BAB 2
TINJAUAN PUSTAKA

2.1 *Data Mining*

Praktik memanfaatkan statistik, matematika, kecerdasan buatan, dan teknik pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi berharga dari data dalam jumlah besar dikenal sebagai *data mining* (Riska dan Farokhah, 2023). *Data mining* mempunyai beberapa tujuan (Sachrrial dan Iskandar, 2023), yaitu:

1. Pola tersembunyi dalam data, seperti pola korelasi, urutan, klasifikasi, atau anomali, dapat ditemukan dengan menggunakan penambangan data. Akan layak untuk memahami hubungan antara variabel dan kualitas lain dengan mengidentifikasi pola-pola tersembunyi ini.

2. Prakiraan dan proyeksi berdasarkan tren data masa lalu dimungkinkan untuk memperkirakan kejadian atau perilaku di masa depan dengan memeriksa pola dan tren historis oleh *data mining*.

3. *Data mining* memungkinkan melakukan pengelompokan berdasarkan karakteristik dan perilaku yang serupa.

4. Pengambilan keputusan yang lebih baik didukung oleh informasi mendalam yang ditawarkan *data mining*. Keputusan dapat didasarkan pada informasi yang andal dan benar dengan memeriksa data masa lalu, melihat tren, dan memperkirakan hasil potensial.

5. Analisis data yang mendalam menggunakan *data mining* dapat mengarah pada penemuan informasi baru. Kemajuan sains, penciptaan barang atau teknologi baru, atau pemahaman yang lebih dalam tentang fenomena rumit semuanya dapat memperoleh manfaat dari ini.

Dalam *data mining*, ada lima (5) pengelompokan *data mining* yang dapat dilakukan oleh peneliti ataupun pihak yang ingin melakukan pengolahan data agar memperoleh suatu informasi yang tersembunyi. Kelima pengelompokan *data mining* tersebut ialah *clustering*, prediksi, klasifikasi, asosiasi dan estimasi (Sachrrial dan Iskandar, 2023).

2.2 Pengelompokan (*Clustering*)

Teknik umum dalam *data mining* adalah pengelompokan, yang membagi objek data menjadi beberapa kelompok sesuai dengan seberapa mirip atau berbeda mereka. Algoritma khusus yang mengelompokkan objek data sesuai dengan kualitas atau fiturnya digunakan dalam proses pengelompokan. Algoritma ini mengidentifikasi kelompok yang paling cocok untuk setiap objek data dengan menghitung jarak atau tingkat kesamaan di antara mereka (Sachrrial dan Iskandar, 2023).

Metode *clustering* banyak dikembangkan oleh para ahli, sehingga memiliki karakter, kelebihan, dan kekurangan masing-masing (Riska dan Farokhah, 2023). *K-Means, Hierarchical Clustering*, dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) adalah beberapa jenis algoritma pengelompokan yang sering digunakan. Jenis data dan tujuan yang perlu Anda capai akan menentukan algoritma pengelompokan mana yang terbaik untuk Anda karena masing-masing memiliki fitur dan metode yang unik (Sachrrial dan Iskandar, 2023).

2.3 Algoritma *Agglomerative Hierarchical Clustering* (AHC)

Dalam analisis *cluster*, teknik atau algoritma yang disebut *Agglomerative Hierarchical Clustering* (AHC) mencoba mengatur data ke dalam struktur *cluster* hierarkis. Dengan menggunakan metrik atau jarak yang telah ditentukan sebelumnya, metode ini pertama-tama memperlakukan setiap kumpulan data sebagai satu *cluster* sebelum secara progresif menggabungkan *cluster* dengan kesamaan tertinggi. Langkah pertama dalam metode *Agglomerative Hierarchical Clustering* (AHC) adalah mengukur jarak antara setiap pasangan data. Dua *cluster* dengan jarak terdekat kemudian digabungkan untuk membuat *cluster* baru. *Cluster* yang lebih besar digabungkan dalam metode ini hingga semua data digabungkan menjadi satu *cluster* utama. Kerangka kerja hierarkis yang menjelaskan koneksi antar *cluster* dibuat selama prosedur ini. Biasanya, struktur hierarkis ini ditampilkan sebagai grafik atau pohon dendogram (Sachrrial dan Iskandar, 2023).

