

Implementasi K-Means Clustering Dalam Menentukan Kualitas Biji Kelapa Sawit (Kasus PPKS Marihat)

Yuda Hawari^{1,*}, Jaya Tata Hardinata², Rizki Alfadillah Nasution³

^{1,2,3}Program Studi Sistem Informasi, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
Email: ^{1*}yudahawari123@gmail.com, ²jayatatahardinata@amiktunasbangsa.ac.id, ³rizkialfadillah@amiktunasbangsa.ac.id
(* : yudahawari123@gmail.com)

Abstract

In determining the quality of oil palm seeds in Marihat PPKS, laboratory tests are still being carried out. This way of grouping really requires a lot of resources such as manpower, time and cost as well as other laboratory equipment. With the development of science, later data mining can be used to determine the quality of palm seeds based on quality standards. With the existence of data mining, the author hopes to help PPKS Marihat in determining the quality of oil palm seeds.

Keywords: Marihat PPKS, Oil Palm Seed Quality, K-Means Method.

Abstrak

Dalam menentukan kualitas biji kelapa sawit pada PPKS Marihat masih dilakukan cara uji laboratorium. Cara pengelompokan tersebut sangat membutuhkan sumber daya yang sangat banyak seperti tenaga, waktu dan biaya serta peralatan laboratorium lainnya. Dengan perkembangan ilmu pengetahuan nantinya data mining dapat digunakan dalam menentukan kualitas biji sawit berdasarkan standar kualitasnya. Dengan adanya Data mining maka penulis mengharapkan dapat membantu PPKS Marihat dalam menentukan kualitas kelapa biji sawit.

Kata Kunci: PPKS Marihat, Kualitas Biji Kelapa Sawit, Metode K-Means.

1. PENDAHULUAN

Data Mining merupakan suatu metode untuk menemukan pengetahuan dalam suatu tumpukan data yang cukup besar dengan proses menggali dan menganalisa sejumlah data yang sangat besar dalam memperoleh sesuatu yang benar, baru dan bermanfaat sehingga dapat ditemukan suatu corak atau pola dalam data tersebut [1], [2]. “Data Mining adalah salah satu cabang ilmu kecerdasan buatan dalam penggalian pola-pola untuk mengubah data menjadi informasi. Data Mining merupakan proses yang mempekerjakan satu atau lebih teknik pembelajaran computer untuk menganalisa dan mengekstraksi pengetahuan (knowledge) secara otomatis” [3], [4]. Istilah Data Mining mulai dikenal sejak tahun 1990, “Munculnya Data Mining didasarkan pada jumlah data yang tersimpan dalam basis data yang semakin besar” [5]. Data mining merupakan salah satu cabang artificial intelligence yang bekerja dengan cara mencari pola atau informasi menarik dengan menguraikan penemuan pengetahuan di dalam database dengan menggunakan teknik tertentu [6]. Ada banyak teknik yang terkenal dalam data mining diantaranya klastering, klasifikasi, estimasi, asosiasi dan peramalan [7], [8]. Diantara teknik tersebut, klastering dan klasifikasi merupakan metode yang sering digunakan. Beberapa metode terkenal dari klastering adalah k-medoids dan k-means [9], [10]. Berbicara metode k-means dan k-medoids, menurut [11] bahwa metode k-means efisien untuk kumpulan data yang lebih kecil dan k-medoids berperforma lebih baik untuk dataset besar. Selain itu menurut [12] k-medoids muncul sebagai penanggulangan kelemahan metode k-means yang sensitif terhadap outlier. Sementara metode klasifikasi yang sering digunakan adalah C4.5 [13].

Pusat Penelitian Kelapa Sawit (PPKS) unit Marihat Pematangsiantar merupakan salah satu perusahaan yang melakukan penelitian terhadap kecambah kelapa sawit untuk mendapatkan biji kelapa sawit yang unggul. Dalam mendukung pelaksanaan tugasnya sebagai lembaga riset [14]. Kelapa sawit merupakan tanaman perkebunan yang menjadi faktor penting dalam perekonomian rakyat, penyerapan tenaga kerja dan sumber devisa bagi negara. Kelapa sawit juga mampu menghasilkan minyak nabati yang sering dimanfaatkan sebagai bahan dasar untuk produk-produk industri. Tingginya penggunaan minyak kelapa sawit membuat pengelola perkebunan kelapa sawit mengembangkan perkebunan yang dikelolanya. Dalam memproduksi kelapa sawit di PPKS Marihat terus meningkat tetapi peningkatan tersebut tidak diikuti dengan pengawasan mutu biji kelapa sawit. Dalam menentukan mutu dari biji kelapa sawit dipengaruhi oleh beberapa faktor diantaranya adalah jenis tanah dan umur tanaman [15], [16].

