

# PENGELOMPOKAN KECAMATAN MENGGUNAKAN METODE *K-MEANS CLUSTER*

Amelia Ayu Anggraini<sup>1</sup>  
Lutfi Ali Muharom, S. Si, M. Si<sup>2</sup>

## Abstrak

Kabupaten Bondowoso sebagai salah satu Kabupaten di Propinsi Jawa Timur, memerlukan suatu kebijakan pengaturan ruang sebagai pengarah perkembangan wilayah di masa mendatang. Kabupaten Bondowoso memiliki banyak potensi, khususnya sektor pertanian, perkebunan, dan pariwisata, tetapi keberadaan potensi tersebut belum dapat dimanfaatkan secara optimal.

*K-Means Clustering* merupakan salah satu metode data *clustering non hirarki* yang mengelompokkan data dalam bentuk satu atau lebih *cluster*/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil.

Dari hasil penentuan kelompok kecamatan dengan menggunakan Algoritma *K-Means Clustering* berhasil dilakukan dan pada cluster ke-1 (C1) memiliki anggota kecamatan “Sukosari, Binakal, Pakem, Taman Krocok dan Sempol”, pada kluster ke-2 (C2) terdiri kecamatan “Tlogosari, Sumber wringin, Tapen, Wonosari, Tenggarang, Bondowoso dan Tegalampel” sedangkan pada kluster ke-3 (C3) terdiri kecamatan “Maesan, Grujugan, Tamanan, Jambesari, Pujer, Curahdami, Wringin, Klabang, Botolinggo, Prajekan dan Cerme”. Proses *clustering* dapat menggunakan beberapa pusat *cluster (centroid)* yang diperoleh secara acak.

**Kata Kunci :** Bondowoso, memerlukan, k-means, cluetering, centroid, acak.

## 1. PENDAHULUAN

### 1.1. Latar Belakang

Saat ini, konsep Data Mining semakin dikenal sebagai tools penting dalam manajemen informasi karena jumlah informasi yang semakin besar jumlahnya. *Data Mining* sendiri sering disebut sebagai *knowledge discovery in database (KDD)* adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola hubungan dalam set data berukuran besar. *Output* dari *Data Mining* ini dapat digunakan untuk pengambilan keputusan di masa depan (Alfina, et al. 2012).

Adapun tujuan dari *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi antar *cluster* (Agusta, 2007). *K-Means* merupakan salah satu metode data *clustering non hirarki* yang dalam prosesnya mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok.

Kabupaten Bondowoso sebagai salah satu Kabupaten di Propinsi Jawa Timur, memerlukan suatu kebijakan pengaturan ruang sebagai pengarah perkembangan wilayah di masa mendatang. Kabupaten Bondowoso memiliki banyak potensi, khususnya sektor pertanian, perkebunan, dan pariwisata, tetapi keberadaan potensi tersebut belum dapat dimanfaatkan secara optimal. Dengan menyusun suatu rencana tata ruang yang mampu mengakomodasi potensi, dan memperhatikan kendala serta limitasi wilayah, maka diharapkan dapat tercipta suatu pemanfaatan ruang yang optimal, resposif terhadap perkembangan dan tetap menjaga keseimbangan dengan lingkungan. Perkembangan Kabupaten Bondowoso secara umum cukup pesat. Berbagai kebijakan program pembangunan yang diambil juga menyesuaikan dengan dinamika dan tututan kebutuhan situasi yang terus berubah, sehingga secara keseluruhan

rencana tata ruang yang ada atau yang telah disusun sebelumnya memerlukan beberapa penyesuaian. Salah satu ketentuan terbaru adalah setiap wilayah harus memiliki atau menetapkan kawasan strategis.

Oleh karena itu perlu dilakukan sebuah penelitian yang menggunakan metode dengan memperhitungkan factor jarak kedekatan antar kecamatan. Sehingga diperoleh pengelompokan kecamatan tertinggal yang tepat dan akurat. Salah satu metode yang cukup populer untuk menjawab permasalahan ini adalah *clustering* atau pengelompokan. Atas dasar ide tersebut, peneliti memilih topik penelitian yang berjudul “Pengelompokan Kecamatan Menggunakan Metode K-Means Cluster”.

