

Optimization of Social Assistance Recipient Determination Using K-Means Clustering Algorithm and K-Nearest Neighbour Algorithm

Optimasi Penentuan Penerima Bantuan Sosial menggunakan Algoritma K-Means Clustering dengan K-Nearest Neighbor

Graha Prakarsa^{1*)}, Nira Rahadiyanti²⁾, Reni Nursyanti³⁾ dan Ratih Hadianitini⁴⁾
(Cambria math, 11 pt, Bold, Centered)

¹⁾Universitas Islam Nusantara

Email: gprakarsa@gmail.com

^{2,3,4)}Universitas Informatika dan Bisnis Indonesia

*) *Corresponding author*

Abstract: *Determining the status of families as recipients of social assistance is very important in order to ensure that assistance is distributed accurately. Data mining techniques such as clustering and classification can be used to improve the quality of social assistance distribution policies. The purpose of this study is to determine the recipients of social assistance accurately. K-Means Clustering and K-Nearest Neighbour are two data mining problem-solving algorithms used in this study to select social assistance recipients. Although both algorithms are reliable, in order to be used widely, it is necessary to examine which algorithm has higher accuracy in determining social assistance recipients. Based on this, this study compares the K-Means Clustering and K-Nearest Neighbour algorithms in selecting social assistance recipients. The comparison was conducted using 1,760 data points. Based on the tests conducted, recipients of assistance using K-Means Clustering accounted for 65.145%, while K-Nearest Neighbour accounted for 99.650%. The study concluded that the K-Nearest Neighbour problem-solving algorithm has higher accuracy.*

Keywords: *Aid Recipients, Data Mining, K-Means, K-Nearest Neighbor*

Abstrak: Penentuan status keluarga sebagai penerima bantuan sosial merupakan hal sangat penting, agar bantuan mampu disalurkan secara akurat. Teknik data mining seperti klastering & klasifikasi dapat digunakan untuk meningkatkan kualitas penentuan kebijakan penyaluran bantuan sosial. Tujuan penelitian ini adalah untuk menentukan penerima bantuan sosial dengan tepat sasaran. K-Means Clustering dan K-Nearest Neighbor merupakan 2 algoritma pemecahan permasalahan data mining yang dipakai dalam penelitian ini untuk memilih penerima bantuan sosial. Meskipun kedua algoritma tersebut sudah handal, namun, untuk bisa dipakai secara luas, perlu diteliti algoritma mana yang mempunyai akurasi lebih tinggi pada permasalahan penentuan penerima bantuan sosial. Berdasarkan hal tersebut, pada penelitian ini dilakukan perbandingan algoritma K-Means Clustering dan K- Nearest Neighbor pada masalah memilih penerima bantuan. Perbandingan dilakukan menggunakan 1760 data. Berdasarkan pengujian yang sudah dilakukan, penerima bantuan menggunakan k-means clustering mendapatkan sebanyak 65,145% sementara K- Nearest Neighbor sebanyak 99,650%. Hal penelitian tersebut menyimpulkan bahwa algoritma pemecahan masalah K-Nearest Neighbor mempunyai akurasi yang lebih tinggi.

Kata Kunci: Penerima Bantuan, Data Mining, K-Means, K-Nearest Neighbor

DOI: <https://doi.org/10.37577/sainteks.v7i02.952>

Received: 07, 2025. Accepted: 08, 2025.

Published: 09, 2025

PENDAHULUAN

Sejalan dengan kemajuan teknologi informasi, kemampuan teknologi untuk mengatasi masalah juga semakin meningkat dengan mudah karena penerapan teknologi itu. Salah satu bentuk penerapan teknologi tersebut adalah penggunaan teknik data mining. Data mining adalah langkah menggunakan cara-cara matematika, statistik, kecerdasan buatan, serta machine

learning untuk menemukan dan mengenali informasi yang berguna serta pengetahuan yang berkaitan dari berbagai basis data besar. (Noviana et al., 2019).

