

Implementasi Metode *Mcquitty Linkage* dan *K-Means Clustering* dengan Pendekatan *Rule-of-Thumb* pada Data Kemiskinan di Jawa Tengah

Yessy Dwi Ramadani^{a,*}, Bowo Winarno^b

^{a, b} Program Studi Matematika FMIPA Universitas Sebelas Maret, Jl. Ir. Sutami 36A, Surakarta, 57126, Indonesia

* Alamat Surel: ramadaniyesy@students.uns.ac.id

Abstrak

Kemiskinan merupakan salah satu penyakit perekonomian mendasar untuk setiap negara tidak terkecuali di Indonesia. Badan Pusat Statistik Jawa Tengah mencatat bahwa penduduk Jawa Tengah yang hidup di bawah garis kemiskinan bertambah sejumlah 128,85 ribu jiwa menjadi 4,11 juta jiwa pada Maret 2021 dibanding Maret 2020. Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk mengelompokkan penduduk miskin di Jawa Tengah agar pemerataan bantuan di setiap daerah lebih tepat sasaran. Penelitian ini menerapkan metode *K-Means* dengan *Mcquitty Linkage* sehingga dapat mengatasi masalah penentuan pusat *cluster K-Means* yang dilakukan secara acak. Jumlah *cluster* dalam penelitian ini ditentukan menggunakan pendekatan *Rule-of-Thumb*. Hasil penelitian menunjukkan bahwa *clustering* metode *K-Means* dengan *Mcquitty Linkage* lebih baik daripada *K-Means* saja. Selain itu, pendekatan *Rule-of-Thumb* menunjukkan bahwa jumlah *cluster* yang optimal untuk pengelompokan data kemiskinan di Jawa Tengah adalah sebanyak 4 *cluster*.

Kata kunci:

Kemiskinan, *Clustering*, *Mcquitty Linkage*, *K-Means*, *Rule-of-Thumb*, *Data Mining*.

© 2023 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

1. Pendahuluan

Kemiskinan menjadi penyakit perekonomian yang sering dihadapi setiap negara termasuk Indonesia. Provinsi Jawa Tengah mengalami peningkatan kemiskinan sebesar 128,85 ribu jiwa sehingga berjumlah 4,11 juta jiwa pada Maret 2021 menurut data Badan Pusat Statistik. Untuk menuntaskan kemiskinan, Jawa Tengah mengupayakan memberikan bantuan sosial melalui program keluarga harapan, kartu Jateng sejahtera, Bantuan Operasional Sekolah (BOS) daerah, kegiatan rehabilitasi rumah tidak layak huni, dan lain sebagainya. Kendala utama yang sering dihadapi pemerintah adalah persoalan penargetan pemberian bantuan untuk penduduk yang membutuhkan. Oleh karena itu, perlu dilakukan upaya yang lebih optimal, adil, dan menyeluruh dalam perbaikan kondisi kemiskinan di Jawa Tengah.

Pengelompokan data keluarga miskin dengan tepat dan efisien sangat diperlukan untuk mengetahui kondisi kemiskinan di setiap wilayah. Hasil pengelompokan berupa gambaran dan evaluasi kondisi sosial ekonomi di Jawa Tengah berdasarkan indikator kemiskinan.

Penerapan *data mining* pada data kemiskinan di Jawa Tengah tahun 2021 menggunakan gabungan metode *Mcquitty Linkage* dan *K-Means clustering*. Kelebihan metode *K-Means* yaitu dapat mengelompokkan *database* yang cukup besar dan waktu yang diperlukan dalam proses komputasi relatif cepat dan efisien. Selain mempunyai kelebihan, metode *K-Means* juga mempunyai kelemahan yaitu pusat *cluster K-Means* dipilih secara acak dan subjektif sehingga menghasilkan hasil yang kurang akurat. Oleh karena itu, penelitian menggunakan metode *Mcquitty Linkage* yang menghasilkan *output* pusat *cluster*, kemudian dilanjutkan dengan perhitungan metode *K-Means*. Jumlah *cluster* dalam penelitian ini ditentukan menggunakan pendekatan *Rule-of-Thumb*. Hasil *clustering* yang diperoleh menjadi informasi yang bermanfaat dalam pengambilan kebijakan pada data penduduk miskin Jawa Tengah.

