

**ALGORITMA K-MEANS DENGAN METODE ELBOW UNTUK MENGELOMPOKKAN
KABUPATEN/KOTA DI JAWA TENGAH BERDASARKAN KOMPONEN PEMBENTUK
INDEKS PEMBANGUNAN MANUSIA**

**K-MEANS ALGORITHM WITH ELBOW METHOD TO GROUPING DISTRICT/CITY IN
CENTRAL JAVA BASED ON COMPONENTS OF HUMAN DEVELOPMENT INDEX**

Rina Yuliana Sari¹, Hardian Oktavianto^{2*}, Henny Wahyu Sulisty³

¹Mahasiswa Fakultas Teknik, Universitas Muhammadiyah Jember
email: Rinayulianasari907@gmail.com

²Dosen Fakultas Teknik, Universitas Muhammadiyah Jember *Koresponden Author
email: hardianoktavianto@unmuhjember.ac.id

³Dosen Fakultas Teknik, Universitas Muhammadiyah Jember
email: hennywahyusulistyo@unmuhjember.ac.id

Abstrak

IPM (Indeks Pembangunan Manusia) merupakan pengukur perbandingan keberhasilan pembangunan manusia yang didasarkan pada indikator kesehatan, pendidikan, dan hidup layak. Pada tahun 2010-2018, pencapaian angka IPM provinsi Jawa Tengah mengalami kenaikan sebesar 5.04 poin dari 66.08 menjadi 71.12. Jawa Tengah di sisi lain, saat ini berada di peringkat ke-13 dari 34 provinsi di Indonesia. Oleh karena itu, diperlukan suatu metode untuk mengelompokkan kabupaten/kota di Jawa Tengah berdasarkan komponen Pembentuk IPM menggunakan metode *clustering* yaitu algoritma *K-Means*. Metode *Elbow* digunakan untuk mencari ukuran *cluster* yang optimal untuk menentukan ukuran *cluster* terbaik. Data yang digunakan yaitu data Komponen Pembentuk IPM di Jawa Tengah tahun 2018. SSE (*Sum of Squares Error*) jarak pada metode *Elbow cluster* yang dihasilkan terbaik yang berada pada 2 *cluster* dari pengujian dimulai dari 2 *cluster* hingga 10 *cluster*. *Cluster* 1 memiliki 29 anggota kabupaten/kota, sedangkan *cluster* 2 hanya memiliki 6 anggota kabupaten/kota. Berdasarkan hasil karakteristik data empat komponen indeks pembangunan manusia tahun 2018 *cluster* 1 memiliki komponen pembentuk lebih rendah dibandingkan dengan *cluster* 2.

Keywords: *Clustering, Elbow, Indeks Pembangunan Manusia, K-Means.*

Abstract

The Human Development Index (HDI) is a comparison estimator accomplishment of human development based on Indicators of Health, Education, and Living Standard. In 2010-2018, the achievement of HDI in Central Java province increased by 5.04 points from 66.08 to 71.12. However, from these developments Central Java is still ranked 13th out of 34 provinces in Indonesia. Therefore, a method is needed to grouping the district / city in Central Java based on HDI components using grouping method that is K-Means algorithm. For measurement optimum cluster in determine as the best cluster, the method used is Elbow method. The data used is data of HDI form component in a Central Java in 2018. From the series of tests started from 2 clusters to 10 clusters, the best clusters resulted in 2 clusters based on range of SSE (Sum of Squares Error) in the Elbow method. In cluster 1 consist of 29 members of district/city and cluster 2 consist of 6 members of district/city. Based on the characteristic result fourth data component of The Human Development Index (HDI) in 2018, cluster 1 has a lower composition than cluster 2.

Keywords: *Students, Semester Achievement Index, Modified K-Nearest Neighbor Method*

1. PENDAHULUAN

IPM (Indeks Pembangunan Manusia) menjelaskan cara penduduk mengakses hasil pembangunan untuk memperoleh pendapatan, kesehatan, pendidikan, dan hidup layak. Guna

menunjang sasaran program pemerintah penting untuk dilakukan bahan perencanaan dan evaluasi, maka wilayah kabupaten/kota di Jawa Tengah perlu dikelompokkan berdasarkan karakteristik 4 komponen Indeks Pembangunan Manusia agar meningkatkan angka

pembangunan manusia berdasarkan komponen Indeks Pembangunan Manusia (IPM). (Ada beberapa Algoritma pengelompokan yang bisa digunakan, salah satu diantaranya adalah *k-means* dengan metode *elbow*).

