

Clusterisasi Perkebunan Pinang Menggunakan Metode K-Means

Muhammad Amin

Program Studi Teknik Industri, Universitas Islam Indragiri
Jln. Provinsi Parit 1 Tembilahan - Indragiri Hilir – Riau
Email: ma618152@gmail.com

Abstract

Indragiri Hilir Regency which has several districts, one of which is the Batang Tuaka sub-district. Batang Tuaka sub-district is one of the districts with the largest crop yields in Indragiri Hilir and even in Riau. There are differences in the yields of areca nut plantations, which are different in number, so it is necessary to group which villages produce the most or little areca nuts. The k-means method can be used as a method for grouping areca producing villages based on land area and yields. The data used were 13 data with criteria of fertility, infertility and infertility. The test results are divided into 3 clusters, namely cluster 0 for the fertile category amounting to 3, cluster 1 for the infertile category amounting to 1 and cluster 2 for the infertile category amounting to 9. The K-means method can be applied in the data processing using the data mining concept.

Keywords: Data Mining, K-Means, Clustering, Rapid Miner, Pinang Plantation

Abstrak

Kabupaten Indragiri Hilir yang memiliki beberapa kecamatan salah satu diantaranya kecamatan Batang Tuaka. Kecamatan Batang Tuaka termasuk kecamatan dengan hasil panen perkebunan terbesar di Indragiri Hilir bahkan di Riau. Adanya perbedaan dari hasil panen perkebunan pinang yang jumlahnya berbeda-beda, maka diperlukan pengelompokan desa mana yang menghasilkan pinang yang banyak atau sedikit. Metode k-means dapat digunakan sebagai metode untuk pengelompokan desa penghasil pinang berdasarkan luas lahan dan hasil panen. Data yang digunakan sebanyak 13 data dengan kriteria subur, kurang subur dan tidak subur. Hasil pengujian dibagi menjadi 3 cluster yaitu cluster 0 untuk kategori subur berjumlah 3, cluster 1 untuk kategori kurang subur berjumlah 1 dan cluster 2 untuk kategori tidak subur berjumlah 9. Metode K-means dapat diterapkan dalam proses pengolahan data menggunakan konsep data mining.

Katakunci: Data Mining, K-Means, Clustering, RapidMiner, Perkebunan Pinang

I. PENDAHULUAN

Kabupaten Indragiri Hilir yang memiliki jumlah kecamatan sebanyak 20 kecamatan. Salah satu diantaranya yaitu kecamatan Batang Tuaka dengan jumlah desa sebanyak 13 desa. Kecamatan Batang Tuaka termasuk kecamatan dengan hasil panen perkebunan terbesar di kabupaten Indragiri Hilir. Pada sektor-sektor perkebunan kebanyakan yang diusahakan oleh para petani yang ada di kecamatan Batang Tuaka adalah perkebunan pinang.

Dalam pencapaian keberhasilan hasil panen perkebunan pinang di kecamatan Batang Tuaka, maka perlu adanya suatu tindakan atau metode untuk mengelompokkan dengan tujuan desa mana yang memiliki potensi hasil perkebunan pinang yang bagus. Metode yang digunakan adalah metode *K-means Clustering*. Adapun acuan dalam melakukan pengelompokan ini diperlukan data-data desa perkebunan pinang yang ada di kecamatan Batang Tuaka.

Manfaat pada penelitian diataranya mendapatkan suatu informasi mengenai pengelompokan desa-desa mana yang dakatagorikan perkebunan yang subur, kurang subur dan tidak subur. Hasil penelitian ini dapat dijadikan pertimbangan pemerintah untuk meningkatkan lagi hasil perkebunan pinang yang ada di kecamatan Batang Tuaka.

Penelitian yang sudah dilakukan sebelumnya dengan judul segmentasi citra paru menggunakan metode *k-means clustering* dengan hasil penelitian bahwa hasil terbaik pada jumlah cluster 8 karena pada $k=8$ citra dengan hasil clustering mendapatkan penayangan batas warna yang jelas. Citra hasil dari clustering dapat dijadikan sebagai data awal (Atina, 2017).