Pada industri kelapa sawit, ketersediaan kelapa sawit harus dipertahankan, kuantitas dan kualitasnya. Terdapat tiga subsistem utama dalam kegiatan produksi biji kelapa sawit, yakni pemanenan, pengangkutan dan pengolahan. Di antara ketiganya terdapat saling keterkaitan, satu hambatan di dalam satu subsistem berpengaruh terhadap kinerja subsistem yang lain. Misalnya hambatan di pengangkutan biji kelapa sawit dari kebun ke pabrik menyebabkan keterlambatan, yang kemudian mengganggu produksi biji kelapa sawit. Dalam hal mempermudah serta menentukan hasil pengelompokan biji kelapa sawit yang ada di PPKS Marihat Pematangsiantar maka penelitian ini akan menggunakan Algoritma Data Mining Metode K-Means. Berdasarkan latar belakang masalah di atas, Diharapkan penelitian ini dapat membantu pihak PPKS Marihat dalam mengcluster biji kelapa sawit sehingga dapat mempermudah proses dalam mendapatkan biji kelapa sawit yang baik untuk digunakan dan diolah lebih lanjut sehingga dapat lebih

meningkatkan perekonomian dari perusahaan. Penelitian ini juga diharapkan dapat menjadi referensi bagi peneliti selanjutnya yang berkaitan dengan pengguna Algoritma yang digunakan.

2. METODOLOGI PENELITIAN

2.1. Metode Penelitian

Metode penelitian merupakan langkah yang dimiliki dan dilakukan oleh peneliti dalam rangka untuk mengumpulkan informasi atau data serta melakukan investigasi pada data yang telah didapatkan. Adapun metode penyelesaian yang digunakan dalam penelitian ini yaitu metode Data Mining dengan Algoritma K-Means.

2.1.1. Metode Pengumpulan Data

Dalam melakukan penelitian untuk mendapatkan data dan informasi, maka metode yang digunakan dalam proses pengumpulan data sebagai berikut:

1. Metode Observasi, melihat serta mempelajari permasalahan yang ada di lapangan yang berkaitan dengan objek yang diteliti.
2. Metode Kajian Literatur, mencari bahan yang mendukung dalam pendefinisian masalah melalui buku-buku, makalah, *internet*, yang erat kaitannya dengan objek permasalahan.

2.1.2. Pengolahan Data

Pengolahan data yang digunakan penelitian ini menggunakan data kuantitatif dengan teknik analisis data yang menggunakan jenis statistik *Deskriptif*. Data yang diperoleh kemudian diolah dengan *RapidMiner* untuk mencari keakuratan data. Adapun data penelitian yang digunakan sebagai berikut :

Tabel 1. Data Penelitian

Bulan	Tahun			
	2017	2018	2019	2020
Januari	5.119.497	5.737.271	5.647.313	5.776.781
Februari	1.302.952	1.248.269	1.253.294	1.312.253
Maret	8.113.852	8.496.029	9.513.208	9.984.215
April	1.849.969	2.691.270	2.884.406	3.022.565
Mei	853.648	900.318	815.667	843.047
Juni	1.933.721	1.464.226	1.665.397	1.561.147
Juli	2.849.710	3.786.477	3.987.883	3.823.221
Agustus	456.608	383.617	381.661	371.717
September	2.784.180	3.086.889	5.235.299	5.471.407
Oktober	568.719	386.113	348.356	348.015
November	911.697	1.037.402	1.133.347	1.134.606
Desember	3.199.481	3.793.622	4.049.156	4.267.023

Data penelitian terdiri dari 12 Bulan dan digunakan data dari tahun 2017- 2020 (dalam satuan biji). Kualitas biji kelapa sawit dapat dikelompokkan kedalam 3 *cluster* yaitu *cluster* Baik, Sedang, dan Buruk.

