

PENGELOMPOKAN KEMISKINAN DI PROVINSI RIAU DENGAN MENGGUNAKAN METODE *K-MEDOIDS*

¹ Putra Heru Wijoanda, ²Alfirman

¹Mahasiswa Program Studi S1 Sistem Informasi

²Dosen Program Studi S1 Sistem Informasi

Jurusan Ilmu Komputer

Fakultas Matematika dan Ilmu Pengetahuan Alam

Kampus Bina Widya Pekanbaru, 28293, Indonesia

putra.heru3968@student.unri.ac.id, alfirman@lecturer.unri.ac.id

ABSTRACT

Poverty is a long-standing problem in Indonesia, and it has been an issue during almost all periods of government. Poverty can cause serious problems in the future when the government does not pay special attention to it. The inability of people to meet their standard of living, lack of access to education and employment will lead to recession. The quality of human resources affects low income productivity. Efforts to fight poverty must be carried out comprehensively, covering all areas of people's lives and coordinating their implementation. Poverty has always been a problem for Indonesia, until now there is no sign that it will stop. Therefore, research was conducted on poverty grouping in Riau province using the K-medoids method. It is expected that the results can provide additional information in optimizing poverty reduction in Riau Province. The research was conducted by forming 2 to 4 clusters, then the best results were obtained where the value close to 1 is the best value, namely in the formation of 2 clusters with 0.314477 in cluster 1 consisting of 3 districts Kuantan singingi, Indragiri Hulu, and Indragiri Hilir with high poverty rates while in cluster 2 consisting of Pelalawan, Siak, Kampar, Rokan Hulu, Bengkalis, Rokan Hilir, Kepulaun meranti, Pekanbaru city, and Dumai city with poverty rates.

Keywords: *Clustering, Data Mining, Poverty, Silhouette coefficient*

ABSTRAK

Kemiskinan merupakan salah satu persoalan yang sudah lama terjadi pada Indonesia, terus menjadi isu selama hampir seluruh periode pemerintahan. Kemiskinan bisa mengakibatkan persoalan serius pada masa depan ketika pemerintah tidak memberikan perhatian khusus untuk itu. Ketidakmampuan seseorang buat memenuhi standar hidup mereka, kurangnya akses ke pendidikan dan pekerjaan akan menyebabkan resesi. Kualitas sumber daya manusia berpengaruh pada produktivitas pendapatan yang rendah. Usaha melawan kemiskinan wajib dilakukan secara komprehensif. mencakup seluruh bidang kehidupan masyarakat dan mengkoordinasikan pelaksanaannya. Kemiskinan selalu menjadi persoalan bagi negara Indonesia, hingga kini belum terdapat tanda-tanda akan berhenti. Oleh karna itu dilakukan penelitian mengenai pengelompokan kemiskinan diprovinsi Riau dengan menggunakan metode *K-medoids*. Diharapkan hasilnya dapat memberikan informasi tambahan dalam optimalisasi penanggulangan kemiskinan di Provinsi Riau. Penelitian dilakukan dengan membentuk 2 hingga 4 *cluster*, maka diperoleh hasil

terbaik dimana nilai yang mendekati 1 adalah nilai yang terbaik yaitu pada pembentukan 2 *clustering* dengan sebesar 0,314477 pada *cluster* 1 terdiri dari 3 kabupaten Kuantan singingi, Indragiri hulu, dan Indragiri hilir dengan tingkat kemiskinan yang tinggi sedangkan pada *cluster* 2 terdiri dari kabupaten Pelalawan, Siak, Kampar, Rokan hulu, Bengkalis, Rokan hilir, Kepulaun meranti, kota Pekanbaru, dan kota Dumai dengan tingkat kemiskinan yang rendah.

Kata Kunci: *Clustering, Data Mining, Kemiskinan, Silhouette coefficient*

PENDAHULUAN

Kemiskinan merupakan salah satu persoalan yang sudah lama terjadi pada Indonesia, terus menjadi isu selama hampir seluruh periode pemerintahan. Kemiskinan bisa mengakibatkan persoalan serius pada masa depan ketika pemerintah tidak memberikan perhatian khusus untuk itu. Ketidakmampuan seseorang buat memenuhi standar hidup mereka, kurangnya akses ke pendidikan dan pekerjaan akan menyebabkan resesi. Kualitas sumber daya manusia berpengaruh pada produktivitas pendapatan yang rendah. Usaha melawan kemiskinan wajib dilakukan secara komprehensif. mencakup seluruh bidang kehidupan masyarakat dan mengkoordinasikan pelaksanaannya. Kemiskinan selalu menjadi persoalan bagi negara Indonesia, hingga kini belum terdapat tanda-tanda akan berhenti.

