BAB II

TINJAUAN PUSTAKA

2.1. Konsep Kemiskinan di Indonesia

Kemiskinan merupakan persoalan yang sering menjadi pusat perhatian di berbagai negara terutama di Indonesia. Secara etomologi, kemiskinan berasal dari kata miskin yang berarti tidak berharta-benda. Sedangkan kemiskinan dalam kamus besar bahasa Indonesia mempunyai persamaan yakni kefakiran. Dimana dua kata yang biasa disebutkan yaitu fakir miskin yang berarti orang yang sangat kekurangan (Mu'is, 2021).

Kemiskinan dalam secara umum, merupakan kondisi kurangnya tingkat pendapatan seseorang untuk memenuhi kebutuhan pokok dan kebutuhan lainnya yang dapat terjamin terpenuhi standart kualitas hidupnya. Dalam suatu negara tingkat kemiskinannya cukup tinggi, maka negara tersebut digolongkan kedalam negara yang sedang berkembang (Wini, 2010).

Berdasarkan Undang-Undang No. 24 Tahun 2004, kemiskinan adalah kondisi sosial ekonomi seseorang atau sekelompok orang yang tidak terpenuhinya (Wini, 2010). Hak-hak dasarnya untuk mempertahankan dan mengembangkan kehidupan yang bermartabat. Kemiskinan Menurut Ahli Gunawan Sumodiningrat, kemiskinan dibedakan dalam tiga pengertian : Pertama kemiskinan yaitu apabila pendapatan seseorang tidak mencukupi dari kebutuhan hidup minimum, antara lain kebutuhan pangan, sandang, kesehatan, perumahan dan pendidikan yang diperlukan untuk hidup dan bekerja. Rendahnya tingkat pendapatan ini disebabkan oleh keterbatasan sarana dan prasarana fisik serta kelangkaan modal atau miskin karena

sebab alami (natural). Kedua kemiskinan yaitu mengacu pada sikap masyarakat yang disebabkan oleh faktor budaya yang tidak mau berusaha untuk memperbaiki tingkat kehidupan meskipun ada usaha dari pohak luar untuk membantunya. Yang terakhir kemiskinan relatif yaitu kaitannya dengan masalah pembangunan yang belum seimbang menyebabkan ketimpangan pendapatan (Itang, 2015).

Pandemi Covid-19 terhadap kehidupan manusia berdampak pada berbagai aspek kehidupan, juga terkena penyakit Virus Covid-19. Di Indonesia, dampak pandemi Covid 19 menjadi masalah utama di bidang kesehatan. Penyebarannya yang cepat ke berbagai dunia dan menyebabkan penderitanya meninggal dengan efek yang parah. Dari sisi ekonomi adanya peningkatan jumlah pengangguran dan pembangunan kemiskinan (Wibisono, Salma, Hasanah, Mutiah, & Heni, 2022).

Secara garis besar faktor yang menyebabkan kemiskinan di Indonesia yaitu:

1. Tingkat pendidikan yang rendah

   Pendidikan menjadi salah satu wadah pengembangan diri bagi individu untuk membentuk diri. Kualitas diri ini dapat dilihat dari keterampilan, wawasan pengetahuan, moral, dan perilaku individu.

2. Mengalami peningkatannya populasi

   Tingginya jumlah penduduk merupakan salah satu faktor penyebab faktor kemiskinan yang terjadi karena pertambahan jumlah penduduk tanpa diikuti dengan kualitas sumber daya manusia (SDM) yang memadai akan mengakibatkan ketimpangan.

2.2. Data Mining

Data mining adalah istilah yang digunakan untuk mendeskripsikan proses penggalian nilai berasal dataset. Data mining ialah salah satu solusi untuk menjelaskan proses penggalian informasi pada suatu basis data yang berskala besar dan proses klasifikasi otomatisasi kasus sesuai pola data yang diperoleh dari data set (Dwilestari, Mulyawan, Martanto, & Ali, 2021). Data mining sangat perlu dilakukan terutama dalam mengelola data yang sangat besar untuk memudahkan aktivitas recording suatu transaksi dan untuk proses data worehousing supaya bisa memberikan informasi yang akurat bagi penggunanya alasan utama mengapa data mining sangat menarik perhatian industri informasi dalam beberapa tahun belakangan ini adalah karena tersedianya data dalam jumlah yang sangat besar dan semakin besarnya kebutuhan untuk mengubah data tersebut menjadi informasi dan pengetahuan yang berguna karena sesuai fokus bidang ilmu yaitu melakukan kegiatan mengekstrasi atau menambang pengetahuan dari data yang berukuran/berjumlah besar, informasi inilah yang nantinya sangat berguna untuk pengembangan (Arianto, 2019).