To cite this article:

Ramadani, Y. D. & Winarno, B. (2023). Implementasi Metode *Mcquitty Linkage* dan *K-Means Clustering* dengan Pendekatan *Rule-of-Thumb* pada Data Kemiskinan di Jawa Tengah. *PRISMA, Prosiding Seminar Nasional Matematika* 6, 751-756.

2. Metode

2.1. Data Penelitian

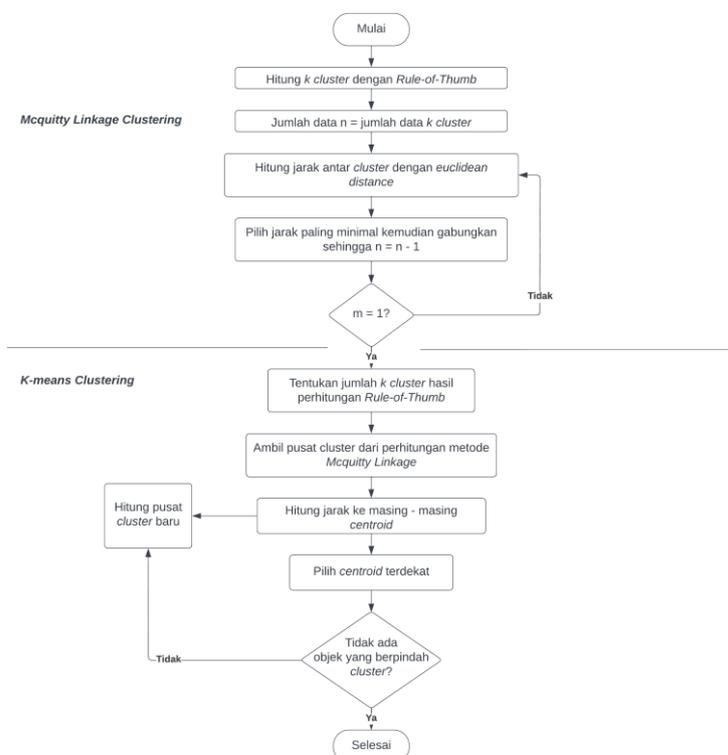
Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik Jawa Tengah. Data terdiri atas sembilan variabel dan 35 kabupaten/kota di Jawa Tengah pada tahun 2021. Variabel kemiskinan yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan
X_1	Persentase penduduk miskin
X_2	Persentase penduduk usia 15 tahun ke atas tidak/belum tamat SD
X_3	Persentase angka melek huruf usia 15-55 tahun
X_4	Persentase angka partisipasi sekolah usia 13-15 tahun
X_5	Persentase penduduk usia 15 tahun ke atas tidak bekerja
X_6	Persentase penduduk usia 15 tahun ke atas bekerja di sektor pertanian
X_7	Persentase penduduk usia 15 tahun ke atas bekerja di sektor informal
X_8	Persentase pengeluaran per kapita per bulan untuk komoditi makanan
X_9	Persentase rumah tangga yang menggunakan jamban sendiri/bersama

2.2. Langkah Penelitian

Berikut merupakan *flowchart* penelitian yang digambarkan seperti pada Gambar 1.



Gambar 1. Flowchart Langkah Penelitian

- (1) Menentukan jumlah *cluster* menggunakan *Rule-of-Thumb*.
- (2) Melakukan perhitungan *Mcquitty Linkage*.
- (3) Visualisasi hasil *Mcquitty Linkage* dengan dendrogram, lalu dicari rata-rata setiap *cluster*.
- (4) Melakukan *clustering* menggunakan *K-Means* dengan *centroid* awal merupakan hasil langkah 3.
- (5) Analisis hasil *cluster* yang diperoleh.