3.2. Analisis Data

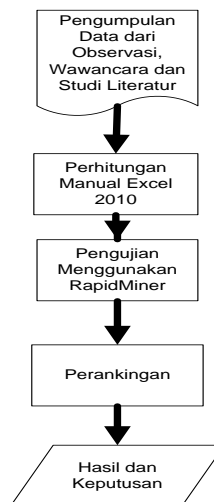
Di dalam analisis dan metode perancangan sistem terdapat rancangan yang digunakan dalam membangun sistem atau perhitungan-perhitungan *Algoritma* yang digunakan.

3.2.1. Alat Analisis Data

Alat yang digunakan dalam menganalisis data dalam penelitian menggunakan *Microsoft Excel 2010*. *Microsoft Excel* banyak digunakan untuk survei penelitian, terutama dalam pengumpulan data. Untuk pengujian data yang digunakan dari tempat penelitian menggunakan *Microsoft Excel*

3.2.2. Instrumen Penelitian

Penelitian ini menggunakan penelitian dengan metode kuantitatif. Metode kuantitatif merupakan penelitian yang menggunakan data berupa angka sebagai alat menganalisis keterangan dengan menggunakan teknik wawancara. Alur instrumen penelitian yang digambarkan sebagai berikut :



Gambar 1. Alur Instrumen Penelitian

Diagram alir model penelitian seperti yang disajikan pada Gambar 1 menjelaskan rancangan penelitian yang dilakukan menggunakan metode *K-Means* yang terdiri dari :

- a) Analisis Masalah
Menganalisis masalah yang terkait dengan mengelompokkan biji kelapa sawit dan menentukan parameter atau kriteria apa yang digunakan.
- b) Mempelajari Literatur
Penelitian ini harus didasari rujukan yang digunakan untuk memperoleh informasi dalam penelitian.
- c) Menetapkan Metode
Menetapkan metode untuk memecahkan masalah. Pada penelitian ini metode yang digunakan yaitu *Data Mining* dengan *Algoritma K-means*.
- d) Mengumpulkan Data
Pengumpulan data dalam penelitian ini dilakukan dengan mengambil data yang telah ada di instansi terkait dan melakukan wawancara dengan pihak PPKS Marihat Pematangsiantar selama 1 minggu yang dimulai pada 1 Februari 2021 sampai dengan 8 Februari 2021.
- e) Mengelolah Data
Melakukan pengolahan data dengan menggunakan *Algoritma K-Means*. Pengolahan data disini menggunakan Microsoft Excel 2010.
- f) Menguji data
Pengujian data dilakukan dengan menggunakan aplikasi *RapidMiner 5.3*.
- g) Kesimpulan
Kesimpulan yang didapatkan dalam mengelompokkan hasil biji kelapa sawit. Jika diperoleh kualitas biji kelapa sawit yang rendah lebih banyak maka dapat menjadi informasi bagian PPKS Marihat Pematangsiantar agar dapat mengelola kembali biji kelapa sawit untuk mendapatkan hasil kualitas yang lebih baik.

3.2.3. Pemodelan Metode

Permasalahan yang sudah dianalisa, selanjutnya penulis menganalisa data menggunakan *Data Mining Algoritma K-Means Clustering*. *Algoritma K-Means* selanjutnya akan melakukan pengulangan langkah-langkah berikut sampai terjadi kestabilan:

- a) Tentukan berapa jumlah klaster yang ingin ditetapkan pusat *cluster k*.
- b) Menggunakan jarak euclidean dan kemudian menghitung setiap data ke pusat cluster.
- c) Kategorikan data ke dalam *cluster* dengan jarak yang terpendek dengan menggunakan persamaan.
- d) Menghitung pusat *cluster* dengan menggunakan persamaan.
- e) Silahkan ulangi langkah dua sampai empat sehingga sudah tidak ada lagi data yang berpindah ke *cluster* yang lain.

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Hasil dan pembahasan pada bab ini disajikan sesuai penelitian yang dilakukan. Data yang digunakan dalam penelitian ini adalah data biji kelapa sawit. Pada penelitian ini digunakan tools RapidMiner sebagai pengujian data yang telah di analisa. Berikut adalah bagian-bagian dari RapidMiner beserta isi dari data yang dibuat. Data penelitian terdiri dari 12 Bulan dan digunakan data dari tahun 2017- 2020 dan Kualitas biji kelapa sawit dapat dikelompokkan

kedalam 3 cluster yaitu cluster Baik, Sedang, dan Buruk. Berikut adalah data penelitian yang digunakan. Dapat dilihat pada Tabel 2 berikut.

