

**KOMPARASI KINERJA ALGORITMA *K-MEANS* DAN *K-MEDOIDS*
DALAM PENGELOMPOKAN KABUPATEN/KOTA DI SUMATERA
BERDASARKAN INDIKATOR KEMISKINAN**

SKRIPSI



Oleh
MARDHIATUL AZMI
NIM. 18337054

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2023**

**KOMPARASI KINERJA ALGORITMA *K-MEANS* DAN *K-MEDOIDS*
DALAM PENGELOMPOKAN KABUPATEN/KOTA DI SUMATERA
BERDASARKAN INDIKATOR KEMISKINAN**

SKRIPSI

*Diajukan sebagai salah satu persyaratan guna memperoleh gelar Sarjana
Statistika*



Oleh
MARDHIATUL AZMI
NIM. 18337054

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2023**

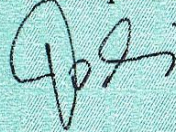
PERSETUJUAN SKRIPSI

KOMPARASI KINERJA ALGORITMA K-MEANS DAN K-MEDOIDS DALAM PENGELOMPOKAN KABUPATEN/KOTA DI SUMATERA BERDASARKAN INDIKATOR KEMISKINAN

Nama : Mardhiatul Azmi
NIM : 18337054
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Padang, 01 Februari 2023

Mengetahui:
Ketua Departemen Statistika



Dodi Vionanda, M.Si., Ph.D
NIP. 197906112005011002

Disetujui Oleh:
Pembimbing



Drs. Atus Amadi Putra, M.Si
NIP. 196308291992031001




PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Mardhiatul Azmi
NIM : 18337054
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

KOMPARASI KINERJA ALGORITMA K-MEANS DAN K-MEDOIDS DALAM PENGELOMPOKAN KABUPATEN/KOTA DI SUMATERA BERDASARKAN INDIKATOR KEMISKINAN

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, 01 Februari 2023

Tim Penguji	Nama	Tanda Tangan
Ketua	: Drs. Atus Amadi Putra, M.Si	
Anggota	: Dodi Vionanda, M.Si., Ph.D	
Anggota	: Admi Salma S.Pd., M.Si	

SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertandatangan di bawah ini:

Nama : Mardhiatul Azmi
NIM : 18337054
Program Studi : S1 Statistika
Departemen : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul **“Komparasi Kinerja Algoritma K-Means dan K-Medoids dalam Pengelompokan Kabupaten/Kota di Sumatera Berdasarkan Indikator Kemiskinan”** adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan.

Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,
Ketua Departemen Statistika,



Dodi Vionanda, M.Si., Ph.D
NIP. 197906112005011002

Saya yang menyatakan,



Mardhiatul Azmi
NIM. 18337054

Komparasi Kinerja Algoritma *K-Means* dan *K-Medoids* dalam Pengelompokan Kabupaten/Kota di Sumatera Berdasarkan Indikator Kemiskinan

Mardhiatul Azmi

ABSTRAK

K-Means adalah metode non hirarki yang membagi data menjadi beberapa kelompok berdasarkan jarak terdekat antara objek dengan *centroid* data. *K-Medoids* adalah metode non hirarki *clustering* yang membagi data menjadi beberapa kelompok berdasarkan jarak terdekat antara objek dengan *medoid* data. Untuk mengetahui metode terbaik diantara kedua metode tersebut, maka dilakukan pengelompokan pada data kemiskinan di Sumatera tahun 2021, dimana pada saat itu angka kemiskinan naik dibandingkan tahun sebelumnya dikarenakan masih terjadinya pandemi Covid-19. Kemiskinan merupakan ketidakmampuan seseorang dalam memenuhi kebutuhan hidupnya, baik itu berupa makanan maupun bukan makanan. Pengelompokan dilakukan agar dapat mengetahui karakteristik dari masing-masing kabupaten/kota, yang mana nantinya diharapkan dapat menjadi acuan bagi pemerintah dalam pengambilan keputusan.