3. Hasil dan Pembahasan

Pada bab ini dibahas mengenai penentuan jumlah *cluster* dengan *Rule-of-Thumb*, *clustering* metode *K-Means* dengan *Mcquitty Linkage*, dan hasil *cluster* pada data Kemiskinan di Jawa Tengah pada tahun 2021.

3.1. Penentuan Jumlah Cluster dengan Rule-of-Thumb

Jumlah *cluster* dalam penelitian ini ditentukan menggunakan perhitungan *Rule-of-Thumb* dengan persamaannya adalah

$$k = \sqrt{\frac{n}{2}} \quad (1)$$

dengan n adalah jumlah objek yang akan di kelompokkan dan k jumlah *cluster*. Sintaksis persamaan *Rule-of-Thumb* adalah

```
#menghitung jumlah cluster optimal dengan rule of thumb
import math
k = math.sqrt(35/2)
print(k)
```

Berdasarkan hasil perhitungan menggunakan pendekatan *Rule-of-Thumb*, jumlah *cluster* yang dihasilkan adalah empat *cluster*.

3.2. Mcquitty Linkage

Salah satu metode hierarki *Agglomerative* (pemusatan) dan sering disebut juga dengan metode *Pair Group Method with Arithmetic Mean* (WPGMA).

$$d_{aj} = \frac{d_{kj} + d_{lj}}{2} \quad (2)$$

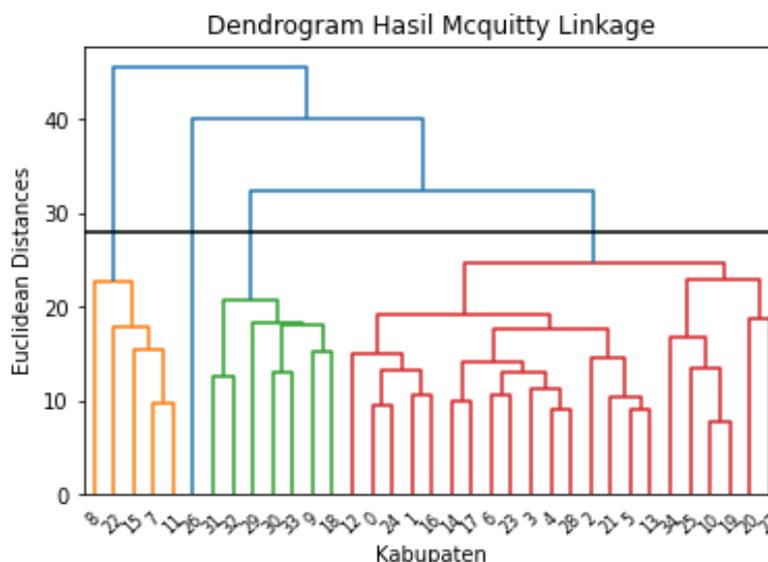
dengan

d_{aj} = jarak antara *cluster a* dan *j*
 d_{kj} = jarak antara *cluster k* dan *j*
 d_{lj} = jarak antara *cluster l* dan *j*
 a = *cluster gabungan yang termasuk dari cluster k dan l*.

Sintaksis yang digunakan untuk perhitungan *Mcquitty Linkage* ini adalah

```
import scipy.cluster.hierarchy as sch
dendrogram = sch.dendrogram(sch.linkage(dataku, method = 'weighted',
metric='euclidean'))
plt.title('Dendrogram Hasil Mcquitty Linkage')
plt.xlabel('Kabupaten')
plt.ylabel('Euclidean Distances')
max_d = 28
plt.axhline(y=max_d, c='black')
plt.show()
```

hasil dari perhitungan *Mcquitty Linkage* ini dapat digambarkan menggunakan dendrogram. Dendrogram kemudian dipotong dengan menggunakan *threshold* tertentu sehingga terbentuk empat *cluster* sesuai perhitungan *Rule-of-Thumb*. Gambar dendrogram hasil *Mcquitty Linkage* dan *threshold* yaitu