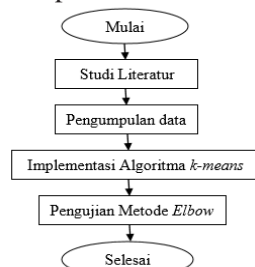
Algoritma *K-means* adalah algoritma untuk mengelompokkan data dengan meningkatkan data dalam satu *cluster*. Penelitian ini membantu pemerintah mengetahui permasalahan dan mempertimbangkan pengambilan kebijakan pada wilayah kabupaten/kota di provinsi Jawa Tengah berdasarkan variabel-variabel IPM dengan memanfaatkan pengukuran dalam menentukan *cluster* terbaik. Angka Harapan Hidup, Rata-rata Lama pengelompokan wilayah kabupaten/kota.

Peneliti akan menggunakan metode *k-means*, yang merupakan algoritma efektif untuk menganalisis data dalam jumlah besar. Akibatnya, sementara penelitian ini akan mencari ukuran *cluster* terbaik, ada berbagai cara untuk melakukannya, yang paling umum adalah menggunakan metode *Elbow*. Metode *elbow* diimplementasikan dengan cara menentukan data optimal dan melihat grafik dari nilai *k* yang akan diinputkan.

2. METODOLOGI PENELITIAN

a. Rancangan Penelitian

Jenis penelitian yang dilakukan dalam penelitian ini adalah penelitian kuantitatif, yang analisisnya data-data *numerik* (angka), Pada metodologi penelitian terdapat langkah-langkah prosedur penelitian, prosedur tersebut dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

b. Studi Literatur

Studi literatur merupakan langkah dalam proses pengumpulan semua informasi

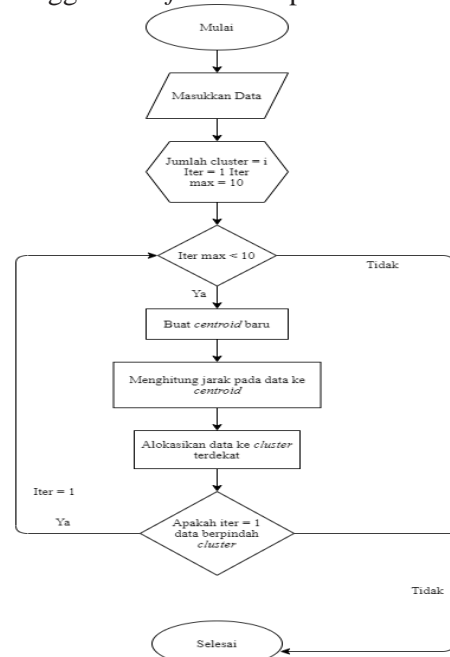
yang dibutuhkan untuk membangun sebuah sistem. Informasi ini akan dikumpulkan dari berbagai sumber. Membaca literatur yang ada di jurnal, artikel, buku, dan skripsi adalah salah satu caranya. Persamaan atau perbedaan terhadap penelitian yang dilakukan oleh penulis maupun peneliti lainnya akan diketahui oleh penulis maupun peneliti lainnya yang akan diketa oleh penulis maupun peneliti lainnya.

c. DataSet

Dataset untuk penelitian ini adalah data Indeks Pembangunan Manusia di provinsi Jawa Tengah pada tahun 2018. Terdiri dari 35 kabupaten/kota di Jawa Tengah. Dimana data tersebut akan di *cluster* menggunakan algoritma *k-means* dan menggunakan metode *elbow*.

