

TUGAS AKHIR
PENGELOMPOKKAN DATA KEMISKINAN PROVINSI PAPUA DENGAN
ALGORITMA *K-MEANS CLUSTERING*



Disusun oleh :

Diki Fernandi	082011233072
Muhammad Sodikur Rifki	082011233075

Dosen Pengampu :

Auli Damayanti, S.Si., M.Si.

PROGRAM STUDI S-1 MATEMATIKA
DEPARTEMEN MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS AIRLANGGA

2023

KATA PENGANTAR

Puji syukur kepada Tuhan Yang Maha Esa atas semua anugerah, kasih sayang, dan berkat-Nya yang telah memungkinkan kami menyelesaikan tugas akhir ini. Makalah ini mengangkat permasalahan *clustering* dengan judul “**Pengelompokkan Data Kemiskinan Provinsi Papua dengan Algoritma *K-Means Clustering***”.

Makalah ini disusun sebagai bagian dari penyelesaian tugas akhir dalam mata kuliah Data Mining, dengan harapan memberikan pembelajaran baik bagi penulis maupun pembaca. Penulis ingin mengucapkan terima kasih kepada Ibu Auli Damayanti S.Si., M.Si., dosen mata kuliah Data Mining, orang tua, dan teman-teman yang telah memberikan bantuan dan dukungan.

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam penyusunan makalah ini. Oleh karena itu, kritik dan saran dari pembaca sangat diharapkan untuk membantu memperbaiki dan menyempurnakan makalah ini.

Harapannya, makalah ini dapat memberikan manfaat dan inspirasi bagi pembaca dalam mempelajari permasalahan *clustering*, khususnya dengan menggunakan *metode K-means clustering*. Penulis juga ingin memohon maaf atas kesalahan dan kekurangan yang ada dalam makalah ini.

Surabaya, Juni 2023

Penulis

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI	ii
DAFTAR GAMBAR	iii
DAFTAR TABEL	iv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	1
1.3. Tujuan	1
1.4. Manfaat	2
BAB II TINJAUAN PUSTAKA.....	3
2.1. Kemiskinan	3
2.2. Data Mining	3
2.3. Clustering	4
2.4. K-Means Clustering	4
2.5. C++	6
BAB III METODE PENELITIAN	7
BAB IV PEMBAHASAN.....	9
4.1. Algoritma K-Means Clustering.	9
4.2. Data.	9
4.3. Penyelesaian Secara Manual Pengelompokan Data Kemiskinan Provinsi Papua.	10
4.4. Penyelesaian Menggunakan Program C++ Pengelompokan Data Kemiskinan Provinsi Papua.	12
4.5. Analisis Hasil Clustering.	16
BAB V PENUTUP.....	17
5.1. Kesimpulan	17
5.2. Saran	17
DAFTAR PUSTAKA.....	18
LAMPIRAN.....	20

DAFTAR GAMBAR

Gambar 1. <i>Flowchart</i> Algoritma <i>K-Means Clustering</i>	5
Gambar 2. <i>Flowchart</i> Metode Penelitian.	8
Gambar 3. <i>Output</i> Iterasi 1.....	13
Gambar 4. <i>Output</i> Iterasi 2.....	13
Gambar 5. <i>Output</i> Iterasi 3.....	14
Gambar 6. <i>Output</i> Iterasi 4.....	14
Gambar 7. <i>Output</i> Iterasi 5.....	15
Gambar 8. <i>Output</i> Lanjutan Iterasi 5.	15

DAFTAR TABEL

Tabel 1. Data rata-rata lama sekolah, data pengeluaran Per kapita riil, dan data jumlah penduduk miskin kabupaten/kota di Provinsi Papua.	10
Tabel 2. Data rata-rata lama sekolah, data rata-rata pengeluaran Per kapita riil, dan rata-rata data jumlah penduduk miskin kabupaten/kota di Provinsi Papua pada 2017-2021. .	10
Tabel 3. Nilai <i>Centroid</i> Awal.....	10
Tabel 4. Perhitungan <i>Euclidean Distance</i>	11
Tabel 5. Hasil Pengelompokan.	12

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kemiskinan merupakan salah satu tantangan yang menghambat pertumbuhan nasional maupun regional. Kemiskinan dianggap sebagai masalah multidimensi yang terkait dengan keterbatasan ekonomi, sosial, budaya, politik, dan partisipasi dalam masyarakat (Nurwati, 2018). Indonesia, sebagai negara dengan populasi keempat terbanyak di dunia, memiliki total penduduk sebanyak 273.523.615 jiwa pada tahun 2022. Salah satu indikator keberhasilan pembangunan suatu negara dapat dilihat dari penurunan jumlah penduduk miskin (Irmanita Nasution, 2020). Menurut Badan Pusat Statistik (BPS), pada bulan September 2021, jumlah penduduk miskin di Indonesia mencapai 26,56 juta orang atau sekitar 9,71%, yang menjadikannya sebagai negara dengan jumlah penduduk miskin terbesar di dunia (BPS, 2022).

Menurut Badan Pusat Statistik (BPS) terjadi kenaikan tingkat angka kemiskinan di Indonesia pada September 2022 dibandingkan periode sebelumnya Maret 2022. Dari segi jumlah penduduk miskin jumlahnya naik sebesar 0,20 juta orang mencapai 26,36 juta orang (Indonesiabaik.id). Hal ini disebabkan oleh berbagai faktor seperti tingginya pertumbuhan penduduk, tingkat pengangguran yang tinggi, pertumbuhan ekonomi yang lambat, tingkat pendidikan yang rendah, dan distribusi yang tidak merata. Kemiskinan dapat memiliki dampak yang signifikan terhadap perkembangan negara karena menghambat daya saing di masa depan.

Analisis *clustering* merupakan teknik *multivariate* yang bertujuan untuk mengelompokkan objek berdasarkan karakteristik yang dimiliki. Dalam algoritma klaster, fokus utamanya adalah mencari pusat klaster secara iteratif, di mana pusat klaster ditentukan berdasarkan jarak minimum dari setiap data ke pusat klaster (Prajna, 2019).

Data yang digunakan dalam penelitian ini berasal dari situs *bps.go.id*. Topik yang dibahas dalam penelitian ini adalah pengelompokan data kemiskinan di provinsi Papua menggunakan algoritma *clustering*. Hasil dari analisis klaster ini dapat menjadi masukan bagi pemerintah provinsi Papua, terutama bagi klaster dengan tingkat kemiskinan rendah, guna memberikan perhatian lebih terhadap perekonomian masyarakat Papua.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dibahas sebelumnya, dapat dirumuskan permasalahan sebagai berikut :

1. Bagaimana menerapkan *K-means clustering* dalam pengelompokan data kemiskinan di provinsi Papua?
2. Bagaimana mengimplementasikan program C++ *K-means clustering* untuk mengelompokkan data kemiskinan di provinsi Papua?

1.3. Tujuan

Berdasarkan rumusan masalah di atas, dapat diambil tujuan dari penulisan makalah ini :

1. Menerapkan *K-means clustering* dalam pengelompokan data kemiskinan di provinsi Papua.
2. Mengimplementasikan program C++ *K-means clustering* untuk mengelompokkan data kemiskinan di provinsi Papua.

