BRAHMANA: Jurnal Penerapan Kecerdasan Buatan
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 5, No. 2, Juni (2024), pp. 334-340

Penerapan Algoritma *K-Means* Dalam *Clustering* Data Seleksi Benih Kelapa Sawit Di PPKS Unit Marihat

Fazli Nugraha Tambunan
Program Studi Komperisasi Akuntansi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
E-mail: fazli@amiktunasbangsa.ac.id

Abstract

Oil palm is one of the most useful gardens in daily life. In addition to producing vegetable oil, oil palm seeds can also be marketed. In this study, palm oil seed data will be clustered using the K-Means method to determine the results of good seeds. Clustering means a method of analyzing data whose purpose is to group data with the same characteristics and characteristics in an area. While K-Means is a data analysis method or Data Mining method that performs unsupervised modeling process and is one of the methods that classify data with a partition system (division of an object into several parts with a specific purpose). The K-Means method seeks to group existing data into several groups, where data in one group has similar characteristics to each other and has different characteristics from data in other groups. In other words, this method tries to minimize the variation between data in a cluster and maximize the variation with data in other clusters.

Keywords: *Oil Palm Seeds, Clustering, Data Mining, K-Means Algorithm*

Abstrak

Kelapa sawit merupakan salah satu taman yang sangat bermanfaat dalam kehidupan seharihari. Selain menghasilkan minyak nabati, benih kelapa sawit juga dapat di pasarkan. Untuk menghasilkan minyak nabati yang baik maka di perlukan benih yang berkualitas.Dalam penelitian ini data benih kelapa sawit akan di cluster menggunakan metode K-Means untuk mengetahui hasil benih yang baik. Clustering bermakna metode penganalisaan data yang tujuannya untuk mengelompokkan data dengan cirikhas dan karateristik yang sama dalam suatu wilayah. Sedangkan K-Means adalah suatu metode penganalisaan data atau metode Data Mining yang melakukan proses pemodelan tanpa supervisi (melihat atau meninjau/ unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi(pembagian suatu objek kedalam beberapa bagian dengan tujuan tertentu). Metode K-Means berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama satu sama lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain. Dengan kata lain, metode ini berusaha untuk meminimalkan variasi antar data yang ada di dalam suatu cluster dan memaksimalkan variasi dengan data yang ada di cluster lainnya.

Kata Kunci*: Benih Kelapa Sawit, Klustering, Data Minig, Algoritma K-Means*

1. Pendahuluan

Tanaman kelapa sawit (*Elaeis guineensis Jacq*) berasal dari benua Afrika, kelapa sawit benyak dijumpai di hutan hujan tropis Negara Kamerun, Pantai Gading, Ghana, Liberia, Togo, Angola, Liberia, Nigeria, Sierre Leone dan Kongo [1]. Selain mampu menciptakan kesempatan kerja yang luas, industri sawit menjadi salah satu devisa terbesar bagi Indonesia. Fenomena yang terjadi adalah sejak beberapa tahun belakangan , harga salah satu jenis minyak nabati ini cenderung stagnan dan menurun , namun disisi lain dapat menjadi waktu yang tepat untuk meningkatkan kembali harga CPO (*Crude Palm Oil*)

ISSN: 2715-9906

salah satunya dengan melakukan peramalan jumlah produksi tanaman kelapa sawit untuk periode kedepan sehingga dapat menjadi acuan dalam menentukan kebijakan dan strategi bisnis, hal ini perlu agar setiap permintaan akan kebutuhan kelapa sawit tersedia [2]. Di Sumatera Utara Pusat Penelitian Kelapa Sawit (PPKS) berada di Medan dan memiliki cabang di daerah Kabutan Simalungun yaitu PPKS Marihat yang dimana Pusat Penelitian Kelapa Sawit (PPKS) ini memiliki peran strategis dalam riset dan pengembang industri perkebunan kelapa sawit nasional. Misi PPKS adalah menunjang industri kelapa sawit melalui penelitian untuk menciptakan benih kelapa sawit terbaik.