Fleksibilitas pendekatan AHC (*Agglomerative Hierarchical Clustering*) dalam memilih metrik atau jarak untuk mengukur kesamaan data menjadi salah satu manfaatnya. Jarak Manhattan, jarak Mahalonibis, dan jarak *Euclidean* adalah beberapa contoh metrik yang dapat diterapkan. Properti data dan tujuan analitik yang dimaksudkan menentukan bagaimana jarak ini digunakan. Selanjutnya, den-

gan memilih jumlah klaster yang sesuai atau dengan menggunakan kriteria *cut-off* untuk menghentikan proses penggabungan klaster, pengguna AHC (*Agglomerative Hierarchical Clustering*) dapat dikelola. Rumus yang sesuai dapat digunakan untuk menentukan jarak Manhattan atau *Euclidean* antara dua gugus (Sachrrial dan Iskandar, 2023).

1. Menghitung Matriks Jarak

   Euclidean Distance

   $$d(x,y) = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2} \tag{2.1}$$

   Manhattan Distance

   $$D_{man}(x,y) = \sum_{i=1}^{n} |x_i - y_i| \tag{2.2}$$

2. Pengelompokan *Agglomerative Hierarchical Clustering* (AHC)

   Pengelompokan AHC memiliki tiga mode yang berbeda dalam prosesnya, yaitu *single linkage, complete linkage* dan *average linkage.*

   Single Linkage

   Single Linkage merupakan teknik dengan menggunakan strategi keterkaitan tunggal, klaster digabungkan berdasarkan seberapa jauh anggota terdekat mereka satu sama lain.

   $$d_{(ab)c} = min\{d_{a.c}; d_{b.c}\} \tag{2.3}$$

   Complate Linkage

   Complate Linkage merupakan teknik yang menggabungkan *cluster - cluster* menurut jarak rata-rata pasangan anggota masing-maisng pada himpunan antara dua *cluster.*

   $$d_{(ab)c} = max\{d_{a.c}; d_{b.c}\} \tag{2.4}$$

   Average Linkage

   Average Linkage meruupakan teknik yang menggabungkan *cluster - cluster* menurut jarak antara anggota-anggota terjauh di antara dua *cluster.*

   $$d_{(ab)c} = average\{d_{a.c}; d_{b.c}\} \tag{2.5}$$

2.4 Algoritma *K-Means*

2.4.1 Sejarah Singkat Algoritma *K-Means*

Awalnya dikembangkan oleh Llyod (1957, 1982), metode *K-Means* sebenarnya diciptakan oleh sejumlah individu, termasuk McQueen (1967), Friedman and Rubin (1967), dan Forgey (1965). Baru pada tahun 1982 konsep pengelompokan menggunakan Algoritma *K-Means*, yang awalnya diidentifikasi oleh Lloyd pada tahun 1957, diterbitkan. Forgey, yang juga menerbitkan penemuannya, datang berikutnya. Akibatnya, algoritma *Llyod-Forgey* adalah nama lain untuk pendekatan *K-Means* ini (Wanto, at al., 2020).

2.4.2 Definisi Algoritma *K-Means*

Salah satu teknik pengelompokan yang menjadi anggota kelompok *Unsupervised Learning* adalah *K-Means*. Pendekatan kecerdasan buatan ini digunakan untuk menemukan pola dalam kumpulan data yang terdiri dari titik data yang tidak berlabel atau tidak diklasifikasikan. Data dapat dikelompokkan menjadi dua atau lebih kelompok menggunakan *K-Means* itu sendiri. Komputer mengurutkan data input yang dapat berupa data atau objek ke dalam *cluster* yang dimaksudkan menggunakan Algoritma *K-Means* tanpa terlebih dahulu mengetahui kelas target. Data kemudian akan dikelompokkan ke dalam kategori yang diperlukan menggunakan teknik *K-Means* (Wanto, at al., 2020).

