Journal of Machine Learning and Data Analytics (MALDA)
Volume 01, No.01, Januari 2022 Page: 45-54

Pengelompokan Hasil Panen Kelapa Sawit Dalam Produksi Per Blok Menggunakan Algoritma *K-Means*

Agustina Lili[1], Suhada[2], Saputra Widodo[3]

[123]Program Studi Teknik Informatika, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia

[1]liliagustina098@gmail.com

Abstrak− Hasil panen kelapa sawit merupakan hal yang sangat penting bagi sebuah perusahaan kelapa sawit terutama pada PT. Surya Intisari Raya Sei Mandau. Hal ini dibuktikan ketika penulis mencoba menentukan secara acak titik awal pusat cluster dari salah satu objek permulaan perhitungan. Jumlah keanggotaan cluster yang dihasilkan berjumlah sama ketika menggunakan data yang lain sebagai titik awal pusat cluster tersebut. Namun, hal ini hanya berpengaruh pada jumlah literasi yang dilakukan. Pengelompokan data (objek *Clustering*) adalah salah satu proses dari objek mining yang bertujuan untuk memartisi data yang ada ke dalam suatu atau lebih cluster data berdasarkan karakteristiknya. Algoritma *K-Means* dalam hasil panen kelapa sawit dalam produksi per blok berdasarkan variabel yang dibentuk per blok tiap PT. Surya Intisari Raya Sei Mandau. Pengelompokan hasil panen kelapa sawit di PT. Surya Intisari Raya Sei Mandau menggunakan metode *K-Means* dilakukan melalui 4 tahapan yaitu : nama per bloknya, tahun tanaman, afdeling dan luas (Ha) dan dibagi menjadi 3 *cluster* yaitu dengan *cluster* 1 tinggi, *cluster* 2 sedang dan *cluster* 3 rendah.

Kata kunci : Hasil Panen kelapa sawit; K-Means.

Abstract−*The yield of oil palm is very important for a plam oil company, especially at PT. Surya Intisari Raya Sei Mandau. This is proven when the author tries to randomly determine the starting point of the cluster center from one of the cluster center from one of the computational starting object. The number of cluster memberships generated. The same amount when using other data as the starting point for the center of the cluster. However, this only affects the amount of literacy performed. Data grouping (object clustering) is one of the processes of object mining that aims to partition existing data into one or more data cluters based on their chacrateristic. K-Means algoritm in palm oil yields n production per block based on the variables formed per block each PT. Surya Intisari Raya Sei Mandau the grouping of oil plam yields at PT. Surya Intisari Raya Sei Mandau using the K-Means method is carried out in 3 stages, namely the name per block, plant year, afdeling and the area (Ha). And divided into 3 clusters, with cluster 1 high, cluster 2 being medium, and cluster 3 being low.*

Keyword : *Plam Oil Harvest; K-Means.*

1. PENDAHULUAN

Panen merupakan salah satu kegiatan penting dalam pengolahan tanaman kelapa sawit menghasilkan. Pengolahan tanaman yang baik dan potensi produksi yang tertinggi tidak akan ada artinya jika panen tidak di laksanakan secara optimal. Jika hasil panen tidak mencapai hasil produksi maka upaya untuk mengoptimalkan produksi kegiatan panen, angkut harus dikelola dengan baik, karena kegiatan panen, angkut, olah merupakan kegiatan yang tidak terpisah dan menentukan kualitas panen. Dan faktor-faktor yang diduga berpengaruh terhadap penurunan produktivitas tanaman kelapa sawit adalah curah hujan, topografi, jenis pupuk, umur tanaman, jumlah populasi tanaman per bloknya, serta faktor penyebab turunnya produksi yaitu buah mentah dipanen dan buah busuk.

Curah hujan juga sangat berpengaruh terhadap produktivitas tanaman kelapa sawit. Cara yang tepat akan mempengaruhi kuantitas produksi, sedangkan waktu yang tepat akan mempengaruhi kualitas produksi, kegiatan panen kelapa sawit meliputi persiapan panen, cara pelaksanaan panen, sistem panen, pemeriksaan panen, serta pengangkutan tandan buah segar ke pabrik. Maka dapat dilakukan pengelompokan data jumlah panen kelapa sawit dalam produksi per bloknya dengan luas (Ha), Tahun tanam, dan berdasarkan masing-masing afdelingnya. Metode *k-means* dapat diterapkan pada pengelompokan hasil panen kelapa sawit dalam produksi per blok berdasarkan dari data tersebut. Dan dari data dapat dilihat karakteristik sehingga diketahui *cluster* tinggi, *cluster* sedang dan *cluster* rendah dengan presentasi hasil produksi pada setiap blok nya.