Oleh karena itu, Untuk menghasilkan minyak nabati yang baik maka di perlukan benih yang berkualitas. Maka tujuan dari penelitian ini adalah menganalisis bentuk benih kelapa sawit yang bagus dan yang baik, melakukan pengclusteran jumlah benih kelapa sawit menggunakan metode *k-means clustering.* Algoritma *k-means* merupakan algoritma pengelompokkan data yang dapat digunakan untuk membentuk beberapa cluster dan dapat diselesaikan tepat waktu [3]. K-Means adalah sebuah metode Clustering yang termasuk dalam pendekatan partitioning. Algoritma K-Means merupakan model Centroid. Mode Centroid adalah model yang menggunakan Centroid untuk membuat Cluster [4]. Metode penganalisaan data atau metode *Data Mining* yang melakukan proses pemodelan tanpa supervisi (melihat atau meninjau/ unsupervised) dan merupakan salah satu metode yang melakukan pengelompokan data dengan sistem partisi(pembagian suatu objek kedalam beberapa bagian dengan tujuan tertentu).

Hasil penelitian [5] berhasil merancang model Algoritma K-Means Clustering menggunakan bahasa pemrograman python untuk mengelompokkan Obat/Alkes yang memiliki permintaan yang tinggi dengan penjualan obat. Metode K-Means cocok digunakan karena dapat mengelompokkan data berdasarkan kesamaan variabel dan mengidentifikasi kecamatan dengan produktivitas serupa [6]. Penerapan algoritma k-means dapat menghasilkan pengelompokkan pake barang berdasarkan cover area pengiriman sehingga tidak terjadi kendala dalam proses pengiriman yang membutuhkan waktu yang lama [7]. Berdasarkan latar belakang, maka penulis melakukan penelitian ini dengan tujuan untuk melakukan pengklasteran bemih kepala sawit. Hasil dari penelitian ini diharapkan dapat menjadi informasi tambahan bagi PPKS Marihat untuk memilih benih kelapa sawit yang baik.

2. Metodologi Penelitian

Penelitian ini menggunakan landasan teori guna mendukung argument teoritis yang digunakan dalam penelitian

2.1. Data Mining

Data mining sendiri merupakan salah satu kegiatan penambangan data dengan mengekstraksi pola yang memerlukan data dalam jumlah besar, dan strategi menjadi menarik apabila tidak sederhana, implisit, dan tidak diketahui sebelumnya, serta strategi yang dihasilkan harus mudah dipahami dan berguna digunakan untuk data yang diprediksi dengan tingkat kepastian tertentu [8]. K-Means merupakan metode penganalisaan data pada Data mining dimana proses pemodelan tanpa supervisi dan merupakan salah satu metode yang mengelompokan data secara partisi [9]. Dari penelitian sebelumnya menyebutkan bahwa Algoritma K-Means adalah teknik clustering yang membagi kumpulan data menjadi kluster K yang tidak tumpang tindih [10]. Dimana data dalam satu kelompok mempunyai karakteristik yang sama satu dengan yang lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada di dalam kelompok yang lain.

2.2. *Clustering*

Salah satu ilmu dari data mining adalah clustering, yang melibatkan pengelompokan data atau objek ke dalam cluster (group) sehingga setiap cluster memiliki data yang

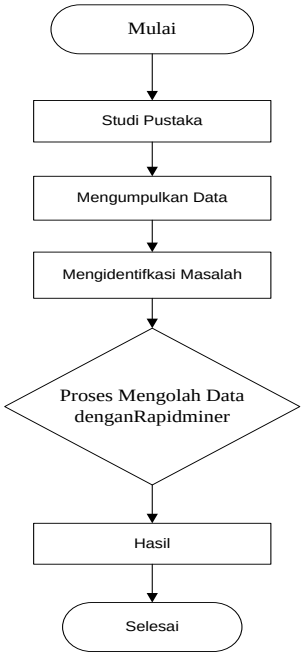
sama dengan cluster lain yang layak namun berbeda dari mereka [11]. Clustering dapat dikatakan juga sebagai identifikasi kelas objek yang memiliki kemiripan. Dengan menggunakan teknik clustering kita bisa lebih lanjut mengidentifikasi kepadatan dan jarak daerah dalam objek ruang dan dapat menemukan secara menyeluruh pola distribusi dan korelasi antara atribut [12]

2.3. Algoritma *K-Means*

K-means clustering adalah metode analisis data atau metode data mining yang melakukan pemodelan tanpa pengawasan dan merupakan salah saat metode pengelompokan data menggunakan sistem partis [13]. Algoritma K-Means adalah model pusat massa. Model cendroid adalah model yang membuat cluster dengan memanfaatkan cendroid. Kami mencoba menggunakan metode clustering non-hierarchical K-Means untuk memisahkan data yang ada menjadi satu atau lebih cluster [14]. Berikutnya *K-Means* menguji masing-masing komponen didalam populasi data dan menandai komponen tersebut ke salah satu pusat *cluster* yang telah di definisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap pusat *cluster*.