2.4.3 Konsep *K-Means*

Terdapat beberapa langkah yang dilakukan dalam algoritma *k-means* (Wanto, at al., 2020), yaitu sebagai berikut:

1. Tentukan k sebagai jumlah *cluster* yang di bentuk pada data
2. Tentukan nilai pusat (*centroid*). Penentuan nilai pusat pada tahap awal dilakukan secara acak atau random. Sedangkan pada tahap iterasi digunakan rumus sebagai berikut:

$$V_{ij} = \frac{1}{n} \sum_{k=0}^{N_i} X_{kj} \tag{2.6}$$

Keterangan :

V_{ij} : *Centroid* rata-rata *cluster* ke-i untuk variabel ke-j

N_i : Jumlah anggota *cluster* ke-i

i, k : Indeks dari *cluster*

J : Indeks dari variabel

X_{kj} : Nilai data ke-k variabel ke-j untuk *cluster* tersebut

3. Pada masing-masing *record* hitung jarak terdekat dengan *centroid*. Jarak *centroid* yang digunakan ialah *Euclidean Distance*, dengan rumus berikut ini:

$$De = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \tag{2.7}$$

Keterangan :

De : *Euclidean Distance*

I : Banyaknya objek

x, y : Koordinat objek

s, t : Koordinat *centroid*

4. Kelompokkan objek berdasarkan jarak ke *centroid* terdekat.

5. Ulangi langkah ke-3 dan ke-4 hingga *centroid* bernilai optimal.

2.4.4 Karakteristik *K-Means*

K-Means memiliki beberapa karakteristik yang berupa kelebihan serta kekurangannya dalam pengklasteran (Wanto, at al., 2020), yaitu sebagai berikut:

1. Dalam proses pengelompokan, *K-Means* merupakan salah satu teknik pengelompokan langsung yang mudah diterapkan dan cepat.

2. *K-Means* tidak dapat mengelompokkan data dengan baik di beberapa jenis himpunan data, di mana hasil segmentasi tidak menghasilkan pola grup yang secara akurat menangkap fitur bentuk alami data.

3. Saat mengelompokkan data dengan *outlier, K-Means* mungkin tidak berfungsi dengan baik.

2.5 Validasi *Clustering*

Uji validitas *Davies Boulding Index* (DBI) adalah salah satu metode untuk memvalidasi klaster yang telah terbentuk untuk menentukan klaster mana yang paling cocok untuk kumpulan data. David L. Davies dan Donald W. Bouldin

(1997) pertama kali mengusulkan DBI dengan maksud untuk melakukan evaluasi klaster. Dengan menentukan jumlah kumpulan data dan kriteria turunan, validitas internal dari proses pengelompokan dievaluasi.. Adapun persamaan yang digunakan ialah sebagai berikut:

$$DBI = \frac{1}{K} \sum_{i=1}^{K} \max_{i=j} (R_{i,j}) \tag{2.8}$$

Keterangan :

K : Jumlah klaster yang digunakan

$R_{i,j}$: Rasio perbandingan antara klaster ke-I dan klaster ke-j

Cluster dengan pemisahan terbesar dan jumlah kohesi paling sedikit dianggap baik. $R_{i,j}$ di formulasikan pada persamaan berikut:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \tag{2.9}$$

Jarak antara *centroids* c_i dan c_j diukur menggunakan persamaan berikut untuk menentukan SSB (*Sum of Square Between Cluster*), metrik untuk pemisahan antara dua *cluster*, seperti *cluster* i dan j:

$$SSB_{i,j} = d(c_i, c_j) \tag{2.10}$$

SWW (*Sum of Square Within Cluster*) adalah sebagai metrik kohesi dalam satu klaster ke-i, seperti berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d\,(x_j, c_i) \tag{2.11}$$

Keterangan:

m_i : Jumlah data di dalam klaster ke-i

c_i : *centroid* klaster ke-i

$d\,(x_j, c_i)$: Dapat dicari menggunakan euclidean *distance*.

