

## KLUSTERING JUMLAH PENDUDUK KOTA BANDUNG BERDASARKAN JENIS KELAMIN PER KECAMATAN PADA TAHUN 2012 DENGAN METODE K-MEANS

Sarah Ashari, Salsabilah Khansa, Calvin Habib Maulana Surudin, Ika Nurlaili  
Isnainiyah

Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

email : [azharsarah250@gmail.com](mailto:azharsarah250@gmail.com), [salsabilahkhansa@ymail.com](mailto:salsabilahkhansa@ymail.com),

[kelvinhabibmaulana@gmail.com](mailto:kelvinhabibmaulana@gmail.com), [nurlailika@upnvj.ac.id](mailto:nurlailika@upnvj.ac.id)

Jl. Rs. Fatmawati, Pondok Labu, Jakarta Selatan, DKI Jakarta, 12450, Indonesia

### Abstrak

Upaya pengendalian permasalahan jumlah penduduk dari setiap wilayah kota menjadi salah satu tugas pemerintah di Indonesia. Setiap tahunnya banyak pendatang yang berasal dari berbagai penjuru daerah akan berdatangan ke suatu kota yang dipercaya memiliki tingkat pendapatan ekonomi yang lebih baik, seperti Kota Bandung. Penelitian ini bertujuan untuk melakukan pengolahan jumlah data penduduk Kota Bandung di tahun 2012 yang dipisahkan berdasarkan jenis kelamin yaitu laki-laki dan perempuan. Pengolahan dilakukan dengan menerapkan proses klustering dengan metode K-Means sesuai jumlah penduduk di Kota Bandung tersebut dengan maksud untuk mendapatkan klaster atas padat atau tidaknya jumlah penduduk di Kota Bandung per kecamatan.

Kata Kunci: Klustering, Penduduk, K-Means

### 1. PENDAHULUAN

Peningkatan jumlah penduduk di suatu kota ataupun kabupaten apabila tidak diselesaikan dengan baik dan sistematis akan mengakibatkan dampak-dampak yang buruk seperti terjadinya pencemaran lingkungan, berkurangnya lahan terbuka hijau dikarenakan untuk pembangunan pemukiman penduduk, tingkat pencemaran yang semakin meningkat, dan dampak-dampak lainnya yang akan ditimbulkan dari peningkatan jumlah penduduk. Pertumbuhan ekonomi sebagai salah satu dari proses pembangunan ekonomi menjadi penting, karena pembangunan ekonomi tidak terlepas dari pertumbuhan ekonomi (*economic growth*), pembangunan ekonomi mendorong pertumbuhan ekonomi dan sebaliknya, pertumbuhan ekonomi itu sendiri dapat memperlancar proses pembangunan (HM, 2013).

Jumlah penduduk di dalam data ini diklasifikasikan berdasarkan jenis kelamin yaitu laki-laki dan perempuan per kecamatan di Kota Bandung pada tahun 2012. Data ini membantu untuk melakukan proses data dan pengolahannya dengan cara klustering berdasarkan pengelompokkan tingkat kepadatan atau tidaknya suatu penduduk per kecamatan di Kota Bandung. Penggalan informasi pada sebuah data yang berukuran besar (mempunyai jumlah *record* dan jumlah *field* yang cukup banyak) tidak dapat dilakukan dengan mudah. Teknologi data mining merupakan salah satu alat bantu untuk penggalan data pada basis data berukuran besar dan dengan spesifikasi tingkat kerumitan yang telah banyak digunakan pada banyak domain aplikasi seperti perbankan maupun bidang telekomunikasi

(Arief, 2010). Metode K-Means adalah salah satu metode cluster non hirarki yang bertujuan untuk membantu pengelompokan variabel untuk dimasukkan kedalam kelas-kelas yang terletak pada hasil akhir perhitungan.

K-Means merupakan salah satu metode *data clustering* non-hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik sama dikelompokkan ke dalam satu cluster yang sama (Agusta, 2007). Tujuan penelitian *cluster* memakai metode K-Means untuk membantu menentukan *centroid* 1(C1) dan *centroid* 2(C2) yang telah ditentukan letaknya secara acak, dalam mencari kepadatan jumlah penduduk pada suatu Kecamatan di Kota Bandung apakah banyak penduduk tersebut berpengaruh terhadap kepadatan atau tidaknya di suatu kecamatan tersebut berdasarkan jenis kelamin yang telah diklasifikasikan.

