

Model Pengelompokan Penerima Bantuan Sosial PKH dengan Teknik Data Mining

Dhoni Brilliant*, Lili Tanti

Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia

Email: ¹*brilliantisiahaan@gmail.com, ²lili@potensi-utama.ac.id

Email Penulis Korespondensi: brilliantisiahaan@gmail.com *

Submitted: 22/10/2025; Accepted: 26/11/2025; Published: 31/12/2025

Abstrak–Program Keluarga Harapan (PKH) merupakan bantuan sosial yang diberikan kepada keluarga kurang mampu, namun proses penentuan penerima masih sering menghadapi kendala karena banyaknya data dan belum optimalnya analisis terhadap karakteristik calon penerima. Penelitian ini menerapkan teknik data mining *K-Means clustering* untuk mengelompokkan penerima bantuan sosial PKH di Kantor Camat Medan Labuhan berdasarkan sejumlah indikator sosial-ekonomi. Metode *K-Means* digunakan untuk menghasilkan pola pengelompokan yang dapat membantu perangkat kecamatan dalam memetakan tingkat kelayakan penerima bantuan secara lebih objektif. Kontribusi penelitian ini adalah menyediakan model pengelompokan yang dapat dijadikan alat bantu pendukung keputusan dalam proses penentuan prioritas penerima PKH. Hasil penelitian menunjukkan bahwa data dapat terbagi ke dalam beberapa kelompok yang menggambarkan variasi tingkat kebutuhan dan kondisi ekonomi masyarakat. Keterbatasan penelitian ini terletak pada cakupan data yang hanya berasal dari satu kecamatan dan variabel yang digunakan masih terbatas pada data administratif yang tersedia. Hasil pengujian menunjukkan bahwa model menghasilkan akurasi sebesar 92%, presisi 96,67%, recall 90,63%, dan F1-score 93,52%. Dengan penerapan sistem ini, proses seleksi penerima bantuan dapat dilakukan secara lebih cepat, transparan, dan tepat sasaran.

Kata Kunci: PKH; Data Mining; K-Means; Pengelompokan; Bantuan Sosial; Sistem Pendukung Keputusan.

Abstract– The Family Hope Program (PKH) is a social assistance program provided to underprivileged families, but the process of determining recipients often faces obstacles due to the large amount of data and suboptimal analysis of the characteristics of prospective recipients. This study applies the *K-Means clustering* data mining technique to group PKH social assistance recipients at the Medan Labuhan District Office based on several socio-economic indicators. The *K-Means* method is used to generate *clustering* patterns that can help sub-district officials map the eligibility level of aid recipients more objectively. The contribution of this study is to provide a *clustering* model that can be used as a decision-support tool in the process of determining the priority of PKH recipients. The results show that the data can be divided into several groups that reflect variations in the level of need and economic conditions of the community. The limitations of this study lie in the data coverage which only comes from one sub-district and the variables used are still limited to available administrative data. The test results show that the model produces an accuracy of 92%, a precision of 96.67%, a recall of 90.63%, and an F1-score of 93.52%. By implementing this system, the selection process for aid recipients can be carried out more quickly, transparently, and on target.

Keywords: PKH; Data Mining; K-Means; *Clustering*; Social Assistance; Decision Support System.

1. PENDAHULUAN

Persoalan kemiskinan adalah keadaan tidak adanya sumber daya atau barang yang signifikan yang dialami oleh individu atau kelompok. Karena ketiadaan sumber daya atau barang, individu atau kelompok merasa kurang siap untuk mendukung kebutuhan hidup mereka secara tepat [1]. Ketidakterdayaan ini bisa hanya pada tingkat kebutuhan sosial atau pada tingkat pemenuhan kebutuhan sosial atau pada tingkat kepuasan kebutuhan esensial. Konsekuensinya, kemelaratan adalah keadaan tidak terpenuhkannya kebutuhan-kebutuhan pokok [2]. Kemiskinan merupakan salah satu persoalan sosial yang secara konsisten menjadi perhatian pemerintah, terutama karena keterbatasan sumber daya membuat banyak keluarga tidak mampu memenuhi kebutuhan dasar secara layak. Untuk mengurangi beban tersebut, pemerintah menjalankan Program Keluarga Harapan (PKH) sebagai bantuan sosial bersyarat yang ditujukan bagi keluarga miskin [3]. Namun dalam praktiknya, penyaluran bantuan ini masih menghadapi berbagai kendala, seperti ketidaktepatan sasaran, data penerima yang tidak mutakhir, perpindahan penduduk, serta subjektivitas dalam penentuan kelayakan. Kondisi ini juga terjadi di Kantor Camat Medan Labuhan, di mana proses pemilihan penerima PKH masih dilakukan secara manual sehingga memerlukan waktu panjang, rentan kesalahan, dan sering kali tidak mencerminkan kondisi ekonomi masyarakat secara objektif.

PKH (Program Keluarga Harapan) merupakan program penanggulangan kemiskinan dan pengembangan sistem perlindungan sosial bersyarat bagi masyarakat miskin yang ditujukan untuk mempercepat pencapaian, dengan memberikan bantuan tunai bersyarat kepada masyarakat miskin. Namun pada Kantor Camat Medan Labuhan proses pemberian bantuan Program Keluarga Harapan (PHK) masih mengalami kendala yaitu tidak meratanya pembagian bantuan kepada masyarakat miskin, sehingga tidak jarang masyarakat merasa kecewa karena seharusnya mereka mendapatkan bantuan tetapi justru tidak mendapatkan bantuan tersebut dan begitu pulak sebaliknya [4].

Selain itu, dikarenakan banyaknya jumlah keluarga, membuat pemeringkatan penentuan penerima bantuan PKH akan memerlukan waktu yang lama. Dan kendala lain yang sering ditemukan adalah perpindahan penduduk, sehingga Camat tidak dapat mengambil data penduduk lama dan harus mensurvei kembali penduduk

di Kantor Camat Medan Labuhan. Hal ini berdampak ada subjektifitas di dalam penentuan peserta Program Keluarga Harapan (PKH), terutama jika beberapa calon peserta yang miskin atau kurang mampu memiliki tingkat kelayakan yang tidak jauh berbeda. Dalam hal ini sistem pendataan masih menggunakan data lama yang belum ter update sedangkan setiap tahun penduduk selalu mengalami perubahan pola status sosial dan pengolahan data masih secara manual.