2.6 Kemiskinan

Ketika seseorang atau sekelompok individu tidak dapat menggunakan hak-hak dasar mereka untuk menegakkan atau memajukan keberadaan yang layak, mereka dikatakan dalam kemiskinan. Rasa aman dari perlakuan atau ancaman tanpa kekerasan, kemampuan untuk terlibat dalam kehidupan sosial-politik, penyediaan kebutuhan pangan, kesehatan, pendidikan, pekerjaan, perumahan, air bersih, pertahanan, sumber daya alam, dan lingkungan adalah contoh dari hak-hak dasar ini. (BPS Sumut, 2023). Dengan mengukur kemiskinan menggunakan pendekatan kebutuhan dasar, yang didasarkan pada gagasan untuk mampu memenuhi kebutuhan dasar, BPS mendefinisikan kemiskinan secara absolut. Menurut metode ini, kemiskinan didefinisikan sebagai ketidakmampuan untuk memenuhi permintaan pangan mendasar daripada pangan seperti yang didefinisikan oleh pengeluaran. Oleh karena itu, orang yang berpenghasilan rata-rata kurang dari ambang batas kemiskinan setiap bulan dianggap miskin. Membandingkan kemiskinan secara umum dan mengevaluasi dampak jangka panjang dari kebijakan pada inisiatif untuk menguranginya adalah tujuannya.

2.7 Kemiskinan di Sumatera Utara

Perkembangan tingkat kemiskinan di Sumatera Utara dari tahun ke tahun relatif berfluktuasi (naik turun), namun jika dilihat antar kabupaten/kota terdapat beberapa kota dengan tingkat kemiskinan yang sangat tinggi salah satunya ialah kota Medan. Kota Medan merupakan peringkat tertinggi di Sumatera Utara dengan jumlah penduduk miskin 187,28 ribu jiwa berdasarkan BPS Sumatera Utara, Maret 2023. Apabila dibandingkan dengan provinsi lainnya di Indonesia maka Sumatera Utara menduduki peringkat ke-17, dimana peringkat ini menunjukkan perlunya dilakukan penanganan agar penduduk miskin berkurang.

2.8 Variabel Penelitian

2.8.1 Tingkat Pengangguran Terbuka

Rasio individu yang menganggur terhadap angkatan kerja dikenal sebagai Tingkat Pengangguran Terbuka (TPT). Orang-orang yang menganggur dan mencari pekerjaan, bersiap-siap untuk memulai bisnis mereka sendiri, atau tidak mencari pekerjaan karena mereka percaya tidak mungkin untuk mendapatkannya semuanya dianggap dalam pengangguran terbuka. Pada agustus 2023, TPT di Sumatera

Utara adalah 5,89 persen. Jumlah angkatan kerja pada bulan tersebut adalah 8,02 juta orang, naik 352 ribu orang dibandingkan agustus 2022.

2.8.2 Persentase Penduduk Miskin

Selain itu, proporsi penduduk yang hidup dalam kemiskinan turun 0,27 poin persentase, dari 8,42 persen menjadi 8,15 persen, antara Maret 2022 dan Maret 2023. Jika dipecah berdasarkan wilayah, jumlah orang miskin di daerah perkotaan dan pedesaan menurun antara Maret 2022 dan Maret 2023. Dari 8,76 persen menjadi 8,23 persen, penduduk miskin perkotaan turun 0,53 poin. Demikian pula, proporsi individu miskin di daerah pedesaan turun dari 7,98 persen menjadi 8,03 persen, turun 0,05 poin.