Terdapat beberapa penelitian terdahulu yang menggunakan algoritma K-Means dalam memprediksi suatu hal. Penelitian yang pertama dalam kasus Clustering Kelompok Swadaya Masyarakat (KSM) dalam Menentukan Kebijakan Bantuan Badan Pemberdayaan Masyarakat di Kota Surabaya menggunakan Self-Organizing Map (SOM) dan K-Means mendapatkan hasil presentase 80,56% (Hafidz, 2012). Penelitian kedua dalam kasus Penerapan Data Mining Menentukan Kelompok Prioritas Penerima Bantuan Beras Rastra dengan K-Means Clustering mendapatkan hasil akurasi dari 3 cluster, cluster pertama mendapatkan hasil presentase 21%, cluster kedua mendapatkan hasil presentase 66%, dan cluster ketiga mendapatkan hasil presentase 13% (Noviana et al., 2019).

Penelitian ketiga dalam kasus pengkombinasian Pohon Keputusan dan K-Means Clustering untuk penentuan Bantuan Langsung Tunai di Indonesia mendapatkan hasil perhitungan K-Means hasil akurasi 3 cluster, cluster pertama 3,334, cluster kedua 6 dan cluster ketiga 8,667 dengan tingkat error sebesar 2,47. Sedangkan hasil perhitungan Decision Tree menunjukkan nilai entropi 1 (Kusanti & Sutanto, 2021). Penelitian keempat dalam kasus Clustering Daerah Miskin di Provinsi Riau Menggunakan Algoritma K-Means mendapatkan hasil yang sangat baik dan mencapai tingkat keakuratan 85% (Isyarah, et al., 2020). Penelitian kelima dalam kasus Pengelompokan Tingkat Kesejahteraan Keluarga sebagai upaya Program Kartu Indonesia Pintar dengan menggunakan Algoritma K-Means Clustering, mendapatkan hasil akurasi sebesar 69% (Eric Fammaldo, 2012). Adapun pengujian tentang algoritma K-Means menjelaskan dalam kasus pengkalsteran menggunakan K-Means Algorithm, bahwa algoritma K-Means tidak hanya menjaga efisiensi tinggi K-Means, tetapi juga meningkatkan kecepatan konvergensi secara efektif dengan meningkatkan cara memilih titik fokus cluster awal. Keuntungannya akan semakin jelas ketika sampai pada titik masalah cluster yang memiliki data berskala besar dan terdistribusi acak sepenuhnya (Li & Wu, 2012).

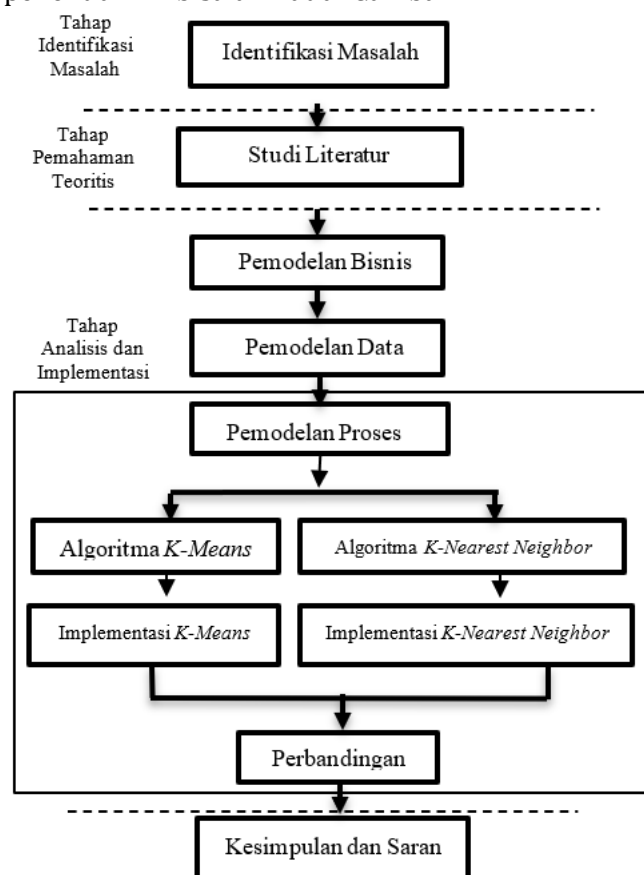
Selain itu, adapun penelitian keenam terkait prediksi sesuatu hal menggunakan algoritma K-Nearest Neighbor, diantaranya dalam kasus pengklasifikasian Penerima Dana Bantuan Desa hasilnya menunjukkan kasus tersebut menggunakan jumlah K adalah 5 mendapatkan hasil akurasi dengan K1 dan K2 pada kategori "Tidak Layak" sebesar 100%. Selanjutnya tingkat akurasi dengan K3, K4 dan K5 memiliki kategori "Layak" sebesar 81,25% (Hasanah et al., 2019). Penelitian ketujuh dalam kasus pengkalsifikasian Bidikmisi pada Calon Mahasiswa menggunakan Algoritma K-Nearest Neighbor dengan menghasilkan nilai akurasi tertinggi sebesar 97.21% (Jaman, 2019). Penelitian kedelapan dalam kasus penentuan Kelayakan Keluarga Penerima Bantuan Pangan Non Tunai Studi Kasus Kelurahan Karya Jaya menggunakan Algoritma K-Nearest Neighbor mendapatkan hasil akurasi 75,79% (Yani et al., 2020). Penelitian kesembilan dalam kasus Penentuan Penerima Dana Bantuan Operasional Pendidikan Lembaga PAUD dengan K-Nearest Neighbor dan K-Means menghasilkan perhitungan yang akurat serta tepat untuk memutuskan prioritas dana bantuan operasional pendidikan satuan PAUD (Soeleman, 2019). Berdasarkan penelitian-penelitian yang disebutkan di atas, dapat disimpulkan bahwa hasil penilain kinerja yang baik ditunjukkan oleh penggunaan algoritma K-Means dan K-Nearest Neighbor karena menunjukkan tingkat keakuratan besar karena mendekati 100% dan memiliki tingkat error sangat minimum karena mendekati 0%.

