BIOS : Jurnal Teknologi Informasi dan Rekayasa Komputer
Vol. 2, No. 1, Maret 2021, hlm. 11-20
ISSN: 2722-0850
Laman Jurnal: http://bios.sinergis.org

Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat

Deny Franata Pasaribu[1], Irfan Sudahri Damanik[2], Eka Irawan[3], Suhada[3], Heru Satria Tambunan[5]
[1,2,3,4,5]Jurusan Sistem Informasi, STIKOM Tunas Bangsa Pematangsiantar
[1]denyfranata2016@gmail.com, [2]irfansudahri@amiktunasbangsa.ac.id, [3]eka.irawan@amiktunasbangsa.ac.id,
[1]suhada.atb@gmail.com, [1]heru@amiktunasbangsa.ac.id

Keywords:

Grouping,
Data Mining,
Cluster,
K-Means Algorithm,
Palm,

ABSTRACT

Based on data on the results of oil palm production in PTPN IV Marihat displays several locations with fruit yields that vary in number. For this reason, grouping of potential fruit-producing locations is needed to know which locations produce large or small numbers of palm fruit. The production sharing is usually done based on the location or block of harvesting oil palm fruit. Therefore, a method is needed to facilitate the grouping of fruit producing locations. With the K-Means clustering approach, the division of location groups can be done based on harvested area (Ha), production realization (kg) and harvest year. In this research, clustering of potential fruit-producing areas was carried out using the K-Means algorithm. By using K-Means aims to facilitate the grouping of a block with a lot of fruit production, and low. The result of this research is that C1 (highest) is 14 Harvest Block data, and C2 (lowest) is 11 Harvest Block data.

Kata Kunci

Pengelompokan,
Data Mining,
Cluster,
Algoritma K-Means,
Kelapa Sawit,

ABSTRAK

Berdasarkan data hasil produksi buah sawit di PTPN IV Marihat menampilkan beberapa Lokasi dengan hasil panen buah yang bervariasi jumlahnya. Untuk itu diperlukan pengelompokan lokasi potensial penghasil buah untuk mengetahuni lokasi mana saja yang menghasilkan buah sawit dengan jumlah banyak ataupun sedikit. Pembagian hasil produksi biasanya dilakukan berdasarkan loaksi atau blok panen buah kelapa sawit. Oleh karena itu, dibutuhkan metode untuk memudahkan dalam pengelompokan lokasi penghasil buah. Dengan pendekatan pengklasteran *K-Means*, pembagian kelompok lokasi dapat dilakukan berdasarkan luas panen (Ha), realisasi produksi (kg) dan tahun panen. Pada penelitian ini dilakukan pengklasteran daerah potensial penghasil buah menggunakan algoritma *K-Means*. Dengan menggunakan *K-Means* bertujuan dalam memudahkan pengelompokan suatu blok dengan hasil produksi buah banyak, dan rendah. Hasil dari penelitian ini didapakan C1 (tertinggi) ialah 14 data Blok Panen, dan C2 (terendah) ialah 11 data Blok Panen.

Korespondensi Penulis:

Deny Franata Pasaribu,
STIKOM Tunas Bangsa,
Jalan Sudirman Blok A Nomor 1, 2, 3 Pematangsiantar
Telepon: +62 851 5501 1979
Email: denyfranata2016@gmail.com

1. PENDAHULUAN

Kelapa sawit *(Elaeis guineensis Jacq)* merupakan tanaman tahunan golongan palma yang dapat tumbuh didaerah tropis**.** Tanaman kelapa sawit berasal dari Nigeria, Afrika Barat. Saat ini perkebunan kelapa sawit menjadi bagian yang sangat penting dari rantai suplai minyak kelapa sawit global dan ada sekitar tiga juta petani perkebunan

kelapa sawit rakyat di seluruh dunia yang memproduksi sekitar 4 juta ton minyak kelapa sawit. Tanaman kelapa sawit merupakan salah satu penghasil minyak tertinggi per hektar.

PT. Perkebunan Nusantara IV (PTPN IV) merupakan Perusahaan yang bergerak dibidang perkebunan kelapa sawit. Namun dalam pengelolaannya sering kali terjadi penurunan hasil minyak sawit dan inti sawit. Hal ini terjadi karena Belum dilakukan Pemetaan terhadap lokasi buah yang berpotensi baik untuk di produksi oleh pihak terkait dikarenakan terlalu luasnya daerah perkebunan kelapa sawit yang ada pada wilayah Marihat, kemudian pihak terkait masih belum dapat memberikan planning untuk melakukan panen buah berdasarkan blok mana yang lebih berpotensi dengan buah yang baik.

Teknologi pemetaan merupakan teknik membuat data-data di lapangan menjadi sebuah data analisis hasil pengelompokan yang dapat memudahkan pengguna untuk mengetahui sebuah posisi atau letak suatu tempat. Dari permasalahan tersebut penulis ingin melakukan penelitian yang dapat mengetahui lokasi buah kelapa sawit yang berpotensi baik berdasarkan data hasil panen. Data mining adalah teknik pengolahan yang penulis pilih untuk meneliti permasalahan tersebut. Data mining adalah suatu bidang ilmu yang dapat menghasilkan sebuah informasi melalui perhitungan dari data-data lama yang sudah tidak digunakan. Data mining adalah campuran dari statistic, kecerdasan buatan, dan riset basis data yang masih berkembang [1]. Penelitian ini menggunakan Data mining algoritma *K-Means* dalam mengelompokkan lokasi buah kelapa sawit yang berpotensi di PT. Perkebunan Nusantara IV. Algoritma *K-Means* merupakan salah satu metode pengelompokan data *non*-hierarki yang mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok [2].