2.3. Clustering

Clustering adalah proses pengelompokan benda serupa ke dalam kelompok yang berbeda, atau yang diketahui sebagai partisi dari sebuah data set kedalam subset, sehingga data dalam setiap subset memiliki arti yang bermanfaat. Dimana sebuah cluster terdiri dari kumpulan benda-benda seperti antara satu dengan yang lainnya serta berbeda dengan benda yang terdapat pada cluster lainnya. Algoritma clustering terdiri dari dua bagian yaitu secara hirarkis dan secara partitional. Algoritma hirarkis menemukan cluster secara berurutan dimana cluster ditetapkan

sebelumnya, sedangkan algoritma partitional menentukan semua kelompok pada waktu tertentu (Merliana, Ernawati, & Santoso). Adapun beberapa metode clustering Self Organizing Map, K-Means, dan K-Medoids.

Clustering disebut juga dengan suatu proses yang mengelompokan dan membagi pola data menjadi beberapa jumlah data set sehingga membentuk pola yang sama yang kemudian dikelompokkan pada cluster yang sama dan terbentuk dengan pola yang berbeda di cluster yang berbeda (Merliana, Ernawati, & Santoso).

Adapun peneliti sebelumnya yang menggunakan Metode Algortima K-Medoids seperti pada penelitian Gifthera Dwilestari, Mulyawan, Martanto, dan Irfan Ali, 2021 yang berjudul "Analisis *Clustering* Menggunakan *K-Medoids* Pada Data Penduduk Miskin Indonesia" menggunakan *clustering* untuk mengelompokkan objek-objek ke dalam *cluster-cluster.*

2.3.1. Algoritma K-Means Clustering

Dalam penyelesaiannya, algoritma K-Means akan menghasilkan titik centroid yang dijadikan tujuan dari algoritma K-Means. Setelah iterasi K-Means berhenti, setiap objek dalam dataset menjadi anggota dari suatu cluster. Nilai cluster ditentukan dengan mencari seluruh objek untuk menemukan cluster dengan jarak terdekat ke objek . Algoritma K –means akan mengelompokan item data dalam suatu dataset ke suatu cluster berdasarkan jarak terdekat (Syam, 2017).

Data clustering menggunakan metode *K-Means Clustering* ini secara umum dilakukan dengan algoritma dasar sebagai berikut:

1. Tentukan jumlah *cluster*
2. Alokasikan data ke dalam *cluster* secara *random*
3. Hitung *centroid*/rata-rata dari data yang ada di masing-masing *cluster*

4. Alokasikan masing-masing data ke centroid/rata-rata terdekat
5. Kembali ke Step 3, apabila masih ada data yang berpindah cluster atau apabila perubahan nilai *centroid*, ada yang di atas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada *objective function* yang digunakan di atas nilai *threshold* yang ditentukan

Ada beberapa kelebihan pada algoritma k-means, yaitu:

1. Mudah untuk diimplementasikan dan dijalankan.
2. Waktu yang dibutuhkan untuk menjalankan pembelajaran ini relatif cepat.
3. Mudah untuk diadaptasi.
4. Umum digunakan.

Algoritma k-means memiliki beberapa kelebihan, namun ada kekurangannya juga. Kekurangan dari algoritma tersebut yaitu :

1. Sebelum algoritma dijalankan, k buah titik diinisialisasi secara *random* sehingga pengelompokkan data yang dihasilkan dapat berbeda-beda. Jika nilai *random* untuk inisialisasi kurang baik, maka pengelompokkan yang dihasilkan pun menjadi kurang optimal.
2. Dapat terjebak dalam masalah yang disebut *curse of dimensionality*. Hal ini dapat terjadi jika data pelatihan memiliki dimensi yang sangat tinggi (Contoh jika data pelatihan terdiri dari 2 atribut maka dimensinya adalah 2 dimensi. Namun jika ada 20 atribut, maka akan ada 20 dimensi).
3. Jika hanya terdapat beberapa titik sampel data, maka cukup mudah untuk menghitung dan mencari titik terdekat dengan *k* titik yang diinisialisasi secara *random.* Namun jika terdapat banyak sekali titik data (misalnya satu milyar buah data), maka perhitungan dan pencarian titik terdekat akan

membutuhkan waktu yang lama. Proses tersebut dapat dipercepat, namun dibutuhkan struktur data yang lebih rumit seperti *kD-Tree* atau *hashing* (Arni, 2019).