Tabel 2. Data Penelitian

Bulan	Tahun			
	2017	2018	2019	2020
Januari	5.119.497	5.737.271	5.647.313	5.776.781
Februari	1.302.952	1.248.269	1.253.294	1.312.253
Maret	8.113.852	8.496.029	9.513.208	9.984.215
April	1.849.969	2.691.270	2.884.406	3.022.565
Mei	853.648	900.318	815.667	843.047
Juni	1.933.721	1.464.226	1.665.397	1.561.147
Juli	2.849.710	3.786.477	3.987.883	3.823.221
Agustus	456.608	383.617	381.661	371.717
September	2.784.180	3.086.889	5.235.299	5.471.407
Oktober	568.719	386.113	348.356	348.015
November	911.697	1.037.402	1.133.347	1.134.606
Desember	3.199.481	3.793.622	4.049.156	4.267.023

Selanjutnya masuk dalam tahap perhitungan menggunakan metode *k-means*:

1. Penentuan Pusat Cluster Awal

Penentuan titik cluster ini dilakukan dengan mengambil nilai terbesar (Maximum) untuk cluster Baik (C1), nilai rata-rata (Average) untuk cluster Sedang (C2), dan nilai terkecil (Minimum) untuk cluster Buruk (C3) sebagai berikut:

Tabel 3. Nilai Cluster

Cluster	2017	2018	2019	2020
Max (C1)	8.113.852,0	8.496.029,0	5.647.313,0	9.984.215,0
Ave (C2)	2.495.336,2	2.750.958,6	7.311.582,3	3.159.666,4
Min (C3)	456.608,0	383.617,0	348.356,0	348.015,0

2. Perhitungan Jarak Cluster

Untuk menghitung jarak antara data dengan pusat cluster memakai persamaan (1):

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (1)$$

a. Cluster Baik

$$C1, \text{Jan} = \sqrt{(5119497 - 8.113.852,0)^2 + (5737271 - 8.496.029,0)^2 + (5647313 - 5.647.313,0)^2 + (5776781 - 9.984.215,0)^2} = 7016021,1$$

$$C1, \text{Feb} = \sqrt{(1302952 - 8.113.852,0)^2 + (1248269 - 8.496.029,0)^2 + (1253294 - 5.647.313,0)^2 + (1312253 - 9.984.215,0)^2} = 15567514$$

$$C1, \text{Mar} = \sqrt{(8113852 - 8.113.852,0)^2 + (8496029 - 8.496.029,0)^2 + (9513208 - 5.647.313,0)^2 + (9984215 - 9.984.215,0)^2} = 0$$

$$C1, \text{Apr} = \sqrt{(1849969 - 8.113.852,0)^2 + (2691270 - 8.496.029,0)^2 + (2884406 - 5.647.313,0)^2 + (3022565 - 9.984.215,0)^2} = 12858345$$

$$C1, \text{Mei} = \sqrt{(853648 - 8.113.852,0)^2 + (900318 - 8.496.029,0)^2 + (815667 - 5.647.313,0)^2 + (843047 - 9.984.215,0)^2} = 16419914$$

$$C1, \text{Jun} = \sqrt{(1933721 - 8.113.852,0)^2 + (1464226 - 8.496.029,0)^2 + (1665397 - 5.647.313,0)^2 + (1561147 - 9.984.215,0)^2} = 14838345$$

$$C1, \text{Jul} = \sqrt{(2849710 - 8.113.852,0)^2 + (3786477 - 8.496.029,0)^2 + (3987883 - 5.647.313,0)^2 + (3823221 - 9.984.215,0)^2} = 10880172$$

$$C1, \text{Aug} = \sqrt{(456608 - 8.113.852,0)^2 + (383617 - 8.496.029,0)^2 + (381661 - 5.647.313,0)^2 + (371717 - 9.984.215,0)^2} = 17327143$$