Penelitian ini merupakan penelitian terapan yang dimulai dengan mempelajari teori-teori relevan. Data yang digunakan pada penelitian ini yaitu data sekunder yang bersumber dari website BPS mengenai indikator kemiskinan. Penelitian ini bertujuan untuk mengetahui kelompok daerah berdasarkan faktor-faktor yang memengaruhi kemiskinan di Sumatera, serta membandingkan hasil pengelompokan yang telah diperoleh. Untuk mengetahui kinerja terbaik diantara kedua metode tersebut yaitu dengan melihat nilai *Davies Bouldin Index* (DBI) yang terendah.

Hasil dari penelitian ini adalah pada algoritma *k-means* menghasilkan sebanyak 34 kabupaten/kota tergabung pada *cluster* 1, 52 kabupaten/kota pada *cluster* 2, 23 kabupaten/kota pada *cluster* 3, dan 45 kabupaten/kota pada *cluster* 4. Selanjutnya hasil pengelompokan pada *k-medoids* yaitu pada *cluster* 1, 2, 3, dan 4 berturut-turut yaitu sebanyak 53, 40, 37, dan 24 kabupaten/kota. Berdasarkan hasil pengelompokan tersebut, diperoleh DBI *k-means* sebesar 1,584 dan *k-medoids* sebesar 2,359. Hal ini berarti bahwa algoritma *k-means* lebih baik daripada *k-medoids*, karena DBI *k-means* lebih kecil daripada *k-medoids*.

Kata Kunci : *Davies Bouldin Index*, Kemiskinan, *K-Means*, *K-Medoids*

Comparison Performance of K-Means and K-Medoids Algorithms in Grouping Regencies/Cities in Sumatera Based on Poverty Indicators

Mardhiatul Azmi

ABSTRACT

K-Means is a non-hierarchical method that divides data into groups based on the closest distance between objects and the data centroid. K-Medoids is a non-hierarchical clustering method that divides data into groups based on the closest distance between objects and data medoids. To find out the best method between the two methods, clustering was carried out on poverty data in Sumatera in 2021, at which time the poverty rate increased compared to the previous year due to the Covid-19 pandemic. Poverty is the inability of a person to fulfill their needs, both in the form of food and non-food. Grouping is done in order to find out the characteristics of each regencies/cities, which later is expected to be a reference for the government in making decisions.

This research is an applied research that begins with studying relevant theories. The data used in this research is secondary data sourced from the BPS website regarding poverty indicators. This research aims to find out the regional groups based on the factors that affect poverty in Sumatera, as well as compare the grouping results that have been obtained. To find out the best performance between the two methods is by looking at the lowest Davies Bouldin Index (DBI).

The result of this research is that the k-means algorithm produces 34 regencies/cities in cluster 1, 52 regencies/cities in cluster 2, 23 regencies/cities in cluster 3, and 45 regencies/cities in cluster 4. Furthermore, the clustering results in k-medoids are in clusters 1, 2, 3, and 4, namely 53, 40, 37, and 24 regencies/cities respectively. Based on the clustering results, the DBI of k-means is 1,584 and k-medoids is 2,359. This means that the k-means algorithm is better than k-medoids, because the DBI of k-means is smaller than k-medoids.

Keywords : Davies Bouldin Index, K-Means, K-Medoids, Poverty

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Alhamdulillahirabbil'alamin. Puji syukur kehadiran Allah SWT, atas berkat dan rahmat-Nya sehingga skripsi yang berjudul “Komparasi Kinerja Algoritma K-Means dan K-Medoids dalam Pengelompokan Kabupaten/Kota di Sumatera Berdasarkan Indikator Kemiskinan” ini dapat terselesaikan. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana, Program Studi Sarjana Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang. Penulis menyadari bahwa keberhasilan penyusunan skripsi ini tidak terlepas dari bantuan berbagai pihak, baik langsung maupun tidak langsung. Dalam kesempatan ini, penulis menyampaikan rasa terima kasih yang sebesar-besarnya kepada :