Berdasarkan latar belakang yang sudah dijelaskan sebelumnnya maka perumusan masalahnya adalah bagaimana menerapkan proses pengelompokan data hasil produksi sawit dengan algoritma K-Means untuk mengetahui lokasi penghasil buah yang berpotensi, yang berguna bagi pihak PTPN IV Marihat.

Adapun konstrirbusi riset dalam memetakan potensi hasil produksi kelapa sawit ini adalah untuk memberikan tingkat pengamanan pada lokasi yang berpotensi tinggi, sehingga tidak terjadinya pencurian buah serta pihak perkebunan dapat memberi perawatan lebih terhadap tanaman sawit yang berpotensi tinggi untuk di produksi.

Beberapa penelitian yang terkait dengan penelitian ini dapat dilihat pada tabel 1 berikut ini:

Tabel 1. Penelitian Terkait

No	Nama Author	Judul Riset	Hasil Penelitian
1	Lailan Azmi, Mesran dan Kurnia Ulfa	Prediksi Replanting Lahan Kelapa Sawit Dengan Menggunakan Metode K-Nearest Neighbor Pada Ptpn Iv Adolina [3].	Algoritma *K-Nearest Neighbor* (KNN) dapat digunakan dalam menentukan lahan kelapa sawit yang akan diremajakan pada PTPN IV Adolina.
2	Nenci Sirait	Implementasi K-Means Clustering Pada Pengelompokan Mutu Biji Sawit (Studi kasus: Pt. Multimas Nabati Asahan) [4].	Data mining pengelompokan mutu biji sawit di PT. Multimas Nabati Asahan menggunakan metode *K-means Clustering* di mulai dari tingkat kematangan buah sawit berdasarkan standar mutunya.
3	Nobertus Krisandi, Helmi dan Bayu Prihandono	Algoritma *K-Nearest Neighbor* Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada Pt. Minamas Kecamatan Parindu [5].	Berdasarkan hasil penelitian, data diklasifikasikan ke dalam 6 *cluster*. Algoritma *k*-NN dapat digunakan dalam klasifikasi data Hasil Produksi Kelapa Sawit pada PT. Minamas Kec. Parindu.

2. METODE PENELITIAN

2.1 Data Mining

Data Mining merupakan proses kegiatan untuk mengumpulkan data dalam jumlah yang besar untuk menemukan pengetahuan sehingga dapat menjadi informasi yang dapat digunakan [6]. *Data mining* adalah proses menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengetahui informasi yang bermanfaat yang terkait dari database yang besar. Tujuan utama Data mining adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki.

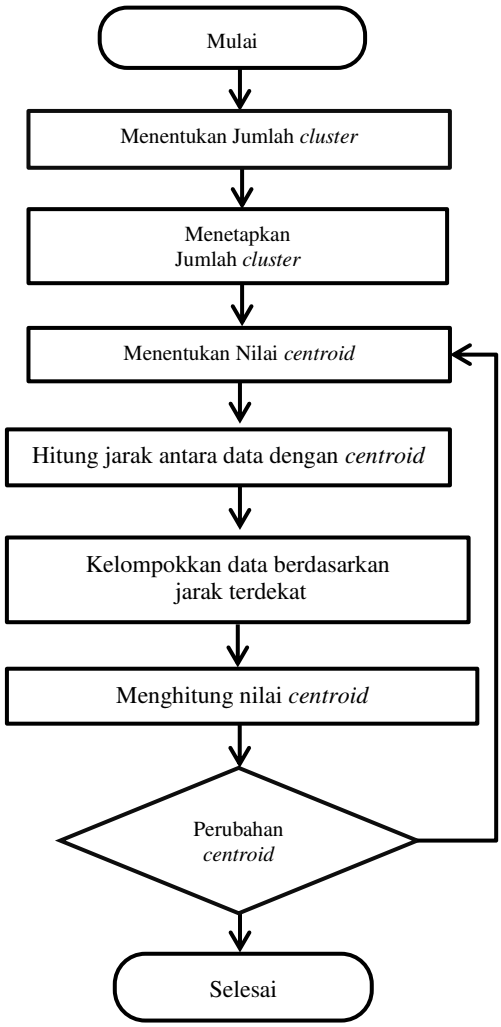
2.2 Clustering

Clustering merupakan proses partisi sekumpulan objek data dari satu set menjadi beberapa kelas. Hal ini dapat dilakukan dengan menerapkan berbagai persamaan dan langkah-langkah mengenai jarak algoritma yaitu dengan Euclidean Distance. Analisis kluster ialah metode yang dipakai untuk membagi rangkaian data menjadi beberapa grup berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya. Dalam menentukan cluster berdasarkan data yang telah tersedia dibutuhkan sebuah flowchart untuk memudahkan dalam menentukan alur perhitungan sebagai alur untuk menemukan hasil dari penerapan cluster terhadap data yang akan diproses [7].

Clustering adalah metode penganalisaan data yang sering dimasukkan sebagai salah satu metode Data mining yang tujuannya adalah untuk mengelompokkan data dengan karakteristik yang sama. Cluster berbeda dari klasifikasi karena cluster tidak memiliki variabel target. Tujuan cluster bukan untuk mengklasifikasikan, memperkirakan, atau memprediksi nilai variabel taget [8].