1.4. Manfaat

Adapun manfaat yang didapat dari adanya penelitian ini, di antaranya :

1. Menghasilkan program *K-means clustering* untuk mengelompokkan data kemiskinan di provinsi Papua.
2. Menambah wawasan kepada pembaca tentang *K-means clustering* dan penerapannya dalam mengelompokkan data kemiskinan di provinsi Papua.

BAB II

TINJAUAN PUSTAKA

2.1. Kemiskinan

Kemiskinan adalah keadaan yang penuh dengan keterbatasan yang terjadi bukan karena keinginan individu yang terlibat. Seseorang dianggap miskin jika ditandai dengan rendahnya tingkat pendidikan, produktivitas kerja, pendapatan, kesehatan, gizi, serta kesejahteraan hidup yang menunjukkan lingkaran ketidakmampuan. Kemiskinan dapat disebabkan oleh keterbatasan sumber daya manusia yang ada, baik melalui pendidikan formal maupun nonformal, yang akhirnya berdampak pada rendahnya pendidikan informal (Supriatna, 1970).

Selanjutnya, menurut laporan Bank Dunia (1990) dalam pertemuan dengan anggota PBB yang berjudul "*Poverty and Human Development*", dikatakan bahwa: "*The case for human development is not only or even primarily an economic one. Less hunger, fewer child deaths, and better chances of primary education are almost universally accepted as important ends in themselves*" (pembangunan manusia bukan hanya atau bahkan utamanya berkaitan dengan aspek ekonomi. Kurangnya kelaparan, penurunan angka kematian anak, dan peluang pendidikan dasar yang lebih baik hampir secara universal diterima sebagai tujuan penting dalam dirinya sendiri) dengan tujuan untuk orang-orang miskin guna meningkatkan kehidupan sosial dan ekonomi mereka.

Lebih lanjut, menurut Booth dan McCawley (Dalam Moeljarto T., 1993), mereka menyatakan bahwa "*di banyak negara, memang terjadi peningkatan tingkat kesejahteraan masyarakat yang diukur berdasarkan pendapatan per kapita, namun hanya sebagian kecil masyarakat yang dapat menikmatinya, sedangkan sebagian besar masyarakat miskin tidak memperoleh manfaat apa pun, bahkan merugi*" (Kadji, 2019).

2.2. Data Mining

Data mining adalah proses penemuan pola atau informasi yang menarik dari data yang dipilih menggunakan teknik atau metode tertentu. Ada berbagai teknik, metode, atau algoritma yang digunakan dalam *data mining*. Pemilihan metode atau algoritma yang tepat sangat tergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan (Mardi).

Secara umum, *data mining* dapat dikelompokkan sebagai berikut:

- Deskripsi, digunakan untuk menggambarkan pola dan tren yang ada dalam data dengan jelas sesuai dengan interpretasi data.
- Prediksi, digunakan untuk memperkirakan nilai yang belum diketahui dan juga untuk memproyeksikan nilai di masa depan.
- Estimasi, digunakan untuk menduga atau memperkirakan apa yang akan terjadi di masa depan berdasarkan informasi yang relevan dari masa lalu melalui metode ilmiah (Wanto, 2018).
- Klasifikasi, dalam klasifikasi terdapat variabel target kategori, misalnya penggolongan pendapatan menjadi tiga kelas: tinggi, menengah, dan rendah. Model *data mining* mengamati serangkaian catatan yang berisi informasi tentang variabel target serta variabel *input* atau prediksi.
- Clustering*, *clustering* melibatkan pengelompokan rekaman, observasi, atau objek berdasarkan kemiripannya. *Clustering* adalah kumpulan rekaman yang serupa satu sama lain dan berbeda dari rekaman dalam

kelompok lain. Perbedaan antara *clustering* dan klasifikasi adalah bahwa *clustering* tidak memerlukan variabel target untuk pengelompokan.

- f. Asosiasi, asosiasi dalam *data mining* bertujuan untuk menemukan atribut-atribut yang sering muncul bersama. Dalam konteks bisnis, ini dikenal sebagai analisis afinitas atau analisis keranjang belanja. Asosiasi mencoba mencari aturan yang mengukur hubungan antara dua atau lebih atribut.

2.3. *Clustering*

Menurut Tan (2006), *clustering* adalah proses pengelompokan data ke dalam beberapa klaster atau kelompok, di mana data dalam satu klaster memiliki tingkat kemiripan maksimum, sementara data antar klaster memiliki kemiripan minimum. *Clustering* merupakan metode yang digunakan untuk membentuk grup objek atau klaster, di mana objek dalam satu klaster memiliki kesamaan ciri yang tinggi, sedangkan objek dalam klaster yang berbeda memiliki kesamaan ciri yang rendah.

Tujuan dari klasterisasi data adalah mengelompokkan data yang memiliki kesamaan ciri dan memisahkan data ke dalam klaster yang berbeda berdasarkan ciri yang berbeda pula. Sebaliknya, dalam klasifikasi, kelas-kelas telah ditentukan sebelumnya. Klaster terbentuk berdasarkan ciri objek dan kriteria pengelompokan yang telah ditetapkan sebelumnya (Thanamani, 2013).

Tujuan dari data *clustering* adalah meminimalkan fungsi objektif yang telah ditetapkan dalam proses *clustering*, yang pada umumnya bertujuan untuk meminimalkan variasi dalam satu klaster dan meminimalkan variasi antara klaster (Thanamani, 2013).

2.4. *K-Means Clustering*

K-Means merupakan salah satu teknik *clustering* dalam proses *Data Mining* yang dilakukan secara *unsupervised* dan menggunakan metode partisi untuk mengelompokkan data. Metode *K-Means* membagi data menjadi beberapa kelompok, di mana setiap kelompok memiliki karakteristik yang mirip atau serupa, tetapi berbeda dengan kelompok lain. Tujuan utama dari *K-Means* adalah meminimalkan perbedaan antara data dalam satu kelompok (*cluster*) dan memaksimalkan perbedaan dengan kelompok lainnya (P. Sari, 2017).

Dalam algoritma *K-Means Clustering*, terdapat beberapa istilah yang digunakan, antara lain:

- a. *Cluster* : *Cluster* adalah kelompok atau grup.
- b. *Centroid* : Titik pusat untuk menentukan *euclidian distance*.
- c. Iterasi : Pengulangan proses, berhenti ketika hasil iterasi telah konvergen.

