

## Optimalisasi K-Means Berbasis Particle Swarm Optimization untuk Hasil Produksi Tanaman Sayuran di Indonesia

Budiman<sup>\*1</sup>, Titan Parama Yoga<sup>2</sup>

<sup>1</sup>Informatika, Fakultas Teknologi dan Informatika

<sup>2</sup>Sistem Informasi, Fakultas Teknologi dan Informatika

Universitas Informatika dan Bisnis Indonesia

E-mail: <sup>\*1</sup>[budiman1982@gmail.com](mailto:budiman1982@gmail.com), <sup>2</sup>[titanparama@unibi.ac.id](mailto:titanparama@unibi.ac.id)

### Abstrak

*The agricultural sector plays a vital role in meeting the community's basic needs, thereby it can increase economic, social, and trade fields. The Indonesian Central Statistics Agency in September 2020 recorded a population of 270.20 million people. The population increases every year, along with the increase of basic needs. Meanwhile, the production of the agricultural sector does not increase every year. This study aims to classify areas that have a high level of vegetable production or a low level of vegetable production. The application of K-Means clustering makes it possible to optimize K-Means clustering with PSO to produce higher quality information. The results of the PSO-based K-Means clustering modeling for vegetable crop production in Indonesia in cluster 0 (medium) are 8.82%, cluster 1 (low) was 88.2%, and cluster 2 (high) was 2.94%. West Java Province is in cluster 2, so West Java Province dominates the production of vegetable crops in Indonesia. The test results based on the average silhouette coefficient, PSO-based K-Means, is better than pure K-Means even though these values are not significantly different, namely the average silhouette coefficient value of 0.714114. The average computation time for pure K-Means is faster, with an average of 0.748 seconds, compared to PSO-based K-Means. The time required is longer because PSO-based K-Means improves data clusters from K-Means clustering results.*

**Keywords:** *K-Means, Particle Swarm Optimization, Clustering, Vegetable Crops, Agriculture*

### Abstract

Sektor pertanian memegang peranan penting dalam memenuhi kebutuhan pokok masyarakat, dengan demikian akan mendorong laju perekonomian, sosial, dan perdagangan. BPS pada September 2020 mencatat jumlah penduduk sebesar 270,20 juta jiwa. Bertambahnya jumlah penduduk pada setiap tahun, berbanding lurus dengan permintaan kebutuhan pokok yang semakin meningkat. Sedangkan produksi sektor pertanian tidak setiap tahun meningkat. Penelitian ini bertujuan untuk mengelompokkan daerah-daerah yang memiliki tingkat hasil produksi tanaman sayuran yang tinggi atau tingkat hasil produksi tanaman sayuran yang masih rendah. Penerapan *K-Means clustering* memungkinkan untuk mengoptimasi *K-Means clustering* dengan PSO sehingga dihasilkan informasi yang lebih bermutu. Hasil pemodelan *K-Means clustering* berbasis PSO untuk hasil produksi tanaman sayuran di Indonesia *cluster 0* (sedang) sebanyak 8,82%, *cluster 1* (rendah) sebanyak 88,2%, dan *cluster 2* (tinggi) sebanyak 2,94%. Provinsi Jawa Barat berada dalam *cluster 2*, sehingga Provinsi Jawa Barat mendominasi dalam produksi tanaman sayuran di Indonesia. Hasil pengujian berdasarkan rata-rata *silhouette coefficient*, *K-Means* berbasis PSO lebih baik daripada *K-Means* murni meskipun nilai tersebut tidak berbeda signifikan yaitu dengan rata-rata nilai *silhouette coefficient* sebesar 0,714114. Rata-rata waktu komputasi waktu komputasi *K-Means* murni lebih cepat dengan rata-rata 0,748 detik dibandingkan *K-Means* berbasis PSO. Waktu yang dibutuhkan lebih lama karena *K-Means* berbasis PSO memperbaiki kelompok data dari hasil bentuk *K-Means clustering*.

**Kata kunci:** *K-Means, Particle Swarm Optimization, Clustering, Tanaman Sayuran, Pertanian*

## 1. PENDAHULUAN

*Big Data* mengacu pada sekumpulan data tidak terstruktur maupun terstruktur yang memiliki volume sangat besar. Data tersebut diperoleh dari berbagai sumber seperti *web*, organisasi bisnis, dan lain-lain. *Data science* merupakan ilmu multidisiplin tentang penggalan pengetahuan dari satu set data dan informasi. *Data science* terdiri dari kombinasi berbagai topik dan dibangun di atas dasar dan metode yang ada di berbagai bidang ilmiah. Beberapa bidang tersebut antara lain matematika, statistika, rekayasa data, pengenalan pola, dan lain-lain. Tujuan dari ilmu ini adalah untuk mengekstrak konsep dari data dan mengembangkan produk berbasis data.

*Data mining* dikenal sebagai *knowledge discovery in data* (KDD) merupakan proses mengungkap pola dan informasi yang berharga dari kumpulan data besar. Mengingat evolusi teknologi *data warehouse* dan pertumbuhan data besar, adopsi teknik data mining telah meningkat pesat selama beberapa dekade terakhir, membantu instansi dengan mengubah data mentah menjadi pengetahuan yang berguna. Namun demikian, terlepas dari kenyataan bahwa teknologi terus berkembang untuk menangani data dalam skala besar, para pemimpin masih menghadapi tantangan dengan skalabilitas dan otomatisasi. *Data mining* telah meningkatkan pengambilan keputusan organisasi melalui analisis data yang mendalam.

Indonesia merupakan negara agraris karena sebagian besar penduduk Indonesia bekerja di sektor pertanian. Indonesia memiliki lahan yang luas dan berbagai sumber daya alam yang melimpah. Sektor pertanian memegang peranan penting dalam memenuhi kebutuhan pokok masyarakat, dengan demikian akan mendorong laju perekonomian, sosial, dan perdagangan. Berdasarkan hasil Sensus Penduduk (SP2020) yang dilakukan BPS pada September 2020 mencatat jumlah penduduk sebesar 270,20 juta jiwa[1]. Bertambahnya jumlah penduduk pada setiap tahun, berbanding lurus dengan permintaan

kebutuhan pokok yang semakin meningkat. Sedangkan produksi sektor pertanian tidak setiap tahun meningkat. BPS mencatat jumlah petani per 2019 mencapai 33,4 juta orang. Adapun dari jumlah tersebut, petani muda di Indonesia yang berusia 20-39 tahun hanya 8% atau setara dengan 2,7 juta orang. Kemudian, sekitar 30,4 juta orang atau 91% berusia di atas 40 tahun, dengan mayoritas usia mendekati 50-60 tahun. Kondisi ini kian diperparah dengan penurunan jumlah regenerasi petani muda. Data yang sama dari periode 2017 ke 2018, penurunan jumlah petani muda mencapai 415.789 orang[2]. Hal ini berdampak langsung pada perekonomian dan kesejahteraan.