2.4. Rancangan Penelitian

Dalam merancang penelitian kali ini, penulis menguraikan alur penelitian yang akan digunakan untuk memecahkan masalah penelitian. Rancangan penelitian ditunjukkan pada Gambar 1 berikut



Gambar 1. Rancangan Penelitian

Dari Gambar 1, menjelaskan rancangan penelitian yang dilakukan untuk menentukan pengklusteran data benih kelapa sawit dengan mengunakan algoritma *k-means* yang terdiri dari

1. Studi Pustaka
   Pada tahap ini merupakan langkah untuk mendapatkan ataupun mengumpulkan informasi yang terkait terhadap topik, metode atau masalah yang terjadi pada penelitian ini. Maka pada tahap ini pengumpulan informasi di peroleh dari berbagai jenis jurnal yang terkait dengan topik penelitian ini.
2. Mengumpulkan Data
   Data di kumpulkan dari PPKS Marihat.

3. Mengidentifikasi Masalah
   Mengindentifikasi masalah yang terkait dengan benih kelapa sawit yang dihasilkan PPKS Marihat. Untuk memproduksi benih yang layak dipasarkan.
4. Mengolah Data Dengan Algoritma *K-Means*
   Maka dilakukan pengolahan ataupun proses penyelesaian masalah dengan menggunakan algoritma *K-means*.
5. Mengolah *Rapidminer*
   Mengolah data dengan menggunakan *rapidminer versi 9.2.*
6. Hasil
   Hasil yang didapatkan dari penelitian ini adalah melakukan pengclusteran jumlah benih kelapa sawit menggunakan metode *K-Means Clustering*.

2.5. Proses Pengumpulan Data set

Dalam melakukan penelitian, terdapat metode pengumpulan data, yaitu adalah:

1. Penelitian kepustakaan yaitu data di peroleh dari perpustakaan dan jurnal yang terkait dengan penelitian ini sebagai bahan refrensi.
2. Penelitian lapangan yaitu penelitian yang dilakukan langsung dengan menggunakan beberapa teknik :
   a. Pengamatan Langsung (*Observasi*)
      Penulis melakukan pengamatan secara langsung ke PPKS Marihat untuk mengambil data yang diperlukan .
   b. Wawancara
      Penulis melakukan wawancara dengan memberikan pertanyaan kepada pegawai di bagian persib (persipan benih) di PPKS Marihat.

2.6. Analsis Data

Penelitian ini menggunakan data benih untuk perhitungan manual yang diperoleh dari proses *clustering* digunakan 99 data dengan total atribut 2 didalamnya. Dataset tersebut dapat dilihat pada Tabel 1 sebagai berikut :

Tabel 1. Benih Yang Akan dicluster

No	Barang	Jlh_Netto	Jlh_Bruto
1	DxP PPKS 540	1230	110
2	DxP Yangambi	2140	20
3	DxP Simalungun	760	0
4	DxP PPKS 540	1070	30
5	DxP PPKS 540	640	10
6	DxP Yangambi	1360	10
7	DxP Simalungun	1840	40
8	DxP Simalungun	880	40
9	DxP Simalungun	1950	10
10	DxP Simalungun	1250	0
11	DxP Yangambi	1630	50
12	DxP Yangambi	1850	7
13	DxP Yangambi	1700	55
14	DxP Simalungun	770	20
15	DxP Yangambi	2600	20
16	DxP PPKS 540	1040	5
17	DxP Yangambi	790	110
18	DxP PPKS 540	2160	0
19	DxP PPKS 540	1370	0
20	DxP Simalungun	1330	40

No	Barang	Jlh_Netto	Jlh_Bruto
….	….	….	…
91	DxP PPKS 540	1590	49
92	DxP Simalungun	1340	110
93	DxP Yangambi	2086	247
94	DxP Simalungun	1080	23
95	DxP Simalungun	1500	55
96	DxP Simalungun	2310	77
97	DxP Yangambi	2740	267
98	DxP Simalungun	2740	247
99	DxP Simalungun	1880	0

3. Hasil dan Pembahasan

Hasil penelitian ini disajikan sesuai dengan penelitian yang dilakukan. Data yang digunakan dalam penelitian ini adalah data nama barang, jlh netto, jlh bruto, data diperoleh dari PPKS Marihat. Data tersebut dikelompokkan menjadi 3 bagian yaitu benih amat baik, benih baik, benih kurang baik Berikut langkah-langkah penyelesaian yang dilakukan penulis dalam *clustering* benih sawit menggunakan *K-means* :