Untuk menyelesaikan masalah tersebut penulis merancang sistem aplikasi dengan data mining dengan menggunakan metode K-Means dapat meningkatkan efisiensi dan akurasi dalam pemilihan penerima. Metode algoritma K-Means *clustering* merupakan metode klasifikasi yang cukup bagus karena tingkat akurasi yang sangat baik dibandingkan dengan metode lain. Algoritma K-Means merupakan algoritma *clustering* yang masuk ke dalam kelompok Unsupervised learning yang dipakai untuk mengelompokkan data kedalam beberapa bagian dengan sistem partisi [5]

Berdasarkan penelitian dari [6] dengan judul “Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means” kasus yang dibahas adalah Pengelompokan Penerima Bantuan Sosial Masyarakat. Variabel Independen: Pekerjaan, Aset, Penghasilan dan Variabel Dependen : Penerima Bantuan Sosial Masyarakat (Yes, No). jumlah dataset 196, label class Penerima Bantuan Sosial Masyarakat (Yes, No), algoritma yang digunakan adalah Metode K-Means, Teknik evaluasi Rasio sebesar 93%. Berdasarkan penelitian dari [7] dengan judul “Implementasi Algoritma *K-Nearest Neighbor* Dalam Menentukan Penerima Bantuan Langsung Tunai” kasus yang dibahas adalah Penerima Bantuan Langsung Tunai. Variabel Independen: Kondisi rumah, Penghasilan, Pekerjaan, Jumlah Tanggungan dan Variabel Dependen : Penerima Bantuan Langsung Tunai (Layak, Tidak layak). Jumlah dataset 170, label class Penerima Bantuan Langsung Tunai (Layak, Tidak layak), algoritma yang digunakan adalah Algoritma *K-Nearest Neighbor*, Teknik evaluasi Akurasi sebesar 68,82 %

Berdasarkan penelitian dari [8] dengan judul “Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan (Pkh) Menggunakan Metode Weighted Naïve Bayes Dengan Laplace Smoothing” kasus yang dibahas adalah Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan (Pkh). Variabel Independen: Penghasilan, Jumlah Tanggungan Anak, Usia, Jumlah Keluarga, pekerjaan, Status Rumah, Jenis Lantai, jenis dinding dan Variabel Dependen : Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan (Pkh) (Layak, Tidak layak). Jumlah dataset 56, label class Klasifikasi Kelayakan Penerima Bantuan Program Keluarga Harapan (Pkh) (Layak, Tidak layak), algoritma yang digunakan adalah Metode Weighted Naïve Bayes, Teknik evaluasi Confusion Matrix sebesar 95,83%. Berdasarkan penelitian dari [9] dengan judul “Algoritma K-Means Untuk Pengelompokan Bantuan Langsung Tunai (BLT)” kasus yang dibahas adalah Pengelompokan Bantuan Langsung Tunai. Variabel Independen : Jumlah keluarga, status tempat tinggal, jenis lantai, jenis dinding, jenis atap, sumber penerangan, jenis bahan bakar, tempat pembuangan, fasilitas buang air besar, sumber air, pembuangan tinja, pekerjaan, penghasilan dan Variabel Dependen : Penerima Bantuan Rumah Akibat Penerima Bantuan Uang Kuliah Tunggal Bagi Mahasiswa Terdampak Covid-19 (Yes, No). Jumlah dataset 1632, label class Pengelompokan Bantuan Langsung Tunai (Yes, No), algoritma yang digunakan adalah K-Means, Teknik evaluasi SSE sebesar 60%. Berdasarkan penelitian dari [10] dengan judul “Algoritma K-Means *Clustering* dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa” kasus yang dibahas adalah Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa. Variabel Independen : Jenis pekerjaan, jumlah tanggungan keluarga, Penghasilan dan kondisi rumah saat sekarang dan Variabel Dependen : Penerima Bantuan Langsung Tunai (BLT) Dana Desa (Yes, No). Jumlah dataset 25, label class Penerima Bantuan Langsung Tunai (BLT) Dana Desa (Yes, No), algoritma yang digunakan adalah K-Means, Teknik evaluasi AUC sebesar 83.33%.

Berdasarkan penelitian dari [11] dengan judul “Implementation of Naïve Bayes Algorithm on Prediction of Acceptance of Keluarga Harapan Program (PKH)” kasus yang dibahas Prediction of Acceptance of Keluarga Harapan Program (PKH). Variabel Independen : Pekerjaan, Penghasilan, Status Tempat Tinggal, Disabilitas, Anak Balita, Anak Sekolah dan Lansia dan Variabel Dependen : Prediction of Acceptance of Keluarga Harapan Program (PKH) (Layak, Tidak layak). Jumlah dataset 200, label class Prediction of Acceptance of Keluarga Harapan Program (PKH) (Layak, Tidak layak), algoritma yang digunakan adalah Naïve Bayes Algorithm, Teknik evaluasi Akurasi sebesar 96%. Berdasarkan penelitian dari [12] dengan judul “Classification of Social Assistance Ricipients Using Machine Learning” kasus yang dibahas Classification of Social Assistance Ricipients. Variabel Independen : Income, Age, Ocupation, Number of Recipients dan Variabel Dependen : Classification of Social Assistance Ricipients (PKH/BLT). Jumlah dataset 176, label class Classification of Social Assistance Ricipients (PKH/BLT), algoritma yang digunakan adalah Naïve Bayes Algorithm, Teknik evaluasi Akurasi sebesar 100%.

Sedangkan penelitian yang dilakukan penulis membahas Pengelompokan penerima bantuan sosial PKH, Variabel Independen : Pekerjaan suami, pekerjaan istri, penghasilan suami, penghasilan istri, jumlah tanggungan, kepemilikan rumah, sumber listrik, Variabel Dependen : pengelompokan penerima bantuan sosial PKH (Layak, Tidak Layak), jumlah data set 1000, label class Pengelompokan penerima bantuan sosial PKH (Layak, Tidak Layak), algoritma yang digunakan adalah Algoritma K-Means dan teknik evaluasi menggunakan elbow. Berdasarkan GAP tersebut, penelitian ini dilakukan untuk membangun model pengelompokan penerima PKH menggunakan algoritma K-Means dengan memanfaatkan dataset yang lebih besar, yaitu 1000 data masyarakat dengan variabel sosial-ekonomi yang lebih lengkap seperti pekerjaan, penghasilan, jumlah tanggungan, kepemilikan rumah, dan sumber listrik. Selain itu, penelitian ini juga menerapkan metode elbow untuk menentukan

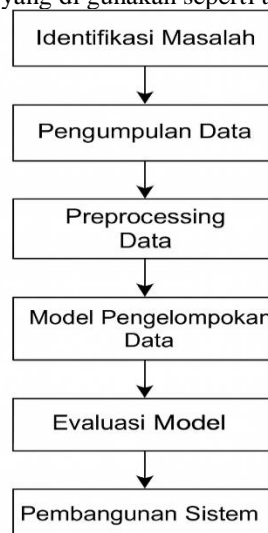
jumlah cluster optimal sehingga hasil pengelompokan lebih tepat dan representatif. Pendekatan ini diharapkan mampu menggambarkan pola kelayakan masyarakat berdasarkan kondisi sosial-ekonomi yang sebenarnya, sekaligus meminimalkan subjektivitas dalam penentuan penerima bantuan.