Permasalahan tersebut memerlukan upaya dalam pengolahan data historis, mengidentifikasi, dan mengelompokkan data sebagai informasi sehingga dapat memberikan rekomendasi kepada para pemangku kepentingan pada sektor pertanian dan masyarakat. Proses mengelompokkan data dapat dilakukan dengan menerapkan metode khusus untuk pada ilmu *data mining*, salah satunya adalah algoritma *K-Means clustering*. Algoritma *K-Means clustering* adalah algoritma yang mempartisi data ke dalam *cluster-cluster* sehingga data yang memiliki kesamaan berada pada satu kelas atau kelompok yang sama dan data yang memiliki ketidak samaan berada pada kelas yang lain. Jadi dapat disimpulkan bahwa metode ini adalah pengelompokan data yang sama ke dalam satu *cluster* atau kelompok[3]. Perbedaan dan persamaannya biasanya nilai atribut dari objek tersebut dan dapat juga berupa perhitungan jarak[4]. *K-Means clustering* belum optimal sehingga perlu menerapkan metode *Particle Swarm Optimization* (PSO). PSO adalah teknik pencarian dan optimasi[5]. PSO dapat dikombinasikan dengan partisi pengelompokan untuk mengatasi keterbatasan partisi *clustering*, dan kombinasi ini dapat menghasilkan kluster yang berkualitas tinggi. Optimasi ini terpercaya juga dalam memastikan pusat *cluster* baru berdasarkan total *cluster* yang ditentukan[6].

Penelitian dengan kajian *data mining* telah banyak dilakukan oleh para peneliti

dalam bidang pertanian, di bawah ini merupakan penelitian terdahulu dalam kurun waktu lima tahun terakhir terkait clustering dan teknik optimasi berbasis *Particle Swarm Optimization*.

Herna Mahulae. Peneliti melakukan pengklasteran daerah potensial penghasil buah menggunakan algoritma *K-Means*. Dengan menggunakan *K-Means* bertujuan dalam memudahkan pengelompokan suatu daerah dalam mengetahui tingkat potensi berdasarkan kabupaten atau kota di Provinsi Sumatera Utara. Hasil penelitian menunjukkan Provinsi Sumatera setiap kabupaten atau kota mempunyai hasil potensi produksi buah-buahan yang bervariasi jumlahnya dari data luas panen (Ha) dan produksi (ton). Data yang digunakan adalah data hasil potensi dari 30 Kabupaten di Sumatera Utara dari tahun 2017 untuk data hasil produksi buah alpukat, duku, dan belimbing[4].

Ariska Dewi Susanti, dkk. Peneliti melakukan penelitian terkait analisis *non-hierarchical partitioning K-Medoid* pada produksi sektor hortikultura tahun 2019 di Indonesia. Hasil penelitian menunjukkan pada nilai *Sillhouette Coefficient* yang diperoleh hasil *cluster* secara keseluruhan memiliki kategori struktur yang buruk karena bernilai 0,24. Hal ini dikarenakan pada *cluster 1* struktur *cluster* yang terbentuk lemah dan pada *cluster 2* struktur *cluster* yang terbentuk buruk[7].

Hengky Andema, dkk. Peneliti menggunakan metode *K-Means Clustering* dengan jumlah 3 *Cluster* yaitu *Cluster 0* (C0) Kurang Berpotensi, *Cluster 1* (C1) Cukup Berpotensi, *Cluster 2* (C2) Sangat Berpotensi untuk penanaman kelapa hibrida. Hasil dari proses *clustering* dengan 2 kali iterasi menyatakan bahwa untuk *Cluster 0* berjumlah 7 data desa, untuk *Cluster 1* berjumlah 1 data desa, dan untuk *Cluster 2* berjumlah 2 data desa[8].

A Supriyatna, dkk. Hasil penelitian menunjukkan terdapat 19 provinsi yang termasuk dalam *cluster 0* (Sedang), terdapat 4 provinsi yang termasuk dalam *cluster 1* (Tinggi), dan terdapat 11 provinsi yang

termasuk dalam *cluster 2* (Rendah). Berdasarkan hasil penelitian, terbukti terdapat 4 provinsi dalam *cluster 1* (Tinggi) yaitu Jawa Barat, Jawa Tengah, Jawa Timur dan Sulawesi Selatan dengan produktivitas padi tertinggi. Tiga di antaranya berada di Pulau Jawa. Hal ini menunjukkan bahwa Jawa masih mendominasi produktivitas tanaman padi[9].

Syaiful Anam. Hasil percobaan yang dilakukan peneliti menunjukkan bahwa *Particle Swarm Optimization-K-means* (PSOK) memiliki performansi yang lebih baik dibandingkan algoritma *K-Means* dalam segmentasi penyakit bercak daun[10].

Darmansah dan Ni Wayan Wardani. Hasil dari pengujian terhadap metode *K-Means* terdapat tiga pengelompokan penyebab kerusakan pada tanaman yaitu *Cluster 0* untuk jenis cabai yang kerusakan sedang, *Cluster 1* kerusakan berat dan *Cluster 2* untuk kerusakan ringan. Kemudian hasil dari per *cluster* yaitu *Cluster 0* terdapat satu jenis hama yaitu Lalat Buah, *Cluster 1* terdapat 3 jenis hama yang terdiri dari Virus kuning, Antraknose dan Thrips, sedangkan *Cluster 2* terdapat 7 jenis hama yang terdiri dari Kutu Daun, Tungau, Layu Fusarium, Layu Bakeri, Virus Keriting, Mati Pucuk, Puru Buah. Analisa ini diharapkan memudahkan petani mengetahui penyebab kerusakan tanaman cabai atau dinas terkait bisa mengambil tindakan mengantisipasi penyebab penyebab kerusakan pada tanaman cabai secepat mungkin[3].

Shashi Mehrotra dan Aditi Sharan. Hasil eksperimen peneliti menunjukkan *recall* dan *F-measure* tertinggi dicapai oleh PSO. Meskipun presisi tertinggi dicapai oleh algoritma *K-Means*, sebagian besar PSO menunjukkan hasil yang lebih baik daripada algoritma *K-Means*[11].

Jemaictry Tamaela, dkk. Penelitian ini bertujuan untuk melakukan analisis *Cluster* dan implementasinya dengan menggunakan algoritma *Fuzzy C-Means* (FCM) dan *K-Means* (KM) untuk mengelola data pertanian dari hasil *data mining* yang dilakukan. *Fuzzy c-means* (FCM) dan *k-means* (KM)

dimplementasikan untuk menemukan dan membentuk klaster-klaster daerah lahan pertanian sesuai dengan jenis komoditi berdasarkan atribut-atribut pendukung yang digunakan. Hasil analisis dan implementasi dapat menyediakan informasi lahan seperti jumlah kluster, luas lahan, luas daerah, letak dan tingkat produktifitas[12].

Ashutosh Mahesh Pednekar. Peneliti mengusulkan penggunaan algoritma optimasi, yaitu PSO untuk menentukan *centroid* awal di *K-Means*, untuk akhirnya mendapatkan akurasi yang lebih baik. Notasi vektor dari *centroid* optimal dapat dianggap sebagai entitas dalam ruang optimasi, di mana akurasi *K-Means* atas subset acak dari data dapat bertindak sebagai *fitness measure*. Vektor optimal yang dihasilkan dapat digunakan sebagai *centroid* awal untuk *K-Means*[13].