d. Proses Clustering

Dataset untuk penelitian ini merupakan data komponen Indeks Pembangunan Manusia. Pengelompokan tersebut berdasarkan data tahun 2018. Pada pengelompokan ini *cluster* validasinya menggunakan teknik *elbow*, proses ini menggunakan 10 data dengan hasil terbaik terdapat pada pengelompokan 2 *cluster*. Berikut gambaran algoritma *k-means* menggunakan *flowchart* pada Gambar 2



Gambar 2. Flowchart Algoritma K-Means Clustering

- e. Proses *Elbow*
 Metode *Elbow* untuk menentukan jumlah *cluster* yang paling optimum atau yang terbaik. Langkah-langkah metode *elbow*:
1. awal nilai *cluster* (2 *cluster*)
 2. Menaikkan nilai *cluster* sampai jumlah *cluster* (4 *cluster*)
 3. Menghitung nilai SSE (*Sum of Squares Error*) dari setiap *cluster*
- $$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} \|x_i - c_k\|_2^2 \quad (1)$$
4. Melakukan perhitungan SSE (*Sum of Squares Error*) sampai *cluster* yang ditentukan.
 5. Melihat hasil SSE (*Sum of Squares Error*) dari nilai *cluster* yang turun secara drastis atau nilai yang berubah signifikan.
 6. Menetapkan nilai *cluster* yang berbentuk siku.

3. HASIL DAN PEMBAHASAN

- A. Data Pengujian
 Hasil yang diperoleh dari perhitungan yang sudah dilakukan ditunjukkan bab ini. Data yang sudah diolah akan di *cluster* menggunakan algoritma *k-means*, kemudian diolah untuk mendapatkan hasil *cluster* terbaik menggunakan metode *Elbow*. Data yang digunakan adalah Indeks Pembangunan Manusia Provinsi Jawa Tengah Tahun 2018 yang terdiri dari 35 kabupaten/kota.
- B. Praproses Data
 Data diolah menggunakan RStudio yang di *cluster* menggunakan algoritma *k-means* dari 2 *cluster* sampai dengan 10 *cluster*. *Output* yang dapat dihasilkan dari eksekusi perintah pada RStudio merupakan jumlah iterasi, pusat *cluster*, fungsi

objektif, dan derajat keanggotaan setiap objek terhadap tiap *cluster*.

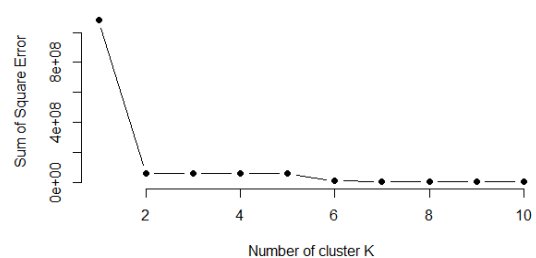
C. Penentuan *Cluster Optimum*

Setelah proses *clustering* menggunakan *k-means*, langkah selanjutnya adalah menggunakan metode *Elbow* untuk mencari *cluster* yang optimal untuk *clustering* terbaik. Hasil metode *Elbow* di RStudio ditampilkan sebagai SSE (*Sum of Squares Error*) dan grafik yang terdiri dari sumbu x dan sumbu y. Berikut ini adalah hasil metode *Elbow*.

Tabel 1. Hasil Nilai *Elbow*

C	SSE	Jarak	Keterangan
1	1080555394	-	-
2	57567926.1	1022987468	Jarak C1 ke C2
3	57563235.3	4690.8	Jarak C2 ke C3
4	57563163.1	72.2	Jarak C3 ke C4
5	57563144.6	18.5	Jarak C4 ke C5
6	9727886.1	47835258.5	Jarak C5 ke C6
7	4813836.9	4914049.2	Jarak C6 ke C7
8	879308.2	3934528.7	Jarak C7 ke C8
9	1595768.6	-716460.4	Jarak C8 ke C9
10	879217.5	716551.1	Jarak C9 ke C10