2.8.3 Indeks Kedalaman Kemiskinan

Jumlah total uang yang dibutuhkan untuk mengangkat semua orang miskin dan rumah tangga ke garis kemiskinan diwakili oleh Indeks Kedalaman Kemiskinan (biasa disingkat P1), yang merupakan rasio terhadap total pendapatan seluruh penduduk di tingkat garis kemiskinan. Ketika biaya transaksi dan hambatan dihilangkan, nilai total indeks kedalaman kemiskinan menggambarkan biaya pengurangan kemiskinan dengan memberikan target transfer yang ideal kepada orang miskin. Berdasarkan identifikasi karakteristik masyarakat miskin dan tujuan bantuan dan program, potensi ekonomi uang pengentasan kemiskinan meningkat dengan nilai Indeks Kedalaman Kemiskinan yang lebih rendah. Intervensi dari program pengentasan kemiskinan relatif lebih berhasil dalam menurunkan tingkat kemiskinan. Indeks Kedalaman Kemiskinan Sumatera Utara turun 0,104 poin antara Maret 2022 dan Maret 2023, dari 1.365 pada Maret 2022 menjadi 1.261 pada Maret 2023. Hal ini menunjukkan bahwa pengeluaran rata-rata orang miskin cenderung mendekati tingkat kemiskinan. Membandingkan kabupaten dan kota di Sumatera Utara, Kabupaten Deli Serdang memiliki indeks kedalaman kemiskinan terendah dari 33 kabupaten dan kota, yaitu 0,34. Pada Maret 2023, Kabupaten Nias Selatan memiliki indeks kedalaman kemiskinan tertinggi, yaitu 3,04 (BPS,2023).

2.8.4 Indeks Keparahan Kemiskinan

Orang miskin diberi bobot lebih dalam Indeks Keparahan Kemiskinan, yang merupakan ukuran kemiskinan yang kadang-kadang diwakili oleh huruf P2. Distribusi pengeluaran di antara orang miskin ditunjukkan oleh Indeks Keparahan Kemiskinan. Kesenjangan pengeluaran antara orang miskin meningkat dengan nilai indeks. Indeks Keparahan Kemiskinan Sumatera Utara turun 0,019 poin antara Maret 2022 dan Maret 2023, dari 0,343 pada Maret 2022 menjadi 0,324 pada Maret 2023. Ini menunjukkan bahwa kesenjangan pengeluaran masyarakat miskin di Sumatera Utara menurun. Dibandingkan dengan kabupaten dan kota lain di Sumatera Utara, Kota Binjai memiliki indeks keparahan kemiskinan terendah (0,06) dari 33. Pada Maret 2023, Kabupaten Nias Selatan memiliki indeks keparahan kemiskinan tertinggi, yaitu 0,85 (BPS,2023).

2.9 Penelitian Terdahulu

Sejumlah tinjauan studi berfungsi sebagai landasan teoritis untuk penelitian saya dan sebagai titik perbandingan dengan penyelidikan sebelumnya. Setelah studi saya tentang beberapa penelitian, saya menemukan sejumlah penelitian yang berfungsi sebagai landasan teoritis untuk penelitian saya dan titik perbandingan dengan studi sebelumnya. Ada sejumlah penelitian yang terkait dengan penelitian yang saya lakukan setelah meninjau sejumlah penelitian lain., diantaranya:

Riska, Suastika Yulia dan Farokha, Lia (2023) menggunakan metode Algoritma *K-Means* dan *K-Medoid* untuk melihat perbandingan kedua metode tersebut berdasarkan kunjungan wisatawan mancanegara ke Indonesia. Hasil dari penelitian ini ialah dijelaskan bahwa Algoritma *K-Means* lebih baik dari Algoritma *K-Medoid* dimana dapat dilihat bahwa *cluster* terbaik dalam kasus ini adalah *K-Means* dengan nilai $k = 5$ dan nilai davies boulding *index* -0,302.