1. Bapak Dodi Vionanda, M.Si., Ph.D., selaku Ketua Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
2. Bapak Drs. Atus Amadi Putra, M.Si., selaku dosen pembimbing yang telah memberikan arahan dan bimbingannya kepada penulis dalam menyusun skripsi ini.
3. Bapak Dodi Vionanda, M.Si., Ph.D., selaku dosen pembahas yang telah memberikan kritik dan saran yang membangun kepada penulis.
4. Ibu Admi Salma S.Pd., M.Si., selaku dosen pembahas yang telah memberikan kritik dan saran yang membangun kepada penulis.
5. Bapak/Ibu dosen departemen statistika, yang telah mengajar, membimbing, dan memberikan ilmu kepada penulis.

6. Terkhusus kepada kedua orang tua penulis, kakak-kakak, adik, serta keluarga penulis yang selalu mendukung penulis dalam penyusunan skripsi ini, baik dukungan secara moril maupun materiil.
7. Teman-teman Program Studi Sarjana Statistika angkatan 2018 untuk dukungan semangat dan kebersamaan selama menjalani masa perkuliahan ini.
8. Dan semua pihak yang juga ikut serta membantu penulis dalam penyusunan skripsi ini hingga selesai.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih banyak kekurangan. Oleh sebab itu penulis sangat terbuka menerima kritik dan saran yang membangun untuk dijadikan sebagai bahan evaluasi. Semoga penyusunan skripsi ini dapat bermanfaat bagi para pembaca.

Padang, Februari 2023

Mardhiatul Azmi

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR TABEL	vi
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN	ix
BAB I PENDAHULUAN	1
A. Latar Belakang Masalah	1
B. Batasan Masalah	6
C. Rumusan Masalah	7
D. Tujuan Penelitian	7
E. Manfaat Penelitian	8
BAB II KAJIAN TEORITIS	9
A. Analisis Multivariat	9
B. Analisis <i>Cluster</i>	12
1. <i>K-Means Clustering</i>	13
2. <i>K-Medoids Clustering</i>	16
C. Ukuran Jarak	17
D. Standardisasi Data	18
E. Validitas <i>Cluster</i>	19
F. Interpretasi Hasil <i>Clustering</i>	21
G. R Packages <i>K-Means</i> dan <i>K-Medoids Clustering</i>	22
H. Kemiskinan	22
I. Penelitian Terdahulu	27
BAB III METODOLOGI PENELITIAN	29
A. Jenis Penelitian	29
B. Jenis dan Sumber Data	29
C. Variabel Penelitian	29
D. Teknik Analisis Data	30

BAB IV HASIL DAN PEMBAHASAN	34
A. Hasil	34
B. Pembahasan	54
BAB V PENUTUP	57
A. Kesimpulan	57
B. Saran	58
DAFTAR PUSTAKA	59
LAMPIRAN.....	60

DAFTAR TABEL

Tabel	Halaman
1. Variabel-Variabel Penelitian	29
2. Struktur Data Penelitian	30
3. Statistika Deskriptif Data Kemiskinan di Sumatera Tahun 2021	34
4. <i>Centroid</i> Awal dari <i>K-Means Clustering</i>	38
5. <i>Centroid</i> Akhir dari <i>K-Means Clustering</i>	40
6. Hasil <i>K-Means Clustering</i>	40
7. Nilai Rata-Rata dari Data Hasil <i>K-Means Clustering</i>	42
8. <i>Medoids</i> Akhir dari <i>K-Medoids Clustering</i>	45
9. Hasil Perhitungan Jarak, Nilai <i>Cost</i> , dan Total <i>Cost</i> pada <i>K-Medoids Clustering</i>	46
10. Hasil <i>K-Medoids Clustering</i>	46
11. Nilai Rata-Rata dari Data Hasil <i>K-Medoids Clustering</i>	48
12. SSW pada <i>K-Means</i> dan <i>K-Medoids</i>	51
13. Hasil DBI	53