K-Means merupakan algoritma *clustering* yang berulang-ulang. Algoritma *K-Means* dimulai dengan pemilihan secara acak K. K merupakan banyaknya *Cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan centroid, mean atau "*means*". Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus *Euclidian* hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil). [1]. Metode ini merupakan metode pengelompokan non hirarki yang bertujuan untuk mengelompokkan objek sehingga jarak-jarak tiap objek ke pusat kelompok di dalam satu kelompok bernilai minimum [2].

Berdasarkan penelitian yang telah dilakukan dalam pemanfaatan metode data *mining* khususnya *K-Means* dalam pengelompokan. Dalam penelitian "Mengelompokkan Potensi Hasil Produksi dengan Metode *K-Means clustering* ". Metode yang digunakan dalam penelitian tersebut adalah *K-Means Clustering* dengan menggunakan data primer. Dan hasil dari peneliti tersebut disimpulkan bahwa kebenaran algoritma *K-Means* dalam mengelompokkan potensi hasil produksi dapat diselesaikan dengan algoritma *K-Means* dapat melakukan pengelompokan data dalam jumlah yang banyak [3]. Data mining merupakan metode yang sering dibutuhkan dalam pengolahan data berskala besar, maka data *mining* mempunyai akses penting pada bidang kehidupan diantaranya yaitu bidang industri, bidang keuangan, cuaca, ilmu, dan teknologi [4].

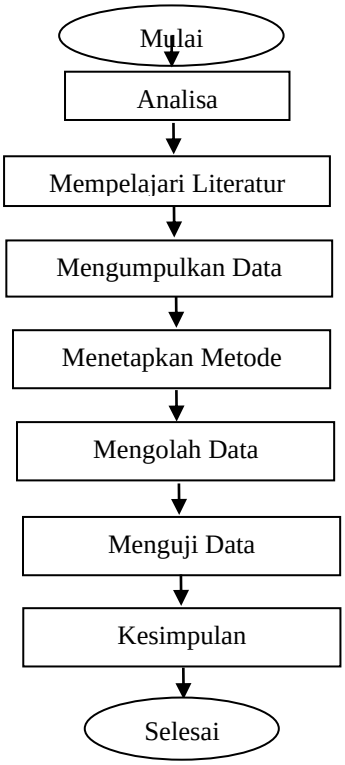
Maka dapat dilakukan pengelompokan data jumlah panen kelapa sawit dalam produksi per bloknya dengan luas (Ha), Tahun tanam, dan berdasarkan masing-masing afdelingnya. Metode *k-means* dapat diterapkan pada pengelompokan hasil panen kelapa sawit dalam produksi per blok berdasarkan dari data tersebut. Dan dari data dapat dilihat karakteristik sehingga diketahui *cluster* tinggi, *cluster* sedang dan *cluster* rendah dengan presentasi hasil produksi pada setiap blok nya. Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasi *objective function* yang ditentukan pada saat proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalisasikan variasi antar *cluster* [5].

Maka dari itu bagaimana agar proses pengelompokan hasil panen kelapa sawit dalam hasil produksi perbloknya di PT. Surya Intisari Raya Sei Mandau dilakukan menggunakan *tools rapidminer* dalam data *mining* dengan algoritma *K-Means* yang pada akhirnya data tersebut dapat dikelompokan sesuai dengan hasil produksinya.

2. METODOLOGI PENELITIAN

2.1 Rancangan Penelitian

Rancangan terhadap penelitian ini dilakukan dengan melakukan sebuah pengamatan selanjutnya mengumpulkan data, setelah data tersebut dimasukkan ke *Microsoft Excel* lalu data akan diolah melalui proses perhitungan dan mengikuti langkah-langkah metode *k-means.* Hasil perhitungan tersebut akan diaplikasikan ke *RapidMiner* untuk melihat hasil yang akurat. Berikut gambar dari perancangan penelitian dapat kita lihat seperti pada gambar 1.

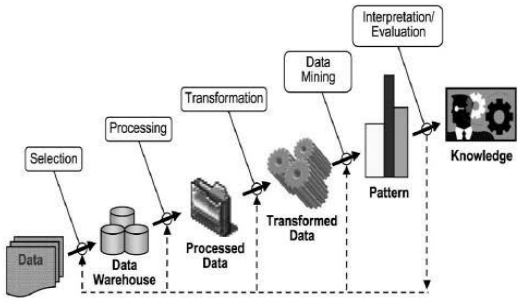


Gambar 1. Perancangan Penelitian

2.2 Data Mining

Data *mining* yang biasa disebut sebagai *knowledge discovery in database* (KDD) merupakan kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam himpunan data yang berukuran besar [6]. Data *mining* juga bagian dari akuisisi pengetahuan basis data. Penguraian data dengan data probabilitas yang tidak diketahui jelas. Data *mining* didefinisikan sebagai proses mengekstrak atau menambah pengetahuan yang dibutuhkan dari sejumlah data besar [7].