Garvishkumar K. Patel, dkk. Peneliti melakukan penelitian clustering menggunakan kombinasi *Particle Swarm Optimization* dan *K-Means*. Perbandingan hasil menunjukkan bahwa dengan bertambahnya ukuran kumpulan data, pendekatan yang diusulkan menghasilkan peningkatan yang signifikan atas teknik pengelompokan partisi *K-Means*[14].

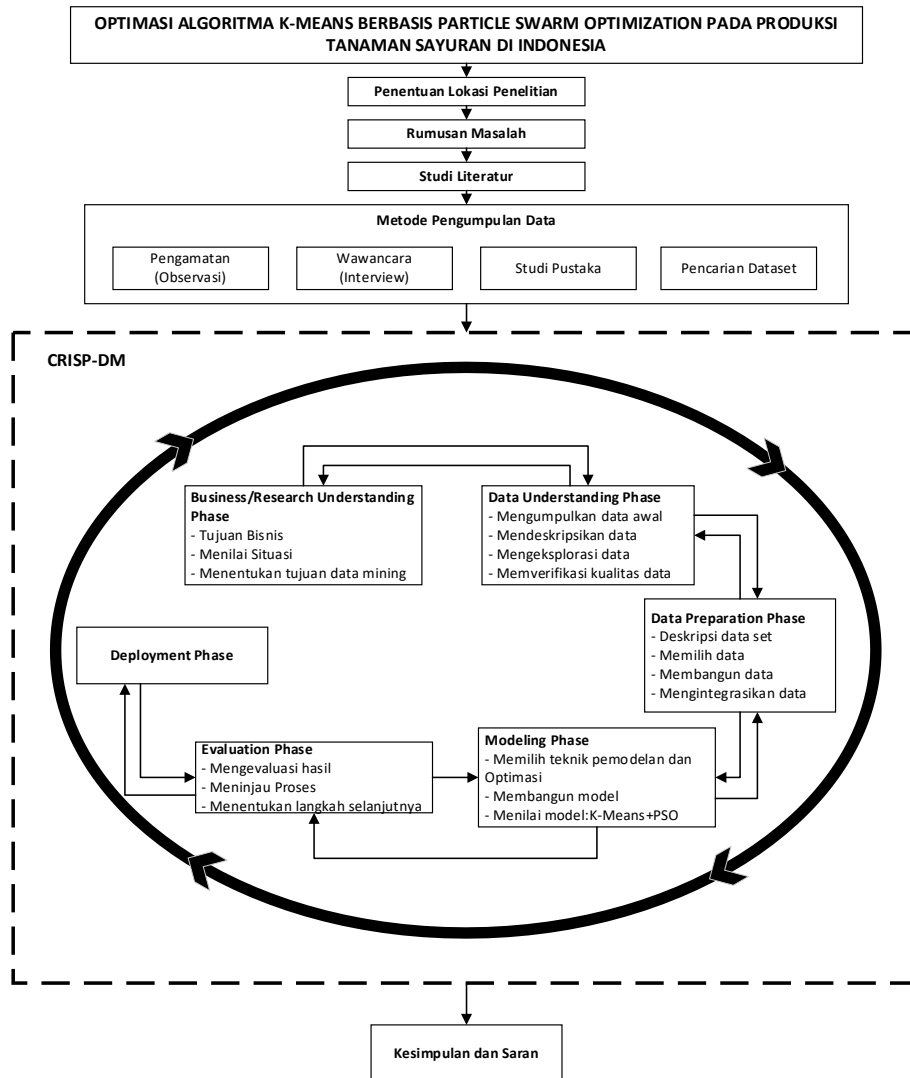
Peneliti mencoba menggabungkan algoritma *K-Means Clustering* dan PSO untuk mendapatkan efektifitas kinerja algoritma yang lebih baik. Algoritma penggabungan ini disebut KPSO atau *K-Means clustering* berbasis PSO. Algoritma ini akan dimaksimalkan dengan metode *K-Means clustering* dalam kelompok awal, selanjutnya PSO memperbaiki kelompok data dari hasil

bentukan *K-Means clustering*. Dataset hasil produksi hortikultura subsektor tanaman sayuran tahun 2000 sampai dengan 2021 diperlukan sebagai acuan untuk melakukan pengelompokan. Penelitian terkait telah dilakukan oleh Ariska Dewi Susanti, dkk. Namun, metode *clustering* yang digunakan dalam pengelompokan data yaitu *K-Medoid*. Selain itu, dataset yang digunakan dalam penelitian tersebut yaitu sektor hortikultura terbagi menjadi subsektor tanaman sayur-sayuran, tanaman buah-buahan, tanaman obat dan tanaman hias pada tahun 2019[7].

Penelitian diharapkan dapat memberikan solusi pemecahan masalah, yaitu pengelompokan daerah-daerah yang memiliki tingkat hasil produksi tanaman sayuran yang tinggi atau tingkat hasil produksi tanaman sayuran yang masih rendah. Penerapan *K-Means clustering* yang memungkinkan untuk mengoptimasi *K-Means clustering* dengan PSO sehingga dihasilkan informasi yang lebih bermutu. Implementasi model *K-Means clustering* dengan PSO pada penelitian ini menggunakan Python dengan *package sklearn* dan *pso*.

## 2. METODE PENELITIAN

Untuk mencapai tujuan yang telah ditetapkan maka diperlukan tahapan penelitian. Metode penelitian yang akan dilakukan menggunakan model *Cross Industry Standard Process for Data Mining (CRISP-DM)* [15]–[17] yang terdiri dari enam fase seperti gambar 1 dibawah ini.



Gambar 1 Tahapan Penelitian

1. Menentukan ide penelitian

Pada tahap ini peneliti menentukan ide dan kajian penelitian yang akan diangkat dalam penelitian.

2. Penentuan lokasi penelitian

Setelah menentukan ide penelitian, kemudian tahap selanjutnya adalah menentukan lokasi penelitian yang akan dijadikan sebagai objek penelitian.

3. Rumusan masalah

Pada tahap rumusan masalah, peneliti akan merumuskan masalah-masalah yang akan diangkat berdasarkan objek penelitian sehingga menjadi fokus penelitian.

4. Studi literatur

Pada tahap ini merupakan tahap pembelajaran pengolahan data mining menggunakan algoritma *K-Means* dan PSO. Pemahaman konsep penelitian didapat dari referensi-referensi yang kredibel diantaranya buku-buku referensi, jurnal penelitian, paper dan lain-lain.

5. Metode pengumpulan data

Pada tahap ini, metode pengumpulan data yang peneliti lakukan adalah pengamatan (observasi), wawancara dan pencarian dataset. Tahap pengamatan (observasi) melakukan pengamatan langsung terhadap objek penelitian. Tahap wawancara melakukan kegiatan tanya jawab langsung dengan pihak-pihak terkait objek yang diteliti. Sedangkan tahap pencarian dataset yaitu mencari dataset



sektor produksi hortikultura subsektor tanaman sayuran pada Badan Pusat Statistik (BPS) untuk beberapa tahun kebelakang.