Tujuan dari penelitian ini adalah membangun model pengelompokan yang efektif dalam menentukan kelompok masyarakat yang layak dan tidak layak menerima PKH dengan menggunakan algoritma *K-Means*. Penelitian ini juga bertujuan menghasilkan model evaluasi berbasis metode *elbow* untuk menentukan jumlah cluster terbaik, sehingga proses analisis data dapat dilakukan secara lebih terukur dan sistematis. Dengan demikian, penelitian ini memberikan gambaran menyeluruh mengenai tingkat kelayakan masyarakat berdasarkan pola-pola yang terbentuk dari data aktual. Manfaat dari penelitian ini tidak hanya dirasakan oleh dunia akademik, tetapi juga oleh Kantor Camat Medan Labuhan. Dari sisi instansi, model yang dihasilkan dapat membantu mempercepat proses seleksi penerima bantuan, mengurangi potensi ketidakakuratan data, serta meningkatkan objektivitas dalam menentukan penerima PKH. Sementara itu, dalam bidang data mining, penelitian ini memberikan kontribusi berupa penerapan metode *unsupervised learning* pada data berskala besar dalam konteks kebijakan publik, serta memperkaya referensi penggunaan metode *elbow* dalam penentuan cluster optimal. Secara keseluruhan, penelitian ini diharapkan dapat mendukung pengambilan keputusan yang lebih baik dan berbasis data dalam proses penyaluran bantuan sosial.

2. METODOLOGI PENELITIAN

2.1 Rancangan Penelitian

Kerangka kerja ini merupakan langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Adapun kerangka kerja penelitian yang di gunakan seperti terlihat pada gambar 1:



Gambar 1. Kerangka Kerja Penelitian

Berdasarkan kerangka kerja penelitian yang telah digambarkan di atas, maka dapat diuraikan pembahasan masing-masing tahap dalam penelitian adalah sebagai berikut:

1. Identifikasi Masalah

Tahap awal penelitian dimulai dengan mengidentifikasi permasalahan pada proses penentuan penerima bantuan PKH di Kantor Camat Medan Labuhan. Permasalahan utama adalah proses seleksi yang masih manual, subjektivitas petugas, data tidak mutakhir, dan tingginya risiko ketidaktepatan sasaran. Identifikasi masalah menjadi dasar bagi penyusunan kebutuhan analisis menggunakan metode data mining, khususnya *K-Means*.

2. Pengumpulan Data

Data dikumpulkan melalui dokumentasi dan wawancara dengan pihak kecamatan. Dataset yang diperoleh berjumlah 1000 data masyarakat dengan variabel terkait kondisi sosial-ekonomi, seperti pekerjaan, penghasilan, jumlah tanggungan, kepemilikan rumah, dan sumber listrik. Variabel inilah yang digunakan dalam membangun model pengelompokan.

3. Preprocessing Data

Tahap preprocessing dilakukan untuk memastikan data siap digunakan dalam algoritma *K-Means*. Tahapannya meliputi:

- Cleaning data: mengatasi *missing value*, menghapus atau memperbaiki data yang tidak konsisten, serta mengidentifikasi *outlier* yang dapat memengaruhi *clustering*.
- Transformasi data: mengonversi data kategorikal menjadi numerik menggunakan skala bobot.
- Seleksi fitur: memilih fitur relevan dan menghapus atribut yang tidak digunakan seperti nama, nomor telepon, dan status orang tua.

Hasil preprocessing menjadi input utama untuk tahap pengelompokan.

4. Model Pengelompokan Data (K-Means)

Pada tahap ini, algoritma K-Means diterapkan untuk mengelompokkan data masyarakat ke dalam dua cluster utama, yaitu Layak dan Tidak Layak.

Tahapan teknis meliputi:

- Menentukan jumlah cluster (k) berdasarkan hasil metode Elbow, yaitu $k = 2$.
- Menetapkan centroid awal secara acak dari dataset.
- Menghitung jarak setiap data ke masing-masing centroid menggunakan Euclidean Distance.
- Mengelompokkan data ke cluster terdekat.
- Menghitung centroid baru berdasarkan rata-rata setiap cluster.
- Mengulangi proses hingga memenuhi kriteria berhenti centroid tidak berubah (konvergen) atau mencapai jumlah iterasi maksimum.

Model K-Means menghasilkan pola pengelompokan berdasarkan kemiripan data.

5. Evaluasi Model

Model yang dihasilkan dievaluasi untuk memastikan kualitas *clustering* dan kesesuaiannya dengan kondisi aktual.

Teknik evaluasi meliputi:

- SSE (Sum of Squared Error): untuk melihat seberapa baik data berada dekat dengan centroid.
- Metode Elbow: digunakan lebih awal untuk menentukan jumlah cluster yang optimal ($k = 2$).
- Confusion Matrix: digunakan untuk membandingkan hasil cluster dengan data aktual pada 50 data uji, menghasilkan metrik evaluasi seperti akurasi, presisi, recall, spesifisitas, dan F1-score.

Evaluasi ini memastikan model memberikan hasil yang reliabel dan valid.

6. Pembangunan Sistem

Tahap ini hanya dijelaskan secara ringkas karena bukan fokus utama jurnal. Model K-Means yang telah diuji kemudian diintegrasikan ke dalam sistem pendukung keputusan yang dirancang menggunakan PHP dan MySQL. Sistem ini memudahkan pihak kecamatan dalam mengklasifikasikan penerima PKH secara otomatis dan objektif.

7. Selesai

Tahap ini menandakan seluruh proses penelitian telah selesai mulai dari identifikasi masalah, pengolahan data, pembangunan model pengelompokan, evaluasi, hingga implementasi dalam sistem pendukung keputusan.

2.2 Data Mining

Data mining merupakan suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari suatu data. Hasil dari pengolahan data dengan menggunakan metode data mining dapat digunakan untuk pengambilan keputusan di masa depan. Data mining kegiatan pengilangan data dengan skala yang besar, sehingga data mining memiliki peranan yang penting dalam berbagai bidang seperti bidang industri, keuangan, cuaca, ilmu dan teknologi. Sebenarnya data mining merupakan suatu langkah dalam *knowledge discovery in databases* (KDD). *Knowledge discovery* sebagai suatu proses terdiri atas data cleaning, data *integration*, data *selection*, data *transformation*, *pattern evaluation knowledge presentation*. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakut dari berbagai database besar. Istilah data mining dan *knowledge discovery in database* (KDD) sering digunakan untuk menjelaskan proses penggalian informasi yang tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, berkaitan antara satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining. [13]

2.3 K-Means

K-Means merupakan algoritma *clustering* yang berulang-ulang. Algoritma k-means menetapkan nilai-nilai *cluster* (k) secara acak, dimana nilai tersebut menjadi pusat dari *cluster* atau disebut sebagai *centroid*, *mean* atau *means*. Algoritma k-means dalam implementasinya sangat mudah, cepat, mudah beradaptasi sederhana untuk diimplementasikan dan dijalankan, relatif cepat, dan mudah beradaptasi serta mempunyai kemampuan yang besar dalam mengolah data yang cukup besar dan waktu lebih efisien. Yang menjadi kelemahan dalam algoritma k-means saat menentukan *cluster* awal, karena bergantung pada inisial data yang diberikan [14]