## 2. METODOLOGI PENELITIAN

### 2.1 Data Mining

Data mining merupakan proses pengekstrakan informasi dari jumlah kumpulan data yang besar dengan menggunakan algoritma dan tehnik gambar dari statistik, mesin pembelajaran dan sistem manajemen database (Han & Kamber, 2001). Data mining yang disebut juga dengan *Knowledge-Discovery in Database* (KDD) adalah sebuah proses secara otomatis atas pencarian data di dalam sebuah memori yang amat besar dari data untuk mengetahui pola dengan menggunakan alat seperti klasifikasi, hubungan (*association*) atau pengelompokan (*clustering*).

### 2.2 Clustering

Menurut Tan, 2006 *clustering* adalah sebuah proses untuk mengelompokan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum. Analisis *cluster* (Clustering) merupakan usaha untuk mengidentifikasi kelompok obyek yang mirip-mirip dan membantu menemukan pola penyebaran dan pola hubungan dalam sekumpulan data yang besar (Narwati, 2012).

### 2.3 K-Means

K-Means clustering merupakan salah satu metode pengelompokan data non-hirarki atau bisa juga disebut metode partisi. Secara umum, cara kerja metode ini adalah mengelompokkan data berdasarkan kesamaan karakteristik yang dimiliki, sehingga suatu *cluster* hanya memiliki tingkat variasi yang rendah. (Ong, 2013) (Merliana, 2010)

Algoritma dasar K-Means adalah sebagai berikut, (Ong, 2013) (Puspitasari dan Haviluddin, 2016)

1. Tentukan banyak *cluster*  $k$ .
2. Beri nilai awal pusat cluster sebanyak  $k$ , biasanya secara random dari data yang tersedia.
3. Masukkan setiap data ke dalam *cluster* yang memiliki *centroid* terdekat. Gunakan *Euclidean distance* untuk menghitung jarak dari setiap data ke setiap *centroid* dari *cluster* dengan rumus sebagai berikut

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

Keterangan :

$d_{ik}$  = jarak antara data ke- $i$  ke *centroid* cluster ke- $k$

$m$  = jumlah atribut

$x_j$  = data ke-j

$c_k$  = centroid cluster ke-k

- Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan sebagai berikut,

$$v = \sum_{i=1}^n \frac{x_i}{n} \quad (2)$$

Keterangan :

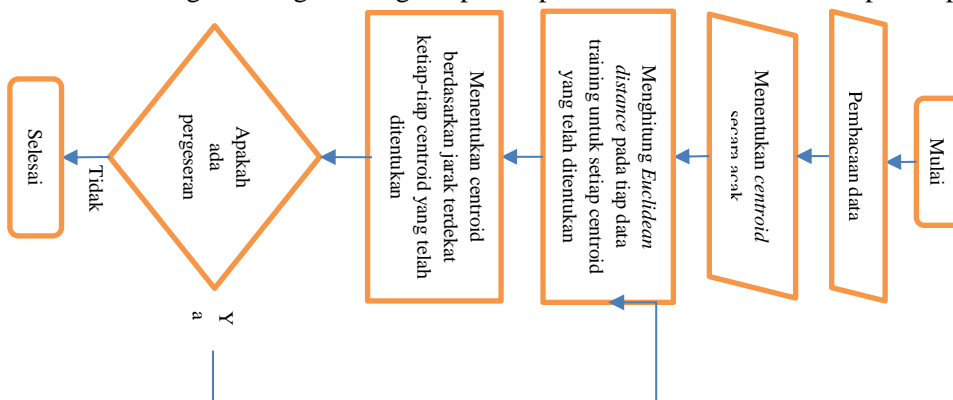
$v$  = centroid dari cluster

$x_i$  = data ke-i ;  $i = 1,2,3\dots n$

$n$  = banyaknya jumlah data yang menjadi anggota cluster

- Kembali ke langkah 3, jika masih ada data yang berpindah cluster atau ada perubahan nilai centroid.

Flowchart mengenai langkah-langkah penerapan metode K-Means ditampilkan pada Gambar 1.