#### 6. CRISP-DM (*Cross Industry Standard Process for Data Mining*)

Pada tahap ini kegiatan yang akan dilakukan terdiri dari enam fase diantaranya Fase Pemahaman Bisnis (*Business Understanding Phase*), Fase Pemahaman Data (*Data Understanding Phase*), Fase Pengolahan Data (*Data Preparation Phase*), Fase Pemodelan (*Modeling Phase*), Fase Evaluasi (*Evaluation Phase*) dan Fase Penyebaran (*Deployment Phase*).

Fase Pemahaman Bisnis (*Business Understanding Phase*) ini akan dilakukan pemahaman tujuan bisnis, menilai situasi, dan menentukan tujuan data mining. Berdasarkan pemahaman tersebut, kemudian menterjemahkan pengetahuan ke dalam pendefinisian masalah data mining. Selanjutnya akan ditentukan strategi untuk mencapai tujuan tersebut.

Fase Pemahaman Data (*Data Understanding Phase*) ini akan dilakukan pengumpulan data awal, mendeskripsikan data, mengeksplorasi data dan memverifikasi kualitas data. Sehingga akan terdeteksi bagian yang menarik dari data yang dapat digunakan untuk menggali informasi yang tersembunyi.

Fase Pengolahan Data (*Data Preparation Phase*) ini akan dilakukan pendeskripsian data set, memilih data, mengintegrasikan data dan membangun data. Proses ini juga mencakup pemilihan tabel, record dan atribut data termasuk pembersihan dan transformasi data. Sehingga akan terbentuk data set akhir sebagai data mentah yang akan diproses pada fase pemodelan.

Fase Pemodelan (*Modeling Phase*) ini akan dilakukan pemilihan teknik pemodelan, membangun model dan menilai model. Saat membangun model dari data set yang digunakan, penelitian ini menggunakan model elbow untuk menentukan jumlah cluster terbaik.

Fase Evaluasi (*Evaluation Phase*) ini akan dilakukan jika model sudah terbentuk dengan harapan memiliki kualitas yang baik

dilihat dari sudut pandang analisa data. Adapun tahapan yang akan dilakukan yaitu evaluasi hasil, meninjau proses dan menentukan langkah selanjutnya. Evaluasi yang dilakukan terhadap keefektifan dan kualitas data model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan berdasarkan fase awal *Business Understanding* yang ditetapkan. Sehingga dapat ditentukan apakah ada masalah bisnis yang belum ditentukan.

Fase Penyebaran (*Deployment Phase*) ini pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna.

#### 7. Kesimpulan dan Saran

Pada tahap ini akan dilakukan penarikan kesimpulan dan saran penelitian berdasarkan pendefinisian masalah data mining yang terdapat pada fase awal yaitu *Business Understanding Phase*.

### 3. HASIL DAN PEMBAHASAN

#### FASE PEMAHAMAN BISNIS (BUSINESS UNDERSTANDING PHASE)

Tujuan bisnis yaitu pengelompokan daerah-daerah yang memiliki tingkat hasil produksi tanaman sayuran yang tinggi, sedang atau masih rendah. Sehingga mendapatkan informasi yang lebih bermutu dan efektifitas algoritma menggunakan penerapan *K-Means clustering* dengan optimalisasi metode *Particle Swarm Optimization*.

#### FASE PEMAHAMAN DATA (DATA UNDERSTANDING PHASE)

Pada fase pemahaman data terdiri dari beberapa proses diantaranya pengumpulan data awal, deskripsi data dan eksplorasi data. Proses pengumpulan data bertujuan untuk mempelajari struktur data hasil produksi tanaman sayuran di Indonesia. Tabel 1 merupakan fitur dataset yang diambil dari

sumber data Badan Pusat Statistik yang akan digunakan dalam proses pemodelan.

Tabel 1 Fitur Dataset Tanaman Sayuran

N	Atribut	Keterangan
1.	Provinsi	Nama provinsi di Indonesia
2.	Bawang Merah	Hasil produksi tanaman sayuran jenis Bawang Merah dalam satuan Ton
3.	Bawang Putih	Hasil produksi tanaman sayuran jenis Bawang Putih dalam satuan Ton
4.	Bawang Daun	Hasil produksi tanaman sayuran jenis Bawang Daun dalam satuan Ton
5.	Kentang	Hasil produksi tanaman sayuran jenis Kentang dalam satuan Ton
6.	Kubis	Hasil produksi tanaman sayuran jenis Kubis dalam satuan Ton
7.	Kembang Kol	Hasil produksi tanaman sayuran jenis Kembang Kol dalam satuan Ton
8.	Petsai/Sawi	Hasil produksi tanaman sayuran jenis Petsai/Sawi dalam satuan Ton
9.	Wortel	Hasil produksi tanaman sayuran jenis Wortel dalam satuan Ton
10.	Lobak	Hasil produksi tanaman sayuran jenis Lobak dalam satuan Ton
11.	Kacang Merah	Hasil produksi tanaman sayuran jenis Kacang Merah dalam satuan Ton
12.	Kacang Panjang	Hasil produksi tanaman sayuran jenis Kacang Panjang dalam satuan Ton
13.	Cabai Besar	Hasil produksi tanaman sayuran jenis Cabai Besar dalam satuan Ton
14.	Cabai Rawit	Hasil produksi tanaman sayuran jenis Cabai Rawit dalam satuan Ton
15.	Tomat	Hasil produksi tanaman sayuran jenis Tomat dalam satuan Ton