### 1.2. Rumusan Masalah

Bagaimana mengimplementasikan sebuah sistem yang dapat digunakan untuk menentukan kelompok kecamatan di Kabupaten Bondowoso berdasarkan bidang pendidikan dengan menggunakan algoritma *K-Means Clustering*.

### 1.3. Batasan Masalah

Luasnya cakupan masalah yang berkaitan dengan penentuan pusat cluster, maka pembatasan masalah yang dilakukan dalam penelitian ini adalah:

1. Metode *K-Means Clustering* digunakan untuk menghitung jarak antara data dan centroid menggunakan rumus *Euclidean Distance Space*. Untuk pusat cluster(Centroid)diambil secara *random*.
2. Data yang digunakan dalam penelitian hanya data Kecamatan di Kabupaten Bondowoso sebanyak 23 kecamatan dan hanya menggunakan kriteria berdasarkan bidang pendidikan seperti berikut “Jumlah Sekolah, Jumlah Guru, dan Jumlah Siswa”.

### 1.4. Tujuan Penelitian

Mengelompokkan kecamatan yang ada di Kabupaten Bondowoso dengan menggunakan algoritma *k-means* sehingga dapat membantu pemerintah untuk memberi kebijakan untuk menanggulangi masalah fasilitas maupun pemberdayaan di daerah tersebut.

### 1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Menambah pengetahuan penulis dalam bidang *data mining* khususnya *clustering*.
2. Membantu pengambilan keputusan oleh pemerintah terkait masalah pemberdayaandalam suatu wilayah.

## 2. TINJAUAN PUSTAKA

### 2.1. Algoritma K-Means

Data *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*). *K-Means* merupakan salah satu metode data *Clustering Non Hirarki* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*.

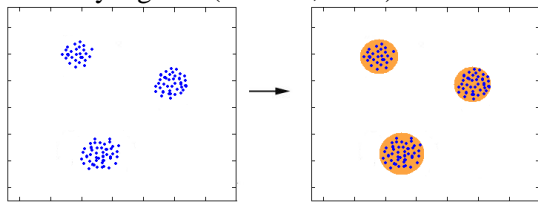
Metode ini mempartisi data ke dalam *cluster* sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diatur dalam proses *clustering*, yang pada umumnya berusaha meminimalkan variasi di dalam suatu cluster dan memaksimalkan variasi antar *cluster* (Heryanto, et al. 2013).

Pada dasarnya *clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *Data Mining* yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) serta tidak memerlukan *target output*. Dalam *Data Mining* ada dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *Hierarchical Clustering* dan *Non Hierarchical Clustering*(Santosa, 2007).

*Cluster* memegang peran penting dalam pengklasifikasian *obyek*. Bergantung pada aplikasinya, obyek biasa berupa sinyal, pelanggan, pasien, berita, tanaman, dan lain-lain. Teknik *clustering* adalah teknik *non parametric* yang sangat banyak diaplikasikan dalam kasus nyata. Teknik *cluster* dikelompokkan ke dalam dua kelas besar :*Partitioning Cluster* dan *Hierarcichal Cluster*. Ada dua macam teknik cluster yang cukup sering dipakai. Yang pertama adalah *K-Means*(mewakili *Partitioning Cluster* atau *Non Hierarcichal*) dan yang berikutnya adalah *hierarcichal Clustering*(Santosa, 2007).

Tujuan utama dari metoda klaster adalah pengelompokan sejumlah data/obyek kedalam *cluster* (*group*) sehingga setiap klaster akan

berisi data yang semirip mungkin. Ini berarti obyek dalam satu kluster sangat mirip satu sama lain dan berbeda dengan obyek dalam kluster-kluster yang lain (Santosa, 2007)



Gambar 2.1. Contoh Proses *Clustering*, Sumber : (Nugraheni, 2011)

Konsep dasar dari algoritma K-Means adalah pencarian pusat *cluster* (*centroid points*) secara *iterative*. Pusat *cluster* ditetapkan berdasarkan jarak setiap data ke pusat *cluster*. Proses *clustering* dimulai dengan mengidentifikasi data yang akan di-*cluster*,  $X_{ij}$  ( $i=1, \dots, j=1, \dots, m$ ) dengan  $n$  adalah jumlah data yang akan di-*cluster* dan  $m$  adalah jumlah *variable*.