Algoritma K-means *Cluster Analysis* merupakan bidang penelitian dalam analisis dan data mining. Pada algoritma ini teknik pengelompokannya berdasarkan kemiripan data yang tidak memiliki acuan apapun (*unsupervised*). Tetapi, akan membagi keseluruhan data yang akan menjadi kelompok atau mempunyai kemiripan yang sama. Pada dasarnya algoritma ini menghitung jarak pada setiap data dengan pusat data (*centroid*) untuk mengukur kemiripan data. Metode ini digunakan bertujuan untuk meminimalisir fungsi objektif yang diatur pada proses *clustering* atau pengklasteran dengan meminimalkan variasi antar data yang terdapat dalam suatu *cluster* dan memaksimalkan variasi dengan data yang terdapat pada *cluster* lain [15]

Adapun prosedur perhitungan Algoritma K-Means *Clustering analysis* diuraikan sebagai berikut:

- Pilih secara acak k buah data sebagai pusat *cluster*
- Jarak antara data dan pusat *cluster* dihitung menggunakan *Euclidian Distance*. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$D_{euclidean}(x,y) = \sqrt{\sum (x_i - y_i)^2} \dots \dots \dots (1)$$

Keterangan:

$d(x,y)$: jarak antara data pada titik x dan y

x : titik data objek

y : titik data centroid

i : jumlah atribut data

3. Data ditempatkan dalam *cluster* yang terdekat, dihitung dari tengah *cluster*
4. Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat. Rumus menghitung titik pusat *cluster* baru:

$$v = \sum_{i=1}^n x_i; i = 1,2,3 \dots \dots \dots n \dots \dots \dots (2)$$

dimana

v adalah *centroid* pada *cluster*

X_i adalah objek ke-i

N adalah banyaknya objek/jumlah objek yang menjadi anggota *cluster*

5. Proses penentuan pusat *cluster* dan penempatan data dalam *cluster* diulangi sampai nilai *centroid* tidak berubah lagi [16]

3. HASIL DAN PEMBAHASAN

Data pada penelitian ini diperoleh dengan mengambil data pada Kantor Camat Medan Labuhan dengan menggunakan Variabel Independen : Pekerjaan suami, pekerjaan istri, penghasilan suami, penghasilan istri, jumlah tanggungan, kepemilikan rumah, sumber listrik. Adapun data yang di dapat dari Kantor Camat Medan Labuhan sebagaimana ditampilkan pada Tabel 1.

Tabel 1. Penggunaan Fitur dalam (PKH)

No	Nama Fitur	Keterangan
1	Nomor	Nomor urutan untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining
2	Nama Masyarakat	Nama masyarakat untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining
3	Jenis kelamin	Jenis kelamin dari masyarakat untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining
4	No Telepon	No telepon untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining
5	Status ayah	Status ayah masyarakat untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining
6	Status ibu	Status ibu masyarakat untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining
7	Nama ibu kandung	Nama ibu kandung untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining
8	Pekerjaan Suami	Pekerjaan suami dari untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining meliputi PNS, TNI/POLRI, Pegawai Swasta, Wirausaha / Wiraswasta, Lainnya, Petani, Tukang Becak dan Tidak Bekerja
9	Pekerjaan istri	Pekerjaan istri dari untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining meliputi PNS, TNI/POLRI, Pegawai Swasta, Wirausaha / Wiraswasta, Lainnya, Petani, Tukang Becak dan Tidak Bekerja
10	Penghasilan suami	Penghasilan suami dari untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining meliputi Tidak Berpenghasilan, 500.000 -1.500.000, 1.500.001-2.500.000, 2.500.001 – 3.500.000, 3.500.001-4.500.000, 4.500.001-7.000.000 dan >7.500.000

No	Nama Fitur	Keterangan
11	Penghasilan istri	Penghasilan istri dari untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining meliputi Tidak Berpenghasilan, 500.000 -1.500.000, 1.500.001-2.500.000, 2.500.001 – 3.500.000, 3.500.001-4.500.000, 4.500.001-7.000.000 dan >7.500.000
12	Jumlah tanggungan	Jumlah tanggungan orang tua dari untuk data masyarakat dalam Model Pengelompokkan Penerima Bantuan Sosial PKH dengan Teknik Data Mining meliputi status anak yg dimiliki.
13	Kepemilikan rumah	Status kepemilikan rumah dari orang tua terdiri dari Tidak Memiliki, Menumpang, sewa bulanan, sewa tahunan dan rumah sendiri
14	Sumber listrik	Sumber listrik untuk penerangan rumah dari orang tua terdiri dari tidak ad, PLN dan genser/ mandiri.

3.1.Hasil Penelitian

1. Preprocessing Data

Tahap preprocessing menghasilkan data yang bersih dan siap digunakan dalam pemodelan. Data kategorikal yang berkaitan dengan pekerjaan, penghasilan, jumlah tanggungan, kepemilikan rumah, dan sumber listrik dikonversi ke bentuk numerik sehingga dapat diproses oleh algoritma K-Means. Outlier yang terdeteksi juga ditangani agar tidak memengaruhi pembentukan centroid secara ekstrem. Setelah preprocessing, dataset menjadi lebih stabil dan representatif.

Tabel 2. Seleksi Fitur

No	Nama Fitur	Keterangan
1	Nomor	Digunakan
2	Nama Masyarakat	Digunakan
3	Jenis kelamin	Tidak digunakan
4	No Telepon	Tidak digunakan
5	Status ayah	Tidak digunakan
6	Status ibu	Tidak digunakan
7	Nama ibu kandung	Tidak digunakan
8	Pekerjaan Suami	Digunakan
9	Pekerjaan istri	Digunakan
10	Penghasilan suami	Digunakan
11	Penghasilan istri	Digunakan
12	Jumlah tanggungan	Digunakan
13	Kepemilikan rumah	Digunakan
14	Sumber listrik	Digunakan

2. Model K-Means

Selanjutnya pemilihan kluster dari *dataset* yang ada untuk penentuan *centroid* secara acak/*random* sebagai awal perhitungan untuk menentukan kedekatan jarak antara data dan pusat *cluster*. Memilih secara acak k buah data sebagai pusat pengelompokan (*cluster*). Penentuan *centroid* secara acak/*random* dapat dilihat pada Tabel 3.