Gambar 2 Flowchart K-Means Clustering

#### 2.4 Datasets

Data yang diproses untuk Klustering adalah data yang berisi jumlah penduduk berdasarkan jenis kelamin per kecamatan di Kota Bandung pada tahun 2012. Data ini diperoleh dari situs dataset portal Kota Bandung yang dikumpulkan pada tahun 2012 dapat dilihat pada Tabel 1 berikut.

Table 1: Data Kependudukan

No	Kecamatan	Jumlah Penduduk		No	Kecamatan	Jumlah Penduduk	
		Laki-laki	Perempuan			Laki-laki	Perempuan
1	SUKASARI	42347	40851	16	KIARACOND ONG	66470	64187
2	COBLONG	66896	63374	17	BOJONGLOA KIDUL	44798	41793
3	BABAKAN	68261	63902	18	CIBEUNYING KALER	38426	36822
4	CIPARAY	64745	60822	19	SUMUR BANDUNG	22216	20829
5	BOJONGLOA	61845	58498	20	ANTAPANI BANDUNG	39977	38413
6	KALER	55515	53319	21	KIDUL	30606	29682
7	ANDIR	55635	53879	22	BUAH BATU	56735	54430
8	CICENDO	29440	27818	23	RANCASARI	43959	42403
9	SUKAJADI	20156	19182	24	ARCAMANIK	38104	36389
10	CIDADAP BANDUNG	42482	41218	25	CIBIRU	42587	39432

11	REGOL BATUNUNGGAL	52170	49889	26	UJUNGBERU NG	44006	41826
12	L	64579	62342	27	GEDEBAGE PANYILEUK	20708	19998
13	LENGKONG CIBEUNYING	43758	42201	28	AN	20307	19422
14	KIDUL BANDUNG	60976	58057	29	CINAMBO MANDALAJA	13470	12763
15	KULON	71473	68543	30	TI	35976	34253

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Analisis Data

Pada data sampel yang digunakan dan dilakukan pemrosesan K-Means ini memiliki jumlah populasi yang dihitung berdasarkan keseluruhan kecamatan ialah sebanyak tiga puluh data laki-laki dan perempuan. Data yang akan diproses seperti yang telah di tampilkan pada Tabel 1.

Selanjutnya dilakukan proses perhitungan K-Means dengan langkah-langkah seperti di bawah ini:

- a. Menentukan letak centroid secara acak
- b. Menghitung jarak centroid dengan data memakai Rumus *Euclidian Distance*
- c. Kelompokkan dataset sehingga membentuk suatu *cluster*, pilih berdasarkan jumlah data dan tentukan titik *centroid* dengan memilih jumlah perhitungan yang paling kecil dari data *centroid* tersebut
- d. Perbaharui nilai perhitungan titik *centroid*
- e. Lakukan pengulangan perhitungan pada langkah b dan c sampai *centroid* tidak ada yang bergeser(stabil) dan tidak terdapat selisih antar *cluster* lama dan *cluster* baru
- f. Selesai

#### 3.2 Iterasi

Menentukan centroid pertama dan centroid kedua, pada perhitungan pertama adalah menentukan kluster pertama dan Kluster kedua yaitu pada  $C_1$  yaitu (42347;40851) dan  $C_2$  yaitu (66896;63374)

1. Melakukan perhitungan jarak ke centroid dengan rumus *Euclidean Distance*.

Perhitungan jarak laki-laki pertama dengan pusat cluster I :

$$C_{11} = \sqrt{(42347 - 42437)^2 + (40851 - 40851)^2} = 0$$

Perhitungan jarak laki-laki kedua dengan pusat cluster I :

$$C_{12} = \sqrt{(66896 - 42437)^2 + (63374 - 40851)^2} = 33315,7$$

Perhitungan jarak perempuan pertama dengan pusat cluster I :

$$C_{11} = \sqrt{(42347 - 66896)^2 + (40851 - 63374)^2} = 33315,7$$

Perhitungan jarak perempuan kedua dengan pusat cluster I :