Penelitian sebelumnya Dini Marlina, Nurelina Fauzer Putri, Andri Fernando, dan Aditya Ramadhan, 2018 yang berjudul "Implemantasi Algoritma K-Medoids dan K-Means Untuk Pengelompokan Wilayah Sebaran Cacat pada anak" menggunakan metode *k-means* untuk perbandingan pengelompokan pada data sebaran anak cacat dengan metode *k-medoids*.

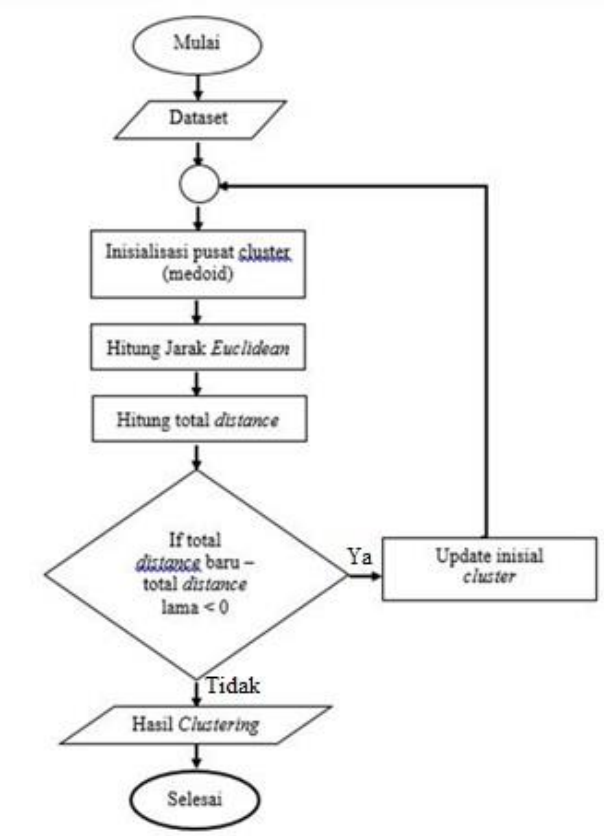
Penelitian sebelumnya Ni Putu Eka Merliana, Ernawati, dan Alb. Joko Santoso, 2015 yang berjudul "Analisa Penentuan Jumlah Cluster Terbaik Pada Metode *K-Means Clustering*" menggunakan metode *k-means* untuk mencari jumlah *cluster* terbaik pada metode *k-means* pada suatu studi kasus yang dilakukan pada lembaga STAHN(*Sekolah Tinggi Agama Hindu Negeri*) Tampung Penyang Palangka Raya.

2.3.2. Algoritma K-Medoids Clustering

K-Medoids adalah salah satu metode partisi, karena menggunakan objek yang paling terpusat (medoids) di cluster menjadi pusat cluster dari nilai rata-rata objek dalam sebuah cluster. Metode K-medoids lebih cocok untuk mengelompokkan data dibandingkan metode K-Means (Marlina, Putri, Fernando, & Aditya, 2018).

Algoritma K-Medoids memiliki kelebihan untuk mengatasi kelemahan pada pada algoritma K-Means yang sensitive terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi

data. Kelebihan lainnya yaitu hasil proses clustering tidak bergantung pada urutan masuk dataset. Berikut Gambar 2.1 Flowchart Algoritma K-Medoids.



Gambar 2. 1 Flowchart Algoritma K-Medoids

Sumber : (Pulungan, Suhada, & Dedi, 2019)

Langkah-langkah menggunakan algoritma K-Medoids:

1. Menentukan Jumlah cluster sebanyak *k* dengan pertimbangan teoritis dan konseptual
2. Mencari jarak tiap objek terhadap *cluster* terdekat dengan menggunakan ukuran *Euclidian distance.*

$$= d_{(x,y)} \sqrt{\sum_{i=1}^{n} (x_i - y_j)^2}$$

Keterangan :

n = Jumlah atribut yang digunakan

$d_{(x,y)}$ = Jarak antara data ke-i serta data ke-j

x_i = Nilai atribut kesatu dari data ke i

x_j = Nilai atribut kesatu dari data ke j

3. Pilih secara acak objek pada masing-masing *cluster* sebagai kandidat *Medoids* baru.

4. Hitung jarak setiap objek yang berada pada masing-masing cluster dengan kandidat medoid baru.

5. Hitung total simpangan (S)dengan menghitung nilai total distancebar total distance lama. Jika S< 0, maka tukar objek dengan data cluster untuk membentuk sekumpulan *k* objek baru sebagai *medoids*.