Tabel 3. Centroid Awal

Centroid	Pekerjaan Suami	Pekerjaan Istri	Penghasilan Suami	Penghasilan Istri	Jumlah Tanggungan	Kepemilikan Rumah	Sumber Listrik	Cluster
1	6	7	6	7	2	5	2	C1 (Layak)

2									C2 (Tidak Layak)
	4	4	3	4	3	2	2		

Jarak data siswa 1 Kevin Sinaga dengan *centroid* 1, 2 :

$$d(ai,bj) = \sqrt{(5-6)^2 + (5-7)^2 + (5-6)^2 + (5-7)^2 + (5-2)^2 + (1-5)^2 + (2-2)^2} = 5.92$$

$$d(ai,bj) = \sqrt{(5-4)^2 + (5-4)^2 + (5-3)^2 + (5-4)^2 + (5-3)^2 + (1-2)^2 + (2-2)^2} = 3.46$$

Jarak data siswa 2 Mhd Irham dengan *centroid* 1, 2 :

$$d(ai,bj) = \sqrt{(6-6)^2 + (8-7)^2 + (6-6)^2 + (7-7)^2 + (2-2)^2 + (1-5)^2 + (2-2)^2} = 4.12$$

$$(ai,bj) = \sqrt{(6-4)^2 + (8-4)^2 + (6-3)^2 + (7-4)^2 + (2-3)^2 + (1-2)^2 + (2-2)^2} = 6.32$$

Hasil perhitungan jarak data pendaftar pada perhitungan- 1 dapat dilihat pada Tabel berikut ini:

Tabel 4. Hasil Perhitungan Jarak Data Dengan *Centroid* Awal Iterasi 1

No	Nama	Cluster 1	Cluster 2	Jarak Terdekat	Cluster
1	Kevin Singa	5,92	3,46	3,46	2
2	MHd Irham	4,12	6,32	4,12	1
3	WINDA MARIA ULFA	4,36	5,83	4,36	1
4	RUDI ALAMSYAH	3,87	3,74	3,74	2
5	MUSLIMAH	4,36	5,10	4,36	1
...
954	Kanaya Aulia Nisa	4,69	5,92	4,69	1
955	Kasih Indah Sembiring	4,69	4,80	4,69	1
956	Kevin delvira	3,87	5,29	3,87	1
957	KHAIRUL	2,00	4,58	2,00	1

Dalam proses K-Means, *centroid* berubah karena data terus berpindah cluster berdasarkan jarak minimal terhadap *centroid*. Setiap iterasi menghasilkan Perbaikan posisi *centroid*, karena dihitung dari rata-rata anggota cluster terbaru, Perubahan keanggotaan cluster, terutama pada awal iterasi ketika jarak antar data dan *centroid* masih jauh dari kondisi optimal. Perubahan *centroid* adalah indikasi bahwa Data memiliki sebaran yang heterogeny, Karakteristik sosial-ekonomi warga berbeda cukup signifikan antara cluster, Algoritma sedang mencari posisi *centroid* yang paling merepresentasikan pola data dan Perubahan berhenti pada iterasi ke-5 dan ke-6, menunjukkan algoritma telah konvergen, yaitu *centroid* tidak berubah lagi dan cluster stabil.

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut sehingga mendapatkan hasil perhitungan *centroid* baru yang akan digunakan untuk iterasi berikutnya. Pada *cluster* 1 yaitu layak terdapat 678 data, *cluster* 2 tidak layak sebanyak 279 data.

Tabel 5. Sample *Centroid*

Centroid	Pekerjaan Suami	Pekerjaan Istri	Penghasilan Suami	Penghasilan Istri	Jumlah Tanggungan	Kepemilikan Rumah	Sumber Listrik	Cluster
1	5,71	7,00	6,03	6,63	3,58	1,64	2,01	C1 (Layak)
2	4,82	5,13	5,68	6,14	4,55	1,19	2,01	C2 (Tidak Layak)

Jarak data siswa 1 Kevin Sinaga dengan *centroid* 1, 2 :

$$d(ai,bj) = \sqrt{(5-5.7)^2 + (5-7)^2 + (5-6.03)^2 + (5-6.63)^2 + (5-3.58)^2 + (1-1.64)^2 + (2-2.01)^2} = 3.26$$

$$d(ai,bj) = \sqrt{(5-4.82)^2 + (5-5.13)^2 + (5-5.68)^2 + (5-6.14)^2 + (5-4.55)^2 + (1-1.19)^2 + (2-2.01)^2}$$

= 1.43

Jarak data siswa 2 Mhd Irham dengan *centroid* 1, 2 :

$$d(a_i, b_j) = \sqrt{(6 - 5.71)^2 + (8 - 7)^2 + (6 - 6.03)^2 + (7 - 6.63)^2 + (2 - 3.58)^2 + (1 - 1.64)^2 + (2 - 2.01)^2}$$

= 2.03

$$(a_i, b_j) = \sqrt{(6 - 4.82)^2 + (8 - 5.13)^2 + (6 - 5.68)^2 + (7 - 6.14)^2 + (2 - 4.55)^2 + (1 - 1.19)^2 + (2 - 2.01)^2} =$$

4.13 dan seterusnya

Tabel 6. Hasil Perhitungan Jarak Data Dengan *Centroid* Awal Iterasi 2

No	Nama	Cluster 1	Cluster 2	Jarak Terdekat	Cluster
1	Kevin Singa	3,26	1,43	1,43	2
2	MHd Irham	2,03	4,13	2,03	1
3	WINDA MARIA ULFA	1,49	3,49	1,49	1
4	RUDI ALAMSYAH	2,73	3,33	2,73	1
5	MUSLIMAH	2,39	2,02	2,02	2
...
954	Kanaya Aulia Nisa	1,69	3,35	1,69	1
955	Kasih Indah Sembiring	1,43	1,61	1,43	1
956	Kevin delvira	3,50	2,99	2,99	2
957	KHAIRUL	3,26	4,12	3,26	1

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut sehingga mendapatkan hasil perhitungan *centroid* baru yang akan digunakan untuk iterasi berikutnya. Pada *cluster* 1 terdapat 612 data, *cluster* 2 sebanyak 345 data.