Penelitian yang sudah dilakukan dengan judul data mining dalam pengelompokan jenis dan jumlah penbagian zakat dengan menggunakan metode clustering (studi kasus: badan amil zakat kota bengkulu) dengan hasil penelitian bahwa pengelompokan *k-means clustering* data 1-2-1-1 dan 3-4-3-4 dengan hasil yang didapat dari pengujian tersebut adalah dapat mengelompokan pembagian zakat dengan hasil pada setiap masing-masing cluster tidak ada sama (Prahasti, 2017).

Penelitian dengan judul klasifikasi data mahasiswa menggunakan metode *k-means* untuk menunjang pemilihan strategi pemasaran dengan hasil penelitian bahwa pengelompokan mahasiswa dengan hasil nilai rerata UN yang nilai rendah mempunyai pengaruh terhadap hasil prestasi akademik terhadap mahasiswa yang rendah pada jenjang D-3 dan S-1. Jika mahasiswa memiliki nilai UN yang tinggi maka prestasi akademik terhadap mahasiswa juga tinggi pada semua jenjang. Dengan hasil pengelompokan berdasarkan dari daerah asal sekolah IPK rerata yang tertinggi yang mana

berasal dari daerah istimewa seperti Yogyakarta dan Jawa Tengah (Suprawanto, 2016).

Penelitian dengan judul klasifikasi penyakit ginjal dengan metode *k-means* dengan hasil penelitian bahwa tingkat akurasi pengclustoran penyakit ginjal menggunakan metode *k-means* dalam penelitian ini cukup tinggi, yaitu 90% (Irtawaty, 2017).

Data mining merupakan salah satu proses yang sangat berguna untuk mendapatkan informasi dari penyimpanan basis data yang besar. Selain itu data mining juga dapat diartikan juga suatu proses pemisah informasi baru dari data besar untuk membantu dalam mengambil suatu keputusan (Rahmawati, et al., 2017).

Data mining yaitu proses ekstraksi atau menemukan data yang sebelumnya data belum diketahui tetapi dapat dipahami dari basis data yang besar dan juga digunakan untuk membuat suatu bisnis yang sangat penting ketika mengambil suatu keputusan (Arta, et al., 2016).

Data mining yaitu suatu proses mencari gambaran atau informasi yang menarik diantara data yang dipilih dengan menggunakan suatu teknik atau metode. teknik atau metode data mining yang digunakan sangatlah bervariasi. Pemilihan teknik atau metode yang akurat tergantung pada proses dan tujuan *knowledge discovery in database*. Metode yang akan digunakan salah satunya metode *k-means clustering* (Muzakir & Wulandari, 2016).

Data mining yaitu suatu proses mencari informasi yang menarik diantara data yang sudah dipilih dengan teknik atau dengan metode tertentu. Beberapa macam teknik atau algoritma yang ada dalam data mining. Metode atau algoritma yang bisa dipilih denga cara yang tepat sangatlah bergantung terhadap tujuan dan proses pada *knowledge discovery in database* secara menyeluruh (Asroni, et al., 2018).

II. METODOLOGI PENELITIAN

Penerapan metode yang digunakan pada Penelitian ini menggunakan metode algoritma *k-means*. Hasilnya dapat mengelompokan desa mana yang subur, kurang subur dan tidak subur berdasarkan dari data perkebunan pinang yang ada di kecamatan Batang Tuaka berdasarkan luas lahan perkebunan dan hasil panen.

Ada beberapa langkah-langkah yang harus diterapkan pada penelitian diantaranya:

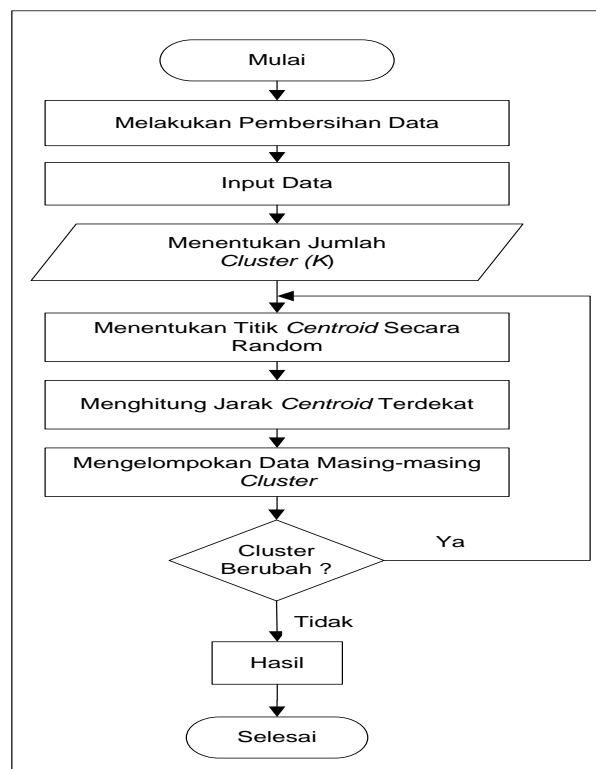
1. Pengumpulan data
Melakukan pengumpulan data yang ada di kecamatan Batang Tuaka atau lembaga yang terkait.
2. Analisa data
Analisa data dilakukan setelah semua data sudah diperoleh kemudian data disusun dan digolongkan kedalam bentuk tabel agar mudah untuk dipahami.