6. Ulangi langkah 3 sampai 5 hingga tidak terjadi perubahan *medoids*, sehingga didapatkan cluster beserta anggota *cluster* masing- masing (Marlina, Putri, Fernando, & Aditya, 2018).

Beberapa peneliti sebelumnya yang menggunakan Metode Algortima K-Medoids seperti pada penelitian Gifthera Dwilestari, Mulyawan, Martanto, dan Irfan Ali, 2021 yang berjudul "Analisis *Clustering* Menggunakan *K-Medoids* Pada Data Penduduk Miskin Indonesia" menghasilkan presentase tingkat kemiskanan menjadi 3 cluster yaitu *cluster* 1 merupakan tingkat kemiskinan tingkat rendah terdapat 19 provinsi yang dihasilkan, *cluster* 2 merupakan tingkat kemiskinan tingkat sedang terdapat 12 provinsi yang dihasilkan dan *cluster* 3 merupakan tingkat kemiskinan tingkat tinggi terdapat 3 provinsi yang dihasilkan. Dalam penelitian Gifthera dan kawan – kawan ini

diharapkan dapat memberikan informasi kepada pemerintah tentang pengelompokan data penduduk miskin diIndonesia yang berdampak pada pemerataan dalam pemberian bantuan di Indonesia.

Pada penelitian sesbelumnya Nurliana Pulungan, Suhada, Dedi Suhendro, 2018 juga menggunakan Metode Algortima K-Medoids yang berjudul "Penelitian Algortima *K-Medoids* Untuk Mengelompokkan Penduduk 15 Tahun Keatas Menurut Lapangan Pekerjaan Utama" yang menghasilkan kesimpulan sebagai berikut: dengan menggunakan metode K-Medoids dapat di terapkan, dengan bantuan aplikasi rapidminer 5.3. sebagai perhitungan perbandingannya, Dengan menggunakan 2 cluster yaitu, cluster tinggi dan cluster rendah di peroleh cluster tinggi sebanyak 3 lapangan pekerjaan utama (Pertanian, perhutanan dan perikanan, Industri Pengolahan, Perdangan besar dan eceran, Reparasi dan perawatan mobil dan sepeda motor). Dan cluster rendah sebanyak 14 lapangan pekerjaan utama (Pertambangan dan penggalian, Pengadaan listrik, gas, uap air panas dan udara dingin, pengadaan air, pengelolaan sampah, dan daur ulang, pembuangan dan pembersihan limbah dan sampah, Konstruksi, Transportasi dan pergudangan, Penyediaan akomodasi dan penyediaan makan minum, Jasa keuangan dan asuransi, Real estate, Jasa perusahaan, Administrasi pemerintahan, pertahanan, dan jaminan sosial wajib, Jasa pendidikan, Jasa kesehatan dan kegiatan sosial, dan Jasa lainnya).

Peneliti sebelumnya Jefri Arianto, 2019 yang berjudul "Penerapan Data Mining Untuk Pengelompokan Penduduk Kurang Mampu Desa Sambirejo Timur Dengan Algoritma K-Medoids (Studi Kasus Kantor Kepala

Desa Sambirejo Timur)" yang menghasilkan kesimpulan bahwa dengan adanya penerapan Algoritma K-Medoids terlebih dahulu membentuk nilai kriteria untuk dilakukan pencarian nilai setiap iterasi untuk mengetahui berapa banyak warga kurang mampu yang layak menerima bantuan yang akan diproses, setelah itu dilakukan perhitungan hasil proses algoritma k-medoids. Dan adapun kriteria yang ditentukan adalah seperti pekerjaan, rumah, jumlah tanggungan dan penghasilan.

Peneliti sebelumnya Disty Wahyuli, Handrizal, Agus Perdana Windarto, Dedi Suhendro, dan Anjar Wanto, 2019 yang berjudul "Mengelompokkan Garis Kemiskinan Menurut Provinsi Mengunnakan Algortima K-Medoids" yang menghasilkan kesimpulan penerapan data mining menggunakan algroritma K-Medoids untuk mengelompokkan garis kemiskinan menurut provinsi dengan data uji sebanyak 34 provinsi dengan menggunakan 2 *cluster* 1 (nilai terendah) sebanyak 11 *cluster* 2 (nilai tinggi) sebanyak 23. Dan pengujian data menggunkan Rapidminier.