Sumber : Hasil Perhitungan



Gambar 3. Hasil metode *Elbow*

Sumber : Hasil Pengamatan

Nilai *cluster* yang diambil sebagai *cluster* optimal dalam metode *Elbow* adalah titik yang membentuk siku. Penjelasan pada titik yang membentuk siku adalah pada titik yang terjadi signifikan penurunan antara 2 titik *cluster* dan kemudian diikuti oleh nilai yang relatif konstan. Tabel di atas menunjukkan nilai SSE (*Sum of Squares Error*) untuk kesalahan

paling signifikan atau terbesar. 1022987468 yaitu pada 2 *cluster* dengan jarak 1 *cluster* ke 2 *cluster*. Nilai jarak 1 *cluster* ke 2 *cluster* ini merupakan nilai jarak yang mengalami penurunan paling signifikan atau paling besar, dengan diikuti oleh nilai jarak yang relatif konstan, sehingga 2 *cluster* merupakan *cluster* yang optimal atau terbaik. titik yang membentuk siku terdapat pada titik 2 *cluster*, seperti terlihat pada Gambar 1.

D. *Profiling Cluster* Optimum

Profiling data *cluster* jumlah Komponen Pembentuk Indeks Pembangunan Manusia di setiap kabupaten/kota yang masuk ke dalam kelompok *cluster* 1 dan *cluster* 2 berdasarkan perhitungan yang telah dilakukan sebelumnya untuk mengetahui karakteristik dari masing-masing *cluster*. pada *cluster* 1 terdiri dari 29 anggota kabupaten/kota dengan komponen pembentuk Indeks pembangunan Manusia yakni PP (Pengeluaran per-Kapita) terendah yaitu 9.02, PP (Pengeluaran per-Kapita) tertinggi yaitu 12830, AHH (Angka harapan Hidup) terendah yaitu 4.25, AHH (Angka harapan Hidup) tertinggi yaitu 76.67, RRLS (Rata-rata Lama Sekolah) terendah yaitu 6.19, RRLS (Rata-rata Lama Sekolah) tertinggi yaitu 8.62, HLS (Angka Harapan Lama Sekolah) terendah yaitu 11.42, , HLS (Angka Harapan Lama Sekolah) tertinggi yaitu 13.48. pada *cluster* 2 terdiri dari 6 anggota kabupaten/kota dengan komponen pembentuk Indeks pembangunan Manusia yakni PP (Pengeluaran per-Kapita) terendah yaitu 11.22, PP (Pengeluaran per-Kapita) tertinggi yaitu 14895, AHH (Angka harapan Hidup) terendah yaitu 76.72, AHH (Angka harapan Hidup) tertinggi yaitu 77.54, RRLS (Rata-rata Lama Sekolah) terendah yaitu 8.51, RRLS (Rata-rata Lama Sekolah) tertinggi

yaitu 10.53, HLS (Angka Harapan Lama Sekolah) terendah yaitu 13.66, HLS (Angka Harapan Lama Sekolah) tertinggi yaitu 15.5. Dari hasil karakteristik data komponen pembentuk Indeks Pembangunan Manusia, *cluster* 1 milik komponen pembentuk lebih rendah dibandingkan dengan *cluster* 2.

4. KESIMPULAN DAN SARAN

a. Kesimpulan

1. Algoritma *k-means* dikembangkan oleh komponen pembentuk Indeks Pembangunan Manusia nilai jarak SSE (Sum of Squares Error) 1 *cluster* ke 2 *cluster* pada Elbow, yaitu 1022987468. Nilai jarak 1 *cluster* ke 2 *cluster* tersebut merupakan nilai yang mengalami penurunan paling signifikan atau paling besar sehingga 2 *cluster* merupakan *cluster* terbaik.
2. Hasil pengelompokan pada *cluster* 1 terdapat 29 kabupaten/kota yaitu Cilacap, Banyumas, Purbalingga, Banjarnegara, Kebumen, Purworejo, Wonosobo, Magelang, Banyolali, Klaten, Wonogiri, Sragen, Grobogan, Blora, Rembang, Pati, Kudus, Jepara, Demak, Semarang, Temanggung, Kendal, Batang, Pekalongan, Pemalang, Tegal, Brebes, Kota Pekalongan, Kota Tegal. Pada *cluster* 2 terdapat 6 kabupaten/kota yaitu Sukoharjo, Karangayar, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang. hasil karakteristik data komponen pembentuk Indeks Pembangunan Manusia, *cluster* 1 memiliki komponen pembentuk lebih rendah disbanding dengan *cluster* 2.
3. Hasil karakteristik dari masing-masing *cluster*, dilakukan *proilling* *cluster* data jumlah Komponen Indeks Pembangunan Manusiadisetiap Kabupaten/kota yang masuk ke dalam kelompok *cluster* 1 dan *cluster* 2. Pada perhitungan karakteristik *cluster* 1