Langkah-langkah pada algoritma *k-means clustering* sebagai berikut:

1. Menentukan jumlah *k* sebagai jumlah klaster yang akan dibentuk. Penentuan jumlah klaster *k* biasanya didasarkan pada faktor-faktor teoritis dan konseptual, yang kemudian digunakan untuk menentukan jumlah klaster yang optimal.
2. Membangkitkan *k centroid* awal (titik pusat klaster) secara acak. Untuk menentukan *centroid* awal, dilakukan secara acak dari beberapa objek yang

tersedia sebanyak jumlah klaster k . Setelah itu, untuk menghitung *centroid* klaster ke- i berikutnya, digunakan rumus sebagai berikut:

$$v = \frac{\sum_{i=1}^n x_i}{n}, \quad i = 1, 2, 3, \dots, n.$$

Dengan :

v = *centroid* pada *cluster*

x_i = objek ke - i

n = banyaknya jumlah objek yang menjadi anggota *cluster*

3. Menghitung jarak antara setiap objek dengan setiap *centroid* klaster, menggunakan rumus matematika *euclidean distance*:

$$d(x, y) = \|x - y\| = \sqrt{\left(\sum_{i=1}^n (x_i - y_i)\right)^2}, \quad i = 1, 2, 3, \dots, n.$$

Dengan :

x_i = objek x ke - i

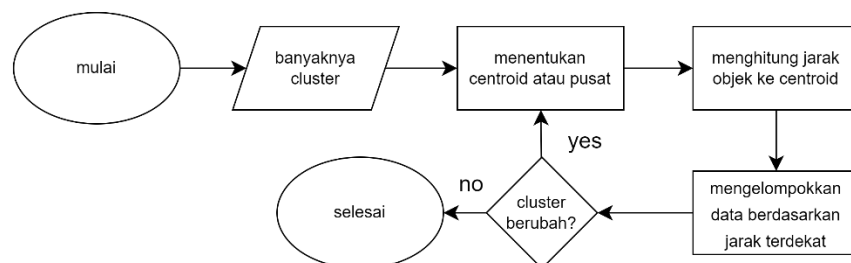
y_i = objek y ke - i

n = banyaknya objek

4. Mengelompokkan setiap objek ke dalam klaster dengan *centroid* terdekat. Setiap objek akan ditempatkan ke klaster yang memiliki *centroid* paling dekat dengannya.
5. Melakukan iterasi ke- i , di mana posisi *centroid* baru ditentukan dengan menggunakan persamaan yang telah disebutkan sebelumnya. Ini akan memperbarui posisi *centroid* dalam setiap iterasi.
6. Mengulangi langkah 3, jika posisi *centroid* baru tidak sama dengan posisi *centroid* sebelumnya. Iterasi akan berhenti jika perbedaan antara posisi *centroid* baru dan posisi *centroid* sebelumnya tidak melebihi nilai ambang yang telah ditentukan sebelumnya, menunjukkan konvergensi dalam penghitungan di setiap data.

(H. Priyatman, 2019)

Algoritma *K-means clustering* dapat dituliskan ke dalam *flowchart* sebagai berikut:



Gambar 1. *Flowchart* Algoritma *K-Means Clustering*.

2.5. C++

Pada sekitar tahun 1970-an, *Bjarne Stroustrup* mengembangkan bahasa pemrograman C++, yang merupakan turunan dari bahasa C, di laboratorium Bell, AT&T. Nama "C++" diberikan oleh *Rick Mascitti* pada tahun 1983, dengan tanda "++" merujuk pada operator penaikan dalam bahasa C. Bahasa C++ telah mendukung paradigma pemrograman berorientasi objek. Menurut *Greg Perry* pada tahun 1993, C++ dapat meningkatkan produktivitas pemrograman lebih dari dua kali lipat dibandingkan dengan bahasa pemrograman prosedural seperti C, Pascal, dan Basic karena memungkinkan penggunaan ulang kode program. Kode program C++ biasanya ditulis menggunakan teks editor dengan format *file* ".cpp". Untuk menjalankan program di komputer, kode program perlu dikompilasi menggunakan kompilator C++ (Jepriana, 2020).

BAB III

METODE PENELITIAN

Langkah-langkah yang digunakan dalam pengelompokan data kemiskinan di Provinsi Papua menggunakan *K-means clustering* adalah sebagai berikut:

1. Mencari data mengenai kemiskinan di Provinsi Papua.

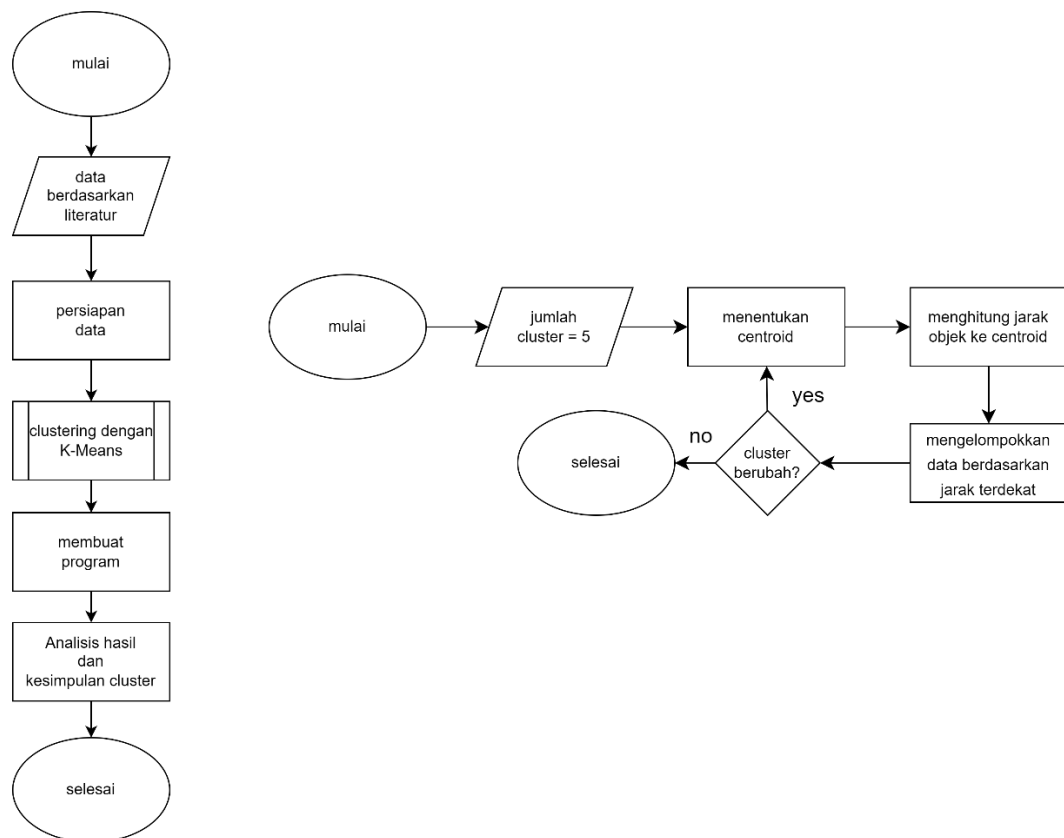
Diambil dari *travel.detik.com*, provinsi Papua menjadi objek pilihan karena menurut data terbaru dari BPS menunjukkan bahwa Papua menduduki peringkat pertama sebagai provinsi termiskin di Indonesia. Tercatat, angka kemiskinan di Papua mencapai 26,80%. Berdasarkan *study literatur* pada Bab 2, penentuan variabel kriteria penilaian terdiri dari tingkat pendidikan, pengeluaran per kapita, dan jumlah penduduk miskin yang masing-masing diambil 5 tahun (2017-2021) tiap kota dari *website* Badan Pusat Statistik.