Larose dalam buku yang ditulis oleh Kusrini dan Luthfi mengelompokan Data Mining dapat dibagi menjadi 6 kelompok yaitu deskripsi, estimasi, prediksi, klasifikasi, clustering (pengelompokan), dan asosiasi [8]. Berikut tahapan proses data *mining* dalam penemuan pengetahuan berulang dalam *database* dapat dilihat pada gambar 2.



Gambar 2. Tahapan Proses Data Mining

Pada gambar 2. merupakan sebuah langkah untuk membentuk proses data yang akan telah disiapkan. Untuk proses penggalian dari sebuah interaksi data. Pola lama yang disajikan dari pengguna yang akan diolah selanjutnya akan di proses dan disimpan sebagai informasi baru.

2.3 Algoritma *K-Means*

Algoritma *K-Means* adalah metode pengelompokkan berbasis jarak yang membagi data ke dalam sejumlah *cluster*, serta memiliki kemudahan dan kemampuan untuk menganalisa data besar dengan cepat dan dapat menemukan data outlier. *K-Means* memiliki konsep dimana elemen data K dipilih sebagai center, kemudian jarak semua data elemen dihitung menggunakan rumus *Euclidean*. Elemen data yang memiliki jarak kurang ke *centroid* dipindahkan ke *cluster* yang sesuai. Proses dilanjutkan hingga tidak ada lagi perubahan yang terjadi dalam *cluster*. Metode ini merupakan metode pengelompokan non hirarki yang bertujuan untuk mengelompokkan objek sehingga jarak-jarak tiap objek ke pusat kelompok di dalam satu kelompok bernilai minimum [9].

Langkah-langkah dalam algoritma *k-means*:

a. Tentukan jumlah *cluster* (k) pada data set.
b. Tentukan niali pusat(*centroid*)
   Penentu nilai *centroid* pada tahap awal dilalukan secara random, sedangkan pada tahap *iterasi* digunakan rumus seperti pada persamaan (1) berikut ini :

$$Vij = \frac{1}{Ni} \sum_{K=0}^{Ni} Xkj \qquad (1)$$

   Keterangan :
   Vij= *centrid* rata-rata *cluster* ke-i untuk variabel ke-j
   Ni = jumlah anggota *cluster* ke-i
   I,k = indeks dari *cluster*
   j = indeks dari variabel
   Xkj = nilai data ke-k variabel ke-j untuk *cluster* tesebut.
c. Pada masing-masing *record*, hitung jaraK
   terdekat dengan *centroid*. Jarak *centroid* yang digunakan adalah *Euclidean distance*, dengan rumus seperti di bawah ini :

$$D(x_2, x_1) = || x_2 - x_1 ||_2 \qquad (2)$$

$$\sqrt{\sum_{j=1}^{p} | X_{2j} - X_{1j} |^2} \qquad (3)$$

   Keterangan :
   D = *Euclidean distance*
   X = banyaknya objek = jumlah data *record*
   Σp = Jumlah data *record*

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \qquad (4)$$

   Keterangan :
   De = *Euclidean distance*
   i = banyaknya objek ²
   (x,y)= koordinasi objek
   (s, t)= koordinasi *centroid*
d. Kelompokan objek berdasarkan jarak ke *centroid* terdekat.

e. Ulangi langkah ke-2, lakukan *iterasi* hingga *centroid* bernilai optimal [10].

3. HASIL DAN PEMBAHASAN

3.1 Perhitungan Dengan Algoritma *K-Means*

Perhitungan algoritma K-Means dimulai dengan memilih data yang akan digunakan dalam *clustering* pada jumlah panen kelapa sawit dalam produksi per blok nya. Data yang akan digunakan sebanyak 50 data. Untuk mendapatkan hasil dari perhitungan manual untuk proses *clustering* menggunakan algoritma *k-means*. Proses *clustering* dilakukan mulai dari penentuan data yang ingin di *cluster*. Pada data tersebut telah dilakukan normalisasi data yang >1.000 =1, >2.000 =2, >3.000 =3, > 4.000 =4, dan >5.000= 5.

Berikut adalah data sampel.