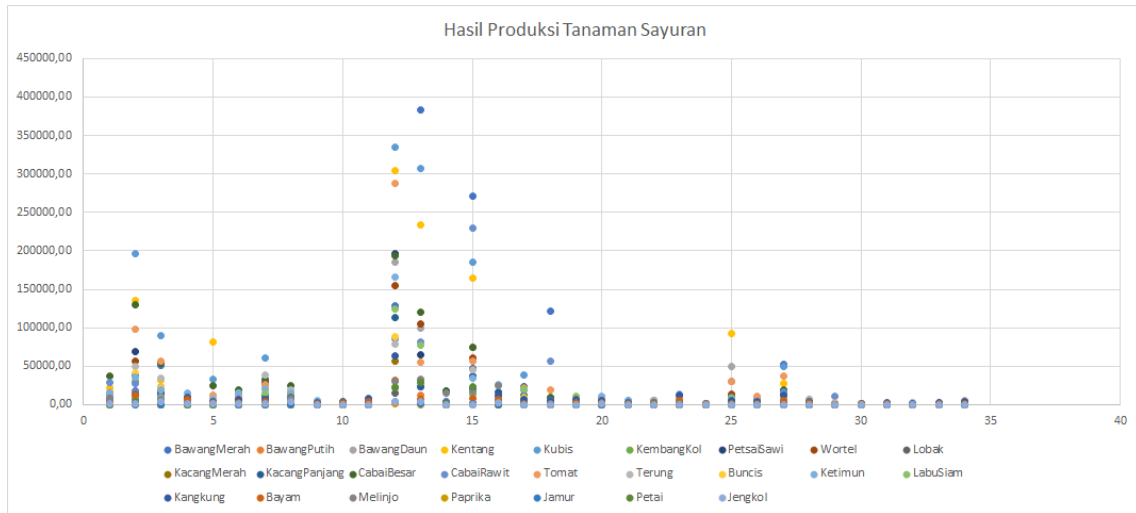
16.	Terung	Hasil produksi tanaman sayuran jenis Terung dalam satuan Ton
17.	Buncis	Hasil produksi tanaman sayuran jenis Buncis dalam satuan Ton
18.	Ketimun	Hasil produksi tanaman sayuran jenis Ketimun dalam satuan Ton
19.	Labu Siam	Hasil produksi tanaman sayuran jenis Labu Siam dalam satuan Ton
20.	Kangkung	Hasil produksi tanaman sayuran jenis Kangkung dalam satuan Ton
21.	Bayam	Hasil produksi tanaman sayuran jenis Bayam dalam satuan Ton
22.	Melinjo	Hasil produksi tanaman sayuran jenis Melinjo dalam satuan Ton
23.	Paprika	Hasil produksi tanaman sayuran jenis Paprika dalam satuan Ton
24.	Jamur	Hasil produksi tanaman sayuran jenis Jamur dalam satuan Ton
25.	Petai	Hasil produksi tanaman sayuran jenis Petai dalam satuan Ton
26.	Jengkol	Hasil produksi tanaman sayuran jenis Jengkol dalam satuan Ton

Proses eksplorasi data bertujuan untuk menentukan sumber data. Data set yang digunakan dalam penelitian ini adalah data set hasil produksi tanaman sayuran yang diperoleh dari Badan Pusat Statistik periode 2000 s.d. 2021. Berdasarkan periode dataset tersebut rekapitulasi dilakukan dengan menghitung rata-rata hasil produksi tanaman sayuran pada wilayah provinsi di Indonesia. Kemudian, Tabel 2 merupakan sampel dataset rata-rata hasil produksi tanaman sayuran dari 34 Provinsi di Indonesia.

Tabel 2 Dataset Hasil Produksi Tanaman Sayuran Bawang Merah, Bawang Putih, Bawang Daun, Kentang, Kubis, Kembang Kol, dan Petsai/Sawi

Provinsi	Bawang Merah	Bawang Putih	Bawang Daun	Kentang	Kubis	Kembang Kol	Petsai/Sawi
Aceh	6.153,36	319,40	1.521,79	22.739,51	6.619,79	808,77	2.613,00
Sumatera Utara	18.382,50	2.737,25	15.402,84	135.149,99	195.785,43	30.609,23	69.446,60
Sumatera Barat	51.698,19	1.250,75	23.410,30	32.351,89	89.169,57	8.186,72	14.549,08
Riau	94,09	0,09	17,96	0,13	1,61	1,22	2.032,43
Jambi	4.493,09	204,20	1.831,37	81.780,72	33.346,37	883,82	3.185,56
Sumatera Selatan	377,58	13,57	2.555,89	835,00	6.678,93	238,70	4.158,16
Bengkulu	737,09	60,70	20.032,17	5.537,84	60.521,45	10.334,84	28.759,36
Lampung	1.182,93	45,78	4.118,86	856,95	12.550,07	340,29	10.984,00
Kep. Bangka Belitung	32,52	0,00	129,43	0,09	0,35	0,65	1.916,67
Kep. Riau	11,13	0,00	156,13	0,00	0,43	0,04	2.634,61
DKI Jakarta	2,13	0,00	0,26	0,00	0,26	0,35	4.195,02
Jawa Barat	128.806,59	1.340,10	185.828,53	303.954,51	335.439,95	23.141,40	196.403,48
Jawa Tengah	382.538,73	12.850,50	99.544,97	233.636,62	307.739,10	23.751,30	64.715,25
DI Yogyakarta	17.430,94	1,11	959,65	61,74	341,04	212,81	4.646,00
Jawa Timur	271.603,85	2.370,92	74.648,20	165.054,85	185.639,06	13.259,23	47.005,72
Banten	687,17	0,13	564,13	4,61	12,30	28,39	10.193,33
Bali	12.751,14	706,73	1.467,71	3.325,67	39.236,70	825,71	23.889,34
Nusa Tenggara Barat	121.171,50	7.926,52	637,46	2.432,37	7.135,59	482,61	2.067,14
Nusa Tenggara Timur	7.039,48	759,17	519,21	1.209,33	1.923,64	472,35	6.406,52
Kalimantan Barat	31,96	0,00	908,80	0,00	209,17	18,74	5.993,03
Kalimantan Tengah	91,22	0,00	832,21	3,13	58,81	16,07	1.763,69
Kalimantan Selatan	392,78	0,09	689,51	0,09	13,04	54,61	1.757,77
Kalimantan Timur	211,08	0,30	685,95	0,78	377,90	84,56	9.359,82
Kalimantan Utara	21,78	0,00	210,35	0,00	8,09	0,70	1.316,78
Sulawesi Utara	3.632,59	121,00	49.548,68	92.301,74	29.752,59	1.001,64	9.699,97
Sulawesi Tengah	6.569,95	63,61	1.627,15	651,63	6.099,31	388,70	3.675,60
Sulawesi Selatan	52.640,25	78,00	17.952,61	27.682,53	49.038,20	1.293,25	11.940,60
Sulawesi Tenggara	527,25	0,57	641,01	21,57	622,68	80,55	1.943,04
Gorontalo	422,60	0,26	61,24	0,60	3,01	1,87	82,57
Sulawesi Barat	465,96	47,78	234,39	80,48	48,91	4,22	675,22
Maluku	807,57	0,30	81,22	198,48	655,09	22,96	1.642,21
Maluku Utara	311,35	3,13	181,59	0,00	181,83	3,13	420,87
Papua Barat	215,91	62,35	536,57	668,65	999,13	22,22	1.978,30
Papua	718,97	126,27	674,35	260,44	2.696,76	148,22	3.893,44