Dimana nilai rata-ratanya dapat dilihat pada dibawah ini:

Tabel 7. Sample Centroid

Centroid	Pekerjaan Suami	Pekerjaan Istri	Penghasilan Suami	Penghasilan Istri	Jumlah Tanggungan	Kepemilikan Rumah	Sumber Listrik	Cluster
1	5,72	7,20	5,95	6,61	3,35	1,54	2,02	C1 (Layak)
2	4,98	5,13	5,89	6,26	4,77	1,45	1,99	C2 (Tidak Layak)

Jarak data siswa 1 Kevin Sinaga dengan *centroid* 1, 2 :

$$d(a_i, b_j) = \sqrt{(5 - 5.72)^2 + (5 - 7.20)^2 + (6 - 5.95)^2 + (6 - 6.61)^2 + (5 - 3.35)^2 + (1 - 1.54)^2 + (2 - 2.02)^2}$$

= 3.44

$d(a_i, b_j) =$

$$\sqrt{(5 - 3.62)^2 + (5 - 5.37)^2 + (6 - 5.56)^2 + (6 - 6.50)^2 + (5 - 4.75)^2 + (1 - 1.69)^2 + (2 - 2.02)^2}$$

= 1.62 dan seterusnya

Tabel 8. Hasil Perhitungan Jarak Data Dengan *Centroid* Awal Iterasi 3

No	Nama	Cluster 1	Cluster 2	Jarak Terdekat	Cluster
1	Kevin Singa	3,44	1,62	1,62	2
2	MHd Irham	1,73	4,21	1,73	1
3	WINDA MARIA ULFA	1,23	3,56	1,23	1
4	RUDI ALAMSYAH	2,66	3,49	2,66	1

No	Nama	Cluster 1	Cluster 2	Jarak Terdekat	Cluster
5	MUSLIMAH	2,65	1,81	1,81	2
...
954	Kanaya Aulia Nisa	1,58	3,38	1,58	1
955	Kasih Indah Sembiring	1,61	1,63	1,61	1
956	Kevin delvira	3,78	2,68	2,68	2
957	KHAIRUL	3,27	4,11	3,27	1

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut sehingga mendapatkan hasil perhitungan *centroid* baru yang akan digunakan untuk iterasi berikutnya. Pada *cluster* 1 terdapat 614 data, *cluster* 2 sebanyak 343 data.

Tabel 9. Hasil Perhitungan Jarak Data Dengan *Centroid* Awal Iterasi 4

No	Nama	Cluster 1	Cluster 2	Jarak Terdekat	Cluster
1	Kevin Singa	3,45	1,67	1,67	2
2	MHd Irham	1,67	4,30	1,67	1
3	WINDA MARIA ULFA	1,25	3,61	1,25	1
4	RUDI ALAMSYAH	2,64	3,56	2,64	1
5	MUSLIMAH	2,74	1,66	1,66	2
...
954	Kanaya Aulia Nisa	1,59	3,42	1,59	1
955	Kasih Indah Sembiring	1,69	1,61	1,61	2
956	Kevin delvira	3,83	2,61	2,61	2
957	KHAIRUL	3,28	4,14	3,28	1

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut sehingga mendapatkan hasil perhitungan *centroid* baru yang akan digunakan untuk iterasi berikutnya. Pada *cluster* 1 terdapat 587 data, *cluster* 2 sebanyak 370 data.

Tabel 10. Hasil Perhitungan Jarak Data Dengan *Centroid* Awal Iterasi 5

No	Nama	Cluster 1	Cluster 2	Jarak Terdekat	Cluster
1	Kevin Singa	3,51	1,72	1,72	2
2	MHd Irham	1,66	4,20	1,66	1
3	WINDA MARIA ULFA	1,27	3,50	1,27	1
4	RUDI ALAMSYAH	2,71	3,45	2,71	1
5	MUSLIMAH	2,86	1,49	1,49	2
...
954	Kanaya Aulia Nisa	1,57	3,36	1,57	1
955	Kasih Indah Sembiring	1,81	1,44	1,44	2

No	Nama	Cluster 1	Cluster 2	Jarak Terdekat	Cluster
956	Kevin delvira	3,86	2,67	2,67	2
957	KHAIRUL	3,31	4,06	3,31	1

Setelah semua data ditempatkan ke dalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut sehingga mendapatkan hasil perhitungan *centroid* baru yang akan digunakan untuk iterasi berikutnya. Pada *cluster* 1 terdapat 542 data, *cluster* 2 sebanyak 415 data.

Dimana nilai rata-ratanya dapat dilihat pada dibawah ini:

Tabel 11. Hasil Perhitungan Jarak Data Dengan *Centroid* Awal Iterasi 6

No	Nama	Cluster 1	Cluster 2	Jarak Terdekat	Cluster
1	Kevin Singa	3,54	1,81	1,81	2
2	MHD Irham	1,65	4,13	1,65	1
3	WINDA MARIA ULFA	1,28	3,44	1,28	1
4	RUDI ALAMSYAH	2,76	3,37	2,76	1
5	MUSLIMAH	2,94	1,37	1,37	2
...
954	Kanaya Aulia Nisa	1,55	3,33	1,55	1
955	Kasih Indah Sembiring	1,89	1,36	1,36	2
956	Kevin delvira	3,90	2,67	2,67	2
957	KHAIRUL	3,36	3,98	3,36	1

Setelah iterasi berhenti, diperoleh Cluster Layak = 542 warga dan Cluster Tidak Layak = 415 warga. Perbedaan jumlah ini menggambarkan bahwa sebagian besar warga Medan Labuhan memiliki karakteristik sosial-ekonomi yang memenuhi kriteria kelayakan PKH, terutama terkait penghasilan rendah, jumlah tanggungan tinggi, dan kondisi perumahan yang kurang layak. Temuan ini dapat membantu kantor kecamatan untuk Menentukan prioritas bantuan berdasarkan pola data, bukan subjektivitas petugas, Mengidentifikasi warga yang rentan tetapi mungkin terlewat dalam pendataan manual dan Melakukan verifikasi ulang pada warga yang berada dekat batas antar cluster.

3.2. Pembahasan

Perhitungan SSE (Sum of Squared Error) dilakukan untuk $k = 1$ sampai $k = 6$. Hasil SSE menunjukkan penurunan paling signifikan dari $k = 1$ ke $k = 2$, sementara mulai $k = 3$ penurunannya relatif kecil. Titik siku (elbow point) terlihat pada $k = 2$, sehingga jumlah cluster optimal untuk pengelompokan warga adalah dua kelompok. Dua cluster tersebut kemudian diinterpretasikan sebagai:

Cluster 1 = Layak

Cluster 2 = Tidak Layak

Penentuan $k = 2$ sejalan dengan konteks kebijakan PKH yang memang membagi warga dalam kategori layak dan tidak layak, sehingga hasil metode elbow memperkuat validitas keputusan pengelompokan.

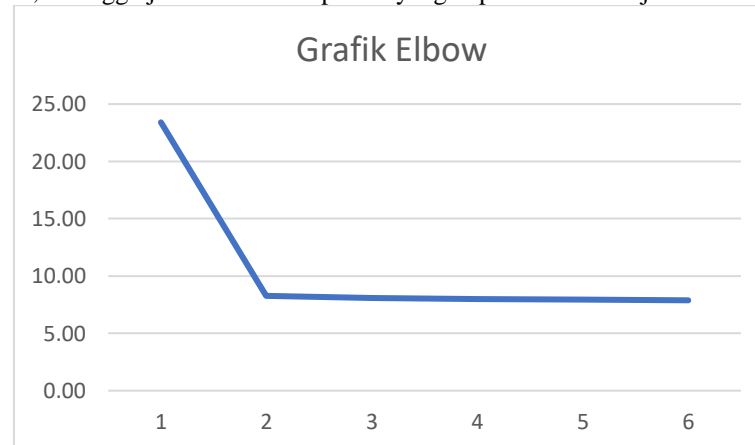
Tabel 12. Hasil SSE

K	Hasil SSE	Selisih
1	23,41	0
2	8,28	15,12
3	8,09	0,20
4	7,99	0,09

5 7,93 0,06

6 7,88 0,05

Dari hasil Tabel diatas, kemudian dilakukan metode optimasi menggunakan metode *elbow* untuk menentukan jumlah *cluster* terbaik. Berikut adalah hasil perhitungan SSE berdasarkan pengujian nilai $k=1$ sampai dengan $k=5$. Dari perhitungan SSE metode Elbow pada Tabel diatas, didapatkan nilai SSE yang memiliki selisih terbesar terdapat pada nilai $k=2$, sehingga jumlah *cluster* optimal yang dapat dibentuk sejumlah 3 *cluster*.