DAFTAR GAMBAR

Gambar	Halaman
1. Persentase Penduduk Miskin di Pulau Sumatera Periode 2020-2021	4
2. Klasifikasi Analisis Multivariat	9
3. <i>Flowchart K-Means Clustering</i>	31
4. <i>Flowchart K-Medoids Clustering</i>	32
5. SSB pada <i>K-Means</i> dan <i>K-Medoids</i>	52
6. Rasio pada <i>K-Means</i> dan <i>K-Medoids</i>	53

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Data Kemiskinan Pulau Sumatera Tahun 2021	62
2. Standardisasi Data	67
3. Hasil Perhitungan Jarak pada <i>K-Means Clustering</i>	71
4. Hasil Perhitungan Jarak pada <i>K-Medoids Clustering</i>	74
5. Sintaks <i>K-Means</i> dan <i>K-Medoids</i> menggunakan RStudio	78



BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Analisis multivariat merupakan analisis yang melibatkan banyak variabel (lebih dari dua variabel). Analisis multivariat dibagi ke dalam dua kategori, yaitu analisis dependensi dan analisis interdependensi. Analisis dependensi adalah analisis yang bertujuan untuk melihat hubungan antara variabel terikat dan variabel bebas. Sedangkan analisis interdependensi adalah analisis yang secara serentak melibatkan banyak variabel, tanpa membedakan variabel terikat dan variabel bebas. Analisis interdependensi terbagi dalam beberapa kategori, salah satunya yaitu analisis *cluster*. Analisis *cluster* merupakan analisis yang digunakan untuk mengelompokkan objek (responden) ke dalam beberapa *cluster*/kelompok yang relatif homogen. Objek dalam setiap kelompok memiliki kemiripan satu sama lain dan berbeda jauh (tidak sama) dengan objek pada kelompok lainnya (Supranto, 2017: 26). Kelebihan dari analisis *cluster* ini yaitu dapat mengelompokkan objek dengan jumlah data yang besar dan variabel yang relatif banyak, sehingga mudah dianalisis.

Hair *et al* (2014: 438) mengemukakan bahwa analisis *cluster* terbagi dalam dua kategori yaitu hirarki dan non hirarki. Hirarki *clustering* adalah metode yang digunakan untuk mengelompokkan objek/pengamatan secara struktural berdasarkan kemiripan sifatnya dan jumlah *cluster* yang diinginkan belum diketahui (Mattjik & Sumertajaya, 2011: 199). Dalam metode hirarki ini dibagi menjadi dua kategori yaitu metode *agglomerative (single linkage, compete*

linkage, average linkage, ward's method, dan centroid method) dan metode *divisive*. Menurut Johnson & Wichern (2002: 694), non hirarki *clustering* adalah metode yang digunakan untuk pengelompokan objek dimana jumlah *cluster* yang akan dibentuk dapat ditentukan terlebih dahulu sebagai bagian dari prosedur pengelompokan. Metode-metode yang terdapat pada non hirarki yaitu *k-means, k-medoids, Clustering Large Applications (CLARA)*, dan metode lainnya. Menurut Kamila, dkk (2019: 121), algoritma yang populer dalam metode non hirarki (partisi) yaitu algoritma *k-means* dan algoritma *k-medoids*. Selain itu, alasan penggunaan *k-means* dan *k-medoids* adalah karena algoritma ini memiliki akurasi yang cukup tinggi, dan mempunyai sifat fleksibel karena *user* dapat menentukan jumlah *cluster* yang akan dibuat (Farissa dkk, 2021: 110).