Pada awal iterasi, pusat setiap *cluster* ditetapkan secara bebas (sembarang),  $c_{kj}$  ( $k=1, \dots, k, j=1, \dots, m$ ). Kemudian dihitung jarak antara setiap data dengan setiap pusat *cluster*. Untuk melakukan perhitungan jarak data ke-1 pada pusat *cluster* ke- $k$  ( $c_k$ ), diberi nama ( $d_{ik}$ ) dapat digunakan *formula Euclidean*, yaitu:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2}$$

Suatu data akan menjadi anggota dari *cluster* ke- $j$  apabila jarak data tersebut ke pusat *cluster* ke- $j$  bernilai paling kecil jika dibandingkan dengan jarak ke pusat *cluster* lainnya. Selanjutnya, kelompokkan data-data yang menjadi anggota pada setiap *cluster*. Nilai pusat *cluster* yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada *cluster* tersebut, dengan rumus:

$$c_{kj} = \sum_{h=1}^P; y_{hj} = x_{ij \in \text{Cluster ke-}k}$$

Menurut Santosa (2007), langkah-langkah melakukan *Clustering* dengan metode K-Means adalah sebagai berikut:

- a. Pilih jumlah *cluster*  $k$ .
- b. Inisialisasi  $k$  pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara *random*. Pusat-pusat *cluster* diberi nilai awal dengan angka-angka *random*.

- c. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke cluster tertentu ditentukan jarak antara data dengan pusat cluster. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat cluster. Jarak paling antara satu data dengan satu cluster tertentu akan menentukan suatu data masuk dalam Cluster mana. Untuk menghitung jarak semua data ke setiap titik pusat Cluster dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D(i, j) = \sqrt{(X1i - X1j)^2 + (X2i - X2j)^2 + \dots + \dots + \dots}$$

dimana:

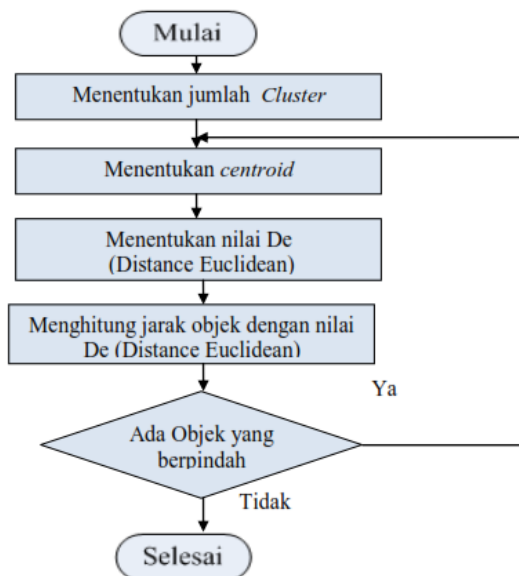
$D(i, j)$  = Jarak data ke I ke pusat cluster j

$X_{ki}$  = Data ke i pada atribut data ke k

$X_{kj}$  = Data ke j pada atribut data ke k

- d. Hitung kembali pusat *Cluster* dengan keanggotaan *Cluster* yang sekarang. Pusat *Cluster* adalah rata-rata dari semua data/objek dalam *Cluster* tertentu. Jika dikehendaki bisa juga menggunakan *median* dari *Cluster* tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang bisa dipakai.
- e. Tugaskan lagi setiap objek memakai pusat *Cluster* yang baru. Jika pusat *Cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

Algoritma K-Means adalah algoritma yang terbaik dalam algoritma *Partitional Clustering* dan yang paling sering digunakan diantara algoritma *Clustering* lainnya karena kesederhanaan dan efesinsinya (Budiman, 2012).



**Gambar 2.1.** Flowchart Algoritma Metode K-Means Clustering, Sumber : (Nugraheni, 2011)

### 3. METODE PENELITIAN

#### 3.1. Lokasi Penelitian dan Data Penelitian

Penelitian dilakukan pada Kabupaten Bondowoso di propinsi Jawa Timur. Lamanya waktu yang dibutuhkan untuk menyelesaikan penelitian ini yaitu 3 bulan yang dimulai pada bulan Mei sampai bulan Juli 2016.