Gambar 2. Visualisasi metode *Mcquitty Linkage*

3.3. *K-Means Clustering*

Setelah dendrogram dipotong, maka *cluster* yang ada di dalam masing-masing potongan tersebut dicari rata-rata dari bobot datanya yang kemudian dijadikan pusat *cluster* perhitungan *K-Means*. Pada penelitian ini, *threshold* yang ditentukan adalah 28, sehingga terdapat empat *cluster* yang akan dijadikan pusat *cluster K-Means*. Hasil perhitungan rata-rata setiap *cluster* hasil *Mcquitty Linkage* yang akan dijadikan pusat *cluster* pada *K-Means* ditunjukkan pada Tabel 2.

Tabel 2. Pusat *cluster K-Means*

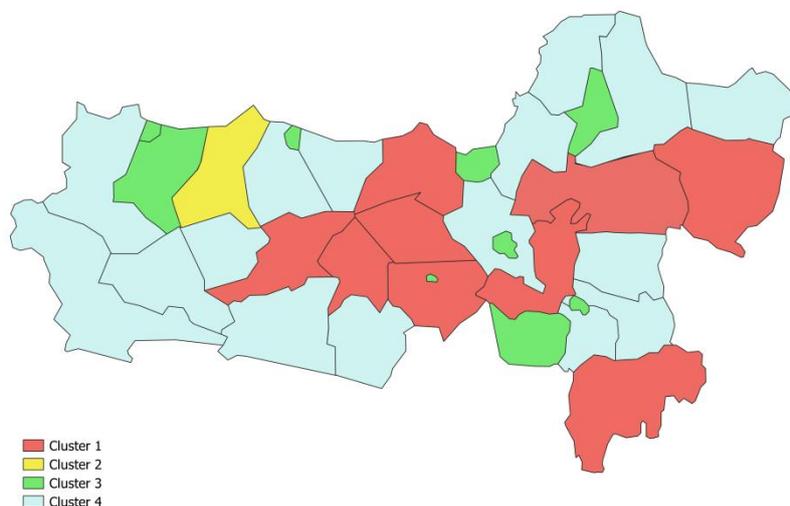
<i>Cluster</i>	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	11,32	30,21	97,31	91,63	32,09	48,308	57,42	60,44	95,77
2	16,56	34,16	95,7	67,21	45,48	23,51	35,03	66,24	90,77
3	7,93	16,45	99,43	98,33	42,31	3,004	22,71	60,51	96,54
4	12,27	28,33	97,65	95,37	41,20	27,04	39,94	62,5	94,40

Sintaksis yang digunakan untuk perhitungan *K-Means* ini adalah

```
centroidawal = np.array([[11.328, 30.21, 97.316, 91.634, 32.094, 48.308, 57.428, 60.444, 95.778], [16.56, 34.16, 95.7, 67.21, 45.48, 23.51, 35.03, 66.24, 90.77], [7.932857, 16.45, 99.43714, 98.33571, 42.31, 3.004286, 22.71, 60.51571, 96.54571], [12.27318, 28.33455, 97.65727, 95.37136, 41.20682, 27.04864, 39.94773, 62.5, 94.40773]], np.float64)
kmeans = KMeans(n_clusters = 4 , max_iter= 900, init = centroidawal, n_init=1, random_state=42)
hasil_clustering = kmeans.fit_predict(dataku)
```

```
print(hasil_clustering)
kmeans.cluster_centers_
```

Hasil akhir *clustering* data kemiskinan kabupaten/kota Jawa Tengah tahun 2021 menggunakan metode *Mcquitty Linkage* dan *K-Means* dapat dipetakan seperti pada Gambar 3.



Gambar 3. Peta *cluster* kemiskinan Jawa Tengah

Berdasarkan peta tersebut dapat dijelaskan untuk setiap anggota *cluster* seperti pada Tabel 3.