Tabel 1. Data Hasil Panen Kelapa Sawit

Blok	Tahun Tanam	AFD	Luas (Ha)	Jan	Peb	Mar	Apr	Mei	Jun	Jul	Agt	Sep	Okt	Nop	Des
S34	1998	1	28.36	1	2	1	2	1	2	2	2	2	2	2	1
S30	1999	1	29.36	2	1	1	2	2	1	2	2	2	2	2	2
S31	1999	1	28.69	1	1	1	2	1	2	2	2	2	2	2	2
S32	1999	1	28.61	1	2	1	2	1	2	2	2	2	2	3	2
S33	1999	1	29.37	1	2	1	2	1	2	2	2	2	2	2	1
S35	1999	1	27.97	1	2	1	2	1	2	2	2	2	2	2	1
S36	1999	1	29.01	1	2	1	1	2	1	2	2	2	2	2	1
T32	2000	1	28.11	1	1	1	2	2	1	4	1	2	2	2	2
T33	2000	1	30.27	3	1	1	3	3	1	3	3	3	3	3	3
T34	2000	1	29.96	2	1	1	2	2	1	2	2	2	2	2	2
T35	2000	1	27.77	1	1	1	1	1	1	2	2	2	2	2	2
T36	2000	1	29.68	1	1	1	1	2	2	2	2	2	2	2	2
U35	2000	1	27.84	2	2	1	1	2	2	2	1	2	2	4	1
U36	2000	1	29.55	2	1	2	2	2	2	2	1	2	4	2	1
T31	2001	1	28.00	2	1	1	1	2	2	2	1	2	2	2	2
U32	2001	1	20.68	1	1	1	1	1	1	1	1	1	1	2	1
U33	2001	1	30.64	3	1	1	3	3	3	3	1	3	3	5	1
U34	2001	1	27.70	2	1	1	2	2	2	2	1	2	2	4	1
P32	1998	2	27.33	1	1	1	1	1	2	2	2	2	4	3	3
P33	1998	2	28.89	1	1	1	1	1	3	2	1	2	3	3	3
P34	1998	2	29.11	1	2	1	1	2	3	3	1	4	4	3	2
P35	1998	2	28.66	1	2	1	1	2	3	3	1	4	4	3	2
P36	1998	2	29.42	1	2	1	1	2	3	3	1	4	4	3	2
R32	1998	2	28.96	2	1	1	2	2	3	3	1	4	4	3	2
R33	1998	2	29.29	2	1	1	2	1	4	2	1	4	4	2	2
R34	1998	2	28.25	1	1	1	2	1	4	2	1	4	4	2	2
R35	1998	2	28.77	1	1	1	2	1	4	2	1	4	4	2	2
R36	1998	2	28.91	1	1	1	2	3	4	4	1	4	4	2	2
N30	1999	2	23.68	2	1	1	1	2	4	4	2	2	4	2	2
P23	1999	2	28.48	2	1	1	1	2	4	4	1	2	4	2	2
P24	1999	2	28.91	2	1	1	1	1	2	2	2	2	4	1	2
P25	1999	2	28.42	2	1	1	1	1	2	2	2	2	4	1	2
P26	1999	2	29.38	2	1	1	1	1	2	2	2	2	4	1	2
R25	1999	2	20.45	1	1	1	2	1	2	2	1	2	4	2	1
R26	1999	2	30.62	3	1	1	2	2	4	2	1	1	2	4	3

R27	1999	2	28.88	2	2	1	2	1	4	2	1	3	2	4	2
R28	1999	2	29.47	2	1	1	2	1	2	2	1	4	2	1	2
R29	1999	2	28.96	2	1	1	2	1	2	2	1	4	2	1	2
B06	2005	3	23.66	1	2	2	1	1	1	2	1	2	2	2	2
B07	2005	3	25.29	1	2	2	1	1	3	3	2	1	1	2	3
B08	2005	3	24.48	1	1	1	1	2	2	3	1	2	1	2	2
B09	2005	3	23.36	1	1	2	1	1	2	2	1	1	2	1	1
C09	2005	3	22.75	1	1	1	2	1	2	1	1	1	3	1	1
C10	2005	3	24.49	1	1	1	1	2	3	1	1	1	2	1	1
C11	2005	3	23.88	1	1	1	1	2	2	1	1	1	3	1	1
C12	2005	3	24.35	1	1	1	1	1	1	3	1	1	3	1	1
C13	2005	3	24.16	1	1	1	1	3	1	3	1	1	3	1	1
C14	2005	3	24.84	1	1	2	1	3	1	3	1	1	3	1	1
C15	2005	3	22.02	1	1	1	1	2	1	2	1	1	2	1	1
D13	2005	3	22.13	1	1	1	1	2	2	1	1	1	2	1	1

a. Menentukan nilai k jumlah *Cluster* jumlah *cluster* sebanyak 3 *cluster*. *Cluster* yang dibentuk yaitu *cluster* tinggi (C1), *cluster* sedang (C2), dan *cluster* rendah (C3).
b. Menentukan nilai *centroid* (pusat *cluster*) secara random, penentuan pusat *cluster* awal ditentukan secara random yang diambil dari data yang ada didalam *range*.