Menurut data Badan Pusat Statistika (BPS), pada bulan januari 2020 sampai september 2022 kemiskinan diprovinsi Riau mengalami fluktuasi, pada september 2020 penduduk miskin di provinsi Riau mengalami kelonjakan. Pada 2021 penduduk miskin diprovinsi Riau mengalami fase tertinggi, ini sebabkan oleh pandemi *Covid-9* yang membuat perekonomian memburuk. Jumlah penduduk miskin Riau pada September 2022 sebanyak 493,130 orang, meningkat 8,100 orang terhadap Maret 2022 dan menurun 3,530 orang terhadap September 2021 (riau.bps.go.id, 2022). Secara umum penyebab kemiskinan itu adalah pembangunan yang tidak merata. Selain itu, kondisi untuk menghasilkan pendapatan bagi masyarakat sangat rendah dan tingkat pengangguran tinggi. Berdasarkan permasalahan di atas maka penulis tertarik melakukan penelitian dengan menggunakan algoritma *k-medoids clustering* pada data kemiskinan yang terdapat di provinsi Riau. Diharapkan hasilnya dapat memberikan informasi tambahan dalam optimalisasi penanggulangan kemiskinan di Provinsi Riau bagi pemangku kepentingan khususnya bagi Pemerintah Provinsi Riau. Sehingga dapat mengetahui variabel mana yang berpengaruh terhadap data kemiskinan.

TINJAUAN PUSTAKA

a. Kemiskinan di Provinsi Riau

Secara umum, tingkat kemiskinan di Provinsi Riau selama periode September 2018 - September 2022 mengalami fluktuasi. Kenaikan jumlah penduduk miskin tertinggi terjadi pada bulan Maret 2021 sebagai akibat dari pandemi Covid-19 yang berdampak pada perekonomian yang memburuk. Namun, pada bulan September 2022, jumlah penduduk miskin di provinsi Riau mengalami peningkatan kembali, mencapai 493,13 ribu orang. Jumlah ini naik sebanyak 8,10 ribu orang dibandingkan dengan Maret 2022,

tetapi turun sebanyak 3,53 ribu orang dibandingkan dengan September 2021. Persentase penduduk miskin pada September 2022 mencapai 6,84%, naik 0,06% dari Maret 2022 dan turun 0,16% dari September 2021. Rokan Hulu merupakan daerah dengan tingkat kemiskinan tertinggi di Provinsi Riau, menurut data dari BPS jumlah penduduk miskin mencapai 73,81 ribu jiwa. Sedangkan Kabupaten Kampar menempati posisi kedua dengan jumlah penduduk miskin sebanyak 63,55 ribu jiwa. Pembatasan kegiatan sosial masyarakat yang diterapkan untuk mengendalikan penyebaran COVID-19 telah berdampak pada perekonomian hampir di seluruh wilayah, termasuk di Provinsi Riau. Banyak kegiatan ekonomi terhenti, yang mengakibatkan meningkatnya pemutusan hubungan kerja, penurunan produktivitas individu dan perusahaan, dan mendorong munculnya orang miskin baru, sehingga secara keseluruhan meningkatkan jumlah penduduk miskin.

b. *Clustering*

Cluster adalah sekelompok atau kumpulan objek data yang mirip satu sama lain dalam *cluster* yang sama dan dengan objek dalam *cluster* yang berbeda. Objek akan dikelompokkan ke dalam satu atau lebih *cluster* sehingga objek dalam *cluster* memiliki kesamaan yang tinggi satu sama lain. *Clustering* adalah metode data mining berdasarkan beberapa karakteristik yang diinginkan, dalam pekerjaan pengelompokan belum jelas label masing-masing data, melalui pengelompokan diharapkan dapat diketahui kelompok datanya, kemudian diberi label sesuai kebutuhan. Jadi dapat disimpulkan bahwa *clustering* adalah suatu metode pengelompokan data yang memiliki kesamaan kemudian diberi label sesuai kebutuhan (Dwilestari et al, 2021).

c. *Metode K-Medoids*

Metode k-medoids dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw pada tahun 1987. Algoritma *k-medoids* sering disebut sebagai algoritma *Partitioning Around Medoids* (PAM). Metode *Partitioning* adalah metode pengelompokan data ke dalam *cluster-cluster* yang tidak memiliki hirarki antara satu sama lain. *Metode k-medoids* memiliki keunggulan dibandingkan metode *k-means*. *K-medoids* memiliki kinerja yang lebih baik jika jumlah data yang digunakan sedikit. Algoritma menggunakan objek dalam kumpulan objek untuk mewakili *cluster*. Objek yang dipilih untuk mewakili *cluster* disebut *medoid*. Namun, metode *k-medoids* memiliki kekurangan saat menginisialisasi *medoid* awal. Titik pusat diperoleh secara acak tanpa mempertimbangkan kesamaan antar titik pusat. Hal ini mempengaruhi hasil *clustering*. Karena kemungkinan *medoid* yang dipilih secara acak memiliki kemiripan yang sangat tinggi. Hasil pengelompokan seperti itu mungkin tidak optimal (Rofiqi, 2017).

Langkah-langkah algoritma *K-Medoids* (Marlina et al, 2018) adalah sebagai berikut:

1. Inisialisasi pusat *cluster* sebanyak k (jumlah *cluster*).
2. Alokasikan setiap data (objek) ke *cluster* terdekat menggunakan ukuran jarak.