1. Menentukan jumlah data yang ingin dicluster, dimana sampel data jumlah benih yang akan digunakan dalam proses *clustering* adalah data netto dan bruto. Jumlah data dapat kita lihat pada tabel 2 diatas yaitu berjumlah 99 data.
2. 2.Menetapkan nilai K jumlah cluster benih sawit sebanyak 3 cluster (k-3). Cluster yang dibentuk yaitu cluster kurang baik, cluster baik, dan cluster amat baik.
3. Menentukan nilai centroid (pusat cluster) awal yang telah di tentukan secara random berdasarkan nilai variabel data yang akan di cluster sebanyak yang ditentukan
   Nilai centroid data awal untuk iterasi 1 adalah:

Tabel 2. Pusat Cluster Iterasi 1

data ke-26 pusat cluster 1	3380	40
data ke-28 pusat cluster 2	1170	143
data ke-5 pusat cluster 3	640	10

4. Mengalokasikan masing-masing data obyek ke dalam centroid yang paling terdekat, bandingkan nilai C0,C1,C2, Yang termasuk dalam clustering 0 sebanyak 12 dan yang termasuk clustering 1 sebanyak 67 dan yang termasuk clustering 2 sebanyak 10.

Tabel 3. Cluster Baru

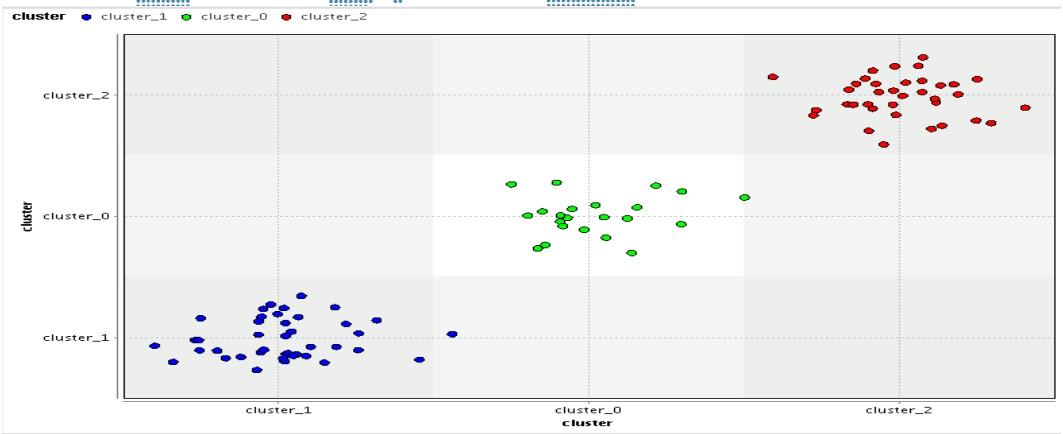
C0	C1	C2
2661,333333	1571,402	778,2
144,4166667	100,9351	55

Yang termasuk dalam clustering 0 sebanyak 22 dan yang termasuk clustering 1 sebanyak 41 dan yang termasuk clustering 2 sebanyak 36. Karena hasil cluster dari kelompok data 10 dengan sebelumnya yaitu data kelompok 9 sama, maka iterasi sudah dapat dihentikan, maka didapatlah hasil cluster terbaik yaitu :

Tabel 4. Hasil Cluster terbaik

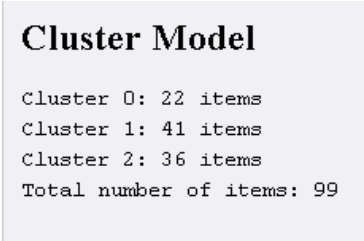
C0	C1	C2
22	41	36

Hasil dari cluster dan performance mana benih yang amat baik, baik, dan kurang baik, benih dapat kita lihat dari jumlah item/data yang besar, sedang dan kecil, besar= benih kurang baik, sedang = benih baik, kecil = benih amat baik, dan juga dapat kita lihat dari grafik clusternya.



Gambar 2. Cluster Model

Hasil pengujian dari aplikasi rapid miner sama dengan hasil dari perhitungan manual yaitu:



Gambar 3. Cluster Model

4. Kesimpulan

Hasil penelitian penulis menyimpulkan bahwa pengujian menggunakan perhitungan manual dan menggunakan Software rapid miner, didapatlah benih yang amat baik, baik, dan kurang baik. Jumlah iterasi yang dihasilkan berbeda-beda untuk setiap dataset sampai menghasilkan pengelompokkan cluster yang sama dengan hasil benih yang amat baik, baik, dan kurang baik yaitu : benih amat baik terdapat 22 benih, benih yang baik terdapat 36 benih, benih yang kurang baik terdapat 41 benih.