2. Melakukan tahap *clustering*.

Langkah-langkah pada algoritma *K-Means clustering* adalah sebagai berikut:

- a. Menentukan jumlah k sebagai *cluster* yang akan dibentuk yaitu $k = 5$, di mana *cluster* terdiri dari sangat tinggi, tinggi, sedang, rendah, dan sangat rendah.
 - b. Bangkitkan k *Centroid* awal (titik pusat *cluster*) secara *random*.
 - c. Menghitung *euclidean distance* dari setiap data terhadap *centroid* dengan rumus berikut.
 - d. Menentukan jarak terpendek pada setiap data.
 - e. Mengelompokkan data berdasarkan jarak terpendek ke pusat *cluster*.
 - f. Mengulangi langkah (b) dengan rumus
 - g. Mengulangi langkah (c) hingga (e). Apabila *centroid* berubah, maka kembali ke langkah (f). Apabila tidak, maka iterasi berhenti.
 - h. Menentukan hasil akhir *clustering*.
 - i. Selesai.
3. Membuat program *K-Means clustering* untuk mengelompokkan data kemiskinan di Provinsi Papua.
 4. Mengimplementasikan contoh kasus pada program.

Berikut adalah *flowchart* pada metode penelitian kali ini:



Gambar 2. *Flowchart* Metode Penelitian.

BAB IV

PEMBAHASAN

Dalam bab ini, akan dijelaskan tentang penerapan *K-Means clustering* untuk mengelompokkan data kemiskinan di Provinsi Papua.

4.1. Algoritma *K-Means Clustering*.

Proses pengelompokan banyaknya data kemiskinan Provinsi Papua dengan menggunakan *K-Means Clustering* adalah sebagai berikut :

1. Menentukan banyak *cluster*.
2. Menentukan titik pusat (*Centroid*).
3. Menentukan jarak (*Euclidean Distance*) antara setiap objek ke titik pusat dengan menggunakan rumus berikut:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$$

Di mana $d(x, y)$ merupakan *euclidean distance*, x_n merupakan koordinat objek, dan y_m merupakan koordinat *centroid*.

4. Menentukan kelompok objek berdasarkan jarak minimum. Jika *cluster* saat ini sama dengan *cluster* sebelumnya proses berhenti. Apabila sebaliknya maka kembali ke Langkah 2.

4.2. Data.

Data yang digunakan pada penelitian ini adalah data rata-rata lama sekolah, data pengeluaran per kapita riil, dan data jumlah penduduk miskin per kabupaten/kota di Provinsi Papua yang diambil dari *website bps.go.id*. Data lengkap dapat dilihat pada lampiran 1. Data lengkap dapat dilihat pada **Lampiran 1**.

No.	Kab/ Kota	Rata-rata Lama Sekolah (Tahun)					Pengeluaran Per Kapita Riil Disesuaikan (Ribu Rupiah)					Jumlah Penduduk Miskin menurut Kabupaten/Kota di Papua (Ribu Jiwa)				
		17 7	18 8	19 9	20 0	21 1	17	18	19	20	21	17 7	18 8	19 9	20 0	21 1
1.	Merauke	8,27	8,49	8,56	8,72	8,73	10277,00	10430,00	10498,00	10097,00	10201,00	10,81	10,54	10,35	10,03	10,06
2.	Jayawijaya	4,99	5,17	5,53	5,51	5,56	7524,00	7637,00	7835,00	7441,00	7545,00	38,26	38,66	38,33	37,22	37,92
3.	Jayapura	9,54	9,96	9,79	9,04	9,05	10055,00	10160,00	10375,00	9898,00	9989,00	13,01	13,44	13,13	12,44	12,43
..
29	Kota Jayapura	11,1	1,3	11,1	11,1	11,1	14781,00	14922,00	15176,00	14763,00	14937,00	11,1	11,1	11,1	11,1	11,1

		1		5	5	5						4	3	4	1	3
		5		5	6	7						6	7	9	6	9

Tabel 1. Data rata-rata lama sekolah, data pengeluaran Per kapita riil, dan data jumlah penduduk miskin kabupaten/kota di Provinsi Papua.

No.	Kab/Kota	Rata-rata Lama Sekolah (Tahun)	Pengeluaran Per Kapita Riil Disesuaikan (Ribu Rupiah)	Jumlah Penduduk Miskin menurut Kabupaten/Kota di Papua (Ribu Jiwa)
1.	Merauke	9	10301	10
2.	Jayawijaya	5	7596	38
3.	Jayapura	10	10095	13
...
29.	Kota Jayapura	11	14916	11

Tabel 2. Data rata-rata lama sekolah, data rata-rata pengeluaran Per kapita riil, dan rata-rata data jumlah penduduk miskin kabupaten/kota di Provinsi Papua pada 2017-2021.

4.3. Penyelesaian Secara Manual Pengelompokan Data Kemiskinan Provinsi Papua.

Penyelesaian secara manual pengelompokan data kemiskinan menggunakan algoritma *K-Means Clustering* sebagai berikut :

- a. Langkah 1 : Menentukan banyak *cluster*.

Pada permasalahan ini akan dibagi menjadi lima *cluster*, yaitu C_1 (*cluster* dengan tingkat kemiskinan sangat tinggi), C_2 (*cluster* dengan tingkat kemiskinan tinggi), C_3 (*cluster* dengan tingkat kemiskinan sedang), C_4 (*cluster* dengan tingkat kemiskinan rendah), dan C_5 (*cluster* dengan tingkat kemiskinan sangat rendah).

- b. Langkah 2 : Menentukan nilai tengah/*centroid*.

Pada permasalahan ini terdapat 5 nilai tengah/*centroid*. Penentuan titik *cluster* dilakukan dengan mengambil nilai minimum untuk *cluster* tingkat kemiskinan sangat rendah dan tingkat kemiskinan rendah (C_1 dan C_2), nilai rata-rata untuk *cluster* dengan tingkat kemiskinan sedang (C_3), dan nilai maksimum untuk *cluster* untuk tingkat kemiskinan tinggi dan tingkat kemiskinan sangat tinggi (C_4 dan C_5),

	A1	A2	A3
C1	1	4047	10
C2	3	4916	25
C3	6,10345	6900,79	28,8621
C4	9	8704	37
C5	11	14916	43

Tabel 3. Nilai *Centroid* Awal.

Keterangan :

A1 : Rata-rata Lama Sekolah (Tahun).

A2 : Rata-rata Pengeluaran Per Kapita Riil Disesuaikan (Ribuan Rupiah).

A3 : Rata-rata Jumlah Penduduk Miskin menurut Kabupaten/Kota di Papua (Ribuan Jiwa).

- c. Langkah 3 : Menentukan jarak (*Euclidean Distance*).

Dengan menggunakan nilai tengah/*centroid* dari Tabel 4.3 maka data pada **Tabel 2**. dapat dibagi menjadi 5 *cluster*. Untuk menentukan jarak/*euclidean distance* dari setiap data maka digunakan rumus berikut.

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_m - y_m)^2}$$

Di mana $d(x, y)$ merupakan *euclidean distance*, x_m merupakan koordinat objek, dan y_m merupakan koordinat *centroid*. Sehingga diperoleh hasil pada **Tabel 4**. adalah sebagai berikut. Data lengkap terdapat pada **Lampiran 1**.