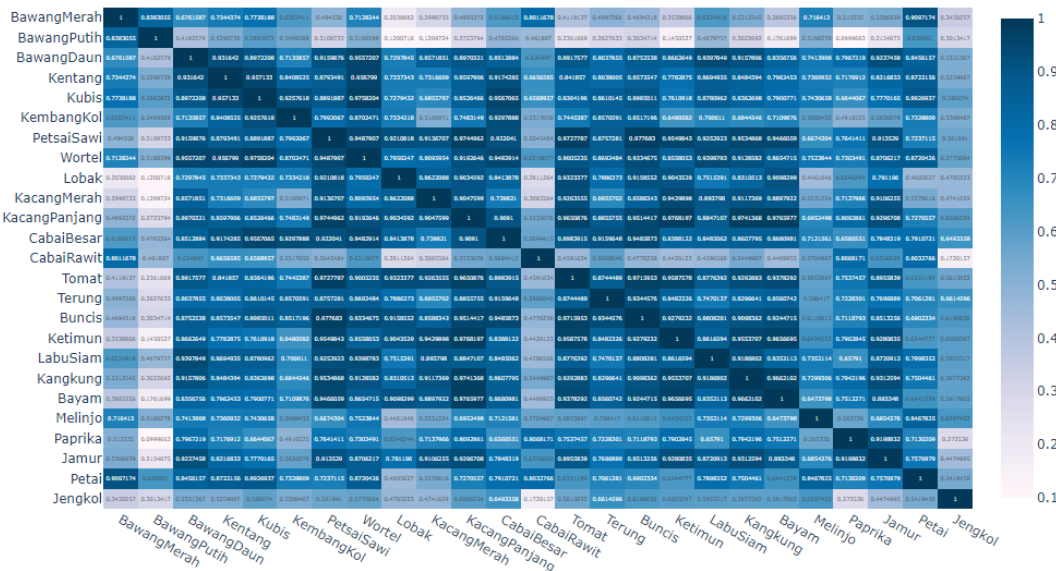
Gambar 2 Persebaran Hasil Produksi Tanaman Sayuran di Indonesia

Persebaran hasil produksi tanaman sayuran di Indonesia tidak merata untuk setiap jenis sayuran di setiap Provinsi. Hal ini dipengaruhi oleh lahan panen yang tersedia, suhu, air dan kelembaban udara. Gambar 2 menunjukkan persebaran hasil produksi tanaman sayuran di Indonesia, berdasarkan gambar 2 menunjukkan bahwa rata-rata hasil produksi tanaman sayuran tertinggi yaitu Bawang Merah sebanyak 382.538,73 Ton per tahun berasal dari Provinsi Jawa Tengah demikian lahan panen bawang merah paling luas dibanding Provinsi lain yaitu seluas 55.250 ha[18]. Kemudian Kubis sebanyak 335.439,95 Ton per tahun dari provinsi Jawa

Barat dengan lahan panen kubis seluas 10.311 ha[18] dan Kentang sebanyak 303.954,51 Ton per tahun dengan lahan panen kentang seluas 10.804 ha[18].

### FASE PENGOLAHAN DATA (DATA PREPARATION PHASE)

Proses verifikasi kualitas data bertujuan untuk memastikan data yang digunakan pada proses data mining tidak ada kesalahan, yaitu melalui langkah sebagai berikut: memastikan sumber record terisi dengan data dan menentukan rata-rata hasil produksi tanaman sayuran.



Gambar 3 Correlation Heatmap

Gambar 3 merupakan *correlation heatmap* dari dataset rata-rata hasil produksi tanaman sayuran di Indonesia. Gambar 3 menunjukkan bahwa pada fitur data terkait variabel tanaman sayuran, mempunyai korelasi yang tinggi satu sama lain. Sementara pada fitur lainnya, korelasi beragam dan ada beberapa yang terlihat korelasi positif cukup signifikan. Dataset tanaman sayuran ini memiliki banyak *outlier* karena nilai rata-rata hasil produksi tanaman sayuran yang berbeda jauh dari nilai rata-rata wilayah lainnya.

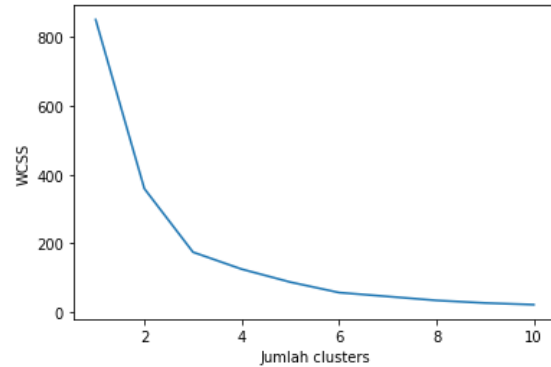
Pada *Unsupervised model* standarisasi dilakukan karena konsep model yang digunakan

menggunakan perhitungan jarak dimana terdapat fitur data numerik dengan *range* nilai besar, dibandingkan dengan *range* data kecil akan sangat mendominasi perhitungan (bias). Serta banyaknya data sampel dan data *outlier* membuat standarisasi dipilih daripada normalisasi. Sehingga normalisasi data yang digunakan tidak memiliki penyimpangan. Standarisasi penelitian ini menggunakan *class StandardScaler* dari *package sklearn*.

## FASE PEMODELAN (MODELING PHASE)

Penelitian yang dilakukan menggunakan model *K-Means Clustering* dengan optimasi *Particle Swarm Optimization*. Uji coba dilakukan sebanyak 10 kali percobaan terhadap model *K-Means* dan *K-Means* berbasis *Particle Swarm Optimization*. Pemodelan *machine learning* menggunakan *package sklearn* dan *ParticleSwarmOptimizedClustering*. Penentuan Jumlah *cluster* yang optimal pada penelitian ini menggunakan metode *Elbow*. Metode *Elbow* ditentukan berdasarkan nilai *Sum of Square Error (SSE)* yang memiliki nilai penurunan drastis. Semakin besar nilai SSE maka *cluster* kualitas *cluster* berkurang. Semakin kecil nilai SSE maka kualitas *cluster* semakin baik. Pada gambar 4 terlihat nilai SSE pada saat k=1 paling tinggi, kemudian saat k=2 nilai SSE mengalami penurunan yang signifikan. Saat k=3 nilai SSE

mengalami penurunan kembali, begitu seterusnya sampai k=10 mengalami penurunan juga.



Gambar 4 Penentuan Jumlah Cluster Optimum

Berdasarkan grafik pada gambar 4 dapat dilihat jumlah *cluster* yang membentuk siku pada k=3, sedangkan pada k=4 hingga k=10 nilai SSE terlihat mulai stabil. Maka ditetapkan jumlah *cluster* optimal pada penelitian ini sebanyak 3 *cluster*. Sebelum melakukan perhitungan PSO maka akan dilakukan inisialisasi terlebih dahulu untuk mendapatkan nilai kecepatan masing-masing partikel dengan jumlah partikel=10, iterasi=20. Selanjutnya akan dilakukan perhitungan untuk perubahan posisi, proses PSO ini akan terus dilakukan sampai dengan maksimal iterasi dengan *gbest position* sebagai solusi yang diberikan. Gambar 4 merupakan *gbest* partikel dari *centroid* yang dihasilkan pada 3 *cluster*.

```
array([[ -0.27927709, -0.33356992, -0.33133575, -0.37921579, -0.4051154 ,  
        -0.39334558, -0.34515776, -0.36218936, -0.2838042 , -0.27527278,  
        -0.39214155, -0.46555442, -0.25291219, -0.43472651, -0.60149107,  
        -0.42406372, -0.32785224, -0.39758835, -0.37561763, -0.19538422,  
        -0.44431145, -0.11225484, -0.19948404, -0.25673635, -0.49880615],  
       [ -0.52132412, -0.45525011, -0.45448522, -0.54355314, -0.56817297,  
        -0.52154079, -0.45692222, -0.50258256, -0.2933163 , -0.28653555,  
        -0.40892531, -0.60348189, -0.49952408, -0.45454048, -0.54359056,  
        -0.47474238, -0.36180396, -0.43129301, -0.58468368, -0.50394315,  
        -0.55605735, -0.32331112, -0.32178294, -0.47628591, -0.37914124],  
       [ -0.43753551, -0.32842783, -0.54191945, -0.58544748, -0.56993699,  
        -0.52322794, -0.51629798, -0.56283265, -0.43832932, -0.45492505,  
        -0.55431574, -0.64942298, -0.44478722, -0.53544256, -0.67857043,  
        -0.50276059, -0.54045616, -0.53804193, -0.54672714, -0.55303555,  
        -0.62503707, -0.39529993, -0.45285423, -0.50978352, -0.74364619]])
```

