

Penerapan Metode Clustering Dengan K-Means Untuk Memetakan Potensi Tanaman Padi Di Kota Semarang

Lianna Felicia

Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

Abstract

Berdasarkan data hasil pertanian padi di Dinas Pertanian Provinsi Jawa Tengah, menampilkan beberapa daerah dengan hasil panen padi yang bervariasi jumlahnya. Untuk itu diperlukan pengelompokan daerah potensial penghasil padi untuk mengetahui daerah mana saja yang menghasilkan padi dengan jumlah banyak ataupun sedikit. Pembagian hasil panen biasanya dilakukan berdasarkan nama kecamatan penghasil padi. Oleh karena itu, dibutuhkan metode untuk memudahkan dalam pengelompokan daerah penghasil padi. Dengan pendekatan pengklasteran k-means, pembagian kelompok daerah dapat dilakukan berdasarkan luas panen (Ha), produksi(ton) dan tahun panen. Pada penelitian ini dilakukan pengklasteran daerah potensial penghasil padi menggunakan algoritma *K-Means*. Dengan menggunakan K-Means bertujuan dalam memudahkan pengelompokan suatu daerah dengan hasil panen padi terbesar, sedang dan rendah. Hasilnya adalah sebuah gambaran yang menunjukkan pengelompokan daerah berdasarkan hasil pertanian padi.

Kata kunci : Pengelompokan, Data Mining, Cluster, Algoritma K-Means

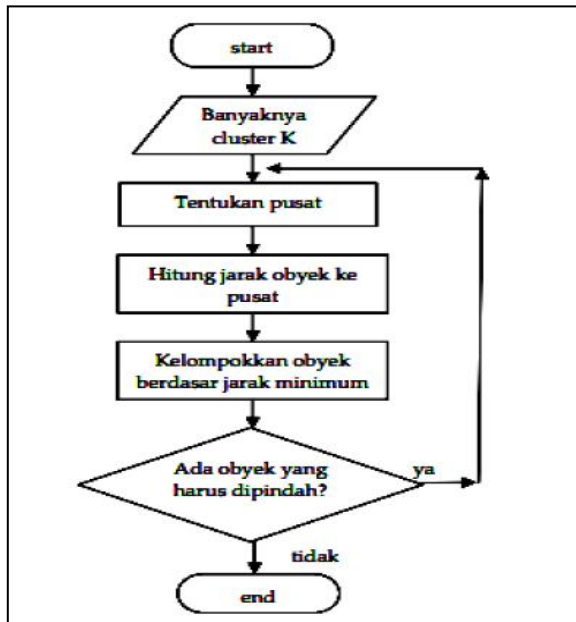
I. PENDAHULUAN

Pangan adalah kebutuhan manusia yang paling mendasar. Menurut UU RI nomor 7 tahun 1996 tentang pangan menyebutkan bahwa pangan merupakan hak asasi bagi setiap individu di Indonesia. Dalam kebutuhan pangan, sektor pertanian digunakan untuk memproduksi beras yang merupakan makanan pokok warga negara Indonesia secara umum. Produksi beras dalam negeri diharapkan dapat memenuhi semua kebutuhan masyarakat Indonesia karena dengan berhasilnya pemenuhan beras dalam negeri berarti pemerintah tidak memerlukan tindakan untuk mengimpor beras dari negara lain. Akan tetapi dalam kenyataannya, Indonesia dalam pemenuhan kebutuhan beras masih bergantung pada impor beras dari negara lain. Harga beras dari tahun ke tahun mengalami peningkatan harga. Kenaikan harga tersebut antara lain disebabkan menipisnya stok beras di beberapa daerah karena belum memasuki panen raya atau juga dikarenakan impor beras yang dilakukan pemerintah untuk memenuhi kebutuhan nasionalnya belum terealisasi. Produksi beras di Indonesia pada tahun ke tahun terus meningkat karena harus memenuhi target yang telah