Euclidean Distance

Jarak *Euclidean* adalah metode pengukuran jarak yang umum digunakan untuk menentukan jarak antara dua titik dalam *Euclidean space*. *Euclidean space* mencakup ruang dua dimensi, tiga dimensi, dan dimensi lebih tinggi. Dalam

mengukur tingkat kesamaan data menggunakan rumus jarak *Euclidean*, berikut adalah persamaan *euclidean distance* (1).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots (1)$$

Keterangan:

d = Jarak antara x dan y (*euclidean distance*)

x = Data pusat *cluster*

y = Data pusat atribut

i = Setiap data

n = Jumlah data

x_i = Data pada pusat *cluster* ke i

y_i = Data pada setiap *cluster* ke i

Manhattan Distance

Manhattan Distance digunakan sebagai metode pengukuran jarak yang dihasilkan berdasarkan jumlah selisih antara dua objek data. Proses perhitungan dapat dilakukan dengan menggunakan rumus pada persamaan (2.2) (Nishom, 2019).

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \dots\dots\dots (2)$$

Keterangan:

d = Jarak antara x dan y

x = Data pusat *cluster*

y = Data pusat atribut

i = Setiap data

n = Jumlah data

x_i = Data pada pusat *cluster* ke i

y_i = Data pada setiap *cluster* ke i

3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *medoid* baru.
4. Hitung jarak setiap objek yang berada pada masing-masing *cluster* dengan kandidat *medoid* baru.
5. Hitung total simpangan (S) dengan menghitung total *distance* baru dikurang dengan total *distance* lama. Jika nilai $S < 0$, maka tukarkan objek dengan data *cluster non-medoids* untuk membentuk kumpulan k objek baru sebagai *medoids*.
6. Ulangi langkah ketiga sampai kelima sampai tidak ada perubahan atau tetap dalam *medoid*, hingga didapat pengelompokan *cluster* beserta anggota *cluster*.

d. *Silhouette Coefficient*

Silhouette coefficient merupakan salah satu metode evaluasi yang digunakan untuk menguji kualitas dan kekuatan dari sebuah *cluster* (Adhe et al, 2020). langkah-langkah algoritma penyelesaian *silhouette coefficient* (Kartikawati, 2022) adalah:

1. Menghitung rata-rata jarak objek dengan semua objek lain yang berada dalam satu *cluster* nilai ini disebut $a(i)$ dengan menggunakan persamaan (3):

$$a(i) = \frac{1}{[A]-1} \sum_{j \in A, j \neq i} d(i, j) \dots\dots\dots (3)$$

Keterangan:

$a(i)$: Rata-rata jarak objek i dengan semua objek lain dalam satu *cluster*.

$d(i, j)$:Jarak objek i terhadap j.

A :Jumlah klaster.

2. Menghitung rata-rata jarak objek dengan semua objek *cluster* lain. Kemudian ambil nilai terkecil dari semua jarak rata-rata tersebut nilai ini disebut $b(i)$ dengan menggunakan persamaan (2.3):

$$b(i) = \frac{1}{|A|} \sum_{j \in C} d(i, j) \dots\dots\dots (4)$$

Keterangan:

$b(i)$:Rata-rata jarak objek i dengan semua objek lain dalam cluster lainnya.
 $d(i,j)$:Jarak objek i terhadap j .
 C :Cluster lain.

3. Menghitung nilai *silhouette coefficient* dengan menggunakan persamaan (5):

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \dots\dots\dots (5)$$

Keterangan:

$a(i)$:Rata-rata jarak objek i dengan semua objek lain dalam *cluster* yang sama.
 $b(i)$:Rata-rata jarak objek i dengan semua objek lain dalam *cluster* yang lainnya.
 $S(i)$:Nilai *silhouette coefficient*.

e. **Python**

Python ialah bahasa pemrograman yang sering dipergunakan oleh *programmer* atau pembuat program. *Python* memiliki fitur *sintaks* yang tidak terlalu sulit buat di pahami. Sehingga *Python* merupakan bahasa pemrograman taraf tinggi yang praktis dipergunakan. Saat menulis kode program pada bahasa pemrograman *Python*, beberapa aturan wajib dipenuhi. Hal ini buat mengantisipasi *error* atau masalah di program yang dirancang. Aturan pertama sintaks *Python* ialah menulis pernyataan atau perintah. Pernyataan ialah instruksi atau baris perintah yang akan dieksekusi oleh komputer. *Python* juga mempunyai perpustakaan komprehensif yang memungkinkan pemrogram untuk membuat perangkat lunak kompleks berasal kode sumber yang solid atau sederhana (Sukerta Wijaya et al, 2021).