3. Pengujian dengan *Cluster K-means*
Setelah analisa data maka dilakukan ketahap proses pengujian menggunakan *cluster k-means* dengan langkah sebagai berikut:
 - a. Menentukan jumlah *cluster*K
 - b. Pilih titik sentroit cluster secara acak atau sembarangan
 - c. Gunakan rumus jarak *euclidean* dalam menghitung jarak data-data yang tertuju pada setiap titik pusat *cluster*.
 - d. Lakukan penghitungan kembali dengan anggota *cluster* yang ada.
 - e. Jika masih dilakukan perhitungan kembali maka setiap objek menggunakan cluster yang baru. Langkah ini akan berakhir jika tidak ada perubahan lagi pada pusat cluster.
4. Implementasi *RapidMiner*
Implementasi dilakukan untuk menguji kebenaran yang didapatkan dari hasil perhitungan algoritma *k-means* sebelumnya menggunakan aplikasi *RapidMiner*
5. Hasil
Hasil yang didapatkan dari penelitian ini adalah pengelompokan dari desa-desa berdasarkan dari hasil cluster pada masing-masing desa. Kemudian hasil dari pengelompokan ini dapat di rekomendasikan kepada pemerintah sebagai pertimbangan untuk mengambil keputusan.

III. HASIL DAN PEMBAHASAN

K-means cluster adalah sistem algoritma yang dapat digunakan dalam model pengelompokan data dalam bentuk set-set data dan juga membagi data menjadi kelompok-kelompok yang sudah ditentukan.

Algoritma *k-means* yang dilaksanakan akan digambarkan proses-prosesnya dalam bentuk *flowchart* berikut ini.



Gambar 1. Flowchart K-means Cluster

Berdasarkan flowchart k-means cluster yang terdapat pada Gambar 1 maka akan dijelaskan uraian langkah-langkahnya diantaranya:

1. Pembersihan data
Proses yang dilakukan untuk pengkoreksian data, pengurangan data atau pembuangan data jika terdapat ada yang ganda. Setelah data di bersihkan tujuannya agar memudahkan dalam proses perhitungan.
2. Input data
Data yang digunakan dalam penginputan adalah data desa yang ada di kecamatan Batang Tuaka dengan jumlah desa 13 desa.
3. Menentukan jumlah cluster
Tentukan jumlah cluster terlebih dahulu. Cluster yang digunakan pada penelitian ini sebanyak 3 cluster. Cluster 1 adalah untuk kategori subur, cluster 2 untuk kategori kurang subur dan cluster yang ke 3 untuk kategori tidak subur.
4. Menentukan centroid awal
Penentuan centroid awal dilakukan dengan cara random. Rumus yang digunakan yaitu rumus Euclidean Distance yang menentukan jarak dari masing-masing titik centroid. Rumusnya yaitu:

$$D_{(k,l)} = \sqrt{(X_{1k} - X_{1l})^2 \dots + (X_{mk} - X_{ml})^2}$$

Dimana:

$D_{(k,l)}$ = Jarak ke k ke pusat cluster l

X_{mk} = Data ke k pada atribut data ke m

X_{ml} = Titik pusat ke l pada atribut ke m

Tabel untuk centroid awal dapat dilihat pada tabel berikut ini.