dicapai pada tahun sebelumnya. Tetapi bukan berarti dapat mencukupi ketersediaan beras karena setiap tahun pula jumlah penduduk meningkat sehingga peningkatan jumlah produksi beras dilakukan untuk mengimbangi tingginya jumlah penduduk Indonesia yang mengkonsumsi beras. Tingkat konsumsi beras di Indonesia saat ini dapat dikatakan tinggi karena setiap orang di Indonesia mengkonsumsi beras setiap tahunnya sebesar 139,5 kg. Untuk itu dalam rangka memenuhi kebutuhan padi, dinas pertanian berupaya untuk mengoptimalkan hasil pertanian dengan memetakan atau mengelompokkan daerah yang menghasilkan tanaman padi di daerah Jawa tengah khususnya di kota Semarang dengan metode *clustering*. Tujuannya adalah untuk mengetahui daerah potensial penghasil padi dan dapat mengetahui daerah tersebut cocok untuk tanaman padi. Pengelompokan tersebut dapat menggunakan metode pengelompokan dengan algoritma *K-Means*. Dengan data yang sudah dikelompokkan menggunakan algoritma *K-Means* diharapkan dapat mempermudah dinas pertanian dalam menghitung hasil pertanian di tiap daerahnya agar mengetahui daerah mana yang menghasilkan padi terbanyak, sedang, dan sedikit.

II. METODE YANG DIUSULKAN

Konsep dasar data mining adalah menentukan informasi tersembunyi dalam sebuah basis data dan merupakan bagian dari Knowledge Discovery in Databases (KDD) untuk menemukan informasi dan pola yang berguna dalam data. Data mining mencari informasi baru, berharga dan berguna dalam sekumpulan data dengan melibatkan komputer dan manusia serta bersifat iteratif baik melalui proses yang otomatis ataupun manual. Clustering membagi data menjadi kelompok-kelompok atau cluster berdasarkan suatu kemiripan atribut-atribut di antara data tersebut. Karakteristik tiap cluster tidak ditentukan sebelumnya, melainkan tercermin dari kemiripan data yang terkelompok di dalamnya. *K-Means* termasuk dalam *partitioning clustering* yaitu setiap data harus masuk dalam *cluster* tertentu dan memungkinkan bagi setiap data yang termasuk dalam *cluster* tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah ke *cluster* yang lain.



Gambar 1. Tahap Algoritma K-Means

algoritma K-Means selanjutnya akan melakukan pengulangan langkah-langkah berikut sampai terjadi kestabilan :

1. Menentukan banyak cluster

Untuk melakukan clustering dengan algoritma k-means langkah yang pertama kali yaitu menentukan banyak cluster yang akan dibentuk. Pada penelitian ini, terdapat tiga cluster yaitu cluster untuk kecamatan dengan produksi padi banyak (C0), produksi sedang (C1) dan produksi sedikit (C2).

2. Menentukan titik pusat (centroid)

Setelah menentukan banyak cluster yang akan dibentuk, langkah selanjutnya yaitu dengan menentukan titik pusat (centroid) dari tiap-tiap cluster. Pengambilan titik pusat secara random (acak). Titik pusat yang diambil pada penelitian ini adalah C0 adalah titik pusat untuk daerah dengan produksi padi tinggi dengan titik pusat (1000;10000), C1 adalah titik pusat untuk daerah dengan produksi sedang dengan titik pusat (500;5000), dan C2 adalah titik pusat untuk daerah produksi padi sedikit dengan titik pusat (50;100).

3. Menghitung jarak setiap objek ke titik pusat (centroid)

Langkah selanjutnya untuk melakukan cluster setelah menentukan titik pusat adalah dengan menghitung jarak setiap data dengan titik pusat yang sudah ditentukan sebelumnya. Rumus untuk menghitung jarak setiap objek ke titik pusat adalah :

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2}$$

Dimana x adalah dataset produksi padi, dataset yang digunakan untuk perhitungan mining ini adalah luas panen dan produksi. Sedangkan y adalah titik pusat yang sudah ditentukan secara acak (random).