No.	Kab/Kota	Rata-rata Lama Sekolah (Tahun)	Pengeluaran Per Kapita Riil Disesuaikan (Ribuan Rupiah)	Jumlah Penduduk Miskin menurut Kabupaten/Kota di Papua (Ribuan Jiwa)	C1	C2	C3	C4	C5	Min
1.	Merauke	6253,605	5384,623	3399,861	2226,759	4615,516	2226,759	6253,605	5384,623	3399,861
2.	Jayawijaya	3549,513	2680,432	695,6703	477,6152	7319,604	477,6152	3549,513	2680,432	695,6703
3.	Jayapura	6048,407	5179,419	3194,652	2021,545	4820,695	2021,545	6048,407	5179,419	3194,652
..
29.	Kota Jayapura	10868,81	9999,813	8015,031	6211,853	31,6295	31,6295	10868,81	9999,813	8015,031

Tabel 4. Perhitungan *Euclidean Distance*.

Setelah melalui beberapa iterasi hingga iterasi 5, diperoleh hasil pengelompokan data kemiskinan menjadi *cluster* sebagai berikut :

No.	Kab/Kota	C1	C2	C3	C4	C5
1.	Merauke	0	0	0	1	0

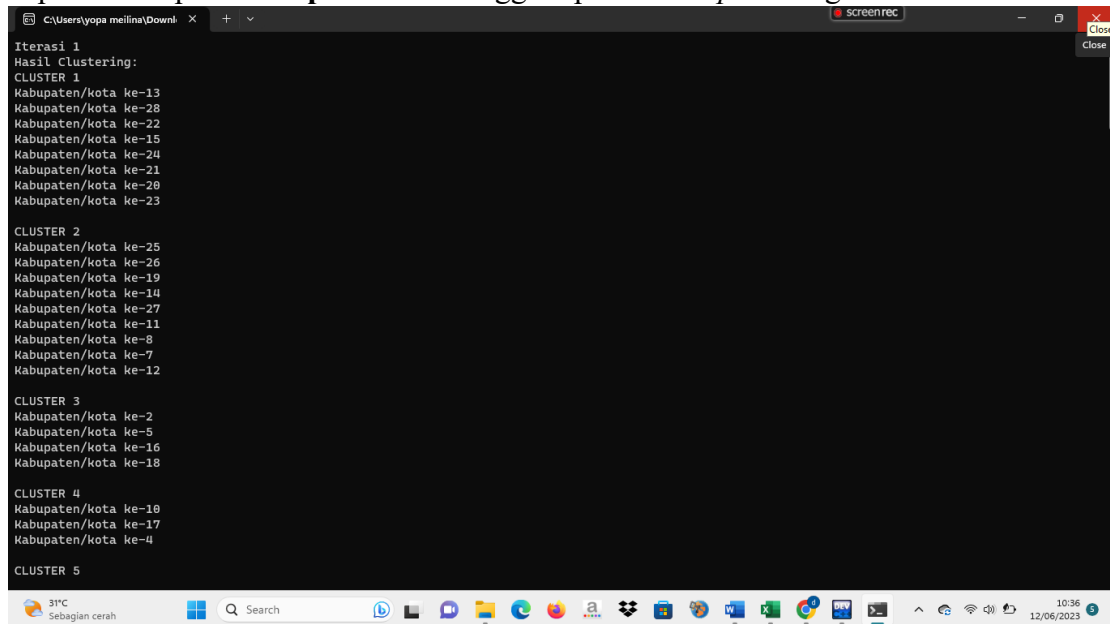
2.	Jayawijaya	0	0	1	0	0
3.	Jayapura	0	0	0	1	0
4.	Nabire	0	0	0	1	0
5.	Kepulauan Yapen	0	0	1	0	0
6.	Biak Numfor	0	0	0	1	0
7.	Paniai	0	0	1	0	0
8.	Puncak Jaya	0	1	0	0	0
9.	Mimika	0	0	0	0	1
10.	Boven Digoel	0	0	1	0	0
11.	Mappi	0	0	1	0	0
12.	Asmat	0	1	0	0	0
13.	Yahukimo	1	0	0	0	0
14.	Pegunungan Bintang	0	1	0	0	0
15.	Tolikara	1	0	0	0	0
16.	Sarmi	0	0	1	0	0
17.	Keerom	0	0	0	1	0
18.	Waropen	0	0	1	0	0
19.	Supiori	0	1	0	0	0
20.	Mamberamo Raya	1	0	0	0	0
21.	Nduga	1	0	0	0	0
22.	Lanny Jaya	1	0	0	0	0
23.	Mamberamo Tengah	1	0	0	0	0
24.	Yalimo	1	0	0	0	0
25.	Puncak	0	1	0	0	0
26.	Dogiyai	0	1	0	0	0
27.	Intan Jaya	0	1	0	0	0
28.	Deiyai	1	0	0	0	0
29.	Kota Jayapura	0	0	0	0	1

Tabel 5. Hasil Pengelompokan.

Dari tabel di atas diperoleh kota/kabupaten yang masuk dalam kategori C_1 (*cluster* dengan tingkat kemiskinan sangat tinggi) adalah Yahukimo, Tolikara, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, dan Deiyai. Kategori C_2 (*cluster* dengan tingkat kemiskinan tinggi) adalah Puncak Jaya, Asmat, Pegunungan Bintang, Supiori, Puncak, Dogiyai, dan Intan Jaya. kategori C_3 (*cluster* dengan tingkat kemiskinan sedang) adalah Jayawijaya, Kepulauan Yapen, Paniai, Boven Digoel, Mappi, Sarmi, dan Waropen. Kategori C_4 (*cluster* dengan tingkat kemiskinan rendah) adalah Merauke, Jayapura, Nabire, Biak Numfor, dan Keerom. kategori C_5 (*cluster* dengan tingkat kemiskinan sangat rendah) adalah Mimika dan Kota Jayapura.