Tabel 1. Titik Centroid Awal Cluster

Centroid	Luas lahan	Hasil Produksi
Cluster 0	106	51,389
Cluster 1	218	115,05
Cluster 2	15	9,204

5. Menghitung jarak

Langkah-langkah perhitungan untuk jarak masing-masing data ke titik pusat cluster dapat dilihat berikut ini:

a. Proses hitung cluster (C0)

$$D1 = \sqrt{(26-106)^2 + (15,34-51,389)^2}$$

$$= 87,7469$$

$$D1 = \sqrt{(14-106)^2 + (8,821-51,389)^2}$$

$$= 101,3707$$

proses perhitungan dilakukan sampai data ke 13

b. Proses hitung cluster (C1)

$$D1 = \sqrt{(26-218)^2 + (15,34-115,05)^2}$$

$$= 216,347$$

$$D1 = \sqrt{(14-218)^2 + (8,821-115,05)^2}$$

$$= 230,001$$

proses perhitungan dilakukan sampai data ke 13

c. Proses hitung cluster (C2)

$$D1 = \sqrt{(26-15)^2 + (15,34-9,204)^2}$$

$$= 12,5956$$

$$D1 = \sqrt{(14-15)^2 + (8,821-9,204)^2}$$

$$= 1,0708$$

proses perhitungan dilakukan sampai data ke 13

Tabel 2. Hasil Iterasi 1

Centroid	Centroid Awal	Iterasi 1	
Cluster 0	106	51,389	87,5 46,979
Cluster 1	218	115,05	218 115,05
Cluster 2	15	9,204	25,6 15,5702

Hasil perhitungan pertama menunjukkan bahwa proses iterasi pertama tidak sama dengan titik pusat awal. Maka langkah selanjutnya akan dilakukan proses perhitungan kembali menggunakan rumus euclidean distance. Berikut ini adalah langkah-langkah perhitungannya.

a. Proses hitung cluster (C0)

$$D1 = \sqrt{(26-87,5)^2 + (15,34-46,979)^2}$$

$$= 69,1612$$

$$D1 = \sqrt{(14-87,5)^2 + (8,821-46,979)^2}$$

$$= 82,8147$$

proses perhitungan dilakukan sampai data ke 13

b. Proses hitung cluster (C1)

$$D1 = \sqrt{(26-218)^2 + (15,34-115,05)^2}$$

$$= 216,347$$

$$D1 = \sqrt{(14-218)^2 + (8,821-115,05)^2}$$

$$= 230,001$$

proses perhitungan dilakukan sampai data ke 13

c. Proses hitung cluster (C2)

$$D1 = \sqrt{(26-25,6)^2 + (15,34-15,5702)^2}$$

$$= 0,4615$$

$$D1 = \sqrt{(14-25,6)^2 + (8,821-15,5702)^2}$$

$$= 13,4205$$

proses perhitungan dilakukan sampai data ke 13

Setelah jarak diperoleh, maka selanjutnya dilakukan pengelompokan berdasarkan titik cluster dengan nilai yang terendah. Perhitungan dihentikan jika titik centroidnya tidak berubah lagi. Berikut ini adalah tabel hasil pengelompokan setiap cluster.

Tabel 3. Hasil Pengelompokan Setiap Cluster

Cluster	Jumlah Data
Cluster 0	3
Cluster 1	1
Cluster 2	9

6. Hasil pengetahuan

Dari hasil proses perhitungan maka didapatkan hasil untuk masing-masing kelompok diantaranya:

a. Kategori subur atau cluster 0 (C0) dengan jumlah 3. Hasil pengelompokan dapat dilihat pada tabel berikut ini.

Tabel 4. Cluster 0 Kategori Subur

No	Nama Desa	Luas lahan	Hasil panen
1	SUNGAI RAYA	102	64,428
2	SUNGAI JUNJANGAN	106	51,389
3	PASIR MAS	69	42,569

b. Kategori kurang subur atau cluster 1 (C1) dengan jumlah 1. Hasil pengelompokan dapat dilihat pada tabel berikut ini.

Tabel 5. Cluster 1 Kategori Kurang Subur

No	Nama Desa	Luas lahan	Hasil panen
1	KUALA SEBATU	218	115,05

c. Kategori tidak subur atau cluster 2 (C2) dengan jumlah 9. Hasil pengelompokan dapat dilihat pada tabel berikut ini.