2.3 Algoritma K-Means Clustering

Algoritma *K-Means* merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah *K cluster* yang sudah ditetapkan di awal [9].



Gambar 1. *Flowchart* Metode K-Means

Berikut adalah langkah-langkah penyelesaian yang dilakukan penulis dalam menggelompokkan Hasil Produksi buah sawit yang Produktif menggunakan algoritma *K-Means* :

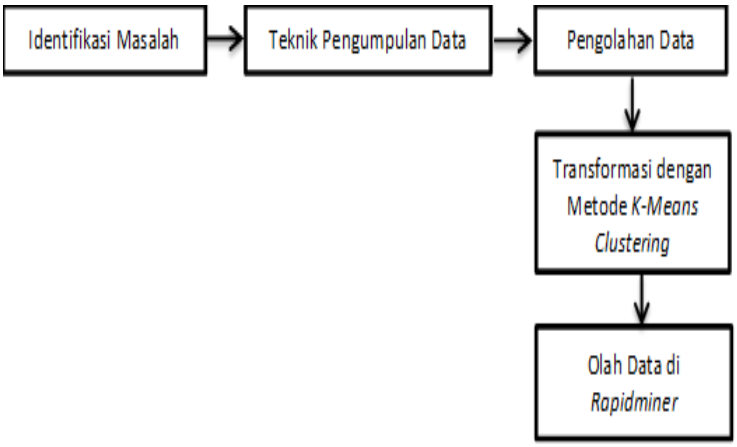
1. Menentukan jumlah data yang akan di *cluster,* dimana sampel data Hasil Produksi yang akan digunakan dalam proses *clustering* adalah Hasil Panen yang paling Produktif.
2. Menetapkan nilai k jumlah *cluster* sebanyak 2 *cluster* (k-2).
3. Menentukan nilai *centroid* (pusat *cluster*) awal yang ditentukan secara *random* berdasarkan nilai variabel data yang di *cluster* sebanyak nilai k yang ditentukan sebelumnya.
4. Menghitung jarak setiap data yang Produktif terhadap pusat *cluster.*
5. Menentukan posisi *cluster* masing-masing data hasil produksi yang produktif berdasarkan jarak minimum data terhadap pusat *cluster*.
6. Menghitung nilai *centroid* hasil iterasi dengan Memeriksa kondisi berhenti proses iterasi, jika nilai *centroid* hasil iterasi dengan nilai *centroid* sebelumnya bernilai sama atau nilai *centroid* sudah optimal serta posisi *cluster* data hasil produksi sawit yang produktif tidak mengalami perubahan lagi maka proses iterasi berhenti. Namun jika nilai *centroid* tidak sama atau belum optimal serta posisi data hasil produksi buah sawit produktif masih berubah maka proses iterasi berlanjut pada iterasi berikutnya

2.4 RapidMiner

Rapidminer adalah platform perangkat ilmu data yang dikembangkan oleh perusahaan bernama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data pembelajaran mesin pembelajaran dalam pengembangan teks dan analisis prediktif.

2.5 Rancangan Penelitian

Rancangan penelitian ini pertama kali penulis melakukan pengamatan selanjutnya mengumpulkan data setelah itu data tersebut dipindahkan ke *exel* (xls) setelah itu data akan diolah melalui proses perhitungan dan mengikuti langkah-langkah proses perhitungan metode *k-means* Selanjutnya hasil perhitungan tersebut akan diaplikasikan ke *RapidMiner* untuk melihat keakuratan hasil yang diperoleh.*Flowchart* Rancangan Penelitian dapat dilihat pada gambar 2.



Gambar 2. Rancangan Penelitian

Berikut penjelasan dari alur rancangan penelitan pada gambar 2 :

1. Identifikasi Masalah
   Merupakan suatu tahap permulaan dari penguasaan masalah dimana suatu objek tertentu dalam situasi tertentu dapat dikenali sebagai suatu masalah. Permasalahan dalam penelitian ini adalah kurangnya pemetaan terhadap blok yang merupakan penghhasil panen sawit yang produktif
2. Teknik Pengumpulan Data
   Pengumpulan data dilakukan dengan cara mencari data Hasil Produksi sawit berdasarkan tahun. Untuk data set pertama penulis mengambil data dari PTPN IV Marihat yang mana data Hasil Produksi sawit dari tahun 2015 - 2019.
3. Pengolahan Data
   Pada tahap pengolahan data, data diolah untuk mendapatkan hasil yang kemudian dapat berguna dalam pengambilan keputusan. Data diolah sehingga menjadi informasi. Data yang didapat dalam bentuk file *excel* sehingga berekstensi .xlsx.
4. Transformasi dengan Metode *K-Means Clustering*
   Transformasi data, data yang tersusun dari urutan blok seperti blok A21, A22, A23, A24 dan A25 harus dilakukan proses inisialisasi data terlebih dahulu ke dalam bentuk angka/numerikal. Kemudian mengelompokkan data yang sudah ada dalam tiga kelompok yaitu banyak, sedang, rendah dengan metode *K-Means Clustering*.
5. Olah Data di *Rapidminer*
   Rapidminer merupakan salah satu *software* pilihan untuk melakukan ekstraksi data dengan metode-metode *data mining*. Pengujian disini kita menggunakan sebuah *software Rapidminer,* dengan pengujian data menggunakan *software* kita akan membandingkan bagaimana hasil pengolahan data secara manual dengan hasil pengolahan data menggunakan sebuah *software*.