Gambar 5 *Gbest* Partikel dari *Centroid*

**FASE EVALUASI (EVALUATION PHASE)**

Performa PSO terhadap K-Means dapat diketahui dengan melakukan pengujian sebanyak 10 kali yaitu:

- Membandingkan hasil pengujian berdasarkan *silhouette coefficient* antara *K-Means* murni dan *K-Means* berbasis PSO.
- Membandingkan rata-rata waktu komputasi dalam satuan detik

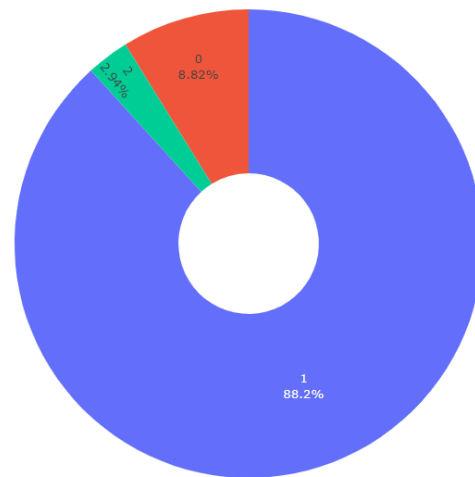
Tabel 3 Perbandingan *Silhouette Coefficient* dan Waktu Komputasi

	K-Means		K-Means+PSO	
	<i>Silhouette Coefficient</i>	Waktu	<i>Silhouette Coefficient</i>	Waktu
1	0,712913	0,531	0,713958	102
2	0,712391	0,781	0,712913	138
3	0,713436	0,787	0,714488	138
4	0,712913	0,798	0,713958	141
5	0,712913	0,794	0,713958	140
6	0,713958	0,762	0,715003	140
7	0,712391	0,746	0,713958	138
8	0,712391	0,774	0,715003	139
9	0,712391	0,754	0,713958	139
10	0,712913	0,754	0,713958	139

Berdasarkan Tabel 3 rata-rata *Silhouette Coefficient* untuk *K-Means* murni sebesar 0,712861 sedangkan rata-rata *Silhouette Coefficient* untuk *K-Means* berbasis PSO sebesar 0,714114. Perbedaan rata-rata *Silhouette Coefficient* pada kedua model tidak signifikan berbeda karena dataset yang digunakan memiliki *record* yang sedikit, nilai *Silhouette* semakin baik jika mendekati nilai 1. Namun demikian, *K-Means* berbasis PSO dapat membentuk model suatu *cluster* menjadi lebih baik dibandingkan *K-Means* murni karena nilai setiap pengujian dan rata-rata *Silhouette Coefficient* lebih besar. Hasil penelitian ini berbeda dengan peneliti sebelumnya yaitu nilai *Silhouette Coefficient*

yang diperoleh hasil *cluster* secara keseluruhan memiliki kategori struktur yang buruk karena bernilai 0,24[7]. Apabila dilihat dari waktu komputasi *K-Means* murni lebih cepat dibandingkan *K-Means* berbasis PSO dengan rata-rata 0,748 detik. *K-Means* berbasis PSO membutuhkan waktu lebih lama karena memperbaiki kelompok data dari hasil bentuk *K-Means clustering*.

Hasil pemodelan *clustering* pada *K-Means* berbasis PSO untuk hasil produksi tanaman sayuran di Indonesia *cluster* 0 sebanyak 8,82%, *cluster* 1 sebanyak 88,2%, dan *cluster* 2 sebanyak 2,94%.



Gambar 6 Persentase Hasil *Clustering*

Gambar 6 menunjukkan hasil *clustering* menggunakan model *K-Means* berbasis PSO. Berdasarkan rata-rata lahan panen pada *cluster* 0 yaitu Provinsi Sumatera Utara memiliki rata-rata lahan panen seluas 3.890,88 ha, Provinsi Jawa Tengah memiliki rata-rata lahan panen seluas 8.504,26 ha, dan Provinsi Jawa Timur memiliki rata-rata lahan panen seluas 10.453,47 ha. Sedangkan *Cluster* 2 yaitu Provinsi Jawa Barat memiliki rata-rata lahan panen seluas 6.574,57 ha.

Tabel 4 Hasil *Clustering K-Means* Berbasis PSO Berdasarkan Provinsi

Cluster	Provinsi
0	Sumatera Utara, Jawa Tengah, Jawa Timur
1	Aceh, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Kepulauan Bangka Belitung, Kepulauan 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, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Maluku, Maluku Utara, Papua Barat, Papua
2	Jawa Barat

Hasil clustering menunjukkan bahwa *Cluster 2* merupakan merupakan wilayah yang memiliki hasil produksi tanaman sayuran yang tinggi. Sedangkan *Cluster 0* memiliki hasil produksi tanaman sayuran yang sedang dan *cluster 1* memiliki hasil produksi tanaman sayuran yang masih rendah.

Tabel 5 Rata-Rata Hasil Produksi Tanaman Sayuran Berdasarkan *Clustering K-Means* Berbasis PSO

Cluster	Rata-Rata Hasil Produksi Tanaman Sayuran (Ton setiap Tahun)
0	40.142,11 – 71.324,68
1	441,91 – 19.124,69
2	107.890,86

Berdasarkan tabel 5 rata-rata hasil produksi tanaman sayuran pada *cluster 0* sebanyak 40.142,11 Ton sampai dengan 71.324,68 Ton setiap tahun. Kemudian *cluster 1* sebanyak 441,91 Ton sampai dengan 19.124,69 Ton setiap tahun dan *cluster 2* sebanyak 107.890,86 Ton setiap tahun. Dengan demikian Provinsi Jawa Barat merupakan Provinsi yang mendominasi dalam produksi tanaman sayuran di Indonesia.