Tabel 1. Dataset produksi padi 2013

no	kecamatan	Luas panen (ha)	produktivitas (kw/ha)	Produksi	tahun
1	Mijen	1081	56,82	6142	2013
2	Gunungpati	2172	56,88	12354	2013
3	Banyumanik	213	56,43	1202	2013
4	Tembalang	828	56,86	4708	2013
5	Pedurungan	135	56,74	766	2013
6	Genuk	176	56,42	993	2013
7	Gayamsari	15	56,67	85	2013
8	Semarang Barat	40	56,00	224	2013
9	Ngaliyan	1014	56,79	5758	2013
10	Tugu	889	56,76	5046	2013

Dibawah ini terdapat contoh untuk perhitungan jarak data dengan titik pusat. Contoh untuk kecamatan mijen tahun 2013 dengan luas panen 1081 dan hasil produksi 6561, maka:

Perhitungan jarak mijen dengan pusat cluster pertama adalah :

$$d_{10} = \sqrt{(1081 - 1000)^2 + (6561 - 10000)^2} = 3858,850217$$

Perhitungan jarak mijen dengan pusat cluster kedua adalah :

$$d_{11} = \sqrt{(1081 - 500)^2 + (6561 - 5000)^2} = 1281,298170$$

Perhitungan jarak mijen dengan pusat cluster ketiga adalah :

$$d_{12} = \sqrt{(1081 - 50)^2 + (6561 - 100)^2} = 6129,333161$$

4. Mengelompokkan obyek berdasarkan jarak minimum

Langkah clustering yang terakhir yaitu dengan mengelompokkan obyek berdasarkan jarak minimum antara data dengan titik pusatnya. Di bawah ini terdapat contoh untuk kecamatan mijen yang telah di hitung di langkah sebelumnya.

$$d_{10} = 3858,850217$$

$$d_{11} = 1281,298170$$

$$d_{12} = 6129,333161$$

Maka, jarak minimum untuk kecamatan mijen pada tahun 2013, terdapat pada **cluster 1**, maka kecamatan mijen pada tahun 2013 dikelompokkan sebagai kecamatan dengan produksi padi sedang.

III. IMPLEMENTASI

Dalam tahap ini akan dijelaskan langkah-langkah pengoperasian program beserta contoh gambar.

Gambar 2. Halaman Utama

Gambar diatas menjelaskan tentang tampilan halaman utama saat program dijalankan. Pengguna menginputkan terlebih dahulu nama kecamatan, luas panen, produksi dan tahun.

Gambar 3. Input Data

Gambar di atas menjelaskan Pengguna menginputkan data pada textfield yang sudah tersedia. Lalu untuk menyimpan data klik pada button simpan.

Gambar 4. Jika Data Berhasil Disimpan

Gambar diatas menjelaskan jika data yang diinputkan berhasil disimpan di database.

Sampel Dataset Padi

No	Kecamatan	Luas Panen	Produksi	tahun
1	Mijen	1081	6142	2013
2	Gunung Pati	2172	12354	2013
3	Banyuwangi	213	1202	2013
4	Tembalang	628	4708	2013
5	Pedharangan	135	766	2013
6	Genek	176	993	2013
7	Gayamsari	15	85	2013
0	mijen	2110	34343	2010