4.4. Penyelesaian Menggunakan Program C++ Pengelompokan Data Kemiskinan Provinsi Papua.

Langkah-langkah penyelesaian pengelompokan data kemiskinan Provinsi Papua menggunakan program C++ yaitu dengan memasukkan data pada lampiran 1 ke dalam *notepad*. Kemudian pada program C++ data tersebut dimasukkan. *Source code* dapat dilihat pada **Lampiran 3**. Sehingga diperoleh *output* sebagai berikut :



```

Iterasi 1
Hasil Clustering:
CLUSTER 1
Kabupaten/kota ke-13
Kabupaten/kota ke-28
Kabupaten/kota ke-22
Kabupaten/kota ke-15
Kabupaten/kota ke-24
Kabupaten/kota ke-21
Kabupaten/kota ke-20
Kabupaten/kota ke-23

CLUSTER 2
Kabupaten/kota ke-25
Kabupaten/kota ke-26
Kabupaten/kota ke-19
Kabupaten/kota ke-14
Kabupaten/kota ke-27
Kabupaten/kota ke-11
Kabupaten/kota ke-8
Kabupaten/kota ke-7
Kabupaten/kota ke-12

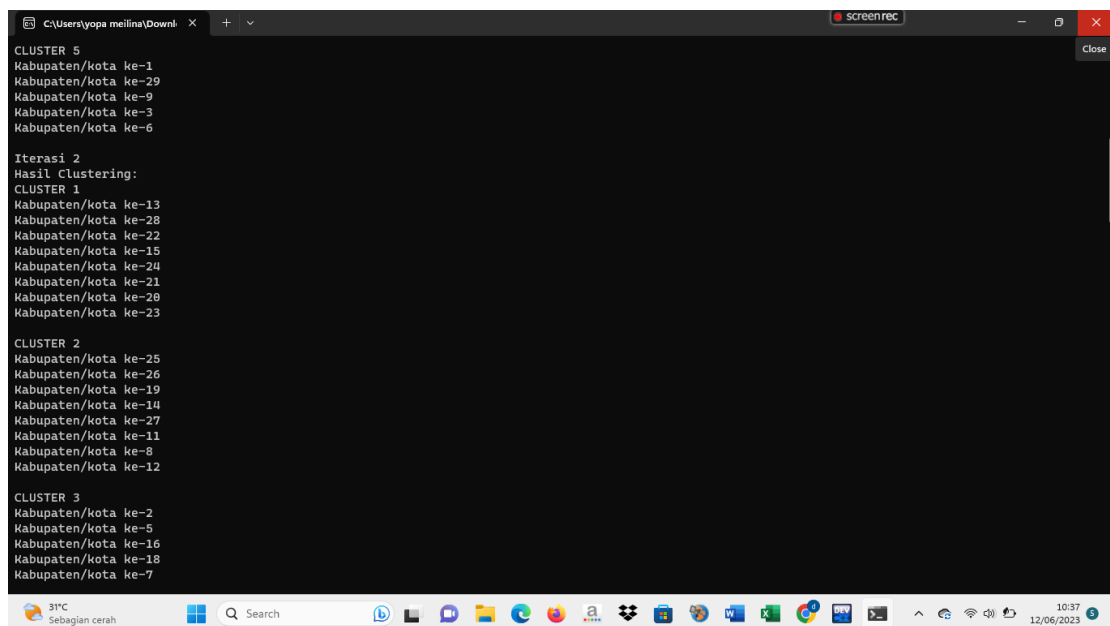
CLUSTER 3
Kabupaten/kota ke-2
Kabupaten/kota ke-5
Kabupaten/kota ke-16
Kabupaten/kota ke-18

CLUSTER 4
Kabupaten/kota ke-10
Kabupaten/kota ke-17
Kabupaten/kota ke-4

CLUSTER 5

```

Gambar 3. *Output* Iterasi 1.



```

CLUSTER 5
Kabupaten/kota ke-1
Kabupaten/kota ke-29
Kabupaten/kota ke-9
Kabupaten/kota ke-3
Kabupaten/kota ke-6

Iterasi 2
Hasil Clustering:
CLUSTER 1
Kabupaten/kota ke-13
Kabupaten/kota ke-28
Kabupaten/kota ke-22
Kabupaten/kota ke-15
Kabupaten/kota ke-24
Kabupaten/kota ke-21
Kabupaten/kota ke-20
Kabupaten/kota ke-23

CLUSTER 2
Kabupaten/kota ke-25
Kabupaten/kota ke-26
Kabupaten/kota ke-19
Kabupaten/kota ke-14
Kabupaten/kota ke-27
Kabupaten/kota ke-11
Kabupaten/kota ke-8
Kabupaten/kota ke-7
Kabupaten/kota ke-12

CLUSTER 3
Kabupaten/kota ke-2
Kabupaten/kota ke-5
Kabupaten/kota ke-16
Kabupaten/kota ke-18
Kabupaten/kota ke-7

```

Gambar 4. *Output* Iterasi 2.

```
C:\Users\yopa mellina\Downl x + v screenrec
CLUSTER 4
Kabupaten/kota ke-10
Kabupaten/kota ke-17
Kabupaten/kota ke-4
Kabupaten/kota ke-6

CLUSTER 5
Kabupaten/kota ke-1
Kabupaten/kota ke-29
Kabupaten/kota ke-9
Kabupaten/kota ke-3

Iterasi 3
Hasil Clustering:
CLUSTER 1
Kabupaten/kota ke-13
Kabupaten/kota ke-28
Kabupaten/kota ke-22
Kabupaten/kota ke-15
Kabupaten/kota ke-24
Kabupaten/kota ke-21
Kabupaten/kota ke-20
Kabupaten/kota ke-23

CLUSTER 2
Kabupaten/kota ke-25
Kabupaten/kota ke-26
Kabupaten/kota ke-19
Kabupaten/kota ke-14
Kabupaten/kota ke-27
Kabupaten/kota ke-11
Kabupaten/kota ke-8
Kabupaten/kota ke-12

CLUSTER 3
```

Gambar 5. *Output* Iterasi 3.

```
C:\Users\yopa mellina\Downl x + v screenrec
CLUSTER 3
Kabupaten/kota ke-2
Kabupaten/kota ke-5
Kabupaten/kota ke-16
Kabupaten/kota ke-18
Kabupaten/kota ke-7

CLUSTER 4
Kabupaten/kota ke-10
Kabupaten/kota ke-1
Kabupaten/kota ke-17
Kabupaten/kota ke-4
Kabupaten/kota ke-3
Kabupaten/kota ke-6

CLUSTER 5
Kabupaten/kota ke-29
Kabupaten/kota ke-9

Iterasi 4
Hasil Clustering:
CLUSTER 1
Kabupaten/kota ke-13
Kabupaten/kota ke-28
Kabupaten/kota ke-22
Kabupaten/kota ke-15
Kabupaten/kota ke-24
Kabupaten/kota ke-21
Kabupaten/kota ke-20
Kabupaten/kota ke-23

CLUSTER 2
Kabupaten/kota ke-25
Kabupaten/kota ke-26
Kabupaten/kota ke-19
```

Gambar 6. *Output* Iterasi 4.

```
C:\Users\yopa mellina\Downl x + v screenrec
Kabupaten/kota ke-19
Kabupaten/kota ke-14
Kabupaten/kota ke-27
Kabupaten/kota ke-11
Kabupaten/kota ke-8
Kabupaten/kota ke-12

CLUSTER 3
Kabupaten/kota ke-2
Kabupaten/kota ke-10
Kabupaten/kota ke-5
Kabupaten/kota ke-16
Kabupaten/kota ke-18
Kabupaten/kota ke-7

CLUSTER 4
Kabupaten/kota ke-1
Kabupaten/kota ke-17
Kabupaten/kota ke-4
Kabupaten/kota ke-3
Kabupaten/kota ke-6

CLUSTER 5
Kabupaten/kota ke-29
Kabupaten/kota ke-9

Iterasi 5
Hasil Clustering:
CLUSTER 1
Kabupaten/kota ke-13
Kabupaten/kota ke-28
Kabupaten/kota ke-22
Kabupaten/kota ke-15
Kabupaten/kota ke-24
Kabupaten/kota ke-21
```