Data yang digunakan dalam penelitian ini adalah data kecamatan yang bersumber dari data Kabupaten Bondowoso tahun 2015 di provinsi Jawa Timur. Setelah pengumpulan data lalu dilakukan analisa data sesuai dengan kebutuhan sistem. Analisis data dilakukan menggunakan algoritma k-means. Total data yang digunakan 23 data kecamatan.

Untuk mendapatkan input yang baik dari *data mining*, dilakukan *preprocessing* terhadap data yang akan digunakan. *Preprocessing* data merupakan tahap prapemrosesan. Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses *cleaning* (pembersihan) pada data yang digunakan. Dalam kasus ini, data yang diambil sebanyak 23 kecamatan pada tahun 2014 dari BPS Bondowoso. Atribut yang digunakan pada penelitian ini berupa Jumlah Sekolah, Jumlah Guru dan Jumlah Siswa.

**Tabel 3.1.** Dataset yang digunakan

KECAMATAN N	Jml Sekolah	Jml Guru	Jumlah Siswa
Maesan	114	746	1266
Grujugan	119	792	1227
Tamanan	103	711	1931
Jambesari	132	783	1506

Pujer	131	780	1407
Tlogosari	161	1067	1227
Sukosari	65	489	1143
Sumber wringin	79	541	1011
Tapen	99	688	1783
Wonosari	122	905	993
Tenggarang	130	887	2599
Bondowoso	138	1605	6656
Curahdami	97	659	1013
Binakal	56	299	156
Pakem	60	337	627
Wringin	102	740	1464
Tegalampel	56	401	687
Taman Krocok	48	270	201
Klabang	61	376	461
Botolinggo	118	749	544
Sempol	30	155	345
Prajeakan	82	589	1622
Cermee	162	982	731

### 4. IMPLEMENTASI SISTEM

#### 4.1. Implementasi

Pada tahap ini, algoritma *k-means clustering* akan diimplementasikan ke dalam sistem dengan menggunakan bahasa pemrograman *PHP* dan *database MySQL* sesuai perancangan yang telah dilakukan.

Adapun implementasi perancangan antarmuka yang telah dilakukan sebelumnya pada sistem adalah:

##### 1. Halaman Home

Halaman *home* merupakan halaman tampilan awal saat sistem dijalankan. Pada halaman ini ditampilkan selamat datang dan definisi algoritma *K-Means Clustering*. Desain tampilan menu halaman home seperti gambar 4.1 dibawah ini.



**Gambar 4.1.** Halaman menu Home

##### 2. Halaman Data Kecamatan

Tampilan data kecamatan berfungsi untuk pendataan kecamatan yang terdiri dari form

pencarian, tombol tambah, rubah dan hapus. Tampilan Data Kecamatan dapat dilihat pada Gambar 4.2.

#	Kode Kecamatan	Nama	Jumlah Sekolah	Jumlah Guru	Jumlah Siswa	Aksi
1	K001	Esukal	10	100	1.234	Tambah Hapus
2	K002	Bondowoso	12	90	78	Tambah Hapus
3	K003	Bekidregan	14	90	1.000	Tambah Hapus
4	K004	Carmes	15	70	700	Tambah Hapus
5	K005	Curandaru	12	10	98	Tambah Hapus

Gambar 4.2. Halaman menu Data Kecamatan

Pada gambar 4.2 jika user memilih tombol tambah kecamatan maka halaman akan membuka form kecamatan seperti gambar 4.3.

Gambar 4.3. Halaman Form Input Kecamatan

Pada halaman form input kecamatan user diminta mengisi data kecamatan yang terdiri dari beberapa atribut yaitu “Kode Kecamatan, Nama Kecamatan, Jumlah Sekolah, Jumlah Guru dan Jumlah Siswa” setelah mengisi data kecamatan user akan memilih tombol simpan untuk menyimpan data ke database data kecamatan dan jika user memilih tombol batal maka halaman akan kembali ke halaman menu data kecamatan.

- Halaman Pengelompokan Kecamatan  
Halaman menu pengelompokan user diminta mengisikan jumlah *cluster* yang akan digunakan untuk mengelompokan data kecamatan, tampilan halaman menu pengelompokan kecamatan dapat dilihat pada gambar 4.4.

Gambar 4.4. Halaman menu Pengelompokan Kecamatan

Pada halaman pengelompokan kecamatan setelah user menginputkan jumlah *cluster* maka sistem akan membuat pusat *cluster*

(*centroid*) yang didapatkan dari data kecamatan secara acak atau *random*. Desain tampilan halaman *cluster* kecamatan dapat dilihat pada gambar 4.5.