K-Means merupakan salah satu algoritma pengelompokan data non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan ke dalam kelompok yang lain (Prasetyo, 2012: 178). Sedangkan *k-medoids clustering* adalah algoritma *clustering* non hirarki yang merupakan varian dari algoritma *k-means*, dimana setiap *cluster* diwakili (*medoid*) oleh objek sebagai pusat *cluster*. Berbeda dengan *k-means* yang menggunakan nilai rata-rata (*mean*) sebagai pusat *cluster*, *k-medoids* memilih *k* objek secara sembarang dari data sebagai *medoid* untuk dijadikan pusat *cluster* (Pujari, 2010: 117). Untuk menguji kinerja dari kedua metode tersebut dapat dilihat dari nilai *Davies Bouldin Index (DBI)*. Hasil

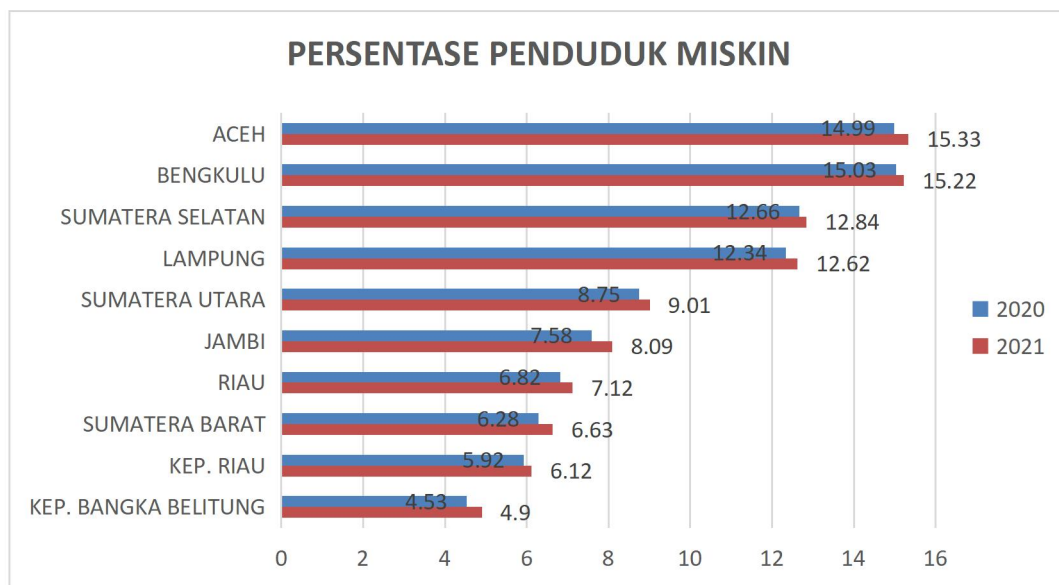
clustering dapat dikatakan baik apabila nilai DBI yang diperoleh semakin rendah, tetapi tidak bernilai negatif.

Berdasarkan pemaparan sebelumnya, maka penerapan metode akan dilakukan pada permasalahan kemiskinan di Pulau Sumatera pada tahun 2021. Pengelompokan dilakukan berdasarkan variabel-variabel yang memengaruhi kemiskinan tersebut, yang dilihat dari tingkat kemiripan masing-masing objek. Hal ini bertujuan untuk mengetahui karakteristik kemiskinan dari masing-masing daerah, sehingga diharapkan dapat menjadi referensi bagi pemerintah dalam merancang kebijakan dalam rangka menanggulangi kemiskinan yang terjadi di Sumatera. Selain itu, pengelompokan dilakukan karena belum ada literatur yang membahas mengenai pengelompokan kemiskinan di Pulau Sumatera menurut kabupaten/kota pada tahun 2021, yang ditentukan berdasarkan variabel-variabel yang memengaruhi kemiskinan tersebut.