$$C_{12} = \sqrt{(66896 - 66896)^2 + (63374 - 63374)^2} = 0$$

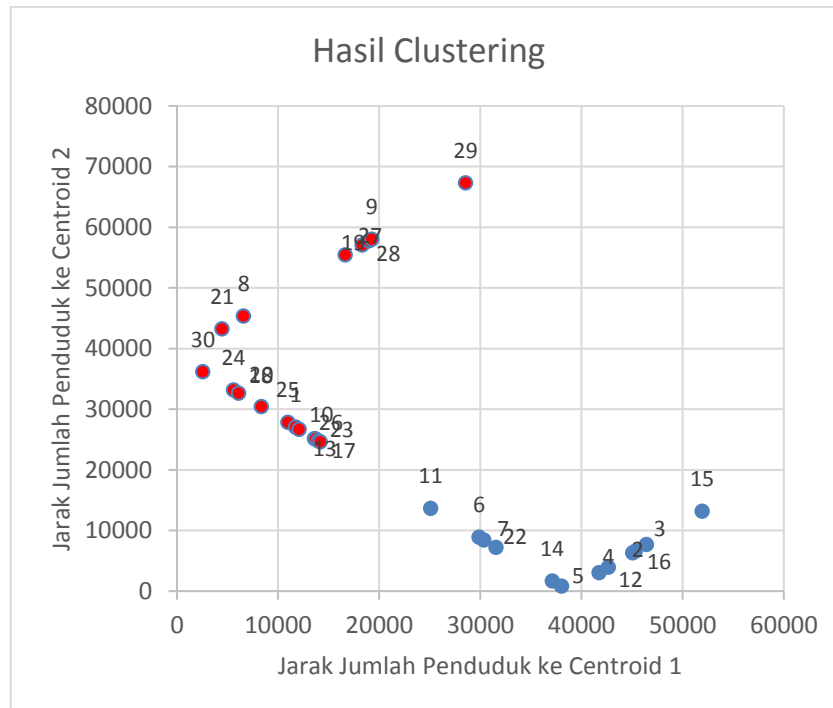
Menghasilkan hasil perhitungan seperti pada Tabel 3.

Pada penelitian ini terjadi pemrosesan iterasi sebanyak tiga kali. Maka dari itu, dilakukan lah iterasi kembali dan berhenti pada proses iterasi ketiga. Pengehentian proses dikarenakan tidak adanya pergeseran centroid sebelumnya yaitu iterasi kedua dengan centroid baru yaitu iterasi ketiga.

Table 2: Iterasi ke-3

No	Kecamatan	Jarak dengan C1	Jarak dengan C2	Kelas	No	Kecamatan	Jarak dengan C1	Jarak dengan C2	Kelas
1	SUKASARI	11743,8	27014,4	C1	16	KIARACONDO	45305,2	6572,6	C2
2	COBLONG	45050,1	6305,8	C2	17	BOJONGLOA	14179,9	24598,8	C1
3	BABAKAN	46409,5	7701,2	C2	18	KIDUL	6122,3	9	C1
4	CIPARAY	41736,7	3059,4	C2	19	CIBEUNYING	16649,2	55399,7	C1
5	BOJONGLOA	38030,3	815,8	C2	20	BANDUNG	8343,9	30410,8	C1
6	KALER	29874,5	8881,9	C2	21	ANTAPANI	4478,5	43218,7	C1
7	ANDIR	30352,3	8424,3	C2	22	BANDUNG	31523,8	7231,9	C2
8	CICENDO	3	45348,2	C1	23	BUAH BATU	13981,0	24777,0	C1
9	SUKAJADI	6598,8	58026,2	C1	24	RANCASARI	0	33162,6	C1
10	CIDADAP	19274,4	26666,4	C1	25	ARCAMANIK	5589,5	27832,10968,	C1
11	BANDUNG	12100,5	13666,6	C1	26	CIBIRU	4	3	C1
12	WETAN	25085,7	6	C2	27	UJUNGBERUN	13613,4	25139,5	C1
13	ASTANA	42663,2	3942,1	C2	28	G	18312,5	57063,9	C1
14	ANYAR	13696,0	25061,7	C1	29	GEDEBAGE	18999,5	57751,4	C1
15	REGOL	37095,4	1659,5	C2	30	PANYILEUKAN	28543,0	67294,9	C1
16	BATUNUNGGAL	51929,2	13178,9	C2		CINAMBO	0	36176,8	C1
17	CIBEUNYING	2	9	C2		MANDALAJATI	2575,9	8	C1