f. **Jupyter Notebook**

Jupyter Notebook adalah aplikasi web sumber terbuka yang memungkinkan pengguna membuat dan berbagi dokumen interaktif menggunakan kode langsung, persamaan, visualisasi, dan teks naratif. *Jupyter notebook* digunakan untuk pembersihan data dan transformasi data, simulasi numerik, pemodelan statistik, visualisasi informasi, pembelajaran mesin, dan banyak tujuan lainnya. *Jupyter Notebook* memudahkan pemrogram untuk membuat, memodifikasi, menjalankan, dan berbagi kode. *Interface Jupyter notebook* sederhana dan mudah digunakan, sangat memudahkan pengguna yang baru mengenal pemrograman (Wang et al, 2020).

g. **Silhouette Coefficient**

Kemiskinan dipandang sebagai ketidakmampuan ekonomi untuk memenuhi kebutuhan pokok pangan dan non pangan. Kemiskinan merupakan masalah yang dihadapi oleh semua negara, dan pertumbuhan ekonomi merupakan salah satu indikator untuk mengatasi kemiskinan, diantaranya pertumbuhan ekonomi merupakan konsep pembangunan ekonomi. Berbagai studi empiris telah menunjukkan bahwa harapan pembangunan ekonomi terletak pada perbaikan ekonomi, seperti pengentasan kemiskinan, tingkat pendidikan yang lebih tinggi, atau peningkatan kesehatan. Pertumbuhan ekonomi itu sendiri dapat menjadi pendorong penciptaan kekayaan, yang

secara bertahap akan menghilangkan kemiskinan dan semua masalah yang menyertainya (riau.bps.go.id, 2022).

h. Data Mining

Data mining ialah proses menemukan pola atau informasi yang menarik pada data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma pada data mining sangat bervariasi. Menentukan metode atau algoritma yang tepat bergantung pada tujuan serta proses *Knowledge Discovery in Database* (KDD) pada semua *Database*. *Data mining* dapat disebut juga proses mengekstraksi serta mengidentifikasi informasi yang berguna serta pengetahuan yang relevan dari basis data besar menggunakan statistik, matematika, kecerdasan buatan, serta teknik pembelajaran mesin. *Data mining* ialah serangkaian proses untuk mengeksplorasi nilai tambah dari kumpulan data dalam bentuk pengetahuan yang tidak diketahui secara manual (Mardi, 2017).

METODE PENELITIAN

1. Tahapan Penelitian

Tahapan penelitian ini merupakan salah satu tahap penting. Tahap ini bertujuan untuk menganalisis dan merancang sistem yang dibutuhkan sehingga dapat membantu pengelompokan tingkat kemiskinan di provinsi riau dengan menggunakan yang didapatkan dari algoritma K-Medoids Clustering. Untuk melakukan penelitian ini diperlukan beberapa proses yaitu:

a. Pengumpulan Data

Sebelum proses perhitungan data, terlebih dahulu dilakukan pengumpulan data. Tahapan pengumpulan data merupakan tahapan yang paling penting dalam penelitian ini. Tahap ini berkaitan dengan masalah yang akan diteliti serta digunakan sebagai bahan analisa kebutuhan sistem yang akan dirancang. Pengumpulan data dilakukan secara tidak langsung atau data diperoleh dari pihak lain yaitu BPS Provinsi Riau data dan informasi kemiskinan kabupaten/kota tahun 2022. Data yang diperoleh harus mencakup variabel-variabel yang diperlukan untuk analisis, seperti pendapatan per kapita, tingkat pendidikan, Ketenagakerjaan, Fasilitas perumahan dan Program pemerintah penanggulangan kemiskinan. Studi literatur adalah metode untuk mendapatkan teori pendukung penelitian dan menjadi dasar penelitian ini, yang dapat diperoleh dari buku, jurnal, e-book dan dokumen proyek. Bagian kajian teks ini mencakup teori yang akan digunakan dalam penelitian, mencakup teori diantaranya adalah *data mining*, dan *metode K-Medoids Clustering*.

b. Data Cleaning

Data yang telah dikumpulkan dilakukan pembersihan untuk memastikan kualitas data yang digunakan. Pada data kemiskinan yang telah dikumpulkan dilakukan pemeriksaan apakah ada data yang kosong (*missing value*), duplikasi ataupun salah ketik.

c. *Processing Data*

Selanjutnya yaitu menentukan pola pengelompokan data menggunakan algoritma K-Medoids Cluster. Jumlah kelompok yang akan dibentuk sebagai nilai K (jumlah cluster) adalah satu, dua dan tiga. Adapun langkah-langkah kerja algoritma K-Medoids Clustering yaitu:

1. Inisialisasi pusat cluster sebanyak k (jumlah cluster).
2. Alokasikan setiap data (objek) ke cluster terdekat menggunakan ukuran jarak, dapat menggunakan persamaan rumus euclidean distance dan manhattan distance.
3. Pilih secara acak objek pada masing-masing cluster sebagai kandidat medoid baru.
4. Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoid baru.
5. Hitung total simpangan (S) dengan menghitung total distance baru dikurang dengan total distance lama. Jika nilai $S < 0$, maka tukarkan objek dengan data cluster non-medoids untuk membentuk kumpulan k objek baru sebagai medoids.
6. Ulangi langkah ketiga sampai kelima sampai tidak ada perubahan atau tetap dalam medoid, hingga didapat pengelompokan cluster beserta anggota cluster.