Hasil perhitungan jarak setiap data dengan titik pusat

No	Kecamatan	Luas Panen	Produksi	tahun	Cluster 0	Cluster 1	Cluster 2
1	Mijen	1081	6142	2013	3858.8502173575	1281.2981696262	6129.3331611196
2	Gunung Pati	2172	12354	2013	3629.6197443737	7541.6775322205	12436.3740696359
3	Banyuwangi	213	1202	2013	8833.1292869515	3808.8282975214	1113.9896767924
4	Tembalang	628	4708	2013	5294.7944247157	439.14462310269	4673.2160232542
5	Pedharangan	135	766	2013	9274.4261817107	4249.706367257	671.40226392251
6	Genek	176	993	2013	9044.6130376042	4020.077335668	901.8453304198
7	Gayamsari	15	85	2013	9963.8070033497	4938.8713285527	38.07886552932
0	mijen	2110	34343	2010	24368.293928792	29387.135774008	34304.907068815

Gambar 5. Halaman Lihat Data (1)

Hasil Clustering obyek dengan memasukkan setiap obyek ke dalam cluster berdasarkan jarak minimum

No	Kecamatan	Cluster
1	Mijen	C1
2	Gunung Pati	C0
3	Banyuwangi	C2
4	Tembalang	C1
5	Pedharangan	C2
6	Genek	C2
7	Gayamsari	C2
0	mijen	C0

Gambar 6. Halaman Lihat Data (2)

Gambar diatas menjelaskan bagaimana untuk melihat data dan hasil clustering.

Penghapusan Dataset					
#	No	Kecamatan	HA	Produksi	Tahun
<input type="checkbox"/>	1	Mijen	1081	6142	2013
<input type="checkbox"/>	2	Gunung Pati	2172	12354	2013
<input type="checkbox"/>	3	Banyumanik	213	1202	2013
<input type="checkbox"/>	4	Tembalang	828	4708	2013
<input type="checkbox"/>	5	Pedurungan	135	766	2013
<input type="checkbox"/>	6	Genuk	176	993	2013
<input type="checkbox"/>	7	Gayamsari	15	85	2013
<input checked="" type="checkbox"/>	0	mijen	2110	34343	2010

delete [Back](#)

Gambar 7. Halaman Hapus Data

Gambar diatas menjelaskan bahwa jika pengguna hendak menghapus data yang sudah diinputkan sebelumnya. Caranya dengan mengklik pada checkbox yang ada pada baris yang akan dihapus, setelah itu mengklik button delete. Untuk kembali ke halaman utama, pengguna mengklik pada button 'back'.

IV. HASIL & PEMBAHASAN

A. Pembahasan

Pada analisis proses mining menggunakan data produksi padi di Kota Semarang yang telah diolah menjadi dataset pada periode tahun 2009-2013 mempunyai sebanyak 10 transaksi per tahunnya yang merupakan hasil produksi kecamatan-kecamatan yang ada di Kota Semarang, sehingga banyak jumlah transaksi yaitu sebanyak 50 transaksi.

B. Hasil

Hasil penelitian menggunakan data produksi padi di Kota Semarang menghasilkan :

1. Produksi 2009

Tabel 2. Hasil Cluster 2009

Nama kecamatan	Cluster
Mijen	C0
Gunungpati	C1
Banyumanik	C2
Tembalang	C1
Pedurungan	C2
Genuk	C2
Gayamsari	C2
Semarang Barat	C2
Ngaliyan	C1
Tugu	C1

2. Produksi 2010

Tabel 3. Hasil Cluster 2010

Nama kecamatan	Cluster
Mijen	C0
Gunungpati	C1
Banyumanik	C2
Tembalang	C1
Pedurungan	C2
Genuk	C2
Gayamsari	C2
Semarang Barat	C2
Ngaliyan	C1
Tugu	C1

3. Produksi 2011

Tabel 4. Hasil Cluster 2011

Nama kecamatan	Cluster
Mijen	C0
Gunungpati	C0
Banyumanik	C2
Tembalang	C2
Pedurungan	C2
Genuk	C2
Gayamsari	C2
Semarang Barat	C2
Ngaliyan	C1
Tugu	C0