2.6 Analisa Data

Proses analisis data dapat dilakukan setelah adanya pengumpulan data dan setelah data dikumpulkan dan diolah ke *excel*. Setelah hasil yang telah dapat di *excel* akan diaplikasikan ke *RapidMiner* untuk menyesuakian hasil yang didapatkan. Penulis akan melakukan analisa data statistik deskriptif. Data deskriptif adalah metode-metode yang berkaitan dengan pengumpulan data penyajian suatu gugus data sehingga memberikan informasi yang berguna informasi yang dapat diperoleh dari statistika deskriptif ini. Jenis data yang digunakan dalam penelitian ini merupakan

data sekunder, Data sekunder adalah data yang diperoleh tidak dari sumbernya langsung melainkan sudah dikumpulkan oleh pihak lain dan sudah di olah serta memiliki keterkaitan dengan permasalahan yang di teliti.

2.7 Alat Analisa Data

Dalam melakukan penelitian ini penulis menggunakan Microsoft excel dan *RapidMiner* sebagai alat untuk memproses data , *Microsoft Excel* digunakan untuk membantu dalam proses perhitungan yang akan dilakukan setelah proses perhitungan selesai maka selanjutnya diaplikasikan ke *RapidMiner* untuk melihat keakuratan hasil yang didapat dari perhitungan di *excel serta* akan dicocokan dengan hasil yang sudah diaplikasikan ke *RapidMiner* tersebut.

3. HASIL DAN ANALISIS

3.1 Pengolahan Data

Penelitian ini menggunakan Data Mining metode *K-Means*, Pada tahapan ini dilakukan analisis data hasil Produksi Kelapa sawit berdasarkan blok masing-masing dengan tool *RapidMiner. RapidMiner* adalah sebuah *software* pilihan untuk melakukan ekstraksi data dengan metode-metode *Data Mining.* Berikut adalah data hasil produksi kelapa sawit yang diambil dari PTPN IV Marihat.

Berikut adalah data hasil Produksi Kelapa sawit yang diambil penulis langsung dari PTPN IV Marihat :

Tabel 2. Data Hasil Produksi Kelapa Sawit

Blok Panen	Tahun	Luas Areal (Ha)	Berat Janjang (Kg)	Realisasi Produksi (Kg)
A21	2015	34,35	8,1	14720
A22		22,65	7,5	10760
A23		31,56	8,7	19350
A24		32,19	8,4	16720
A25		33,12	8,2	15100
A21	2016	34,35	8,1	14520
A22		22,65	7,5	10260
A23		31,56	8,7	19150
A24		32,19	8,4	16220
A25		33,12	8,2	14900
A21	2017	34,35	8,1	14620
A22		22,65	7,5	10790
A23		31,56	8,7	19990
A24		32,19	8,4	16810
A25		33,12	8,2	15300
A21	2018	34,35	8,1	14780
A22		22,65	7,5	10800
A23		31,56	8,7	19850
A24		32,19	8,4	16620
A25		33,12	8,2	15230
A21	2019	34,35	8,1	14770
A22		22,65	7,5	10860
A23		31,56	8,7	19890
A24		32,19	8,4	16560
A25		33,12	8,2	15380

Langkah- Langkah perhitungan metode *K-Means Clustering:*

1. Menentukan Jumlah *cluster* dan *Centorid*
   Data tersebut akan masuk ke tahapan *clustering* dengan menerapkan algoritma *K-Means* menggunakan *RapidMiner* untuk meng*cluster* data menjadi dua *cluster*. Data yang sudah dipilih akan dimasukkan kedalam *tool RapidMiner*. Kemudian Dalam penerapan algoritma *K-Means* dihasilkan nilai titik tengah atau *centroid* dari data yang didapat dengan ketentuan bahwa clusterisasi yang diinginkan adalah 2, penentuan *cluster* dibagi atas 2 bagian

yakni *cluster* tingkat tinggi (C1), *cluster* tingkat rendah (C2). Penentuan titik *cluster* ini dilakukan dengan mengambil nilai terbesar (*maksimum*) untuk *cluster* tingkat tinggi (C1), nilai terkecil (*minimum*) untuk *cluster* tingkat rendah (C2).
Untuk mendapatkan nilai tertinggi dan terendah (C1) dan (C2) dari data tabel yang ada maka kita harus mengambil nilai berdasarkan atribut nya masing masing. Untuk Luas Areal yang terbesar (C1) adalah 34,35 Ha sedangkan yang terendah (C2) adalah 22,65 Ha. Kemudian Untuk Berat Janjang yang terbesar (C1) adalah 8,7 Kg sedangkan yang terendah (C2) adalah 7,5 Kg. Kemudian Untuk Realisasai Produksi yang terbesar (C1) adalah 19990 Kg sedangkan yang terendah (C2) adalah 10260 Kg. Sehingga di dapat nilai untuk C1 dan C2 seperti tabel berikut.

