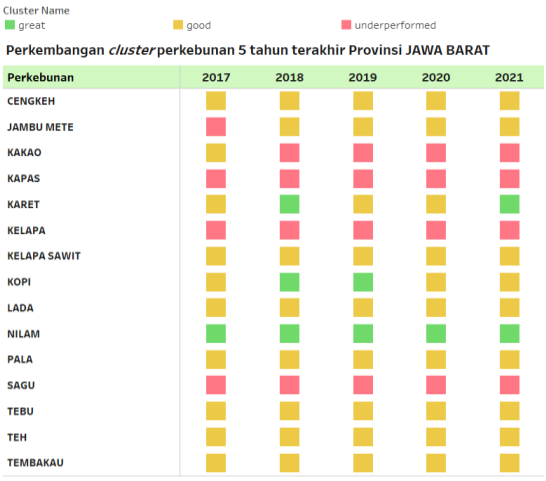
Pada gambar 4 dan gambar 5 berikut ini adalah gambaran dari hasil pengelompokan data perkebunan pada daerah-daerah di Indonesia yang divisualisasikan dalam bentuk peta penyebaran serta bentuk tabel yang berisi informasi tentang masing-masing daerah berada ke dalam *cluster* apa untuk tahun dan perkebunan yang terfilter.



Gambar 4. Informasi Peta Sebaran Kelompok Daerah untuk Setiap Perkebunan

Gambar 5. Visualisasi Hasil *Clustering*

Selanjutnya untuk gambar 6 adalah hasil visualisasi dari kelompok produktivitas semua perkebunan selama lima tahun terakhir untuk masing-masing daerah terpilih. Dengan kata lain, tampilan ini merupakan hasil lebih detail dari visualisasi pada gambar 4 dan gambar 5.



Gambar 6. Informasi produktivitas 5 tahun terakhir

Tahun: 2021 | Perkebunan: KARET

Tabel Produktivitas *Year on Year*

Provinsi	2020	2021	YoY (%)
BALI	0	286	28600.00%
KALIMANTAN UTARA	433	1,121	158.89%
SULAWESI SELATAN	653	1,144	75.19%
SULAWESI TENGGARA	358	624	74.30%
MALUKU	374	619	65.51%
PAPUA	880	1,157	31.48%
SULAWESI TENGAH	860	1,068	24.19%
ACEH	967	1,073	10.96%
RIAU	992	1,066	7.46%
JAWA BARAT	968	1,032	6.61%
KEPULAUAN RIAU	1,040	1,098	5.58%
BANTEN	912	957	4.93%
DI YOGYAKARTA	629	650	3.34%
LAMPUNG	963	990	2.80%
SUMATERA BARAT	1,103	1,125	1.99%
SUMATERA UTARA	1,030	1,047	1.65%
JAMBI	862	876	1.62%
JAWA TENGAH	1,196	1,214	1.51%
JAWA TIMUR	1,253	1,267	1.12%
KEPULAUAN BANGKA BELITUNG	1,167	1,176	0.77%
KALIMANTAN TIMUR	1,189	1,195	0.50%
KALIMANTAN SELATAN	956	960	0.42%
KALIMANTAN BARAT	913	916	0.33%
BENGKULU	1,196	1,199	0.25%
KALIMANTAN TENGAH	806	808	0.25%
SUMATERA SELATAN	1,153	1,154	0.09%
SULAWESI UTARA	0	0	0.00%
SULAWESI BARAT	0	0	0.00%
PAPUA BARAT	0	0	0.00%
NUSA TENGGARA TIMUR	0	0	0.00%
NUSA TENGGARA BARAT	0	0	0.00%
MALUKU UTARA	0	0	0.00%
GORONTALO	0	0	0.00%
DKI JAKARTA	0	0	0.00%

Gambar 7 . Informasi *Growth* produktivitas

Gambar di atas ini menampilkan informasi dari *growth* produktivitas pada masing-masing daerah dalam basis *yearly* untuk tahun dan perkebunan terpilih. Di mana untuk mendapatkan *growth year on year* adalah sebagi berikut:

$$YoY = \frac{Pn-(Pn-1)}{(Pn-1)} \text{ x } 100\%$$

Keterangan:
P_n : Produktivitas di tahun terpilih
P_{n-1} : Produktivitas di tahun sebelumnya

Perhitungan nilai *year on year* pada proses visualisasi ini bertujuan untuk rekapitulasi prosentase *growth* produktivitas dari masing-masing daerah setiap tahunnya.