d. *Proses Evaluasi*

Setelah cluster terbentuk, langkah selanjutnya adalah mengevaluasi hasil clustering. Evaluasi dilakukan dengan mengukur kualitas cluster menggunakan metode Silhouette Coefficient. Hasil evaluasi akan membantu dalam menentukan apakah cluster yang dibentuk sudah optimal atau perlu disesuaikan.

e. *Analisa Hasil*

Setelah diketahui hasil clustering terbaik menggunakan metode evaluasi silhouette coefficient, maka selanjutnya hasil tersebut akan dianalisis untuk mengetahui golongan disetiap kelompok.

2. *Peralatan*

Dalam penelitian dibutuhkan beberapa perangkat pendukung yang berupa perangkat keras dan perangkat lunak untuk berjalannya perancangan dan implementasi sistem. Adapun perangkat yang akan digunakan dalam penelitian ini yaitu:

a. *Perangkat Keras*

Perangkat keras komputer merupakan komponen komputer yang dapat dilihat secara fisik dan rasakan secara langsung yang berguna untuk pengoperasian komputer. Perangkat keras yang digunakan dalam perancangan sistem di penelitian ini yaitu Laptop Acer Z476 dengan spesifikasi processor Intel core i3, RAM 4 GB, Harddisk 500 Megabyte dan Printer canon pixma.

b. *Perangkat Lunak*

Perangkat lunak komputer merupakan sebuah program yang diciptakan untuk menjalankan perangkat keras komputer dan sebagai tempat berinteraksi antara manusia dan komputer. Perangkat lunak yang berfungsi untuk pengolah data dalam penelitian ini yaitu OS (*Operating System*) Windows 11 64-bit (untuk penghubung perangkat lunak (*software*) dengan perangkat keras (*hardware*)). *Sistem operasi* akan memberikan perintah dan sebagai mediator antara manusia dan komputer, *Microsoft Office Word* 2019 (untuk membuat penulisan skripsi), *Microsoft Office Excel* 2019 (untuk melakukan perhitungan) dan *Python* (bahasa pemrograman yang dipakai didalam program).

HASIL DAN PEMBAHASAN

a. Pengumpulan data

Dalam penelitian ini data yang digunakan merupakan data yang telah dipublikasikan oleh BPS yaitu Data dan Informasi Kemiskinan Kabupaten/Kota Tahun 2020-2022. Variabel yang digunakan dalam penelitian ini adalah sebagai berikut yang dapat dilihat pada Tabel 1.

Tabel 1 Variabel Penelitian

Variabel	Keterangan
X1	Presentase Penduduk Miskin Usia 15 keatas Tamat Pendidikan dibawah SD
X2	Presentase Penduduk Miskin Usia 15 keatas Tamat Pendidikan tamat SD/SMP
X3	Presentase Penduduk Miskin Usia 15 keatas Tamat Pendidikan diatas SMA
X4	Angka Melek Huruf Usia 15-24 Tahun
X5	Angka Melek Huruf Usia 15-55 Tahun
X6	Angka Partisipasi Sekolah Penduduk Miskin Usia 7-12 Tahun
X7	Angka Partisipasi Sekolah Penduduk Miskin Usia 13-15 Tahun
X8	Penduduk Miskin Usia 15 Tahun keatas Tidak Bekerja
X9	Penduduk Miskin Usia 15 Tahun keatas Bekerja di Sektor Informal
X10	Penduduk Miskin Usia 15 Tahun keatas Bekerja di Sektor Formal
X11	Penduduk Miskin Usia 15 Tahun keatas Bekerja di Sektor Pertanian
X12	Penduduk Miskin Usia 15 Tahun keatas Bekerja di Sektor Bukan Pertanian
X13	Presentase Pengeluaran Perkapita untuk Makanan rumah tangga miskin
X14	Presentase Pengeluaran Perkapita untuk Makanan rumah tangga tidak miskin
X15	Presentase Pengeluaran Perkapita untuk Makanan rumah tangga miskin dan tidak miskin
X16	Presentase Rumah Tangga Miskin yang Menggunakan Air Layak
X17	Persentase Rumah Tangga Miskin yang Menggunakan Jamban Sendiri/Bersama
X18	Persentase Rumah Tangga Miskin yang Menerima penerima BPNT/program sembako

Adapun data yang akan digunakan dalam penelitian ini adalah data kabupaten/kota yang ada di Provinsi Riau dengan data rata-rata variabel pada Tabel 4.1 dari tahun 2020-2022. Tabel 4.2 merupakan rata-rata data kemiskinan di provinsi Riau dari 2020-2022 yang telah dibuat dalam bentuk tabel *excel*.