Tabel 3. Nilai Centroid

C1	34,35	8,7	19990
C2	22,65	7,5	10260

2. Menghitung Jarak dari *Centorid*
Untuk menghitung jarak antara titik *Centroid* dengan titik tiap objek menggunakan *Euclidean Distance*. Rumus untuk menghitung jarak dari *Centroid* adalah:

$$De = \sqrt{(xi - si)^2 + (yi - ti)^2} \quad (1)$$

Maka perhitungan untuk jarak dari *Centroid* ke-1 adalah sebagai berikut :

$D_{A1,c1} = \sqrt{(34,35- 32,23)^2 + (8,1- 8,45)^2 + (14720-1990)^2} = 5270$

$D_{A2,c1} = \sqrt{(22,65- 32,23)^2 + (7,5- 8,45)^2 + (10760-1990)^2} = 9230$

$D_{A3,c1} = \sqrt{(31,56- 32,23)^2 + (8,7- 8,45)^2 + (19350-1990)^2} = 640$

$D_{A4,c1} = \sqrt{(32,19- 32,23)^2 + (8,4- 8,45)^2 + (16720-1990)^2} = 3270$

$D_{A5,c1} = \sqrt{(33,12- 32,23)^2 + (8,2- 8,45)^2 + (15100-1990)^2} = 4590$

$D_{A6,c1} = \sqrt{(34,35- 32,23)^2 + (8,1- 8,45)^2 + (14520-1990)^2} = 5470$

$D_{A7,c1} = \sqrt{(22,65- 32,23)^2 + (7,5- 8,45)^2 + (10260-1990)^2} = 9730$

$D_{A8,c1} = \sqrt{(31,56- 32,23)^2 + (8,7- 8,45)^2 + (19150-1990)^2} = 840$

$D_{A9,c1} = \sqrt{(32,19- 32,23)^2 + (8,4- 8,45)^2 + (16220-1990)^2} = 3770$

$D_{A10,c1} = \sqrt{(33,12- 32,23)^2 + (8,2- 8,45)^2 + (14900-1990)^2} = 5090$

Selanjutnya perhitungan untuk jarak dari *Centroid* ke-2 adalah sebagai berikut:

$D_{A1,c2} = \sqrt{(34,35- 22,65)^2 + (8,1- 7,5)^2 + (14720-10260)^2} = 4460$

$D_{A2,c2} = \sqrt{(22,65- 22,65)^2 + (7,5- 7,5)^2 + (10760-10260)^2} = 500$

$D_{A3,c2} = \sqrt{(31,56- 22,65)^2 + (8,7- 7,5)^2 + (19350-10260)^2} = 9090$

$D_{A4,c2} = \sqrt{(32,19- 22,65)^2 + (8,4- 7,5)^2 + (16720-10260)^2} = 6460$

$D_{A10,c2} = \sqrt{(33,12-- 22,65)^2 + (8,2- 7,5)^2 + (14900-10260)^2} = 4640$

sehingga didapat tabel jarak dari *Centroid* dan mencari nilai minimal dari kedua *centroid*. Tabel jarak dari *centroid* adalah sebagai berikut :

Tabel 4. Nilai Iterasi Ke 1

C1	C2	JARAK TERPENDEK	C1	C2
5270,000034	4460,015387	4460,015387		1
9230,007493	500	500		1
640,0060813	9090,004446	640,0060813	1	
3270,000727	6460,007107	3270,000727	1	
4890,00018	4840,011375	4840,011375	1	
5470,000033	4260,016109	4260,016109		1
9730,007108	0	0		1
840,0046334	8890,004546	840,0046334	1	
3770,000631	5960,007703	3770,000631	1	
5090,000173	4640,011865	4640,011865		1
5370,000034	4360,01574	4360,01574		1
9200,007518	530	530		1
2,79	9730,004154	2,79	1	

3180,000748	6550,007009	3180,000748	1	
4690,000188	5040,010924	4690,000188	1	
5210,000035	4520,015182	4520,015182		1
9190,007526	540	540		1
140,0277976	9590,004214	140,0277976	1	
3370,000706	6360,007219	3370,000706	1	
4760,000185	4970,011078	4760,000185	1	
5220,000034	4510,015216	4510,015216		1
9130,007576	600	600		1
100,0389129	9630,004197	100,0389129	1	
3430,000693	6300,007287	3430,000693	1	
4610,000191	5120,010753	4610,000191	1	

3. Menentukan *Cluster atau Pengelompokan*

Dalam menentukan *Cluster* dengan nilai *Cluster* berdasarkan nilai yang sama dari nilai *Cluster* dan diletakkan pada *Cluster* yang sesuai dengan nilai sama pada Iterasi 1. Selanjutnya dalam metode *K-Means*, perhitungan berhenti apabila *Cluster* pada iterasi yang dihasilkan sama pada iterasi sebelumnya. Maka selanjutnya mencari *Cluster* pada iterasi selanjutnya sampai nilai iterasinya sama. Untuk mencari nilai *Centroid* selanjutnya dengan menggunakan *Centroid* baru pada Iterasi ke-1 dengan menjumlahkan nilai sesuai yang tertera pada *Cluster* di tabel diatas. Adapun *Centroid* baru untuk mencari *Cluster* selanjutnya adalah dengan menjumlahkan nilai yang terpilih pada *Cluster* tersebut kemudian membagikannya sebanyak jumlah nilai sebagai berikut :

$C_{1a} = (31{,}56 + 32{,}19 + 33{,}12 + 31{,}56 + 32{,}19 + 31{,}56 + 32{,}19 + 33{,}12 + 31{,}56 + 32{,}19 + 33{,}12 + 31{,}56 + 32{,}19 + 33{,}12)/14 = 32{,}230$