pengeluaran per-Kapita, Angka Harapan Hidup, Rata-rata Lama Sekolah dan Angka Harapan Lama Sekolah memiliki komponen pembentuk lebih rendah dibandingkan dengan *cluster* 2.

b. Saran

1. Validitas *cluster* digunakan mencari *cluster* terbaik, bisa menggunakan alternatif lain selain metode *Elbow*, seperti metode *Silhouette*, *Gap Statistic*, *Davis Bouldin index*, dll.
2. Perhitungan manual *cluster k-means* di Microsoft Excel di gunakan sebagai pembanding hasil *cluster* pada RStudio. Perhitungan manual di microsoft excel dapat menggunakan bilangan acak/random untuk menentukan *cluster* awal.

5. DAFTAR PUSTAKA

- Agusta, Y.2007. K-Means-Penerapan, Permasalahan dan Metode Terkait. *Jurnal Sistem dan Informatika*. Vol.3,pp:47-60. Denpasar, Bali
- Anonim. Badan Pusat Statistik. 2018. *Angka Harapan Hidup Penduduk Jawa Tengah menurut Kabupaten/Kota*. Semarang: Badan Pusat Statistika Jawa Tengah.
- Anonim.Badan Pusat Statistik. 2018. *Harapan Lama Sekolah Penduduk Jawa Tengah menurut Kabupaten/Kota*. Semarang: Badan Pusat Statistika Jawa Tengah.
- Anonim. Badan Pusat Statistik. 2018. *Rata-rata Lama Sekolah Penduduk Jawa Tengah menurut Kabupaten/Kota*. Semarang: Badan Pusat Statistika Jawa Tengah.
- Anonim.Badan Pusat Statistik. 2018. *Pengeluaran Perkapita yang Disesuaikan Penduduk Jawa Tengah menurut Kabupaten/Kota*. Semarang: Badan Pusat Statistika Jawa Tengah.
- Anonim. Badan Pusat Statistik Provinsi Jawa Tengah. 2018. *Jawa Tengah Dalam Angka 2018*. Semarang: Badan Pusat Statistik Jawa Tengah.
- Astuti, E.W. 2015. Clustering Program Keahlian Pada Pendaftaran Siswa Baru (PSB) Dengan Menggunakan Algoritma K-Means. *Jurnal SPIRIT*, Vol.7, hal. 58-65.
- Han, J., & Kamber, M. 2006. *Data Mining: Concept and Techniques, Second Edition*. Waltham: Morgan Kaufmann Publishers.
- Kasiram, M. 2008. Metode Penelitian Kuantitatif-Kualitatif. Malang: UIN Malang Press.
- Kusrini & Luthfi, Emha Taufiq. 2009. *Algoritma Data Mining*. Yogyakarta: Andi Offset
- Larose, Daniel, T. 2005. *Discovering Knowledge In Data: An Introduce to Data Mining*. Canada: John Willey and Sons, Inc
- Merliana, N. P. E., Ernawati, & Santoso, A. J. 2015. *Analisa Penentuan Jumlah Cluster Terbaik pada Metode K-Means Clustering*. Yogyakarta: Prosiding Seminar Nasional Multi Disiplin Ilmu & *Call for Papers Unisbank* (Sendi_U), 978-979.
- Prasetyo, Eko. 2013. *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Jakarta: Andi Publisher
- Rismawan, T, dan Kusumadewi, S. 2008. *Aplikasi K-Means Untuk pengelompokan Mahasiswa Berdasarkan Nilai Body Mass Index (BMI) & Ukuran Kerangka, SNATI*. Yogyakarta.
- Santosa, B. 2007. *Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu
- Santoso, S. 2010. *Statistik Multivariat*. Jakarta: Elex Media Komputindo