Selain informasi pengelompokkan daerah di Indonesia untuk setiap perkebunan, pada gambar 8 ini juga menunjukkan informasi kelompok daerah berdasarkan produktivitas semua hasil perkebunan. Dari gambar 8 tersebut dapat terlihat kelompok *cluster "good"* mendominasi penyebaran dengan jumlah daerah yang mendapat *cluster* ini sebanyak 17 daerah.

Gambar 8. Peta Pembagian Cluster Secara Keseluruhan

Rincian informasi dari kelompok setiap daerah berdasarkan gambar di atas beserta hasil perhitungan jarak *euclidean* dengan menggunakan fungsi pada bahasa pemrograman *python* adalah sebagai berikut:

Table 3. Rincian Kelompok Daerah

No	Daerah	Produktivitas	Jarak *Euclidean*	*Cluster*
1.	Aceh	10440	1333.95	Good
2.	Sumatera Utara	17343	2681.84	Great
3.	Sumatera Barat	15422	760.94	Great
4.	Riau	14264	397.44	Great
5.	Jambi	10503	1396.80	Good
6.	Sumatera Selatan	15706	1044.86	Great
7.	Bengkulu	10864	1757.65	Good
8.	Lampung	14437	224.42	Great
9.	Kepulauan Bangka Belitung	8642	466.00	Good
10.	Kepulauan Riau	9697	591.17	Good
11.	DKI Jakarta	0	4250.11	Under-performed
12.	Jawa Barat	14236	425.21	Great
13.	Jawa Tengah	12626	2035.20	Great
14.	DI Yogyakarta	12118	2543.20	Great
15.	Jawa Timur	14456	205.31	Great
16.	Banten	6427	2177.15	Under-performed
17.	Bali	5242	992.06	Under-performed
18.	Nusa Tenggara Barat	6876	2230.72	Good
19.	Nusa Tenggara Timur	5043	793.05	Under-performed
20.	Kalimantan Barat	11265	2158.29	Good
21.	Kalimantan Tengah	8676	430.80	Good
22.	Kalimantan Selatan	9963	856.32	Good
23.	Kalimantan Timur	8058	1048.77	Good

24.	Kalimantan Utara	5626	1376.00	Under-performed
25.	Sulawesi Utara	8744	363.14	Good
26.	Sulawesi Tengah	9383	276.89	Good
27.	Sulawesi Selatan	16004	1343.55	Great
28.	Sulawesi Tenggara	7031	2075.82	Good
29.	Gorontalo	10731	1624.41	Good
30.	Sulawesi Barat	8825	282.61	Good
31.	Maluku	6936	2170.89	Good
32.	Maluku Utara	4371	122.61	Under-performed
33.	Papua Barat	5218	968.42	Under-performed
34.	Papua	10860	1753.74	Good

Dari table di atas terdapat informasi sekitar 10 daerah yang termasuk ke dalam *cluster great* dengan produktivitas diantara 12,000 – 17,500 kg/Ha hasil perkebunan. Sedangkan 17 daerah termasuk ke dalaman *cluster good* yang memiliki produktivitas hasil perkebunan diantara 6,500 – 11,500 kg/Ha. Dan 7 daerah termasuk ke dalam *cluster underperformed* dengan produktivitas hasil pertanian diantara 0 – 6,400 kg/Ha.