Kemiskinan merupakan ketidakmampuan masyarakat dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran (BPS). Kemiskinan merupakan masalah yang terjadi di dalam masyarakat yang berhubungan dengan ekonomi, budaya, politik, dan sosial. Lebih luas, kemiskinan juga berkaitan dengan ketidakmampuan dalam mencapai aspek diluar pendapatan seperti akses kesehatan, pendidikan, air bersih dan sanitasi (Bappenas, 2007). Suryawati (2005: 122) menyatakan bahwa kemiskinan adalah keadaan masyarakat yang kekurangan uang dan barang untuk kelangsungan hidup. Makna kemiskinan selain kekurangan uang adalah mengenai tingkat kesehatan, pendidikan rendah, dan perlakuan tidak adil dalam hukum. Terdapat sebanyak 267

Kabupaten/Kota di Indonesia memiliki persentase penduduk miskin diatas angka nasional (10,14%).

Perubahan jumlah dan persentase penduduk miskin tidak terlepas dari nilai garis kemiskinan. Garis kemiskinan (GK) merupakan rata-rata pengeluaran per kapita masyarakat setiap bulan (BPS, 2022). GK terdiri dari Garis Kemiskinan Makanan (GKM) dan Garis Kemiskinan Non Makanan (GKNM). Komoditi yang paling berpengaruh pada GK berupa makanan adalah beras, rokok, dan cabe merah. Sedangkan komoditi bukan makanan yang paling berpengaruh terhadap GK adalah biaya perumahan, bensin, listrik, pendidikan, dan perlengkapan mandi. Masalah kemiskinan ini juga perlu memerhatikan tingkat kedalaman dan keparahan dari kemiskinan. Selain harus mampu memperkecil jumlah penduduk miskin, kebijakan kemiskinan juga sekaligus harus bisa mengurangi tingkat kedalaman dan keparahan dari kemiskinan. Menurut BPS, persentase penduduk miskin di Sumatera pada tahun 2021 mengalami peningkatan dari tahun sebelumnya, dimana dapat ditunjukkan pada Gambar 1.



Gambar 1. Persentase Penduduk Miskin di Pulau Sumatera Periode 2020-2021

Pada Gambar 1, diketahui bahwa setiap provinsi yang ada di Pulau Sumatera memiliki kenaikan persentase penduduk miskin pada tahun 2021 dari tahun sebelumnya. Terdapat empat provinsi di Sumatera pada tahun 2021 yang memiliki persentase di atas angka nasional (10,14%) yaitu provinsi Aceh, Bengkulu, Lampung, dan Sumatera Selatan. Hal ini menunjukkan ketimpangan pertumbuhan ekonomi di semua lapisan masyarakat yang hanya dirasakan oleh sekelompok kecilnya saja.

BPS menyebutkan bahwa terdapat beberapa permasalahan yang menyebabkan meningkatnya angka kemiskinan pada tahun 2021 yaitu karena masih terjadinya pandemi covid-19 yang mengakibatkan perubahan aktivitas ekonomi masyarakat sehingga angka kemiskinan terpengaruhi, dan pengeluaran konsumsi rumah tangga mengalami peningkatan. Peningkatan pengeluaran konsumsi rumah tangga ini disebabkan oleh berkumpulnya anggota keluarga di rumah yang mana kebutuhan akan makanan dan bukan makanan juga akan meningkat, dimana pada masa itu pemerintah memerintahkan masyarakat untuk tetap berada di rumah karena pandemi Covid-19. Selain itu, sistem pendidikan juga dilakukan secara *online/daring* yang mengharuskan peserta didik untuk membeli paket data, serta kebutuhan akan listrik juga meningkat, sehingga hal tersebut juga memengaruhi pengeluaran konsumsi rumah tangga. Pemutusan hubungan kerja dan penutupan Usaha Mikro Kecil Menengah (UMKM) juga terjadi akibat perekonomian tidak menentu. Hal tersebut juga akan memengaruhi peningkatan angka kemiskinan atau kesejahteraan masyarakat yang biasanya bekerja menghasilkan uang harus terhenti.