Tabel 2 Data Kemiskinan Provinsi Riau

No	Kabupaten/Kota	X1	X2	X3	X4	...	X16	X17	X18
1	Kuantan Singingi	27,95	50,71	21,34	99,4	...	66,94	73,03	31,18
2	Indragiri Hulu	33,73	47,65	18,62	98,39	...	42,62	68,14	22,41
3	Indragiri Hilir	28,51	58,89	12,6	100	...	91,91	73,98	12,99
4	Pelalawan	15,36	57,37	27,26	100	...	70,45	89,47	13,34
5	Siak	9,35	50,39	40,26	100	...	87,57	97,37	13,64
6	Kampar	8,58	58,98	32,45	100	...	79	85,06	9,05
7	Rokan Hulu	14,64	60,31	25,05	100	...	73,92	85,62	15,22
8	Bengkalis	20,21	51,06	29,88	100	...	89,42	86,94	22,99
9	Rokan Hilir	19,94	54,2	25,87	100	...	84,41	95,8	18,52
10	Kepulauan Meranti	19,45	57,23	23,32	100	...	86,88	92,87	47,92
11	Kota Pekanbaru	5,7	44,59	49,72	100	...	93,45	80,55	5,27
12	Kota Dumai	13,29	52,58	34,13	97,32	...	86,87	98,31	22,66

b. Data Cleaning

Data yang telah dikumpulkan sebelumnya dilakukan pembersihan dengan tujuan untuk memastikan kualitas data yang digunakan. Pada data kemiskinan yang telah dikumpulkan dilakukan pemeriksaan apakah ada data yang kosong (*missing value*), duplikasi ataupun salah ketik. Terdapat variabel data *missing values* pada variabel X_{11} . Maka solusinya adalah dengan mengeluarkan variabel yaitu pada variabel X_{11} .

Tabel 5 Hasil *Cleaning*

No	Kabupaten/Kota	X1	X2	X3	X4	...	X16	X17	X18
1	Kuantan Singingi	27,95	50,71	21,34	99,4	...	66,94	73,03	31,18
2	Indragiri Hulu	33,73	47,65	18,62	98,39	...	42,62	68,14	22,41
3	Indragiri Hilir	28,51	58,89	12,6	100	...	91,91	73,98	12,99
4	Pelalawan	15,36	57,37	27,26	100	...	70,45	89,47	13,34
5	Siak	9,35	50,39	40,26	100	...	87,57	97,37	13,64
6	Kampar	8,58	58,98	32,45	100	...	79	85,06	9,05
7	Rokan Hulu	14,64	60,31	25,05	100	...	73,92	85,62	15,22
8	Bengkalis	20,21	51,06	29,88	100	...	89,42	86,94	22,99
9	Rokan Hilir	19,94	54,2	25,87	100	...	84,41	95,8	18,52
10	Kepulauan Meranti	19,45	57,23	23,32	100	...	86,88	92,87	47,92
11	Kota Pekanbaru	5,7	44,59	49,72	100	...	93,45	80,55	5,27
12	Kota Dumai	13,29	52,58	34,13	97,32	...	86,87	98,31	22,66

c. Processing Data

Dalam penelitian ini *clustering* dilakukan sebanyak 3 kali percobaan, yaitu dengan membentuk 2 hingga 4 *clustering* dengan perhitungan jarak menggunakan *euclidean distance* dan *manhattan distance* yang menghasilkan *cluster* yang sama. Hasil pembentukan 2 *clustering* adalah *cluster* 1 terdiri dari 3 daerah yaitu Kuantan Singingi, Indragiri Hulu, dan Indragiri Hilir, sedangkan *cluster* 2 terdiri dari 9 daerah yaitu Pelalawan, Siak, Kampar, Rokan Hulu, Bengkalis, Rokan Hilir, Kepulauan Meranti, Kota Pekanbaru, dan Kota Dumai. Hasil pembentukan 3 *clustering* adalah *cluster* 1 terdiri dari 3 daerah yaitu Kuantan Singingi, Indragiri Hulu, dan Indragiri Hilir, *cluster* 2 terdiri dari satu daerah yaitu Kota Pekanbaru, sedangkan *cluster* 3 terdiri dari daerah yaitu Pelalawan, Siak, Kampar, Rokan Hulu, Bengkalis, Rokan Hilir, Kepulauan Meranti, dan Kota Dumai. Hasil pembentukan 4 *clustering* adalah *cluster* 1 terdiri dari 3 daerah yaitu Kuantan Singingi, Indragiri Hulu, dan Indragiri Hilir, *cluster* 2 terdiri dari satu daerah yaitu Kota Pekanbaru, *cluster* 3 terdiri dari 5 daerah yaitu Siak, Bengkalis, Rokan Hilir, Kepulauan Meranti, dan Kota Dumai, sedangkan *cluster* 4 terdiri dari 3 daerah yaitu Pelalawan, Kampar, Rokan Hulu.