$$C1, \text{ Sep} = \sqrt{(2784180 - 8.113.852,0)^2 + (3086889 - 8.496.029,0)^2 + (5235299 - 5.647.313,0)^2 + (5471407 - 9.984.215,0)^2} = 9814791,9$$

$$C1, \text{ Oct} = \sqrt{(568719 - 8.113.852,0)^2 + (386113 - 8.496.029,0)^2 + (348356 - 5.647.313,0)^2 + (348015 - 9.984.215,0)^2} = 17307531$$

$$C1, \text{ Nov} = \sqrt{(911697 - 8.113.852,0)^2 + (1037402 - 8.496.029,0)^2 + (1133347 - 5.647.313,0)^2 + (1134606 - 9.984.215,0)^2} = 16001244$$

$$C1, \text{ Dec} = \sqrt{(3199481 - 8.113.852,0)^2 + (3793622 - 8.496.029,0)^2 + (4049156 - 5.647.313,0)^2 + (4267023 - 9.984.215,0)^2} = 10431003$$

Berikut adalah hasil dari perhitungan jarak cluster:

Tabel 4. Hasil Iterasi 1

Bulan	Cluster			Jarak	Pengelompokan		
	Baik	Sedang	Buruk		C1	C2	C3
Januari	7.016.021,1	5.409.615,6	10.390.119,9	5.409.615,6		1	
Februari	15.567.513,9	3.227.376,2	1.792.369,6	1.792.369,6			1
Maret	0,0	12.352.427,3	17.357.865,0	0,0	1		
April	12.858.345,3	689.682,5	4.566.342,6	689.682,5		1	
Mei	16.419.913,5	4.073.937,5	942.367,7	942.367,7			1
Juni	14.838.345,1	2.552.814,1	2.560.442,3	2.552.814,1			1
Juli	10.880.171,6	1.571.388,1	6.529.135,2	1.571.388,1		1	
Agustus	17.327.142,9	4.979.375,9	40.878,0	40.878,0			1
September	9.814.791,9	3.194.044,7	7.928.217,2	3.194.044,7		1	
Oktober	17.307.531,1	4.964.990,6	112.138,8	112.138,8			1
November	16.001.243,8	3.649.647,5	1.367.288,2	1.367.288,2			1
Desember	10.431.002,9	1.937.976,7	6.943.056,3	1.937.976,7		1	

Berdasarkan matrik yang didapatkan pada tabel di atas maka didapatkan pengelompokan sebagai berikut :

C1 = Maret

C2 = Januari, April, Juli, September, Desember

C3 = Februari, Mei, Juni, Agustus, Oktober, November

Setelah didapatkan hasil dari setiap *cluster* kemudian pusat *cluster* baru dihitung berdasarkan data member tiap – tiap *cluster* yang sudah didapatkan dengan menjumlahkan setiap member dari tiap cluster dan dibagi dengan jumlah dari setiap cluster yang diperoleh. Berikut adalah pusat cluster baru yang diperoleh dan akan digunakan dalam perhitungan cluster 2 :

Tabel 5. Pusat Cluster Baru

Cluster	2017	2018	2019	2020
C1	8.113.852	8.496.029	9.513.208	9.984.215
C2	3.160.567,4	3.819.105,8	4.360.811,4	4.472.199,4
C3	1.004.557,5	903.324,2	932.953,7	928.464,2

Iterasi selanjutnya dilakukan dengan cara yang sama hingga tidak ada perubahan data dalam suatu cluster.

a. Cluster Baik

$$C1, \text{ Jan} = \sqrt{(5119497 - 8.113.852,0)^2 + (5737271 - 8.496.029,0)^2 + (5647313 - 5.647.313,0)^2 + (5776781 - 9.984.215,0)^2} = 7016021,1$$

$$C1, \text{ Feb} = \sqrt{(1302952 - 8.113.852,0)^2 + (1248269 - 8.496.029,0)^2 + (1253294 - 5.647.313,0)^2 + (1312253 - 9.984.215,0)^2} = 15567514$$

$$C1, \text{ Mar} = \sqrt{(8113852 - 8.113.852,0)^2 + (8496029 - 8.496.029,0)^2 + (9513208 - 5.647.313,0)^2 + (9984215 - 9.984.215,0)^2} = 0$$

.....