Gambar 7. *Output* Iterasi 5.

```
C:\Users\yopa mellina\Downl x + v screenrec
Kabupaten/kota ke-21
Kabupaten/kota ke-20
Kabupaten/kota ke-23

CLUSTER 2
Kabupaten/kota ke-25
Kabupaten/kota ke-26
Kabupaten/kota ke-19
Kabupaten/kota ke-14
Kabupaten/kota ke-27
Kabupaten/kota ke-11
Kabupaten/kota ke-8
Kabupaten/kota ke-12

CLUSTER 3
Kabupaten/kota ke-2
Kabupaten/kota ke-10
Kabupaten/kota ke-5
Kabupaten/kota ke-16
Kabupaten/kota ke-18
Kabupaten/kota ke-7

CLUSTER 4
Kabupaten/kota ke-1
Kabupaten/kota ke-17
Kabupaten/kota ke-4
Kabupaten/kota ke-3
Kabupaten/kota ke-6

CLUSTER 5
Kabupaten/kota ke-29
Kabupaten/kota ke-9

-----
```

Gambar 8. *Output* Lanjutan Iterasi 5.

Dari *output* di atas diperoleh hasil yang sama pada perhitungan manualnya, yakni kota/kabupaten yang masuk dalam kategori C_1 (*cluster* dengan tingkat kemiskinan sangat tinggi) adalah Yahukimo, Tolikara, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, dan Deiyai. Kategori C_2 (*cluster* dengan tingkat kemiskinan tinggi) adalah Puncak Jaya, Asmat, Pegunungan Bintang, Supiori, Puncak, Dogiyai, dan Intan Jaya. kategori C_3 (*cluster* dengan tingkat kemiskinan sedang) adalah Jayawijaya, Kepulauan Yapen, Paniai, Boven Digoel, Mappi, Sarmi, dan Waropen. Kategori C_4 (*cluster* dengan tingkat kemiskinan rendah) adalah Merauke, Jayapura, Nabire, Biak Numfor, dan Keerom. kategori C_5 (*cluster* dengan tingkat kemiskinan sangat rendah) adalah Mimika dan Kota Jayapura.

4.5. Analisis Hasil *Clustering*.

Setelah melakukan perhitungan, diperoleh pengelompokan sebagai berikut yakni untuk kategori C_1 (*cluster* dengan tingkat kemiskinan sangat tinggi) adalah Yahukimo, Tolikara, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, dan Deiyai. Kategori C_2 (*cluster* dengan tingkat kemiskinan tinggi) adalah Puncak Jaya, Asmat, Pegunungan Bintang, Supiori, Puncak, Dogiyai, dan Intan Jaya. kategori C_3 (*cluster* dengan tingkat kemiskinan sedang) adalah Jayawijaya, Kepulauan Yapen, Paniai, Boven Digoel, Mappi, Sarmi, dan Waropen. Kategori C_4 (*cluster* dengan tingkat kemiskinan rendah) adalah Merauke, Jayapura, Nabire, Biak Numfor, dan Keerom. kategori C_5 (*cluster* dengan tingkat kemiskinan sangat rendah) adalah Mimika dan Kota Jayapura.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian kami , maka didapatkan kesimpulan sebagai berikut :

1. Pengelompokan data kemiskinan Provinsi Papua diperoleh 5 *cluster*, yakni C_1 (*cluster* dengan tingkat kemiskinan sangat tinggi, C_2 (*cluster* dengan tingkat kemiskinan tinggi), C_3 (*cluster* dengan tingkat kemiskinan sedang), C_4 (*cluster* dengan tingkat kemiskinan rendah, dan C_5 (*cluster* dengan tingkat kemiskinan sangat rendah).
2. Implementasi program untuk menyelesaikan pengelompokan banyaknya data kemiskinan di provinsi Jawa Timur menggunakan K-means clustering dapat dibuat dengan mengubah bahasa *pseudocode* setiap prosedur menjadi bahasa pemrograman C++ dengan bantuan *software* Dev-C++. Diperoleh hasil pengelompokan dari implementasi program *K-means clustering* sebagai berikut:
 - *Cluster* 1 (Tingkat Kemiskinan Sangat Tinggi) :
Yahukimo, Tolikara, Mamberamo Raya, Nduga, Lanny Jaya, Mamberamo Tengah, Yalimo, dan Deiyai.
 - *Cluster* 2 (Tingkat Kemiskinan Tinggi) :
Puncak Jaya, Asmat, Pegunungan Bintang, Supiori, Puncak, Dogiyai, dan Intan Jaya.
 - *Cluster* 3 (Tingkat Kemiskinan Sedang) :
Jayawijaya, Kepulauan Yapen, Paniai, Boven Digoel, Mappi, Sarmi, dan Waropen.
 - *Cluster* 4 (Tingkat Kemiskinan Rendah) :
Merauke, Jayapura, Nabire, Biak Numfor, dan Keerom.
 - *Cluster* 5 (Tingkat Kemiskinan Sangat Rendah) :
Mimika dan Kota Jayapura.

5.2. Saran

Saran dari penulis yakni pemerintah diharapkan agar lebih dapat memperhatikan daerah-daerah dengan tingkat kemiskinan sedang hingga sangat tinggi terutama pada Provinsi Papua mengingat provinsi tersebut merupakan provinsi dengan tingkat kemiskinan sangat tinggi berdasarkan data. Untuk penelitian selanjutnya disarankan untuk menggunakan algoritma lain sebagai pertimbangan hasil algoritma saat ini.

DAFTAR PUSTAKA

- Berry, M. (2004). *Data Mining Techniques*. John Wiley & Sons.
- Fikriansyah, Ilham. 2023. Daftar Terbaru Provinsi Termiskin di Indonesia, Ada di Mana Saja?. Diakses pada 15 Mei 2023 dari <https://travel.detik.com/domestic-destination/d-6646969/daftar-terbaru-provinsi-termiskin-di-indonesia-ada-di-mana-saja#:~:text=Sementara%20itu%2C%20data%20terbaru%20dari,kemiskinan%20mencaipai%2021%2C43%25>.
- H. Priyatman, F. S. (2019). Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi dan Ilmu Komputer*, vol 5, 1.
- Hidayat, Feriawan. 2016. Soal Kemampuan Berbahasa Inggris, Indonesia Dinilai Masih Tertinggal. Diakses pada 2 Maret 2018 dari <http://www.beritasatu.com/pendidikan/403858>.
- Irmanita Nasution, A. P. (2020). Penerapan Algoritma K-Means dalam Pengelompokan Data Penduduk Miskin Menurut Provinsi. *Building of Informatics, Technology and Science (BITS)*, 76-83.
- Jumlah Penduduk Miskin Naik TIPIS. (2023). Diakses pada 15 Mei 2023 dari <https://indonesiabaik.id/infografis/jumlah-penduduk-miskin-naik-tipis#:~:text=0%20Komentar&text=Indonesiabaik.id%20%2D%20Badan%20Pusat%20Statistik,mencapai%2026%2C36%20juta%20orang>
- Jepriana, S. H. (2020). *Konsep Algoritme dan Aplikasinya dalam Bahasa Pemrograman C++*. Yogyakarta: Penerbit ANDI.
- Kadji, Y. (2019). Kemiskinan dan Konsep Teoritisnya. 1-7.
- Mardi, Y. (t.thn.). Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*, 213.
- Nurwati, N. (2018). Kemiskinan : Model Pengukuran , Permasalahan dan Alternatif Kebijakan. 1-11.
- P. Sari, B. P. (2017). Improve K-Means terhadap Status Nilai Gizi pada Balita. *SemanTIK*, vol 3, no 1, 143-148.
- Praja, B. S. (2019). Penerapan Metode K-Means Clustering dalam Pengelompokan Data Penumpang dan Kapal Angkutan Laut di Indonesia. *E-Proceeding of engineering*. Vol.6.
- Pratama, Y. C. (2013). Analisis Faktor-faktor yang Mempengaruhi Kemiskinan di Indonesia. *Jurnal Bisnis dan Manajemen*, 210.
- Tan, P.-N. e. (2006). *Introduction to data mining*. Boston: Pearson Addison Wesley.
- Thanamani, A. S. (2013). Multidimensional Clustering Methods of Data Mining for Industrial Applications. *International Journal of Engineerig Science Invention*: 2, 01-08.