Daftar Pustaka

[1] F. Roosmawati, A. Widjajanto, T. Ningsih, And M. S. Gunawan, “Manajemen Pemupukan Tanaman Belum Menghasilkan Kelapa Sawit (Elaeis Guineensis Jacq) Di Lahan Gambut Pt . Xxx Kabupaten Tapanuli Selatan Provinsi Sumatera Utara,” Vol. 7, Pp. 144–160, 2024.

[2] S. Parlinsa Elvani, A. Rachma Utary, And R. Yudaruddin, “Peramalan Jumlah Produksi Tanaman Kelapa Sawit Dengan Menggunakan Metode Arima (Autoregressive Integrated Moving Average),” *J. Manaj.*, Vol. 8, No. 1, P. 2016, 2016, [Online]. Available: Http://Journal.Feb.Unmul.Ac.Id

[3] Y. P. Anggriani, A. Arif, And T. Informatika, “Implementasi Algoritma K-Means Clustering Dalam Menentukan,” Vol. 8, No. 2, Pp. 1820–1825, 2024.

[4] D. Haryadi, “Penerapan Algoritma K-Means Clustering Pada Produksi Perkebunan

Kelapa Sawit Menurut Provinsi," Vol. 1089, Pp. 1–15, 2021.
[5] E. Budianita, A. Nazir, And W. Mahesa, "Penerapan K-Means Clustering Pada Data Obat / Alkes Di Apotik Rsud Selasih," Pp. 220–229, 2023.
[6] D. Destarna, K. Saputra, K. Auliasari, A. Faisol, And T. Informatika, "Penerapan Metode K-Means Clustering Untuk Pemetaan Pengelompokan Lahan Produksi Jagung Di Kabupaten Pasuruan," Vol. 8, No. 5, Pp. 8364–8372, 2024.
[7] Q. S. Osadi *Et Al.*, "Penerapan Metode K-Means Clustering Untuk Pengelompokan Paket Berdasarkan," Vol. 04, No. 03, Pp. 275–281, 2024.
[8] S. Lestari, R. Nazila, L. Aulia Ulhar, And M. Zidane, "Implementasi Data Mining Dalam Menentukan Penjualan Alat Perabot Dengan Menggunakan Metode K-Means Clustering Pada Pt.Xyz," *J. Sist. Inf. Dan Ilmu Komput.*, Vol. 2, No. 1, Pp. 271–278, 2024, [Online]. Available: Https://Doi.Org/10.59581/Jusiik-Widyakarya.V2i1.2824
[9] M. Melisa, Syaiful Zuhri Harahap, "Implementasi Data Mining Untuk Klustering Stunting Gizi Pada Balita Dipuskesmas Sigambal Meggunakan Metode K-Medoids Dan K-Means," *Sport. Cult.*, Vol. 15, No. 1, Pp. 72–86, 2024, Doi: 10.25130/Sc.24.1.6.
[10] A. W. Aranski, S. Astiti, And R. A. Putra, "Pengaplikasian Data Mining Dalam Mengelompokan Data Penerima Bantuan Subsidi Rumah Dengan Menggunakan Metode K-Means Clustering," Vol. 6, No. 1, Pp. 480–489, 2024, Doi: 10.47065/Bits.V6i1.5366.
[11] A. Pujianti And M. Mulyawan, "Implementasi Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Status Kematian Bayi Di Jawa Barat," *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 7, No. 1, Pp. 459–463, 2023, Doi: 10.36040/Jati.V7i1.6347.
[12] L. P. Naibaho, A. P. Windarto, And A. Info, "Jurnal Jpilkom (Jurnal Penelitian Ilmu Komputer) Penerapan Data Mining Pada Tingkat Penghunian Kamar Pada Hotel Berbintang Berdasarkan Provinsi," Vol. 1, No. 1, 2023.
[13] N. Afiasari, N. Suarna, And N. Rahaningsi, "Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering Dengan Metode K-Means," *J. Saintekom*, Vol. 13, No. 1, Pp. 100–110, 2023, Doi: 10.33020/Saintekom.V13i1.402.
[14] M. P. A. Ariawan, I. B. A. Peling, And G. B. Subiksa, "Prediksi Nilai Akhir Matakuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus : Matakuliah Pemrograman Dasar)," *J. Nas. Teknol. Dan Sist. Inf.*, Vol. 9, No. 2, Pp. 122–131, 2023, Doi: 10.25077/Teknosi.V9i2.2023.122-131.