Tabel 2. Centroid Data Awal

Cluster	Blok	Tahun Tanam	Afd	Luas (Ha)	Jan	Peb	Mar	Apr	Mei	Jun	Jul	Agt	Sep	Okt	Nop	Des
C1	U34	2001	1	27.70	2	1	1	2	2	2	2	1	2	2	4	1
C2	R35	1998	2	28.77	1	1	1	2	1	4	2	1	4	4	2	2
C3	C10	2005	3	24.49	1	1	1	1	2	3	1	1	1	2	1	1

c. Menghitung jarak setiap data terhadap *centroid* (pusat *cluster*). Berikut adalah contoh perhitungan dengan algoritma *k-means;*

$$DS34(28,36)C2 = \sqrt{\begin{matrix}(1-2)^2+(2-1)^2+(1-1)^2+(2-2)^2+(1-2)^2+\\(2-2)^2+(2-2)^2+(2-1)^2+(2-2)^2+(2-2)^2+=2.828427\\(2-4)^2+(1-1)^2\end{matrix}} \quad (5)$$

$$DS34(28,36)C2 = \sqrt{\begin{matrix}(1-1)^2+(2-1)^2+(1-1)^2+(2-2)^2+(1-1)^2+\\(2-4)^2+(2-2)^2+(2-1)^2+(2-4)^2+(2-4)^2+=3.872983\\(2-2)^2+(1-2)^2\end{matrix}} \quad (6)$$

$$DS34(28,36)C3 = \sqrt{\begin{matrix}(1-1)^2+(2-1)^2+(1-1)^2+(2-1)^2+(1-2)^2+\\(2-3)^2+(2-1)^2+(2-1)^2+(2-1)^2+(2-2)^2+=2.828427\\(2-1)^2+(1-1)^2\end{matrix}} \quad (7)$$

Tabel 3. Hasil Perhitungan Jarak Pusat Cluster Iterasi 1

Blok	Tahun Tanam	Afd	Luas (Ha)	C1	C2	C3	Jarak Terdekat
S34	1998	1	28.36	2.828427	3.872983	2.828427	2.828427
S30	1999	1	29.36	2.645751	4.472136	3.316625	2.645751
S31	1999	1	28.69	2.828427	3.605551	2.828427	2.828427
S32	1999	1	28.61	2.44949	3.872983	3.464102	2.44949

S33	1999	1	29.37	2.828427	3.872983	2.828427	2.828427
S35	1999	1	27.97	2.828427	3.872983	2.828427	2.828427
S36	1999	1	29.01	3	4.690416	3	3
T32	2000	1	28.11	3.316625	4.690416	4.123106	3.316625
T33	2000	1	30.27	4	5.196152	5.830952	4
T34	2000	1	29.96	2.645751	4.472136	3.316625	2.645751
T35	2000	1	27.77	3.162278	4.358899	3.162278	3.162278
T36	2000	1	29.68	2.828427	3.872983	2.44949	2.44949
U35	2000	1	27.84	1.414214	4.582576	3.741657	1.414214
U36	2000	1	29.55	3	3.464102	3.316625	3
T31	2001	1	28.00	2.44949	3.872983	2.44949	2.44949
U32	2001	1	20.68	3.316625	5.477226	2.645751	2.645751
U33	2001	1	30.64	2.828427	4.795832	5.830952	2.828427
U34	2001	1	27.70	0	4.358899	3.741657	0
P32	1998	2	27.33	3.605551	3.464102	4.123106	3.464102
P33	1998	2	28.89	3.162278	3	3.464102	3
P34	1998	2	29.11	3.872983	2.44949	4.795832	2.44949
P35	1998	2	28.66	3.872983	2.44949	4.795832	2.44949
P36	1998	2	29.42	3.872983	2.44949	4.795832	2.44949
R32	1998	2	28.96	3.464102	2.236068	4.898979	2.236068
R33	1998	2	29.29	4.242641	1	4.472136	1
R34	1998	2	28.25	4.358899	0	4.358899	0
R35	1998	2	28.77	4.358899	0	4.358899	0
R36	1998	2	28.91	4.795832	2.828427	5.196152	2.828427
N30	1999	2	23.68	4.358899	3.464102	4.358899	3.464102
P23	1999	2	28.48	4.242641	3.316625	4.242641	3.316625
P24	1999	2	28.91	4.123106	3.464102	3.316625	3.316625
P25	1999	2	28.42	4.123106	3.464102	3.316625	3.316625
P26	1999	2	29.38	4.123106	3.464102	3.316625	3.316625
R25	1999	2	20.45	3.162278	3	3.162278	3
R26	1999	2	30.62	3.162278	4.795832	4.472136	3.162278
R27	1999	2	28.88	2.828427	3.316625	4.472136	2.828427
R28	1999	2	29.47	3.872983	3.162278	3.872983	3.162278
R29	1999	2	28.96	3.872983	3.162278	3.872983	3.162278
B06	2005	3	23.66	3.316625	4.472136	3.316625	3.316625
B07	2005	3	25.29	4.242641	5	3.741657	3.741657
B08	2005	3	24.48	3	4.472136	3	3
B09	2005	3	23.36	3.741657	4.582576	2	2