Tabel 3. Hasil *clustering* dengan metode *Mcquitty Linkage* dan *K-Means*

<i>Cluster</i>	Anggota
1	Banjarnegara, Wonosobo, Magelang, Boyolali, Wonogiri, Grobogan, Blora, Temanggung, dan Kendal.
2	Kabupaten Pemalang.
3	Klaten, Kudus, Tegal, Kota Magelang, Kota Surakarta, Kota Salatiga, Kota Semarang, Kota Pekalongan, dan Kota Tegal.
4	Cilacap, Banyumas, Purbalingga, Kebumen, Purworejo, Sukoharjo, Karanganyar, Sragen, Rembang, Pati, Jepara, Demak, Semarang, Batang, Pekalongan, Brebes.

4. Simpulan

Berdasarkan hasil dan pembahasan diperoleh simpulan bahwa pengelompokan data kemiskinan di Jawa Tengah pada tahun 2021 menggunakan metode *Mcquitty Linkage* dan *K-Means* dengan pendekatan *Rule-of-Thumb* menghasilkan empat *cluster*. *Cluster* pertama terdiri atas sembilan kabupaten/kota, *cluster* dua terdiri atas satu kabupaten/kota, *cluster* tiga terdiri atas sembilan kabupaten/kota, dan *cluster* empat terdiri atas 16 kabupaten/kota.

Daftar Pustaka

- Alfina, T., Santosa, B., Barakbah, A. R. (2012). Tahta Alfina, Budi Santosa, dan Ali Ridho Barakbah Jurusan Teknik Industri, Fakultas Teknologi Industri, Institut Teknologi Sepuluh Nopember (ITS) Jl. Arief Rahman Hakim, Surabaya 60111. *Jurnal Teknik POMITS*, 1(1), 1–5
- Andrea, T., Wahyuningsih, S., Nanda, A. (2019). Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu. *Jurnal of Mathematics*, 1(2), 64–78.
- Aprilia, A., Rahmawati, W. M., Hakimah, M. (2019). Penentuan Kategori Status Gizi Balita Menggunakan Penggabungan Metode Klasterisasi Agglomerative Dan K-Means. *Seminar Nasional Sains Dan Teknologi Terapan VII - Institut Teknologi Adhi Tama Surabaya*, 595–600.
- Badan Pusat Statistik (BPS). 2021. *Data dan Informasi Kemiskinan Kabupaten/Kota 2021*. Jakarta: Badan Pusat Statistik.
- Febianto, N. I., Palasara, N. (2019). Analisa Clustering K-Means Pada Data Informasi Kemiskinan Di Jawa Barat Tahun 2018. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 8(2), 130–140. <https://doi.org/10.32736/sisfokom.v8i2.653>
- Handoyo, R., R. Rumani, M., Naution, S. M. (2014). Perbandingan Metode Clustering Menggunakan Metode Single Linkage dan K-Means Pada Pengelompokan Dokumen. *JSM STMIK Mikroskil*, 15(2), 73–82.
- Nishom, M., Fathoni, M. Y. (2018). Implementasi Pendekatan Rule-Of-Thumb untuk Optimasi Algoritme K-Means Clustering. *Jurnal Informatika: Jurnal Pengembangan IT*, 3(2), 237–241. <https://doi.org/10.30591/jpit.v3i2.909>
- Rahmawati, L., Widya Sihwi, S., Suryani, E. (2016). Analisa Clustering Menggunakan Metode K-Means Dan Hierarchical Clustering (Studi Kasus : Dokumen Skripsi Jurusan Kimia, Fmipa, Universitas Sebelas Maret). *Jurnal Teknologi Informasi ITSmart*, 3(2), 66. <https://doi.org/10.20961/its.v3i2.654>
- Wahyuni, S., Jatmiko, Y. A. (2019). Pengelompokan Kabupaten/Kota di Pulau Jawa Berdasarkan Faktor-Faktor Kemiskinan dengan Pendekatan Average Linkage Hierarchical Clustering. *Jurnal Aplikasi Statistika Komputasi Statistik*, 10(1), 1. <https://doi.org/10.34123/jurnalasks.v10i1.197>
- Yuli Mardi. (2019). *Data Mining : Klasifikasi Menggunakan Algoritme C4 . 5* Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD). *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2(2), 213–219.