$$C1, \text{ Sep} = \sqrt{(2784180 - 8.113.852,0)^2 + (3086889 - 8.496.029,0)^2 + (5235299 - 5.647.313,0)^2 + (5471407 - 9.984.215,0)^2} = 9814791,9$$

$$C1, \text{Oct} = \sqrt{(568719 - 8.113.852,0)^2 + (386113 - 8.496.029,0)^2 + (348356 - 5.647.313,0)^2 + (348015 - 9.984.215,0)^2} = 17307531$$

$$C1, \text{Nov} = \sqrt{(911697 - 8.113.852,0)^2 + (1037402 - 8.496.029,0)^2 + (1133347 - 5.647.313,0)^2 + (1134606 - 9.984.215,0)^2} = 16001244$$

$$C1, \text{Dec} = \sqrt{(3199481 - 8.113.852,0)^2 + (3793622 - 8.496.029,0)^2 + (4049156 - 5.647.313,0)^2 + (4267023 - 9.984.215,0)^2} = 10431003$$

Berikut adalah hasil pengelompokan dari iterasi 2, dapat dilihat pada Tabel 6.:

Tabel 6. Hasil Iterasi 2

Bulan	Cluster			Jarak	Pengelompokan		
	Baik	Sedang	Buruk		C1	C2	C3
Januari	7.016.021,1	3.297.541,9	9.275.296,7	3.297.541,9		1	
Februari	15.567.513,9	5.449.941,5	676.711,2	676.711,2			1
Maret	0,0	10.165.537,2	16.242.494,3	0,0	1		
April	12.858.345,3	2.696.459,6	3.479.209,0	2.696.459,6		1	
Mei	16.419.913,5	6.291.263,4	209.368,0	209.368,0			1
Juni	14.838.345,1	4.773.897,3	1.454.206,5	1.454.206,5			1
Juli	10.880.171,6	811.138,4	5.424.884,2	811.138,4		1	
Agustus	17.327.142,9	7.194.551,7	1.088.225,7	1.088.225,7			1
September	9.814.791,9	1.562.355,1	6.861.743,0	1.562.355,1		1	
Oktober	17.307.531,1	7.184.136,2	1.065.897,9	1.065.897,9			1
November	16.001.243,8	5.861.002,5	330.532,7	330.532,7			1
Desember	10.431.002,9	376.018,8	5.833.369,8	376.018,8		1	

Berdasarkan matrik yang didapatkan pada tabel di atas maka didapatkan pengelompokan sebagai berikut :

C1 = Maret

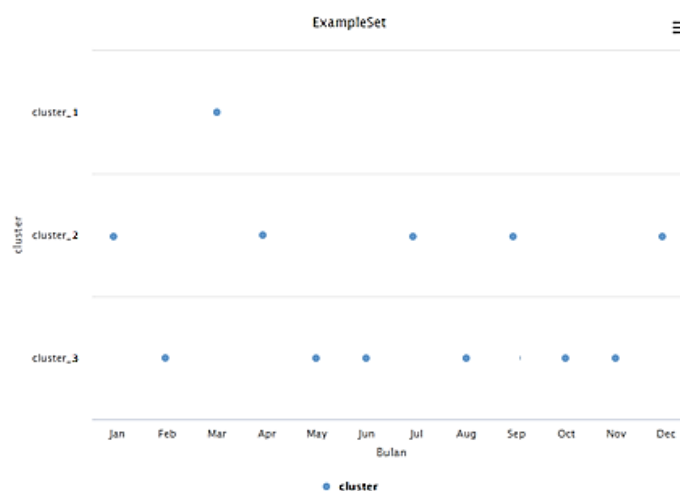
C2 = Januari, April, Juli, September, Desember

C3 = Februari, Mei, Juni, Agustus, Oktober, November

Dari hasil iterasi 2 dapat dilihat bahwa hasil yang diperoleh sama dengan iterasi 1 maka perhitungan dihentikan sampai iterasi 2. Untuk hasil pengelompokan dari kualitas biji kelapa sawit pada cluster Baik berada di bulan (Maret), Cluster Sedang berada di bulan (Januari, April, Juli, September, Desember), dan untuk cluster Buruk berada di bulan (Februari, Mei, Juni, Agustus, Oktober, November)

3.2. Implementasi

Setelah melakukan proses sistem seperti yang tertera pada gambar 4. maka akan muncul hasil dari RapidMiner sebagai berikut:



Gambar 2. Grafik Hail Cluster

Berdasarkan gambar 2 dapat dilihat bahwa cluster 3 adalah cluster rendah dan cluster 1 adalah cluster tinggi. Dari Gambar 3 juga dapat dilihat bahwa terdapat 1 alternatif yang berada pada cluster Baik dan 7 alternatif pada cluster Sedang dan 4 alternatif pada cluster Buruk seperti yang tertera pada gambar 3.