C09	2005	3	22.75	3.741657	4.123106	2	2
C10	2005	3	24.49	3.741657	4.358899	0	0
C11	2005	3	23.88	3.741657	4.358899	1.414214	1.414214
C12	2005	3	24.35	4	4.795832	3.162278	3.162278
C13	2005	3	24.16	4	5.196152	3.162278	3.162278
C14	2005	3	24.84	4.123106	5.291503	3.316625	3.316625
C15	2005	3	22.02	3.605551	5.09902	2.236068	2.236068
D13	2005	3	22.13	3.605551	4.690416	1	1

d. Menghitung *centroid* baru untuk itearsi berikutnya dengan menghitung rata-rata nilai pada masing-masing *cluster*. Berikut ini adalah perhitungan *centroid* baru pada masing-masing *cluster*.

$$D1,C1=\frac{1+2+1+1+1+1+1+3+2+1+1+2+2+2+1+3+2+1}{19}=1.526316 \quad (8)$$

$$D2,C1=\frac{2+1+1+2+2+2+1+1+1+1+1+2+1+1+1+1+1+1}{19}=1.315789 \quad (9)$$

$$D3,C1=\frac{1+1+1+1+1+1+1+1+1+1+1+1+1+2+1+1+1+1+1}{19}=1.053 \quad (10)$$

$$D1,C2=\frac{1+1+1+1+2+2+1+1+1+2+2+2+2+2+1+3+2+2+2}{19}=1.631579 \quad (11)$$

$$D2,C2=\frac{1+2+2+1+1+1+1+1+1+1+1+2+1}{19}=1.210526 \quad (12)$$

$$D3,C2=\frac{1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1+1}{19}=1 \quad (13)$$

$$D1,C3=\frac{1+1+1+1+1+1+1+1+1+1+1+1}{12}=1 \quad (14)$$

$$D2,C3=\frac{2+2+1+1+1+1+1+1+1+1+1+1}{12}=1.166667 \quad (15)$$

$$D3,C3=\frac{2+2+1+2+1+1+1+1+1+2+1+1}{12}=1.333 \quad (16)$$

Jika tahap iterasi telah mencapai hasil yang sama tanpa dan tidak mengalami perubahan lagi. Maka perhitungan dihentikan. Perhitungan manual pada data diatas didapatkan hasil akhir yang dimana pada iterasi 1 dan iterasi 2 pengelompokan data yang dilakukan terhadap 2 cluster didapatkan hasil yang sama. Hasil dari kedua itersai tersebut bernilai C1 =31, C2 =7, dan C3 =12 pada posisi data tiap *cluster*. Sehingga posisi *cluster* pada data tersebut tidak mengalami perubahan lagi maka proses iterasi berhenti sampai iterasi2.

Tabel 4.Hasil Pengelompokan Iterasi Ke 2

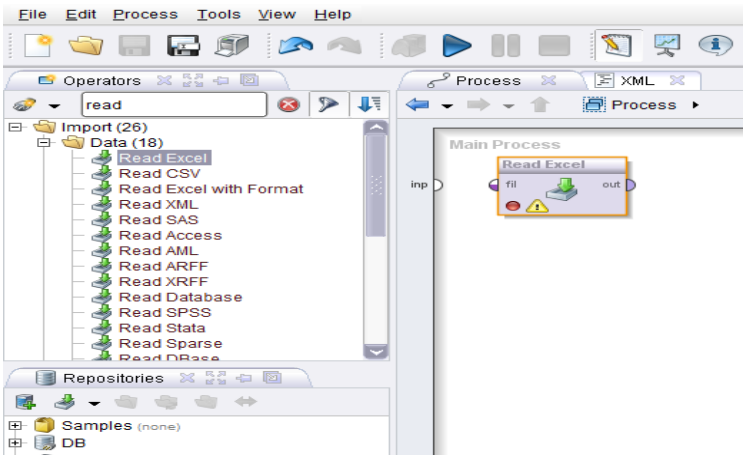
Cluster	Blok	Jumlah
C1	S34,S32,S33,S32,S33,T32,T33,T34,T35,T36,U35,U36, T31,U32,U33,U34.	19
C2	P32,P33,P34,P35,P36,R32,R33,R34,R35,R36,N30,P23 P24,P25,P26,R25,R26,R27,R28,R29	19
C3	B06,B07,B08,B09,C09,C10,C11,C12,C13,C14,C15,D13	12

3.2 Implementasi Pada Aplikasi *RapidMiner*

Dalam pengimplementasian pada aplikasi *RapiMiner* dapat dilihat sebagai berikut :

a. *Import* data ke dalam *RapidMiner* dalam bentuk *Sheet Excel.*

Sistem menjelaskan cara memasukkan data baru yang akan dieksekusi lebih lanjut, pada hal ini data yang akan di eksekusi berupa data *excel.* Klik pada bagian kiri bawah tab *respositories* lalu pilih *"Import Read Ecxel"*. Kemudian akan muncul tampilan seperti gambar di bawah ini.