Menurut penelitian yang dilakukan oleh Fajriyah dan Rahayu (2016), terdapat beberapa variabel yang memengaruhi kemiskinan yaitu persentase penduduk miskin, indeks kedalaman kemiskinan, indeks keparahan kemiskinan, angka melek huruf, penduduk yang tidak mendapatkan akses air bersih, angka partisipasi sekolah usia menengah, penduduk yang tidak mendapatkan akses fasilitas kesehatan, tingkat partisipasi angkatan kerja, pekerja di sektor pertanian, laju pertumbuhan penduduk, dan Produk Domestik Regional Bruto (PDRB) per kapita. Dari penelitian tersebut, diperoleh hasil bahwa variabel angka melek huruf, tingkat partisipasi angkatan kerja, penduduk yang bekerja di sektor pertanian, dan PDRB per kapita memiliki pengaruh negatif dan signifikan terhadap kemiskinan. Sedangkan variabel penduduk yang tidak mendapatkan akses fasilitas kesehatan memiliki pengaruh positif dan tidak signifikan terhadap kemiskinan. Pratama (2014) juga menyebutkan dalam penelitiannya bahwa terdapat beberapa faktor kemiskinan yaitu tingkat pendapatan rumah tangga, tingkat pendidikan, tingkat inflasi, konsumsi rumah tangga, indeks pembangunan manusia, dan tingkat kemiskinan.

Berdasarkan pemaparan di atas, maka penulis melakukan penelitian dengan judul **“Komparasi Kinerja Algoritma *K-Means* dan *K-Medoids* dalam Pengelompokan Kabupaten/Kota di Sumatera Berdasarkan Indikator Kemiskinan”**.

B. Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Data yang digunakan pada penelitian ini adalah data sekunder yaitu data kemiskinan di Pulau Sumatera tahun 2021 berdasarkan kabupaten/kota.

2. Variabel yang digunakan pada penelitian ini yaitu persentase penduduk miskin, indeks pembangunan manusia, pengeluaran per kapita, penduduk yang bekerja di sektor pertanian, angka partisipasi sekolah usia 13-15 tahun, fasilitas perumahan yang memiliki air layak, dan fasilitas perumahan yang memiliki jamban sendiri.
3. Penelitian ini menggunakan analisis *cluster* dengan algoritma *k-means* dan *k-medoids*.
4. Kinerja algoritma *k-means* dan *k-medoids* dilihat dari nilai *Davies Bouldin Index*.
5. Data diolah dengan bantuan *software* R Studio.

C. Rumusan Masalah

Berdasarkan uraian pada latar belakang, rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana hasil pengelompokan kabupaten/kota di Sumatera berdasarkan indikator kemiskinan pada tahun 2021 menggunakan algoritma *k-means* dan *k-medoids*?
2. Bagaimana hasil perbandingan kinerja antara algoritma *k-means* dan *k-medoids* dalam pengelompokan kabupaten/kota di Sumatera berdasarkan indikator kemiskinan pada tahun 2021?

D. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut :

1. Untuk mengetahui hasil pengelompokan kabupaten/kota di Sumatera berdasarkan indikator kemiskinan pada tahun 2021 menggunakan algoritma *k-means* dan *k-medoids*.

2. Untuk mengetahui hasil perbandingan kinerja antara algoritma *k-means* dan *k-medoids* dalam pengelompokan kabupaten/kota di Sumatera berdasarkan indikator kemiskinan pada tahun 2021.

E. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat bagi semua pihak, baik bagi instansi, masyarakat maupun bagi penulis sendiri. Adapun manfaat yang diharapkan adalah sebagai berikut :

1. Diharapkan dapat menambah wawasan dan pengetahuan mengenai pengelompokan indikator kemiskinan menggunakan *k-means clustering* dan *k-medoids clustering*.
2. Diharapkan dapat menjadikan bahan pertimbangan bagi instansi terkait dalam melaksanakan kebijakan di bidang penanggulangan kemiskinan.
3. Diharapkan dapat digunakan sebagai dasar penelitian lanjutan dan dapat memberikan kontribusi dalam pengembangan teori mengenai pengelompokan indikator kemiskinan menggunakan *k-means clustering* dan *k-medoids clustering*, bagi yang ingin melanjutkan penelitian ini.