Dalam studi mereka, Sachrrial, Rifqi Habibi, dan Iskandar, Agus (2023) membandingkan dua pendekatan metode AHC dan *K-Medoids Complate Linkage* untuk mengklasifikasikan data kemiskinan di Indonesia. Tiga klaster dihasilkan dengan lokasi klaster yang berbeda di kedua metodologi, menurut hasil analisis penelitian. Meskipun demikian, setiap klaster dan jumlah provinsi sama.

(Hidayat, et al., 2022) dalam penelitiannya memperoleh hasil penelitian dengan menggunakan metode *silhouette* dimana jumlah *cluster* optimal sebanyak 2 *cluster*. Hasil pada *cluster* 1 terdapat 11 kabupaten/kota, lalu *cluster* 2 terdapat 12 kabupaten/kota.

(Fransiska R, at al., 2022) melakukan penelitian menggunakan algoritma K-Means dengan *Silhouette Coefficient*. Temuan penelitian ini berasal dari perhitungan metode *k-means*. Data baru dikumpulkan dalam bentuk tingkat kemiskinan, yang dipisahkan menjadi dua klaster: klaster 0 memiliki tingkat kemiskinan yang tinggi sebanyak 14 daerah, sedangkan klaster 1 memiliki tingkat kemiskinan yang rendah sebanyak 13 daerah.

(Luchia, at al., 2022) membandingkan dua metode yaitu *k-means* dan *k-medoids* pada pengelompokan data miskin di Indonesia. Dan memperoleh hasil penelitian dimana *k-means* lebih unggul dibanding *k-medoids* dengan nilai DBI terbaik yaitu 0,041 dengan nilai $k = 8$ pada *k-means*.

2.10 Wahdatul Ulum

Wahdatul Ulum terdiri dari dua kata yaitu wahda dan ulum yang artinya kesatuan ilmu-ilmu. Wahdatul Ulum adalah ilmu dari Allah SWT yang mana manusia diberi kesempatan untuk berharap pada cinta-Nya dan dalam rangka bertaqwa kepada-Nya. Fridiyanto (2020) menjelaskan bahwa wahdatul ulum ialah sebuah filsafat ilmu khas UIN Sumatera Utara yang berpandangan bahwa ilmu merupakan satu kesatuan. Pada UIN Sumatera Utara (UIN-SU), konsep, visi, dan paradigma keilmuan yang dikembangkan melalui berbagai bidang keilmuan berupa jurusan atau fakultas, serta program studi dan mata kuliah yang secara alami memiliki hubungan kesatuan sebagai ilmu yang dianggap sebagai karunia dari Allah SWT, merupakan Wahdatul Ulum. Akibatnya, wahdatul ulum yang berkaitan dengan masalah penelitian harus dijelaskan dalam penelitian ini. Dalam setiap penelitian tentunya ada masalah yang harus diselesaikan. Solusi dalam menyelesaikan suatu masalah berhubungan dengan perintah Allah SWT yang terdapat dalam Al-Qur'an Q.S Al-Insyirah (1-8), sebagai berikut:

أَلَمْ نَشْرَحْ لَكَ صَدْرَكَ وَوَضَعْنَا عَنْكَ وِزْرَكَ ٱلَّذِي أَنْقَضَ ظَهْرَكَ وَرَفَعْنَا لَكَ ذِكْرَكَ فَإِنَّ مَعَ ٱلْعُسْرِ يُسْرًا إِنَّ مَعَ ٱلْعُسْرِ يُسْرًا فَإِذَا فَرَغْتَ فَانْصَبْ وَإِلَىٰ رَبِّكَ فَارْغَبْ

Terjemahan:

Bukankah Kami telah melapangkan untukmu dadamu? Dan Kami telah menghilangkan daripadamu bebanmu, yang memberatkan punggungmu dan Kami tinggikan bagimu sebutan (nama) mu karena sesungguhnya sesudah kesulitan itu ada kemudahan, sesungguhnya sesudah kesulitan itu ada kemudahan. Maka apabila kamu telah selesai (dari suatu urusan), kerjakanlah dengan sungguh-sungguh (urusan) yang lain, dan hanya kepada Tuhanmulah hendaknya kamu berharap.