2.4. Metode Elbow

Menurut peneliti sebelumnya Merliana, Ernawati dan Santoso (2015), Metode *Elbow* merupakan yang menghasilkan informasi dengan cara melihat perbandingan hasil antara jumlah *cluster* yang akan membentuk siku suatu titik terakhir, nilai *cluster* akan dijadikan model data untuk menentukan cluster yang terbaik. Adapun nilai *cluster* pertama dengan nilai *cluster* kedua jika nilainnya mengalami penurunan paling besar maka nilai *cluster* tersebut yang terbaik (Pulungan, Suhada, & Dedi, 2019).

Berikut rumus untuk mendapatkan perbandingannya dengan menghitung SSE (*Sum Of Square Error*) dari setiap nilai *cluster.*

$$SSE = \sum_{k=1}^{k} \sum_{xi \in Sk} ||Xi - Ck||^2$$

Keterangan :

k = Jumlah *cluster*

xi = nilai atribut data ke-i

C*k* = nilai atribut titik pusat *cluster* ke-i

Langkah-langkah metode *elbow* menurut Merliana, Ernawati dan Santoso (2015), untuk menentukan nilai *k* sebagai berikut :

1.Inisialisasi awal nilai *k*

2. Menaikkan nilai *k*

3. Perhitung hasil *sum of square error* dari tiap nilai *k*

4. Melihat hasil *sum of square error* dari nilai *k* yang mengalami penurunan secara drastis

5. Menetapkan nilai *k* yang berbentuk siku.

2.5. Google Colaboratory

Sejak 2017 google mengembangkan AI *framework* yang disebut dengan *Tensorflow* dan fitur pengembangan yang disebut dengan *Colaboraotry*. *TensorFlow* bersifat *open source,* dan pada saat itu Google membuat *Colaboratory* yang dapat diakses secara gratis untuk penggunaan publik. *Colaboratory* sekang dikenal dengan Google Colab (Rizal, Kharisma, & Fahrurrozi, 2021). Pada penelitian ini menggunakan Google Colab ini untuk melakukan perhitungan metode *elbow* menentukan nilai *k.*

Pada penelitian sebelumnya Meida Cahyo Untoro, Leslie Anggraini, Maria Andini, Hesti Retnosari, M.Anas Nasrulloh, 2021 yang berjudul "Penerapan metode *k-means clustering* data COVID-19 di Provinsi Jakarta" yang menggunakan Google Colab ini untuk menjalankan algoritma *k-means* dengan pencarian nilai *k* dengan menggunakan metode elbow terhadap *dataset* jumlah penduduk jakarta dan data persebaran COVID-19 di wilayah jakarta.

2.6. KNIME Analytics

KNIME Analytics platfrom ini merupakan *software open source* yang gratis untuk dikases. Sistem alur pada KNIME adalah *software open source* yang bertujuan untuk memecahkan masalah ini dengan menyediakan *platfrom* yang dapat diperluas dengan mudah dengan integrasi alat baru, memiliki sistem data yang di ketik dengan kuat dan memungkikan pembuat alur kerja secara rinci. Adapun salah satu kekuatan KNIME adalah banyaknya node untuk analisis data dan *machine learning* (Fillbrunn, et al., 2017). Pada penilitian ini KNIME Analytics digunakan untuk pengklusteran dengan perhitungan Algoritma K-Medoids.

Adapun penelitian sebelumnya Imaduddin Farih, Lukman fadillah, Nadira, Verry Dina Aromy, Harry Patria, 2022 yang berjudul "Pengelompokan Negara Berdasarkan Indikator Kesejahteraan Dengan Metode *Unsupervised Learinung-Clustering* : Bukti Empiris dari 167 Negara" yang menggunakan tool KNIME Analytics untuk pengujian dengan metode algoritma *K-Medoids* dengan penentuan 3 *cluster* yang cukup bagus karena hasil uji *sillhoette*-nya mendekati 1. Dan menghasilkan *cluster 0* terdiri dari 36 negara, *cluster* 1 terdiri dari 85 negara dan *cluter* 2 terdiri dari 46 negara.