d. Proses Evaluasi

Setelah *processing* data langkah selanjutnya adalah proses evaluasi. Proses evaluasi dilakukan dengan menggunakan metode *silhouette coefficient* terhadap hasil pembentukan 2 hingga 4 *clustering*. Dimana hasil pembentukan 2 hingga 4 *clustering* menggunakan *euclidean distance* dan *manhattan distance* disetiap pembentukannya hasilnya sama. Dari nilai hasil yang didapat dari metode *silhouette coefficient* terletak pada antar nilai -1 hingga 1 yang mana nilai *silhouette coefficient* mendekati nilai 1 maka semakin baik pengelompokan objeknya dalam satu *cluster*. Sebaliknya jika *silhouette coefficient* mendekati nilai -1, maka semakin buruk pengelompokan itemnya didalam satu *cluster*. Nilai *silhouette coefficient* pada hasil evaluasi 2 *clustering* adalah 0,31, hasil evaluasi *silhouette coefficient* 3 *clustering* menunjukkan nilai 0,2596, sedangkan hasil evaluasi *silhouette coefficient* 4 *clustering* nilai 0,1196.

e. Analisis Hasil

Pada perhitungan *clustering* sebelumnya menggunakan metode *k-medoid* kemudian akan dilakukan evaluasi menggunakan *silhouette coefficient*. Berikut merupakan hasil evaluasi menggunakan *silhouette coefficient* pada Tabel 6.

Tabel 6 Hasil Perhitungan *Silhouette Coefficient*

Banyaknya <i>cluster</i>	Nilai <i>silhouette coefficient</i>
2	0,314477
3	0,2596
4	0,1196

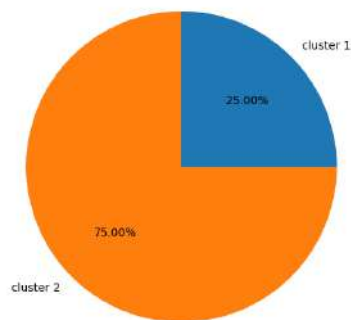
Hasil dari perhitungan *silhouette coefficient* yang menghasilkan 2 *clustering* terbaik, selanjutnya dapat menghitung rata-rata dari variabel/indikator kemiskinan. Sehingga dapat diperoleh indikator kemiskinan tinggi maupun rendah di masing-masing

cluster. Pada *cluster* 1 yang terdiri dari kabupaten/kota Kuantan Singingi, Indragiri Hulu, dan Indragiri Hilir dengan variabel atau indikator kemiskinan yang paling berpengaruh berkaitan dengan penduduk miskin usia 15 keatas tamat pendidikan dibawah SD, penduduk miskin usia 15 tahun keatas tidak bekerja, penduduk miskin usia 15 tahun keatas bekerja di sektor informal, pengeluaran perkapita untuk makanan rumah tangga miskin, pengeluaran perkapita untuk makanan rumah tangga tidak miskin, pengeluaran perkapita untuk makanan rumah tangga miskin dan tidak miskin, dan rumah tangga miskin yang menerima penerima BPNT/program sembako.

Sedangkan pada *cluster* 2 terdiri atas kabupaten/kota Pelalawan, Siak, Kampar, Rokan Hulu, Bengkalis, Rokan Hilir, Kepulauan Meranti, Kota Pekanbaru, dan Kota Dumai dengan variabel atau indikator kemiskinan yang paling berpengaruh berkaitan dengan penduduk miskin usia 15 keatas tamat pendidikan tamat SD/SMP, penduduk miskin usia 15 keatas tamat pendidikan diatas SMA, angka melek huruf usia 15-24 tahun, angka melek huruf usia 15-55 tahun, angka partisipasi sekolah penduduk miskin usia 7-12 tahun, angka partisipasi sekolah penduduk miskin usia 13-15 tahun, penduduk miskin usia 15 tahun keatas bekerja di sektor formal, penduduk miskin usia 15 tahun keatas bekerja di sektor bukan pertanian, rumah tangga miskin yang menggunakan air layak, dan rumah tangga miskin yang menggunakan jamban sendiri/bersama.

Sehingga dari beberapa uraian analisis sektor kemiskinan pada Tabel 4.13 dapat diketahui bahwa untuk cluster 1 memiliki kategori kemiskinan yang tinggi sedangkan cluster 2 dikategorikan kemiskinan yang rendah. Adapun bentuk grafik persentase yang dapat dilihat pada Gambar 4.33.

Persentase Pengelompokan di Provinsi Riau Tahun 2020 - 2022



Gambar 4.1 Grafik Persentase

Keterangan:

1. Warna biru merupakan kabupaten/kota dengan persentase 25% atau berjumlah 3 Kabupaten diprovinsi Riau yang masuk dalam cluster 1 atau kabupaten yang memiliki kategori kemiskinan yang tinggi yaitu Kuantan Singingi, Indragiri Hulu, dan Indragiri Hilir.
2. Warna jingga merupakan kabupaten/kota dengan persentase 75% atau berjumlah 9 Kabupaten diprovinsi Riau yang masuk dalam cluster 2 atau kabupaten/kota yang memiliki kategori kemiskinan yang rendah yaitu Pelalawan, Siak, Kampar, Rokan Hulu, Bengkalis, Rokan Hilir, Kepulauan Meranti, Kota Pekanbaru, dan Kota Dumai.