Pada penelitian ini yang menjadi pokok bahasan ialah kemiskinan. UUD 1945 mengamanatkan bahwa pemerintahan Republik Indonesia memiliki salah satu tugas yaitu untuk memajukan kesejahteraan umum, mencerdaskan kehidupan bangsa dan menegakkan keadilan sosial bagi seluruh rakyat Indonesia. Hal ini berarti seluruh rakyat Indonesia harus bebas dari kemiskinan atau mempunyai kehidupan yang layak dan ini merupakan salah satu tugas pemerintah yang dapat dikategorikan sebagai suatu nafkah. Q.S Al-Baqarah (267) menerangkan tentang nafkah, yaitu sebagai berikut:

يَاَيُّهَا ٱلَّذِينَ ءَامَنُو ٓا ٱنفِقُو ٓا مِنْ طَيِّبَتِ مَا كَسَبْتُمْ وَمِمَّا ٱخْرَجْنَا لَكُمْ مِّنَ ٱلَا رْضِ وَ لَا تَيَمَّمُو
ٓا ٱلْخَبِيثَ مِنْهُ تُنْفِقُونَ وَلَسْتُمْ بِئَ خِذِيهِ إِ لَّا ٱنْ تُغْمِضُو ٓا فِيهِ وَاعْلَمُو ٓا أَنَّ ٱللَّهَ غَنِيٌّ حَمِيدٌ

Terjemahan:

Hai orang-orang yang beriman, nafkahkanlah (di jalan Allah) sebagian dari hasil usahamu yang baik-baik dan sebagian dari apa yang Kami keluarkan dari bumi untuk kamu. Dan janganlah kamu memilih yang buruk-buruk lalu kamu menafkahkan daripadanya, padahal kamu sendiri tidak mau mengambilnya melainkan dengan memincingkan mata terhadapnya. Dan ketahuilah, bahwa Allah Maha Kaya lagi Maha Terpuji.

Indonesia memiliki 38 provinsi yang salah satu provinsi dengan tingkat kemiskinan tinggi ialah Sumatera Utara yang terdiri dari beberapa kabupaten/kota. Agar mengoptimalkan usaha dalam menurunkan tingkat kemiskinan maka pemerintah harus mengetahui terlebih dahulu daerah mana yang memerlukan penanganan terbesar hingga terkecil. Oleh karena itu, digunakan sistem *clustering* untuk mengelompokkan daerah-daerah yang ada sesuai dengan tingkat kemiskinan tertinggi sampai terendah. Dalam Al-Qur'an juga terdapat sistem *clustering* atau

pengelompokan yaitu pada Q.S Al-Araf (178) yang berbunyi:

مَنْ يَّهْدِ ٱللَّهُ فَهُوَ ٱلْمُهْتَدِي وَمَنْ يُّضْلِلْ فَأُ۟لَٰٓئِكَ هُمُ ٱلْخَٰسِرُونَ

Terjemahan:

Barangsiapa yang diberi petunjuk oleh Allah, maka Dialah yang mendapat petunjuk; dan Barangsiapa yang disesatkan Allah, maka merekalah orang-orang yang merugi.

Ayat diatas menerangkan bahwasanya terdapat dua kelompok atau dua *cluster* yang digabungkan berdasarkan karakteristik yang sama yaitu kelompok yang mendapat petunjuk dan kelompok yang disesatkan. Sistem *clustering* sendiri melakukan pengelompokan berdasarkan karakteristik yang sama sehingga ayat diatas sangat relevan dengan sistem pengelompokan.