Gambar 3. *Import* data

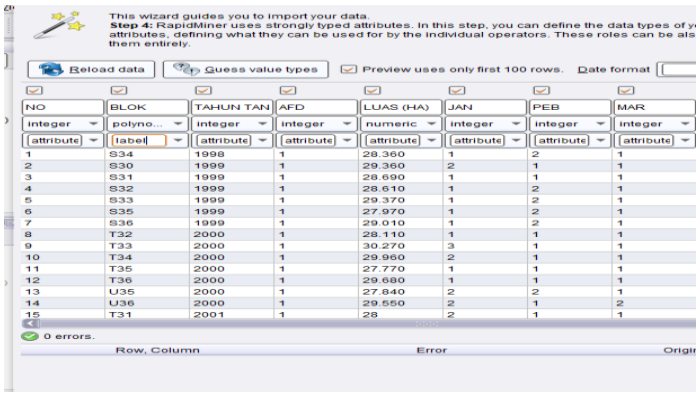
b. Import Configurasi Wizard

Kemudian akan muncul data *import confirguration wizard* kemudian pilih tempat kita menyimpan data yang akan digunakan. Selanjutnya pilih *file name* data yang akan digunakan. Lalu klik *next* pada bagian kanan bawah seperti dapat dilihat pada gambar dibawah ini.

Gambar 4. Konfigurasi *wizard*

c. Setelah konfigurasi wizard, pilih file sesuaKlan dengan atribut dan tipe data

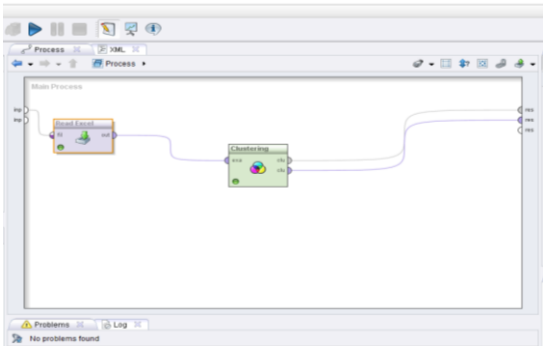
Pada tahap ini dilakukan pemilihan tipe data dimana pada bagian attribute yang di awal diubah tipe "*label", " attribute"*.



Gambar 5. Penyesuaian Atribut Dan Tipe Data

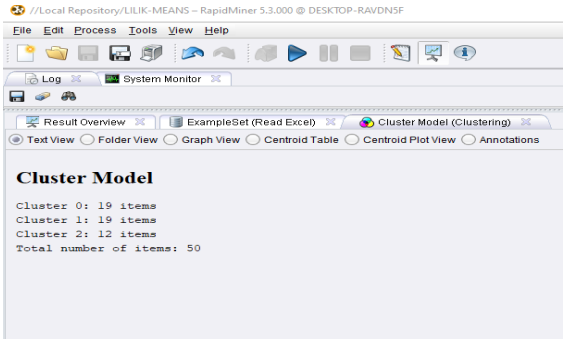
d. Pemrosesan sistem *RapidMiner*

Pada tahap ini akan dijelaskan tahapan-tahapan proses penggunaan *k-means* di dalam *Rapidminer* data yang telah di *impor*. Tahapan pertama dengan mengklik *clustering and segmentation* lalu pilih *k-means* dapat dilihat pada gambar 6 Lalu hubungkan antara *Read excel* dengan *clustering* seperti gambar di bawah ini:



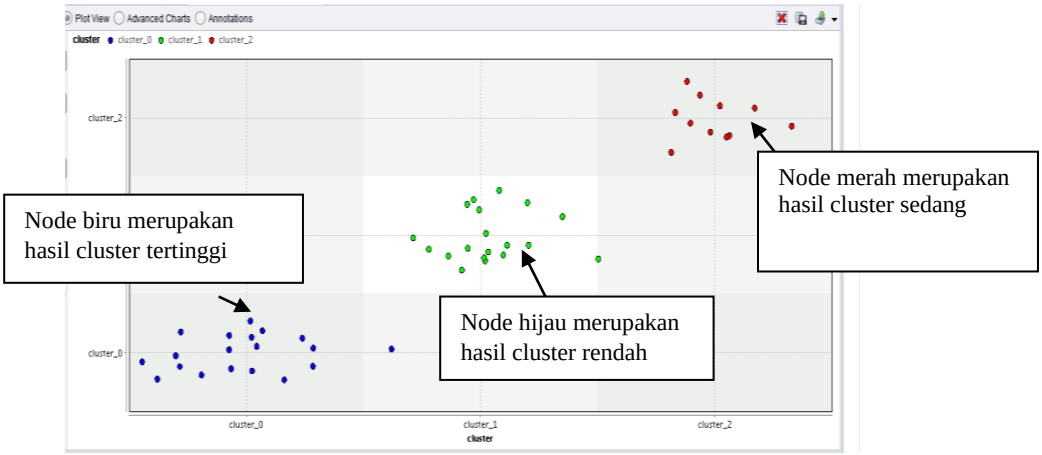
Gambar 6. Pemprosessan *RapidMiner*

e. Hasil pengelompokan dalam *RapiMiner*
Maka akan menampilkan hasil akhir serta langkah terakhir dalam penggunaan *tools rapidminer* ini. Dapat dilihat pada gambar di bawah ini:



Gambar 7. Hasil Pengelompokkan

f. Tampilan plot View
Berdasarkan pada gambar di bawah dapat diketahui bahwa pada kelompok rendah memiliki *node* hijau yaitu 12, dan kelompok sedang memiliki *node* merah yaitu 19, sedangkan kelompok tertinggi memiliki *node* biru yaitu 19.



Gambar 8. Hasil Tampilan Plot View

4. KESIMPULAN

Berdasarkan pembahasan dan hasil *Implementasi Ssoftware Rapidminer* pengelompokan hasil panen kelapa sawit berdasarkan produksi per blok nya, maka dapat ditarik kesimpulan bahwa data yang diolah untuk memperoleh nilai dari hasil panen kelapa sawit dalam produksi per blok yang memiliki hasil panen kelapa sawit menurut per bloknya menerapkan metode algoritma *k-means.* Data tersebut diolah menggunakan *Microsoft Excel* untuk ditentukan nilai *centroid* dalam 3 *cluster* yaitu *cluster* tinggi, *cluster* sedang dan *cluster* rendah. Dan hasil yang diperoleh dari metode *k-means* yang di implementasikan ke dalam *Rapidminer* memiliki nilai yang sama yaitu menghasilkan 3 *cluster* yaitu *cluster* tinggi, *cluster* sedang, dan *cluster* rendah.

Dengan *cluster* tinggi memiliki 19 blok, *cluster* sedang memiliki 19 blok dan *cluster* rendah memiliki 12 blok. Hasil yang didapat dari penelitian dapat menjadi masukan kepada PT. Surya Intisari Raya Sei Mandau yang menjadi perhatian lebih pada blok yang memiliki hasil panen kelapa sawit berdasarkan produksi per bloknya berdasarkan cluster yang telah dilakukan..

REFERENCES

[1] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.

[2] S. S. Helma, R. R. R, and E. Normala, "Clustering pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma K - Means," no. November, pp. 131–137, 2019.

[3] M. A. W. K. MURTI, "Penerapan Metode K-Means Clustering Untuk Mengelompokan Potensi Produksi Buah – Buahan Di Provinsi Daerah Istimewa Yogyakarta," *Skripsi*, 2017.

[4] L. Maulida, P. Studi, and M. Informatika, "KUNJUNGAN WISATAWAN KE OBJEK WISATA UNGGULAN DI PROV . DKI JAKARTA DENGAN K-MEANS," vol. 2, no. 3, pp. 167–174, 2018.

[5] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.

[6] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.

[7] M. Lailil, ratnawati eka Dian, and putri mardi regasari Rekyan, *Data Mining*. UB Press, 2018.

[8] W. Dhuhita, "Clustering Menggunakan Metode K-Mean Untuk Menentukan Status Gizi Balita," *J. Inform. Darmajaya*, vol. 15, no. 2, pp. 160–174, 2015.

[9] S. S. Helma, R. R. R, and E. Normala, "Clustering pada Data Fasilitas Pelayanan Kesehatan Kota Pekanbaru Menggunakan Algoritma K - Means," no. November, pp. 131–137, 2019.

[10] I. Parlina, W. A. Perdana, W. Anjar, and L. Ridwan.M, "MEMANFAATKAN ALGORITMA K-MEANS DALAM MENENTUKAN PEGAWAI YANG LAYAK MENGIKUTI ASESSMENT CENTER," vol. 3, no. 1, pp. 87–93, 2018.