Gambar 2. Grafik SSE Metode Elbow

Dari grafik SSE metode *Elbow* pada Gambar diatas terlihat bahwa pada $k=2$ mengalami penurunan paling besar jika dibandingkan jumlah k yang lain. Jumlah *cluster* selain $k=2$ mengalami penurunan yang stabil. Maka dapat disimpulkan jumlah *cluster* optimal yang terbentuk sejumlah 3 *cluster*.

Hasil pengujian sistem dapat ditampilkan dari nilai elbow untuk tiap iterasinya memiliki selisih yang cukup signifikan, untuk iterasi 1 dengan kedua menampilkan selisih 15.12, iterasi 2 dengan 3 memiliki 0.2 dan iterasi 3 dan 4 memiliki selisih 0.9 dan kemudian 0.06 dan terakhir adalah 0.05 dan menampilkan aturan dari hasil Analisa K-Means sebanyak 957 aturan dalam penentuan kelayakan penerimaan PKH. Berikut ini merupakan hasil evaluasi untuk data baru sebanyak 50 data masyarakat :

Tabel 13. Evaluasi data Baru

No Nama Alternatif	Data actual	Keputusan
1 Tiara Sirait	Layak	Layak
2 Latika Agustina	Tidak Layak	Tidak Layak
3 Uli Purwanti	Layak	Layak
4 Purwadi Budiman	Layak	Layak
5 Maida Wulandari	Layak	Layak
6 Zulfa Utama	Layak	Layak
7 Mursinin Kurniawan	Tidak Layak	Tidak Layak
8 Faizah Winarsih	Tidak Layak	Tidak Layak
9 Shania Lailasari	Tidak Layak	Tidak Layak
10 Farah Aryani	Tidak Layak	Tidak Layak
11 Rangga Pertiwi	Layak	Layak
12 Jaiman Uyainah	Layak	Tidak Layak
13 Lala Marpaung	Layak	Layak
14 Hesti Rahimah	Layak	Layak
15 Kamila Saragih	Layak	Layak
16 Ida Iswahyudi	Layak	Layak
17 Padmi Winarsih	Tidak Layak	Tidak Layak
18 Legawa Laksita	Tidak Layak	Tidak Layak
19 Unggul Zulaika	Layak	Layak
20 Kiandra Pratiwi	Tidak Layak	Tidak Layak
21 Muhammad Natsir	Layak	Layak
22 Umay Uyainah	Layak	Layak
23 Adhijarja Budiman	Tidak Layak	Layak

No Nama Alternatif	Data actual	Keputusan
24 Yance Wibowo	Layak	Layak
25 Adikara Oktaviani	Layak	Layak
26 Mahesa Waskita	Tidak Layak	Tidak Layak
27 Hasim Adriansyah	Tidak Layak	Tidak Layak
28 Bakidin Mustofa	Tidak Layak	Tidak Layak
29 Mariadi Kusmawati	Tidak Layak	Tidak Layak
30 Queen Mardhiyah	Layak	Layak
31 Yahya Nuraini	Layak	Layak
32 Purwadi Sitompul	Layak	Layak
33 Johan Hutagalung	Layak	Layak
34 Marwata Pudjiastuti	Tidak Layak	Tidak Layak
35 Kayla Prayoga	Layak	Layak
36 Oni Puspita	Layak	Tidak Layak
37 Yance Hardiansyah	Layak	Layak
38 Nugraha Pradipta	Layak	Layak
39 Vega Maryati	Tidak Layak	Tidak Layak
40 Tomi Wulandari	Layak	Layak
41 Kurnia Megantara	Tidak Layak	Tidak Layak
42 Warta Megantara	Layak	Layak
43 Padmi Napitupulu	Layak	Layak
44 Sidiq Kurniawan	Tidak Layak	Tidak Layak
45 Samiah Yuliarti	Layak	Layak
46 Usman Usada	Layak	Layak
47 Silvia Lailasari	Layak	Layak
48 Nurul Hidayah	Layak	Tidak Layak
49 Rosaria Marbun	Layak	Layak
50 Elissa Tampubolon	Layak	Layak

Dari data diatas dapat diketahui hasil perhitungan confusion matriks sebagai berikut :

Tabel 14. Perbandingan: Data Aktual vs Keputusan Sistem

	Prediksi: Layak	Prediksi: Tidak Layak
Aktual: Layak	29 (True Positive / TP)	3 (False Negative / FN)
Aktual: Tidak Layak	1 (False Positive / FP)	17 (True Negative / TN)

Dari table diatas diketahui hasil perhitungan data evaluasi

- Akurasi

$$= (TP + TN) / \text{Total Data}$$

$$= (29 + 17) / 50$$

$$= 92\%$$
- Presisi (Precision)

$$= TP / (TP + FP)$$

$$= 29 / (29 + 1)$$

$$= 96.67\%$$
- Recall (Sensitivity)

$$= TP / (TP + FN)$$

$$= 29 / (29 + 3)$$

$$= 90.63\%$$
- Spesifisitas (Specificity)

$$= TN / (TN + FP)$$

$$= 17 / (17 + 1)$$

$$= 94.44\%$$
- F1-Score

$$= 2 * (\text{Presisi} * \text{Recall}) / (\text{Presisi} + \text{Recall})$$

$$= 2 * (0.9667 * 0.9063) / (0.9667 + 0.9063)$$

$$\approx 93.52\%$$

Dari perhitungan diatas diambil kesimpulan bahwa Presisi tinggi menandakan sistem sangat baik dalam menandai warga yang benar-benar layak menerima bantuan. Recall tinggi menunjukkan model tidak banyak melewatkan warga yang seharusnya layak. Spesifisitas tinggi berarti model efektif menghindari kesalahan pemberian bantuan kepada warga yang tidak layak. Hasil penelitian ini kompetitif dan tetap kuat mengingat jumlah data lebih besar dan variabel lebih beragam dibanding penelitian sebelumnya.