Berikut ini adalah *range* produktivitas perkebunan untuk masing-masing *cluster* yang diprediksi sebelum melakukan *clustering* menggunakan *k-means*:

Table 4. Prediksi Range Produktivitas

Cluster	**Produktivitas (Kg/Ha)**
Great	10,000 – 17,000
Good	6,000 – 10,000
Underperformed	0 – 5,000

Berdasarkan hasil tersebut, ada sekitar 7 daerah yang kurang sesuai penempatan kelompoknya, sehingga dapat dihitung tingkat akurasi dari algoritma *k-means* pada penelitian ini dengan menggunakan rumus (Sari & Chotijah, 2022) berikut:

$$akurasi = \frac{Jumlah\ data\ yang\ sesuai}{Jumlah\ keseluruhan\ data} \text{ x } 100\%$$

$$akurasi = \frac{27}{34} \text{ x } 100\%$$

$$akurasi = 79.41\%$$

4. SIMPULAN

Berdasarkan hasil evaluasi terhadap percobaan pembentukan *clustering* pada data sektor perkebunan di Indonesia maka didapat kesimpulan sebagai berikut:

a. Penentuan nilai *centroids* (nilai pusat) awal data sangat berpengaruh terhadap keberhasilan pengelompokkan data menggunakan algoritma *k-means.*

b. Proses berapa kali melakukan iterasi juga berpengaruh terhadap kesesuaian hasil *clustering data.*

c. Algoritma *K-Means Clustering* dapat digunakan sebagai dasar untuk memilih data potensial dari data yang memiliki dimensi tinggi.

Pada penelitian ini menunjukkan akurasi yang mencapai 79.41% dengan menghasilkan 3 klasifikasi *cluster* dari masing-masing jenis perkebunan untuk daerah-daerah di Indonesia. Ketiga *cluster* tersebut adalah *Great, Good,* dan *Underperformed*. Dari 34 daerah dapat diklasifikasikan menjadi tiga *cluster* dengan jumlah masing-masing *cluster* sebagai berikut: *great* sebanyak 10 daerah; *good* sebanyak 17 daerah; dan *underperformed* sebanyak 7 daerah. Namun, penelitian ini juga masih terdapat kekurangan yaitu belum adanya perbandingan dengan proses pengolahan data menggunakan algoritma lain sehingga penentuan tingkat akurasi akan lebih baik lagi serta mengenai sumber data daerah yang belum terlalu terlihat sampai level detailnya.

Maka pada penelitian selanjutnya, disarankan untuk proses mengelompokkan *cluster* daerah menggunakan metode *k-means* dapat menambahkan data daerah dengan level lebih detail lagi sehingga monitoring terhadap distribusi perkebunan di Indonesia lebih maksimal. Penggunaan metode *clustering* pun dapat dimodifikasi dengan algoritma yang lebih kompleks dan lebih baik lagi. Dari segi metode *clustering* dapat ditambahkan hasil *compare* dengan metode serupa lain seperti *k-medoids, k-means* dengan *purity,* atau bahkan *dynamic k-means* untuk hasil yang lebih optimal.

5. DAFTAR PUSTAKA

Al., R. T. S. et. (2019). *Cluster analysis identifying clinical phenotypes of preterm birth and related maternal*

and neonatal outcomes from the Brazilian multicentre study on preterm birth. *146*(1), 110–117.

Alves, H. J. de P., Fernandes, F. A., Lima, K. P. de, Batista, B. D. de O., & Fernandes, T. J. (2020). A pandemia da COVID-19 no Brasil: uma aplicação do método de clusterização k-means. *Research, Society and Development*, *9*(10), e5829109059. https://doi.org/10.33448/rsd-v9i10.9059

Amelio, A., & Tagarelli, A. (2019). Data Mining: Clustering. In *Encyclopedia of Bioinformatics and Computational Biology* (pp. 437–448). Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.20489-5

Andrean, R., Fendy, S., & Nugroho, A. (2019). Klasterisasi Pengendalian Persediaan Aki Menggunakan Metode K-Means. *JOINTECS (Journal of Information Technology and Computer Science)*, *4*(1), 5. https://doi.org/10.31328/jointecs.v4i1.998

Darmawan, Didit; Genua, Veronika; Kristianto, Sonny; Murdaningsih; Hutubessy, J. (2019). *Tanaman Perkebunan Prospektif Indonesia*. CV. Penerbit Qiara Media.