#	Kode Kecamatan	Nama	Jumlah Sekolah	Jumlah Guru	Jumlah Siswa
1	K001	Esukal	10	100	1.234
2	K002	Bondowoso	12	90	78
3	K003	Bekidregan	14	90	1.000
4	K004	Carmes	15	70	700
5	K005	Curandaru	12	10	98

#	Cluster	Jumlah Sekolah
1	Cluster 1	8
2	Cluster 2	9
3	Cluster 3	10

Gambar 4.5. Halaman Proses K-Means Clustering

## 5. KESIMPULAN DAN SARAN

### 5.1. Kesimpulan

Dari hasil penelitian yang telah penulis lakukan atas Implementasi Algoritma *K-Means Clustering* Dalam Pengelompokan Kecamatan di Kabupaten Bondowoso maka penulis dapat mengambil kesimpulan sebagai berikut:

- Hasil penentuan kelompok kecamatan dengan menggunakan Algoritma *K-Means Clustering* berhasil dilakukan.
- Proses *clustering* dapat menggunakan beberapa pusat *cluster (centroid)* yang diperoleh secara acak.
- Pengelompokan kecamatan dengan algoritma k-means kedalam bahasa pemrograman PHP dan database MySQL berhasil dibangun.

### 5.2. Saran

Sistem ini dibangun berdasarkan alur pemikiran penulis, maka untuk hasil yang lebih baik dan maksimal diperlukan saran dari semua pihak untuk melengkapi kekurangan yang ada. Saran dari penulis yaitu:

- Untuk mendapatkan hasil yang lebih variatif penelitian ini juga dapat dikembangkan dengan teknik data mining lainnya seperti algoritma K-Nearest Neighbor, algoritma Fuzzy Clustering, atau algoritma lainnya.
- Sistem ini dapat dikembangkan dengan menambah data dan parameter dari potensi kecamatan untuk penentu kelompok kecamatan yang lebih akurat.
- Diharapkan peneliti selanjutnya mampu mempersingkat waktu proses *clustering* sehingga tidak membutuhkan waktu yang lama dalam penghitungan iterasi *centroid* dan jarak *clusternya*.

## DAFTAR PUSTAKA

- Agusta, Y. 2007. K-Means-Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika* (3) : 47-60.
- Alfina, T., Santosa, B. & Barakbah, A. R. 2012. Analisa Perbandingan Metode *Hierarchical Clustering*, K-Means dan Gabungan Keduanya dalam Membentuk *Cluster* Data (Studi Kasus: Problem Kerja Praktek Jurusan Teknik Industri) (1) : 1-5.
- Budiman, I. 2012. Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma. Tesis. Universitas Diponegoro. 2013.
- Hemansyah, A. 2011. Strategi Pengembangan Donatur Lembaga Amil Zakat dengan Pendekatan Data Mining.
- Heryanto, D. N., Rachman, F. H. & Satoto, B.D. 2013. Penerapan Radial Basis *Function Network* Dengan *K-Means Cluster* Untuk Peramalan Kebutuhan Straw1: 1-10.
- Mahrus, M. A & Khotimah, B. K. 2012. Pengelompokan Tingkat Kesehatan Masyarakat Menggunakan K-Means Berbasis OWA (Ordered Weighted Averaging) Vol .1: 47-56.
- Nugraheni, Y. 2011. Data Mining dengan Metode Fuzzy Untuk Customer Relationship Management (CRM) pada Perusahaan Retail. Universitas Udayana
- Paulanda, Z. 2012. Model Profil Mahasiswa Yang Potensial Drop Out Menggunakan Teknik Kernel-K-Mean Clustering Dan Decision Tree. Tesis. Universitas Sumatera Utara.
- Rismawan, T & Kusumadewi, S. 2008. Aplikasi K-Means Untuk Pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) & Ukuran Kerangka:43-48.
- Santoso, S. 2010. Statistik Multivariat. Jakarta: Elex Media Komputindo.
- Santoso, Budi. 2007. Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Graha Ilmu, Yogyakarta.
- Utami, D. D. P & Sutikno. 2010. Pengelompokan Zona Musim (ZOM) dengan Fuzzy K-Means Clustering.