KESIMPULAN

Berdasarkan penelitian yang telah dilakukan mengenai penerapan algoritma K-Medoids clustering dalam pengelompokan kemiskinan di provinsi Riau maka dapat diambil kesimpulan sebagai berikut:

1. Metode K-medoids dapat digunakan dalam pengelompokan kemiskinan di provinsi Riau dengan menggunakan data dari badan pusat statistik tahun 2020 sampai 2022.
2. Dalam analisis data kemiskinan di Provinsi Riau, perbandingan antara Euclidean Distance dan Manhattan Distance dalam membentuk cluster data menghasilkan temuan bahwa meskipun metode perhitungan jarak berbeda, keduanya menghasilkan cluster yang sama. Ini mengindikasikan bahwa pola dan distribusi data kemiskinan di Provinsi Riau dapat direpresentasikan dengan baik menggunakan kedua metode tersebut.
3. Hasil evaluasi cluster metode k-medoids yang dilakukan dengan silhouette coefficient di peroleh hasil terbaik pada pembentukan 2 clustering dengan nilai sebesar 0,314477.
4. Pengelompokan kemiskinan kabupaten-kota di Riau dengan metode k-medoids didapatkan pembentukan 2 clustering, berjumlah 3 Kabupaten diprovinsi Riau yang masuk dalam cluster 1 yang memiliki kategori kemiskinan yang tinggi yaitu Kuantan Singingi, Indragiri Hulu, dan Indragiri Hilir. Sedangkan pada cluster 2 berjumlah 9 Kabupaten diprovinsi Riau yang memiliki kategori kemiskinan yang rendah yaitu Pelalawan, Siak, Kampar, Rokan Hulu, Bengkalis, Rokan Hilir, Kepulauan Meranti, Kota Pekanbaru, dan Kota Dumai.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Bapak Alfirman, S.Kom., M.Kom., yang telah membimbing, memotivasi serta membantu penelitian dan penulisan karya ilmiah ini.

SARAN

Berdasarkan temuan yang telah didapatkan, berikut saran untuk penelitian selanjutnya.

1. Penelitian selanjutnya diharapkan dapat menambah variabel lebih banyak dan dapat menerapkan diprovinsi yang lain di Indonesia.
2. Penelitian selanjutnya diharapkan dapat membangun sebuah sistem dalam bentuk aplikasi *mobile* atau berbasis website sebagai implementasi dari hasil pengelompokan yang terbentuk

DAFTAR PUSTAKA

- Adhe, D., Rachman, C., Goejantoro, R., Deny, F., & Amijaya, T. (2020). Implementasi Text Mining Pengelompokan Dokumen Skripsi Menggunakan Metode K-Means Clustering Implementation Of Text Mining For Grouping Thesis Documents Using K-Means Clustering. *Jurnal EKSPONENSIAL*, 11(2), 167–174.
- Dwilestari, G., Mulyawan, Martanto, & Ali, I. (2021). Analisis Clustering menggunakan K-Medoids pada Data Penduduk Miskin Indonesia. *Jursima: Jurnal Sistem Informasi Dan Manajemen*, 9(3), 282–290.

- Kartikawati, L. (2022). Analisis Kualitas Pengelompokkan Algoritma K-Means di Knime dan Excel untuk PTMT Pasca Vaksinasi Covid-19. *Ideguru: Jurnal Karya Ilmiah Guru*, 7(1), 70–79. <https://doi.org/10.51169/ideguru.v7i1.316>
- Mardi, Y. (2017). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Edik Informatika*, 2(2), 213–219. <https://doi.org/10.22202/ei.2016.v2i2.1465>
- Marlina, D., Lina, N., Fernando, A., & Ramadhan, A. (2018). Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokkan Wilayah Sebaran Cacat pada Anak. *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer Dan Teknologi Informasi*, 4(2), 64. <https://doi.org/10.24014/coreit.v4i2.4498>
- riau.bps.go.id. (2022). Penduduk Kabupaten/Kota (Jiwa), 2021-2023. Riau.Bps.Go.Id.
- Rofiqi, A. Y. (2017). Clustering Berita Olahraga Berbahasa Indonesia Menggunakan Metode K-Medoids Bersyarat. *Jurnal Simantec*, 6(1), 25–32.
- Sukerta Wijaya, I. W., Harjumawan Wiratmaja KS., I. G., Pramana Setya Bintara, I. D. M. A., & Ryan Aditya Permana, I. K. G. (2021). Program Menghitung Banyak Bata pada Ruangan Menggunakan Bahasa Python. *TIERS Information Technology Journal*, 2(1).
- Wang, J., Kuo, T. Y., Li, L., & Zeller, A. (2020). Assessing and Restoring Reproducibility of Jupyter Notebooks. *Proceedings - 2020 35th IEEE/ACM International Conference on Automated Software Engineering, ASE 2020*,