Wanto, A. (2018). Optimasi Prediksi Dengan Algoritma Backpropagation Dan Conjugate Gradient Beale-Powell Restarts. *Jurnal Teknologi dan Sistem Informasi*, vol 3, no 3, 370-380.

LAMPIRAN

Lampiran 1. Data Lengkap dan Perhitungan manual pengelompokan data kemiskinan Provinsi Papua untuk mencari *euclidean distance*.

https://docs.google.com/spreadsheets/d/1Q1hO1kPXXUcgkAd8SjrpTIweO_34kDsh/edit?usp=sharing&ouid=114299096502704238071&rtpof=true&sd=true

Lampiran 2. *Source Code C++*.

```
#include <iostream>

#include <vector>

#include <cmath>

#include <limits>

#include <algorithm>

using namespace std;

struct DataPoint {
    int kabupaten;
    double x;
    int y;
    int z;
};

void printClusters(const vector<DataPoint>& data, const vector<vector<DataPoint>>&
clusters) {
    cout << "Hasil Clustering:" << endl;

    // Menampilkan Cluster 1
    cout << "CLUSTER 1" << endl;
    for (const auto& point : clusters[0]) {
        cout << "Kabupaten/kota ke-" << point.kabupaten << endl;
    }
    cout << endl;
```

```

// Menampilkan Cluster 2
cout << "CLUSTER 2" << endl;
for (const auto& point : clusters[3]) {
    cout << "Kabupaten/kota ke-" << point.kabupaten << endl;
}
cout << endl;

// Menampilkan Cluster 3
cout << "CLUSTER 3" << endl;
for (const auto& point : clusters[1]) {
    cout << "Kabupaten/kota ke-" << point.kabupaten << endl;
}
cout << endl;

// Menampilkan Cluster 4
cout << "CLUSTER 4" << endl;
for (const auto& point : clusters[2]) {
    cout << "Kabupaten/kota ke-" << point.kabupaten << endl;
}
cout << endl;

// Menampilkan Cluster 5
cout << "CLUSTER 5" << endl;
for (const auto& point : clusters[4]) {
    cout << "Kabupaten/kota ke-" << point.kabupaten << endl;
}
cout << endl;
}

double calculateDistance(const DataPoint& point1, const DataPoint& point2) {

```

```

double distX = point1.x - point2.x;
int distY = point1.y - point2.y;
int distZ = point1.z - point2.z;
return sqrt(distX * distX + distY * distY + distZ * distZ);
}

vector<DataPoint> calculateCentroids(const vector<vector<DataPoint>>& clusters) {
    vector<DataPoint> centroids;
    for (const auto& cluster : clusters) {
        double sumX = 0.0;
        int sumY = 0;
        int sumZ = 0;
        for (const auto& point : cluster) {
            sumX += point.x;
            sumY += point.y;
            sumZ += point.z;
        }
        double centroidX = sumX / cluster.size();
        int centroidY = sumY / cluster.size();
        int centroidZ = sumZ / cluster.size();
        centroids.push_back({0, centroidX, centroidY, centroidZ});
    }
    return centroids;
}

int findNearestCluster(const DataPoint& point, const vector<DataPoint>& centroids) {
    int nearestCluster = 0;
    double minDistance = numeric_limits<double>::max();
    for (int i = 0; i < centroids.size(); i++) {
        double distance = calculateDistance(point, centroids[i]);
    }
}

```

```

        if (distance < minDistance) {
            minDistance = distance;
            nearestCluster = i;
        }
    }
    return nearestCluster;
}

void kMeansClustering(vector<DataPoint>& data, int k) {
    vector<DataPoint> centroids(k);
    vector<vector<DataPoint>> clusters(k);

    // Inisialisasi posisi awal centroid secara acak
    random_shuffle(data.begin(), data.end());
    for (int i = 0; i < k; i++) {
        centroids[i] = data[i];
    }

    bool converged = false;
    int iteration = 0;

    while (!converged) {
        // Mengosongkan kluster
        for (int i = 0; i < k; i++) {
            clusters[i].clear();
        }

        // Menentukan kluster tiap data
        for (const auto& point : data) {
            int nearestCluster = findNearestCluster(point, centroids);

```

```

        clusters[nearestCluster].push_back(point);
    }

    // Menghitung posisi centroid baru
    vector<DataPoint> newCentroids = calculateCentroids(clusters);

    // Cek konvergensi
    converged = true;
    for (int i = 0; i < k; i++) {
        if (calculateDistance(centroids[i], newCentroids[i]) > 1e-10) {
            converged = false;
            break;
        }
    }

    // Memperbarui posisi centroid
    centroids = newCentroids;

    // Menampilkan hasil setiap iterasi
    cout << "Iterasi " << iteration + 1 << endl;
    printClusters(data, clusters);

    iteration++;
}

int main() {
    vector<DataPoint> data = {
        {1, 9, 10301, 10},
        {2, 5, 7596, 38},

```

```

{3, 10, 10095, 13},
{4, 10, 9003, 25},
{5, 9, 7621, 27},
{6, 10, 9861, 25},
{7, 4, 6479, 37},
{8, 4, 5379, 36},
{9, 10, 11631, 14},
{10, 9, 8074, 20},
{11, 6, 6321, 26},
{12, 5, 5838, 26},
{13, 4, 4818, 38},
{14, 3, 5511, 30},
{15, 4, 4916, 33},
{16, 9, 6723, 14},
{17, 8, 8943, 17},
{18, 9, 6865, 30},
{19, 9, 5726, 38},
{20, 6, 4668, 29},
{21, 1, 4047, 38},
{22, 3, 4437, 39},
{23, 3, 4548, 37},
{24, 3, 4734, 34},
{25, 2, 5482, 37},
{26, 5, 5479, 30},
{27, 3, 5387, 42},
{28, 3, 4724, 43},
{29, 11, 14916, 11}
};

```

```

int k = 5; // Jumlah cluster

```

```
kMeansClustering(data, k);  
  
return 0;  
}
```