Dwiarni, B. A., & Setiyono, B. (2020). Akuisisi dan Clustering Data Sosial Media Menggunakan Algoritma K-Means sebagai Dasar untuk Mengetahui Profil Pengguna. *Jurnal Sains Dan Seni ITS*, *8*(2). https://doi.org/10.12962/j23373520.v8i2.49815

Heryati, A., & Herdiansyah, M. I. (2020). The Application of Data Mining by Using K-Means Clustering Method in Determining New Students' Admission Promotion Strategy. *International Journal of Engineering and Advanced Technology*, *9*(3), 824–833.

Hussein, A. A. (2018). Improve The Performance of K-means by using Genetic Algorithm for Classification Heart Attack. *International Journal of Electrical and Computer Engineering (IJECE)*, *8*(2), 1256. https://doi.org/10.11591/ijece.v8i2.pp1256-1261

Khairani, N. A., & Sutoyo, E. (2020). Application of K-Means Clustering Algorithm for Determination of Fire-Prone Areas Utilizing Hotspots in West Kalimantan Province. *International Journal of Advances in Data and Information Systems*, *1*(1), 9–16.

Kusuma, A. S., & Aryati, K. . S. (2019). Sistem Informasi Akademik Serta Penentuan Kelas Unggulan Dengan Metode Clusttering Dengan Algoritama K-Means Di Smp Negeri 3 Ubud. *J. Sist. Inf. Dan Komput. Terap. Indonesia*, *1*(3), 143–152.

Mahmudan, A. (2020). Clustering of District or City in Central Java Based COVID-19 Case Using K-Means Clustering. *Jurnal Matematika, Statistika Dan Komputasi*, *17*(1), 1–13. https://doi.org/10.20956/jmsk.v17i1.10727

Marlina, D., Lina, N., Fernando, A., & Ramadhan, A. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, *4*(2), 64. https://doi.org/10.24014/coreit.v4i2.4498

Muliono, R., & Sembiring, Z. (2019). Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen. *Journal of Computer Engineering, System and Science*, *4*(2), 2502–714.

Ninda, R., Ahmad, F., Euis, N., & Pratama, A. R. (2020). Implementasi algoritma k-means dan k-medoids dalam pengelompokan nilai ujian nasional tingkat smk. *Ciastech*, 717–726.

Poerwanto, B., & Fa'rifah, R. Y. (2019). Algoritma k-means dalam mengelompokkan kecamatan di tana luwu berdasarkan produktifitas hasil pertanian. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699.

Sari, M. A. P., & Chotijah, U. (2022). Pengelompokan Anggota Divisi Himpunan Mahasiswa Jurusan pada Universitas XYZ dengan Metode K-Means Clustering. *ANTIVIRUS: Jurnal Ilmiah Teknik Informatika*, *16*(1), 52–62.

Satria, C., & Anggrawan, A. (2021). Aplikasi K-Means berbasis Web untuk Klasifikasi Kelas Unggulan. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(1), 111–124. https://doi.org/10.30812/matrik.v21i1.1473

Sianipar, K. D. R., Siahaan, S. W., Siregar, M., R.H Zer, F. I., & Hartama, D. (2020). Penerapan Algoritma K-Means Dalam Menentukan Tingkat Kepuasan Pembelajaran Online Pada Masa Pandemi Covid-19. *Jurnal Teknologi Informasi*, *4*(1), 101–105. https://doi.org/10.36294/jurti.v4i1.1258

Sugianto Castaka, A., Pratiwi, T., & Riska, O. (2021). K-Means Algorithm For Clustering Poverty Data in Bangka Belitung Island Province Journal of Computer Networks, Architecture and High Performance Computing. *Journal of Computer Networks, Architecture and High Performance Computing*, *3*(1), 58–67.

Yunita, F. (2018). Penerapan Data Mining Menggunkan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru. *SISTEMASI*, *7*(3), 238. https://doi.org/10.32520/stmsi.v7i3.388