#### 4. KESIMPULAN

Penelitian diharapkan dapat memberikan solusi pemecahan masalah, yaitu pengelompokan daerah-daerah yang memiliki tingkat hasil produksi tanaman sayuran yang tinggi atau tingkat hasil produksi tanaman sayuran yang masih rendah. Hasil penelitian yang telah dilakukan maka dapat ditarik kesimpulan sebagai berikut:

- Hasil pemodelan *clustering K-Means* berbasis PSO untuk hasil produksi tanaman sayuran di Indonesia *cluster 0* (sedang) sebanyak 8,82%, *cluster 1* (rendah) sebanyak 88,2%, dan *cluster 2* (tinggi) sebanyak 2,94%. Provinsi Jawa Barat berada dalam *cluster 2*, sehingga Provinsi Jawa Barat mendominasi dalam produksi tanaman sayuran di Indonesia.
- Hasil pengujian berdasarkan rata-rata *silhouette coefficient*, *K-Means* berbasis PSO lebih baik daripada *K-Means* murni meskipun nilai tersebut tidak berbeda signifikan yaitu dengan rata-rata nilai *silhouette coefficient* sebesar 0,714114.
- Rata-rata waktu komputasi waktu komputasi *K-Means* murni lebih cepat dengan rata-rata 0,748 detik dibandingkan *K-Means* berbasis PSO. Waktu yang dibutuhkan lebih lama karena *K-Means* berbasis PSO memperbaiki kelompok data dari hasil bentukan *K-Means clustering*.

#### 5. SARAN

Untuk memperoleh hasil efektifitas clustering yang baik terhadap dataset hasil produksi tanaman sayuran di Indonesia, maka diperlukan penelitian lebih lanjut dengan penambahan algoritma clustering yang lainnya sebagai perbandingan seperti DBSCAN, *Affinity Propagation Agglomerative Clustering*, BIRCH dan lain-lain.

## REFERENCES

- [1] Badan Pusat Statistik Provinsi Jawa Barat, “Jumlah Keluarga Berdasarkan Kelompok Usia sejahtera Kabupaten/Kota di Provinsi Jawa Barat, 2017,” 2019. <https://jabar.bps.go.id/statictable/2019/05/31/606/jumlah-keluarga-berdasarkan-kelompok-usia-sejahtera-kabupaten-kota-di-provinsi-jawa-barat-2017.html> (accessed Feb. 13, 2022).
- [2] Badan Pusat Statistik, “Persentase Tenaga Kerja Informal Sektor Pertanian (Persen), 2019-2021.” <https://www.bps.go.id/indicator/6/1171/1/persentase-tenaga-kerja-informal-sektor-pertanian.html> (accessed Feb. 07, 2022).
- [3] Darmansah and N. Wayan Wardani, “Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 7, no. 2, pp. 126–134, Aug. 2020, doi: 10.35957/JATISI.V7I2.309.
- [4] H. Mahulae, “Pengelompokan Potensi Produksi Buah-Buahan di Provinsi Sumatera Utara dengan Menerapkan K-Clustering (Studi Kasus: Dinas Tanaman Pangan dan Holtikultura),” *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 2, pp. 312–325, Apr. 2020, doi: 10.30865/JURIKOM.V7I2.2122.
- [5] S. Rustam, S. Rustam, H. A. Santoso, and C. Supriyanto, “Optimasi K-Means Clustering Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization Di Kota Semarang,” *Ilk. J. Ilm.*, vol. 10, no. 3, pp. 251–259, Dec. 2018, doi: 10.33096/ilkom.v10i3.342.251-259.
- [6] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques Third Edition*, Third Edit. Morgan Kaufmann, 2012.
- [7] A. Dewi Susanti, S. Sufri, and C. Sormin, “Analisis Non-Hierarchical Partitioning K-Medoid Pada Produksi Sektor Hortikultura Tahun 2019 Di Indonesia,” 2021.
- [8] H. Andema, S. Defit, and U. Putra Indonesia YPTK Padang, “Optimalisasi Penggunaan Lahan Perkebunan Kelapa Hibrida Menggunakan K-Means Clustering,” *J. Inform. Ekon. Bisnis*, vol. 2, no. 2, pp. 32–38, Jun. 2020, doi: 10.37034/INFEB.V2I2.23.
- [9] A. Supriyatna, I. Carolina, W. Widiati, and C. Nuraeni, “Rice Productivity Analysis by Province Using K-Means Cluster Algorithm,” in *IOP Conference Series: Materials Science and Engineering*, 2020, vol. 771, no. 1, doi: 10.1088/1757-899X/771/1/012025.
- [10] S. Al-Wazni, A. Y. Azainul-Abideen - , and S. Anam, “Segmentation of Leaf Spots Disease in Apple Plants Using Particle Swarm Optimization and K-means Algorithm,” *J. Phys. Conf. Ser.*, vol. 1562, no. 1, p. 012011, Jun. 2020, doi: 10.1088/1742-6596/1562/1/012011.
- [11] S. Mehrotra and A. Sharan, “Comparative Analysis of K-Means Algorithm and Particle Swarm Optimization for Search Result Clustering,” *Smart Innov. Syst. Technol.*, vol. 165, pp. 109–114, 2020, doi: 10.1007/978-981-15-0077-0\_12.
- [12] J. Tamaela, E. Sedyono, and A. Setiawan, “Cluster Analysis Menggunakan Algoritma Fuzzy C-means dan K-means Untuk Klasterisasi dan Pemetaan Lahan Pertanian di Minahasa Tenggara,” *J. Buana Inform.*, vol. 8, no. 3, Oct. 2017, doi: 10.24002/JBI.V8I3.1317.
- [13] A. M. Pednekar, “Optimal initialization of K-means using Particle Swarm Optimization,” Apr. 2019, Accessed: Feb. 04, 2022. [Online]. Available:



<https://arxiv.org/abs/1904.09098v1>.

- [14] G. K. Patel, V. K. Dabhi, and H. B. Prajapati, "Clustering Using a Combination of Particle Swarm Optimization and K-means," *J. Intell. Syst.*, vol. 26, no. 3, pp. 457–469, Jul. 2017, doi: 10.1515/JISYS-2015-0099/MACHINEREADABLECITATION/RIS.
- [15] D. T. Larose, *DATA MINING METHODS AND MODELS*. A JOHN WILEY & SONS, INC PUBLICATION, 2006.
- [16] B. Budiman, R. Nursyanti, R. Y. R. Alamsyah, and I. Akbar, "Data Mining Implementation Using Naïve Bayes Algorithm and Decision Tree J48 In Determining Concentration Selection," *Int. J. Quant. Res. Model.*, vol. 1, no. 3, pp. 123–134, Sep. 2020, doi: 10.46336/ijqrm.v1i3.72.
- [17] B. Budiman and Z. Niqotaini, "Perbandingan Algoritma Klasifikasi Data Mining untuk Penelusuran Minat Calon Mahasiswa Baru," *NUANSA Inform.*, vol. 15, no. 2, pp. 37–52, Aug. 2021, doi: 10.25134/NUANSA.V15I2.4162.
- [18] Badan Pusat Statistik, "Luas Panen Tanaman Sayuran Menurut Provinsi dan Jenis Tanaman 2021." [https://www.bps.go.id/indikator/indikator/view\\_data\\_pub/0000/api\\_pub/bXNVb1pmZndqUDhKWElUSjhZRitidz09/da\\_05/1](https://www.bps.go.id/indikator/indikator/view_data_pub/0000/api_pub/bXNVb1pmZndqUDhKWElUSjhZRitidz09/da_05/1) (accessed Aug. 23, 2022).