4. Produksi 2012

Tabel 5. Hasil Cluster 2012

Nama kecamatan	Cluster
Mijen	C0
Gunungpati	C0
Banyumanik	C2
Tembalang	C1
Pedurungan	C2
Genuk	C2
Gayamsari	C2
Semarang Barat	C2
Ngaliyan	C1
Tugu	C1

5. Produksi 2013

Tabel 6. Hasil Cluster 2013

Nama kecamatan	Cluster
Mijen	C1
Gunungpati	C0
Banyumanik	C2
Tembalang	C1
Pedurungan	C2
Genuk	C2
Gayamsari	C2
Semarang Barat	C2
Ngaliyan	C1
Tugu	C1

Tabel 2,3,4,5 dan 6 merupakan tabel hasil cluster produksi padi di kecamatan yang ada di kota semarang dari tahun 2009-2013 dimana hasil yang paling banyak yaitu C2 dikatakan produksi padi rendah.

Keterangan :

C0 : kecamatan dengan produksi padi banyak

C1: kecamatan dengan produksi padi sedang

C2 : kecamatan dengan produksi padi sedikit

V. PENUTUP

A. Kesimpulan

Dari permasalahan diatas tentang data produksi tanaman padi di Kota Semarang dapat diselesaikan dengan teknik data mining menggunakan aturan clustering untuk mengelompokkan daerah potensial tanaman padi di Kota Semarang. Software data mining yang penulis rancang dapat memudahkan penyelesaian tugas data mining dengan menggunakan algoritma K-Means. Hasil analisis kelayakan software yang telah dijelaskan diatas maka dapat disimpulkan bahwa software layak untuk diimplementasikan pada Dinas Pertanian di Kota Semarang.

B. Saran

1. Untuk memaksimalkan proses clustering, data yang digunakan lebih banyak agar mendapatkan hasil clustering yang lebih maksimal.
2. Penelitian selanjutnya dapat dikembangkan pada jenis data yang serupa dengan metode yang lebih baik seperti menggunakan algoritma FUZZY C-MEANS
3. Hasil clustering yang terbentuk dapat dikembangkan menjadi basis pengetahuan untuk sistem pendukung keputusan maupun sistem rekomendasi daerah yang cocok untuk ditanami padi oleh dinas pertanian.

REFERENCES

- [1] Edward Christianto, "Faktor Yang Mempengaruhi Volume Impor Beras Di Indonesia," JIBEKA, 2013.
- [2] Hermanto. (2012) <http://bangka.tribunnews.com/2012/10/31>.
- [3] Budi Santoso, Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu, 2007.
- [4] Singgih Santoso, Statistik Multivariat. Jakarta: Elex Media Komputindo, 2010.
- [5] Heribertus Giyanto, "Penerapan algoritma Clustering K-Means , K-Medoid, Gath Geva," 2008.
- [6] Henri Harianja, "Visualisasi K-Means Clustering Pada Data Potensi Pertanian Desa di Bogor Menggunakan Mapserver," 2008.
- [7] Johan Oscar Ong, "Implementasi Algoritma K-Means clustering Untuk Menentukan Strategi Marketing President University," 2013.
- [8] Andri Syafrianto, "Perancangan Aplikasi K-means Untuk Pengelompokan Mahasiswa STMIK ELRAHMA Yogyakarta Berdasarkan Frekuensi Kunjungan Ke Perpustakaan Dan IPK," 2010.
- [9] Daniel T. Larose, Discovering Knowledge In Data. Canada: Wiley-Interscience, 2005.
- [10] Ian Sommerville, Software Engineering. Boston: Pearson Education, Inc., 2011.
- [11] J. I. M. Kerr and R. Hunter, Inside RAD: How To Build a Fully Functional System In 90 Days Or Less., 1993.
- [12] Lonnie D. Bentley, Victor M. Barlow Jeffrey L. Whitten, System Analysis And Design Methods.: Irwin Professional Publishing, 1989.