$C_{1b} = (8{,}7 + 8{,}4 + 8{,}2 + 8{,}7 + 8{,}4 + 8{,}7 + 8{,}4 + 8{,}2 + 8{,}7 + 8{,}4 + 8{,}2 + 8{,}7 + 8{,}4 + 8{,}2)/14 = 8{,}45$

$C_{1c} = (19350 + 16720 + 15100 + 19150 + 16220 + 19990 + 16810 + 15300 + 19850 + 16620 + 15230 + 19890 + 16560 + 15380)/14 = 17297{,}857$

Maka, data *Centroid* baru Iterasi ke-1 adalah sebagai berikut :

Tabel 5. Centroid Baru

C1	32,23071429	8,45	17297,85714
C2	28,92	7,836363636	12889,09091

Sehingga didapat tabel jarak dari *Centroid* dan mencari nilai minimal dari kedua *centroid*. Tabel jarak dari *centroid* adalah sebagai berikut :

Tabel 6. Nilai Iterasi Ke 2

C1	C2	JARAK TERPENDEK	C1	C2
2577,858038	1830,917162	1830,917162		1
6537,864232	2129,100168	2129,100168		1
2052,142982	6460,909688	2052,142982	1	
577,8571465	3830,910528	577,8571465	1	
2197,857337	2210,91311	2197,857337	1	
2777,857973	1630,918152	1630,918152		1
C1	**C2**	**JARAK TERPENDEK**	**C1**	**C2**
7037,863728	2629,098407	2629,098407		1
1852,142995	6260,909707	1852,142995	1	
1077,857145	3330,910744	1077,857145	1	
2397,857321	2010,91351	2010,91351		1
2677,858004	1730,917628	1730,917628		1
6507,864264	2099,1003	2099,1003		1
2692,142952	7100,909634	2692,142952	1	
487,8571471	3920,910495	487,8571471	1	
1997,857356	2410,912777	1997,857356	1	
2517,858059	1890,916906	1890,916906		1
6497,864275	2089,100345	2089,100345		1
2552,142958	6960,909645	2552,142958	1	
677,8571459	3730,910566	677,8571459	1	

2067,857349	2340,912887	2067,857349	1	
2527,858055	1880,916947	1880,916947		1
6437,864342	2029,100624	2029,100624		1
2592,142956	7000,909642	2592,142956	1	
737,8571457	3670,910591	737,8571457	1	
1917,857365	2490,912658	1917,857365	1	

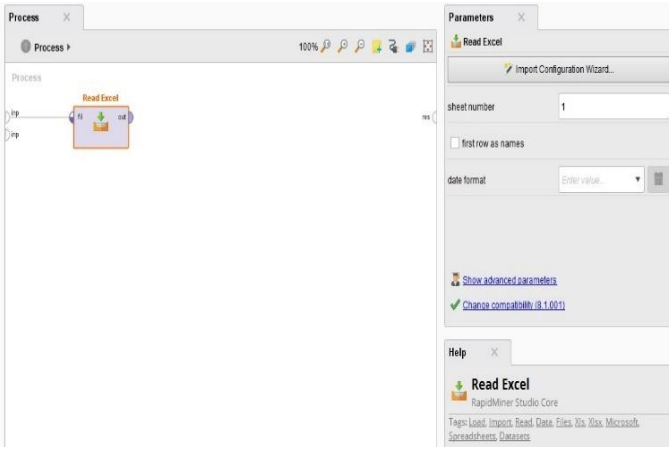
Dari tabel *Cluster* Iterasi ke-1 dan tabel *Cluster* Iterasi ke-2 memiliki nilai *Cluster* yang sama atau tidak berubah pada *cluster* terakhir maka perhitungan dihentikan dan hasil yang diperoleh yaitu:

1. *Cluster* 1 (*C1*) mendapatkan 14 data Blok Panen dalam artian bahwa kelompok pertama termasuk kategori Hasil panen tertinggi Pada Tahun 2015 - 2019.
2. *Cluster* 2 (*C2*) mendapatkan 11 Blok dalam artian bahwa kelompok kedua termasuk kategori hasil panen rendah pada tahun 2015-2019.

Dari tabel *Cluster* Iterasi ke-1 dan tabel *Cluster* Iterasi ke-2 memiliki nilai *Cluster* yang sama atau tidak berubah pada *cluster* terakhir maka perhitungan dihentikan.

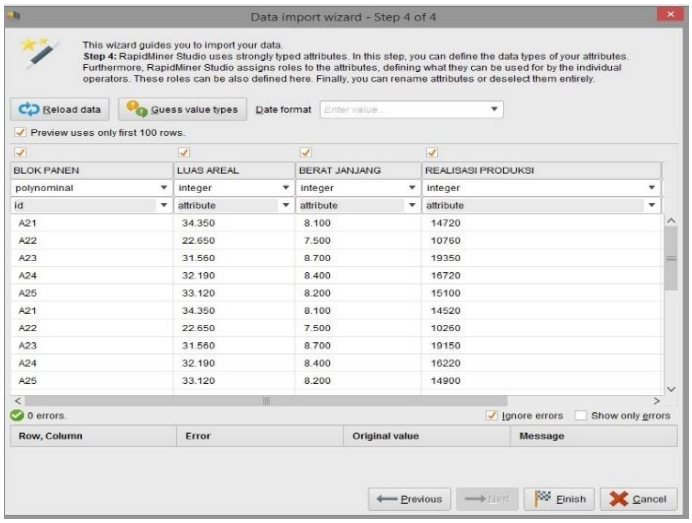
3.2 Pemodelan *RapidMiner*

Dalam menjalankan proses untuk membentuk pemodelan, tahap pertama yang dilakukan adalah *importing* data dari data yang sudah di transformasi terlebih dahulu ke dalam *Microsoft Excel* dengan format .xls agar dapat diakses menggunakan *software RapidMiner*. Atribut yang digunakan sebagai label adalah data hasil panen jagung. Dalam melakukan *importing* data, dibutuhkan operator baru seperti gambar berikut.