Result History

Cluster Model

Cluster 1 : 1 items

Cluster 2 : 5 items

Cluster 3 : 6 items

Total number of items: 12

Gambar 3. Cluster Model

Data di atas akan diuji dengan sampel sebanyak 4 tahun dan diselesaikan dengan menggunakan *tools RapidMiner* sehingga diperoleh hasil cluster seperti pada Gambar 4 berikut :

Row No.	Bulan	cluster	2017	2018	2019	2020
1	Januari	cluster_2	5119497	5737271	5647313	5776781
2	Februari	cluster_3	1302952	1248269	1253294	1312253
3	Maret	cluster_1	8113852	8496029	9513208	9984215
4	April	cluster_2	1849969	2691270	2884406	3022565
5	Mei	cluster_3	853648	900318	815667	843047
6	Juni	cluster_3	1933721	1464226	1665397	1561147
7	Juli	cluster_2	2849710	3786477	3987883	3823221
8	Agustus	cluster_3	456608	383617	381661	371717
9	September	cluster_2	2784180	3086889	5235299	5471407
10	Oktober	cluster_3	568719	386113	348356	348015
11	November	cluster_3	911697	1037402	1133347	1134606
12	Desember	cluster_2	3199481	3793622	4049156	4267023

Gambar 4. Hasil Cluster

Pada Gambar 4 dapat dilihat bahwa hasil dari perhitungan manual adalah sama dengan hasil perhitungan menggunakan *tools RapidMiner*. Dalam melakukan validasi data perhitungan algoritma harus menghasilkan hasil akhir berupa pengelompokan dengan 3 *cluster*, serta data yang digunakan merupakan data yang *valid* dan sama dengan yang dipakai pada *tools RapidMiner*. Berikut ditampilkan hasil yang didapatkan dari perhitungan algoritma dan pengujian pada *RapidMiner*.

Tabel 7. Perbandingan Hasil Manual dan *RapidMiner*

No	Jenis Dagangan	Hasil Perhitungan Manual	Hasil <i>RapidMiner</i>
1	Januari	cluster_2	cluster_2
2	Februari	cluster_3	cluster_3
3	Maret	cluster_1	cluster_1
4	April	cluster_2	cluster_2
5	Mei	cluster_3	cluster_3
6	Juni	cluster_3	cluster_3
7	Juli	cluster_2	cluster_2
8	Agustus	cluster_3	cluster_3
9	September	cluster_2	cluster_2
10	Oktober	cluster_3	cluster_3
11	November	cluster_3	cluster_3
12	Desember	cluster_2	cluster_2

Berdasarkan tabel yang didapatkan pada tabel di atas maka didapatkan pengelompokan sebagai berikut :

C1 = Maret

C2 = Januari, April, Juli, September, Desember

C3 = Februari, Mei, Juni, Agustus, Oktober, November

4. KESIMPULAN

Kesimpulan yang didapat bahwa implementasi *K-Means Clustering* Dalam Menentukan Kualitas Biji Kelapa Sawit (Studi Kasus PPKS Marihat) menjadi pertimbangan pengambil keputusan (*decision maker*) dalam menentukan pengelompokan biji kelapa sawit. Jumlah Cluster atau *k* yang digunakan sebanyak 3 yaitu Cluster Baik, Sedang dan Cluster Buruk. Setelah didapatkan hasil dari setiap *cluster* kemudian pusat *cluster* baru dihitung berdasarkan data member tiap – tiap *cluster* yang sudah didapatkan. Iterasi selanjutnya dilakukan dengan cara yang sama hingga tidak ada perubahan data dalam suatu cluster. Hasil yang diperoleh sama dengan iterasi 1 maka perhitungan dihentikan sampai iterasi 2. Untuk hasil pengelompokan dari kualitas biji kelapa sawit pada cluster baik berada di bulan (Maret), Cluster sedang berada di bulan (Januari, April, Juli, September, Desember), dan untuk cluster buruk berada di bulan (Februari, Mei, Juni, Agustus, Oktober, November).