Dalam penelitian ini, proses iterasi berhenti di iterasi ke-3 yaitu tabel ke-4 dan diperoleh C1 dan C2 yang terdiri dari kecamatan-kecamatan di Kota Bandung yang terklasifikasi berdasarkan jenis kelamin, yaitu C1 sebanyak delapan belas data dan C2 sebanyak dua belas data. Data yang diperoleh dari C1, yaitu Kecamatan Sukasari, Kecamatan Cidadap, Kecamatan Bandung Wetan, Kecamatan Astana Anyar, Kecamatan Lengkong, Kecamatan Bojongloa Kidul, Kecamatan Cibeunying Kaler, Kecamatan Sumur Bandung, Kecamatan Antapati, dan Kecamatan Bandung Kidul. Pada data C2 diperoleh kecamatan yang terdiri atas Kecamatan Coblong, Kecamatan Babakan Ciparay, Kecamatan Bojongloa Kaler, Kecamatan Andir, Kecamatan Cicendo, Kecamatan Sukajadi, Kecamatan Regol, Kecamatan Batununggal, Kecamatan Cibeunying Kidul, Kecamatan Bandung Kulon, Kecamatan Kiaracondong, dan Kecamatan Buah Batu.



Gambar 3 Chart Hasil Akhir Clustering

Hasil akhir penelitian *cluster* yang memakai metode K-Means untuk membantu menentukan *centroid* 1(C1) dan *centroid* 2(C2) yang telah ditentukan letaknya secara acak, bertujuan untuk mencari kepadatan jumlah penduduk pada suatu Kecamatan di Kota Bandung apakah banyak penduduk tersebut berpengaruh terhadap kepadatan atau tidaknya di suatu kecamatan tersebut berdasarkan jenis kelamin yang telah diklasifikasikan berdasarkan jenis kelamin, dan terlihat hasil yang diperoleh bahwa C1 lah yang memiliki jumlah kepadatan penduduk berdasarkan klasifikasi jenis kelamin per kecamatan lebih padat dibandingkan perolehan hasil C2 yang tidak mengalami kepadatan penduduk di kecamatan.

#### 4. Kesimpulan

Metode *K-Means* telah diterapkan untuk melakukan pengolahan data kependudukan kecamatan kota Bandung yang kami ambil dari data.bandung.go.id. Pengolahan data ini menghasilkan hasil *clustering* sebanyak 2 klaster yang dapat digolongkan menjadi pengelompokan padat penduduk dan tidak padat penduduk. Dalam mengolah data, penelitian ini menerapkan perhitungan menggunakan *euclidean distance*. Terdapat tiga kali iterasi hingga *centroid* tidak berubah dan menghasilkan *clustering final*. Sebagai perbaikan penelitian selanjutnya di masa mendatang, penulis menyarankan untuk menerapkan metode *K-Means* pada jumlah datasets yang lebih banyak dan jumlah klaster yang dipakai lebih dari dua agar setiap klaster memiliki anggota yang memiliki kemiripan lebih spesifik. Pada penulisan ini, pengolahan data hanya dilakukan dengan metode *K-Means* dan tidak dilakukan perbandingan jika menggunakan metode *clustering* lain. Besar kemungkinan jika menggunakan metode *clustering* lain akan menghasilkan hasil akhir yang lebih baik.

#### Referensi

Agusta, Yudi. Pebruari (2007). “*K-Means Penerapan, Permasalahan dan Metode Terkait*”. Jurnal Sistem dan Informatika Vol.3 : 47-60.

- Arief Jananto,,(2010). “*Memprediksi kinerja mahasiswa menggunakan teknik data mining (studi kasus data akademik mahasiswa unisbank)*”, Tesis Tidak Terpublikasi, Yogyakarta: Universitas Gajah Mada.
- H. Syamsuddin. HM.,(2013)“*Analisis Pengaruh Faktor Kependudukan Terhadap Pertumbuhan Ekonomi Provinsi Jambi*”, Tesis Tidak Terpublikasi,Jambi : Universitas Jambi.
- Jiawei Han and Micheline Kamber,(2001), *Data mining concepts and techniques second edition*, San Francisco: Morgan Kauffman,.
- Narwati,(2010),“*Pengelompokan Mahasiswa Menggunakan Algoritma K-Means.Jurnal.Fakultas Teknologi Informasi*”.
- John Oscar Ong, (2013), “*Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University*”
- Novianti Puspitasari dan Haviluddin, (2016), “*Penerapan Metode K-Means Dalam Pengelompokan Curah Hujan di Kalimantan Ti*