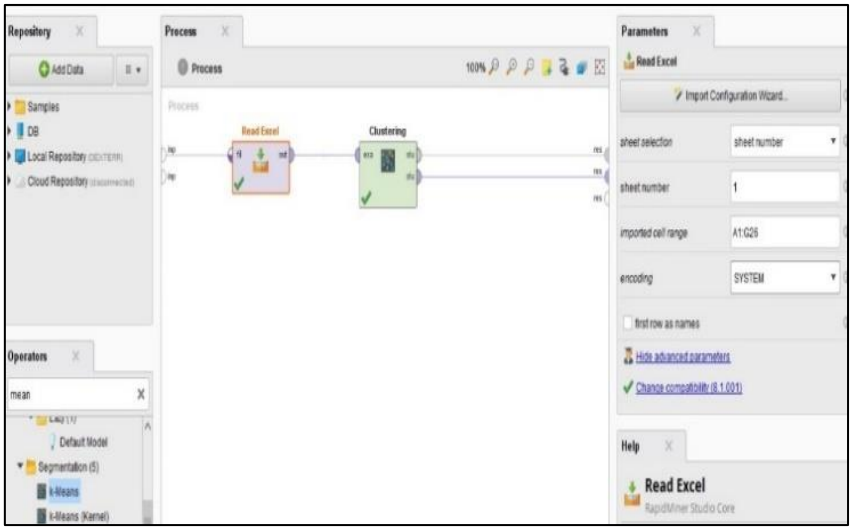
Gambar 3. *New Process* Untuk *Import* Data Excel

Pilih data yang akan digunakan untuk di *import*. Lakukan *drag* and *drop* data yang dipilih kedalam *main process*. Berikut alur proses *import* data yang dilakukan. Pada *import* data *wizard step* 1, pilih data berformat .xls pada folder yang akan di *import*.



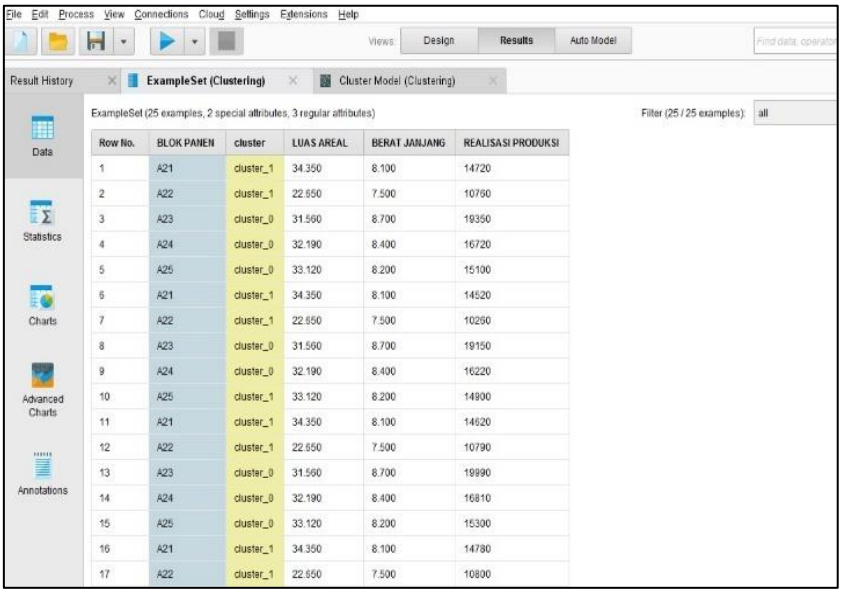
Gambar 4. Alur Proses *Import* Data *Step* 2

Pada data *import wizard step* 2, data sawit.xls dengan *import excel* dimasukan. File *Excel* dapat berisi beberapa lembar, cukup menandai beberapa sel yang akan dimuat. Pada proses kali ini, peneliti memilih sel 1 untuk dilanjutkan prosesnya kemudian klik *next*. Langkah berikutnya *drag* dan *drop* operator *K-Means* lalu hubungkan pada *output.*



Gambar 5. Operator K-Means

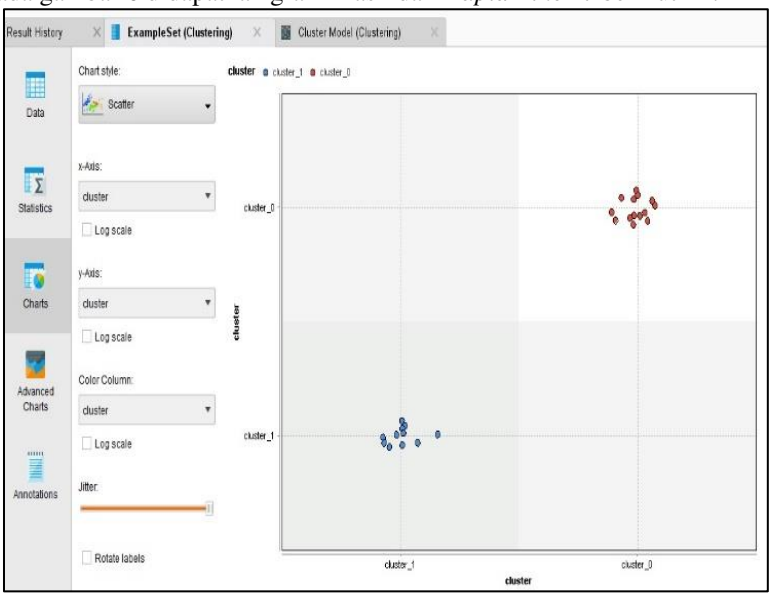
Pada saat *tool run* diklik akan muncul hasil cluster seperti gambar dibawah.