Tabel 6. Cluster 2 Kategori Tidak Subur

No	Nama Desa	Luas lahan	Hasil panen
----	-----------	------------	-------------

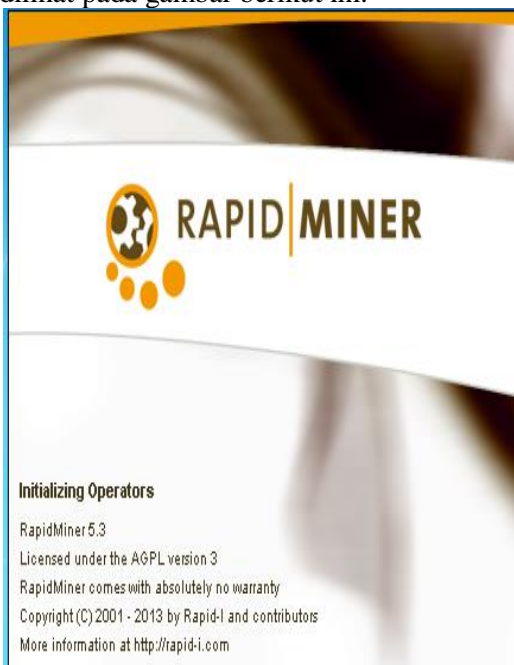
1	SUNGAI DUSUN	26	15,34
2	TASIK RAYA	14	8,821
3	SIALANG JAYA	8	5,369
4	SUNGAI RAWA	7	4,602
5	SUNGAI PIRING	13	6,52
6	GEMILANG JAYA	15	9,204
7	SUNGAI LUAR	25	12,272
8	TANJUNG SIANTAR	42	28,379
9	SIMPANG JAYA	4	0,767

IV. IMPLEMENTASI

Pada pengujian ini peneliti menggunakan *software rapidminer* untuk menguji kebenarannya. Terdapat beberapa langkah dalam menjalankan *software rapidminer* diantaranya sebagai berikut:

1. *Icon RapidMiner*

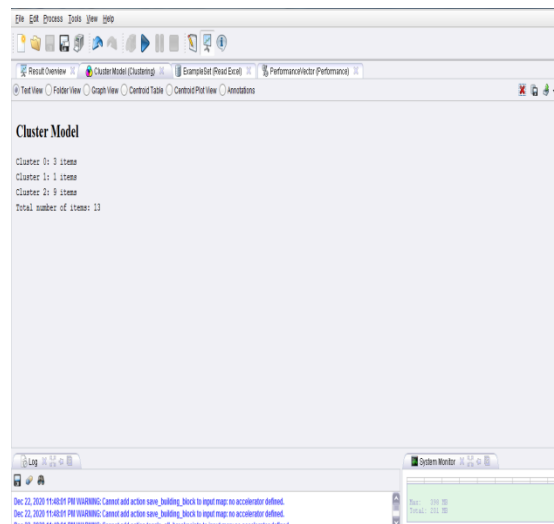
Tampilan utama *software rapidminer* dapat dilihat pada gambar berikut ini.



Gambar 2. Icon RapidMiner

2. *Text view*

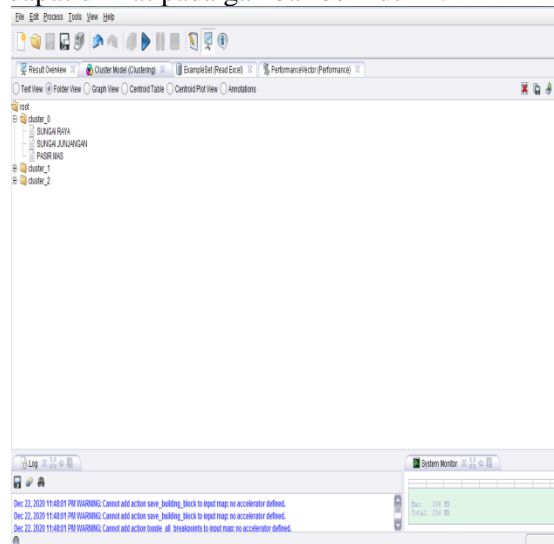
Text view merupakan hasil keseluruhan dari hasil *cluster* yang sudah dilakukan pada *RapidMiner*. Hasil *cluster* dapat dilihat pada gambar berikut ini.



Gambar 3. Tampilan Hasil Cluster

3. *Folder view*

Folder view yang dihasilkan untuk *cluster 0*, dapat dilihat pada gambar berikut ini.



Gambar 4. Folder view cluster 0

Folder view yang dihasilkan untuk *cluster 1* dapat dilihat pada gambar berikut ini.