Tabel 15. Perbandingan: Penelitian Terdahulu

Peneliti	Metode	Akurasi	Keterangan
Mulyadien & Enri (2022)	K-Means (BLT)	Rasio 93%	Dataset kecil, variabel terbatas
Putra & Anggrawan (2021)	K-Means	Rasio cluster	Tidak menggunakan uji akurasi
Utami & Devi (2022)	Naïve Bayes	95.83%	Klasifikasi, bukan <i>clustering</i>
Penelitian ini	K-Means + Evaluasi Confusion Matrix	92%	Dataset lebih besar dan variabel lebih kompleks

4. KESIMPULAN

Penelitian ini dilakukan untuk mengatasi permasalahan ketidaktepatan sasaran dalam penentuan penerima bantuan sosial PKH di Kantor Camat Medan Labuhan, yang selama ini masih dilakukan secara manual dan rentan subjektivitas. Melalui penerapan algoritma K-Means, penelitian ini berhasil mengelompokkan data masyarakat ke dalam dua kategori utama, yaitu layak dan tidak layak, berdasarkan variabel sosial-ekonomi yang relevan. Hasil evaluasi menunjukkan bahwa model memiliki performa yang baik dan mampu menggambarkan pola kelayakan masyarakat secara objektif. Kontribusi dari penelitian ini adalah tersedianya model pengelompokan berbasis data mining yang dapat digunakan sebagai alat bantu pendukung keputusan dalam proses seleksi penerima PKH. Model ini membantu meningkatkan efisiensi, mengurangi potensi kesalahan subjektif, serta mempercepat proses pengolahan data dalam jumlah besar sehingga kebijakan bantuan sosial dapat diterapkan dengan lebih tepat sasaran dan transparan.

REFERENCES

- [1] R. R. A. Aria, S. Susilowati, and I. R. Rahadjeng, "Data Mining Menentukan Cluster Penerima Program Bantuan dengan Metode K-Means," *Remik*, vol. 7, no. 1, pp. 291–300, 2023, doi: 10.33395/remik.v7i1.12030.
- [2] C. O. Putri, D. M. Efendi, and R. Rustam, "Classification of Social Assistance Recipients Using Machine Learning," *PIKSEL Penelit. Ilmu Komput. Sist. Embed. Log.*, vol. 12, no. 2, pp. 241–250, 2024, doi: 10.33558/piksel.v12i2.9550.
- [3] E. Moruk, M. Martanto, and U. Hayati, "Algoritma K-Means Untuk *Clustering* Penerima Program Keluarga Harapan Di Nanaeoe Nusa Tenggara Timur," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1681–1687, 2024, doi: 10.36040/jati.v8i2.9148.
- [4] Y. Filki, "Algoritma K-Means *Clustering* dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa," *J. Inform. Ekon. Bisnis*, vol. 4, pp. 166–171, 2022, doi: 10.37034/infeb.v4i4.166.
- [5] A. I. Purnama, A. Aziz, and A. S. Wiguna, "Penerapan Data Mining Untuk Mengklasifikasi Penerima Bantuan Pkh Desa Wae Jare Menggunakan Metode Naïve Bayes," *Kurawal - J. Teknol. Inf. dan Ind.*, vol. 3, no. 2, pp. 173–180, 2020, doi: 10.33479/kurawal.v3i2.348.
- [6] R. Andrea and N. Nursobah, "Penerapan Algoritma K-Medoids Untuk Pengelompokan Data Penerima Bantuan Uang Kuliah Tunggal Bagi Mahasiswa Terdampak Covid-19," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 632–638, 2022, doi: 10.47065/bits.v3i4.1294.
- [7] L. G. Rady Putra and A. Anggrawan, "Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 205–214, 2021, doi: 10.30812/matrik.v21i1.1554.
- [8] M. Khandava Mulyadien and U. Enri, "Algoritma K-Means Untuk Pengelompokan Bantuan Langsung Tunai (BLT)," *J. Ilm. Wahana Pendidik.*, vol. 2022, no. 12, pp. 198–210, 2022.
- [9] C. J. Siti Mariam, Fitri Handayani, "Penerapan Algoritma *Clustering* K-Means Untuk Menentukan Prioritas Penerima Bantuan Rumah Akibat Bencana Alam," *J. Tek. Inform. dan Sist. Inf.*, vol. 10, no. 2, pp. 231–240, 2023.
- [10] D. Utami, P. Aisyiyah, and R. Devi, "544186584," vol. 07, pp. 1373–1384, 2022.
- [11] D. Bahtiar *et al.*, "Pemetaan Penduduk Penerima Bantuan Sosial Desa Waru Jaya Menggunakan Algoritma K-Means *Clustering*," *Sci. Sacra J. Sains*, vol. 3, no. 2, pp. 29–39, 2023, [Online]. Available: <http://pijarpemikiran.com/index.php/Scientia>
- [12] F. Febriansyah and S. Muntari, "Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 8, no. 1, pp. 66–77, 2023, doi: 10.14421/jiska.2023.8.1.66-77.
- [13] N. A. Maori, "Metode Elbow dalam Optimalisasi Jumlah Cluster pada K-Means *Clustering*," *J. SIMETRIS*, vol. 14, no. 2, pp. 277–287, 2023.
- [14] B. Susarianto and T. Nizami, "Penentuan Penerima Bantuan Sosial Beras untuk Masyarakat Miskin Menggunakan Metode K-Means," *Jutisi J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 12, no. 2, p. 772, 2023, doi:



- 10.35889/jutisi.v12i2.1424.
- [15] N. Siti Paridah and M. Martanto, "Klasterisasi Penerima Dana Bantuan Program Keluarga Harapan Menggunakan Metode K-Means Pada Desa Gereba," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 1036–1043, 2024, doi: 10.36040/jati.v8i1.8873.
- [16] E. Dwiguna and A. Bahtiar, "Penerapan Data Mining Untuk Menentukan Penerima Bantuan Blt Menggunakan Metode Clustering K-Means Pada Desa Pamulihan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1382–1388, 2024, doi: 10.36040/jati.v8i2.9029.
- [17] Renaldi, G., & Tanti, L. (2024, December). Penerapan Data Mining Dalam Sistem Penilaian Kinerja Karyawan Menggunakan Metode Algoritma C4. 5 Pada Pt Raharja Sinergi Komunikasi. In *PROSIDING SEMINAR NASIONAL MULTI DISIPLIN ILMU (SENADIMU)* (Vol. 1, No. 1, pp. 447-459).
- [18] Thanri, Y. Y., Iriani, J., Tanti, L., & Mukti, S. (2025). PENERAPAN ALGORITMA C4. 5 DALAM KLASIFIKASI KUALITAS BUAH PISANG BERDASARKAN KARAKTERISTIK FISIK DAN ORGANOLEPTIK: APPLICATION OF THE C4. 5 ALGORITHM IN CLASSIFYING BANANA FRUIT QUALITY BASED ON PHYSICAL AND ORGANOLEPTIC CHARACTERISTICS. *Journal of Computer Science and Technology Research (J-CSTR)*, 1(2), 64-72.