REFERENCES

- [1] Haviluddin, F. Agus, M. Azhari, And A. S. Ahmar, "Artificial Neural Network Optimized Approach For Improving Spatial Cluster Quality Of Land Value Zone," *Int. J. Eng. Technol.*, Vol. 7, No. 2, Pp. 80–83, 2018, Doi: 10.14419/Ijet.V7i2.2.12738.
- [2] R. A. Malik, "Perbandingan Algoritma K-Means Clustering Dengan Fuzzy C- Means Dalam Mengukur Tingkat Kepuasan Terhadap Televisi Dakwah," Vol. 3, No. 1, Pp. 10–21, 2018.
- [3] M. L. Sibuea And A. Safta, "Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering," *Jurteks*, Vol. 4, No. 1, Pp. 85–92, 2017, Doi: 10.33330/Jurteks.V4i1.28.
- [4] R. Rosmini, A. Fadlil, And S. Sunardi, "Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah," *It J. Res. Dev.*, Vol. 3, No. 1, P. 22, 2018, Doi: 10.25299/Itjrd.2018.Vol3(1).1773.
- [5] Mardalius, "Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means," Vol. IV, No. 2, Pp. 401–411, 2018.
- [6] F. Nasari And C. J. M. Sianturi, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Diare Di Kabupaten Langkat," *Cogito Smart J.*, Vol. 2, No. 2, P. 108, 2018, Doi: 10.31154/Cogito.V2i2.19.108-119.
- [7] S. Harlina, "Data Mining Pada Penentuan Kelayakan Kredit Menggunakan Algoritma K-Nn Berbasis Forward Selection Data Mining On Credit Feasibility Determination Using K-Nn Algorithm Based On Forward Selection," Vol. 11, No. 2, Pp. 236–244, 2018.
- [8] P. Alkhairi, I. S. Damanik, And A. P. Windarto, "Penerapan Jaringan Saraf Tiruan Untuk Mengukur Korelasi Beban Kerja Dosen Terhadap Peningkatan Jumlah Publikasi," *Pros. Semin. Nas. Ris. Inf. Sci.*, Vol. 1, No. September, P. 581, 2019, Doi: 10.30645/Senaris.V1i0.65.
- [9] M. R. Ridlo, S. Defiyanti, A. Primajaya, M. Rosyid Ridlo, S. Defiyanti, And A. Primajaya, "Implementasi Algoritme K-Means Untuk Pemetaan Produktivitas Panen Padi Di Kabupaten Karawang," *Citee 2017*, Pp. 426–433, 2017.
- [10] I. P. Yudi Prayoga, Heru Satria Tambunan, "Penerapan Clustering Pada Laju Inflasi Kota Di Indonesia Dengan Algoritma K-Means," *Brahmana*, Vol. 1, No. 1, Pp. 24–30, 2019.
- [11] F. Nurzaman, "Penerapan Algoritma K-Means Dalam Pengelompokan Lokasi Rumah Sakit Provider Pada Asuransi Kesehatan," Pp. 61–67, 2018.
- [12] T. I. S And F. I. Komputer, "Analisis Data Rawat Inap Rumah Sakit Kota Semarang Untuk Mengetahui Daerah Endemi Penyakit Menggunakan Algoritma K-Means," 2018.
- [13] A. D. I. Suradi, "Penerapan Data Mining Untuk Menentukan Rekomendasi Beasiswa Dengan Metode Algoritma C4.5," 2018.
- [14] I. Romli And A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," *J. Sains Komput. Inform. (J-SAKTI)*, Vol. 4, No. 2, Pp. 694–702, 2020.
- [15] N. Azwanti And E. Elisa, "Analisa Kepuasan Konsumen Menggunakan Algoritma C4.5," *Pros. Semin. Nas. Ilmu Sos. Dan Teknol.*, No. 3, Pp. 126–131, 2020.
- [16] T. R. I. B. Tusarwenda, "Penerapan Data Mining Dengan Algoritma C4.5 Dalam Prediksi Penjualan Botol Pada Cv. Seribukilo," 2018.