Row No.	BLOK PANEN	cluster	LUAS AREAL	BERAT JANJANG	REALISASI PRODUKSI
1	A21	cluster_1	34.350	8.100	14720
2	A22	cluster_1	22.650	7.500	10760
3	A23	cluster_0	31.560	8.700	19350
4	A24	cluster_0	32.190	8.400	16720
5	A25	cluster_0	33.120	8.200	15100
6	A21	cluster_1	34.350	8.100	14520
7	A22	cluster_1	22.650	7.500	10260
8	A23	cluster_0	31.560	8.700	19150
9	A24	cluster_0	32.190	8.400	16220
10	A25	cluster_1	33.120	8.200	14900
11	A21	cluster_1	34.350	8.100	14620
12	A22	cluster_1	22.650	7.500	10790
13	A23	cluster_0	31.560	8.700	19990
14	A24	cluster_0	32.190	8.400	16810
15	A25	cluster_0	33.120	8.200	15300
16	A21	cluster_1	34.350	8.100	14780
17	A22	cluster_1	22.650	7.500	10800

Gambar 6. Data Hasil *Clustering*

Dari gambar 6 dapat dilihat bahwa perhitungan manual yang dilakukan memiliki hasil yang sama dengan menggunakan *software RapidMiner,* dimana dari 25 data hasil produksi kelapa sawit yang diuji hasil yang didapatkan yaitu 14 data termasuk kedalam *cluster* tertinggi dan 11 data termasuk kedalam *cluster* terendah. Sehingga berdasarkan data hasil *clustering* pada gambar 6 didapatkan grafik hasil dari *RapidMiner* 7 berikut ini:



Gambar 7. Grafik *Clustering*

4. KESIMPULAN

Berdasarkan pembahasan sebelumnya dapat disimpulkan bahwa: Metode yang digunakan dalam perhitungan skripsi ini adalah penerapan data mining dengan metode *K-Means Clustering*. Hasil dari penelitian menggunakan metode *K-Means* ini berupa 2 (dua) buah *cluster* yaitu *cluster* tinggi dan *cluter* rendah. Penerapan Algoritma *K-means* dapat diimplementasikan terhadap pemetaan hasil produksi buah sawit. Penerapan Algoritma *K-means* dapat diimplementasikan ke dalam software *Rapidminner.* Hasil dari penelitian ini di dapakan C1 (tertinggi) ialah 14 data Blok Panen, dan C2 (terendah) ialah 11 data Blok Panen.

UCAPAN TERIMA KASIH

Dalam melakukan penelitian ini tentu ada dukungan dari banyak pihak mulai dari anggota keluarga, rekan kerja, dosen dan para pegawai di PTPN IV Marihat. Untuk itu kepada semua pihak yang terlibat kami ucapkan terima kasih.

REFERENSI

[1] Asroni and R. Adrian, “Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang,” vol. 18, no. 1, pp. 76–82, 2016.

[2] A. F. Tanjung, T. M. Diansyah, and Rismayanti, “Pemanfaatan Algortima K-Means Clustering Sebagai Pengamanan Pencurian Buah Kelapa Sawit Se-Distrik Tandun PT . Perkebunan Nusantara V,” vol. 3, no. 4, pp. 351–356, 2019.

[3] A. Lailan and U. Kurnia, “Prediksi Replanting Lahan Kelapa Sawit Dengan Menggunakan Metode K-Nearest Neighbor,” *PELITA Inform.*, vol. 18, pp. 517–522, 2019.

[4] N. Sirait, “Implementasi K-Means Clustering Pada Pengelompokan Mutu Biji Sawit,” *J. Pelita Inform.*, vol. 16, no. 4, pp. 368–372, 2017.

[5] N. Krisandi, Helmi, and B. Prihandono, “Algoritma K-Nearest Neighbor Dalam Klasifikasi Data HAsil Produksi Kelapa Sawit Pada PT. Minamas Kecamatan Parindu,” vol. 02, no. 1, pp. 33–38, 2013.

[6] Gustientiedina, M. H. Adiya, and Y. Desnelita, “Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru,” vol. 01, pp. 17–24, 2019.

[7] L. Maulida, “Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. DKI Jakarta Dengan K-Means,” vol. 2, no. 3, pp. 167–174, 2018.

[8] N. Erlangga, Solikhun, and Irawan, “Penerapan Data Mining Dalam Mengelompokan Produksi Jagung Menurut Provinsi Menggunakan Algoritma K-Means,” vol. 3, pp. 702–709, 2019.

[9] I. Parlina, A. P. Windarto, A. Wanto, and M. R. Lubis, “Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Asessment Center Untuk Clustering Program SDP,” vol. 3, no. 1, pp. 87–93, 2018.