Dikarenakan pada penelitian sebelumnya belum membahas terkait tentang memprediksi penyaluran dana bantuan sosial, untuk itu dalam penelitian ini akan memprediksi penyaluran bantuan sosial dengan menggunakan algoritma K-Means Clustering dengan K-Nearest Neighbor. Namun, untuk melakukan prediksi penerima bantuan, perlu diketahui algoritma mana yang memiliki performa lebih tinggi. Maka, penelitian ini akan membandingkan kedua algoritma tersebut untuk mengetahui akurasi algoritma terbaik diantara K-Means dan K-Nearest Neighbor dalam kasus memprediksi penerima bantuan sosial. Penentuan status keluarga menjadi penerima bantuan adalah hal yang sangat krusial supaya bantuan sosial berdasarkan pemerintah bisa

disalurkan secara akurat. Namun, melihat kondisi masyarakat yang masih banyak benar-benar tidak mampu belum menerima bantuan, sedangkan masyarakat yang mampu malah mendapatkan bantuan dari pemerintah. Kerumitan dalam penyaluran bantuan dikarenakan banyaknya data penduduk di Kecamatan Gedebage, maka dari itu dipertimbangkan dengan beberapa kriteria diantaranya pekerjaan, jumlah penghasilan, dan jumlah tanggungan. Nilai dari setiap kategori itu digunakan sebagai acuan untuk memilih warga yang paling utama untuk menerima bantuan sosial.

METODOLOGI

Penelitian yang dilakukan penulis berjenis engineering research dimana proses penelitian menggunakan model proses RAD (Yogachi et al., 2020), metode pengembangan berorientasi objek (Pramudito., 2024), dan pemodelan data mining (Permana et al., 2020). Dalam kajian ini, penulis melakukan studi berdasarkan rencana penelitian. Langkah-langkah yang diambil untuk mencapai tujuan dari penelitian ini bisa dilihat di Gambar 1.



Gambar 1 Tahapan Penelitian

Tahap mengidentifikasi permasalahan bertujuan untuk mengetahui permasalahan yang sedang dialami. Pada tahap ini, dijelaskan masalah apa yang akan diteliti. Masalah yang diangkat nanti akan dipelajari dan akan diperoleh informasi yang berhubungan dengan masalah itu. Tahap pemahaman teoritis bertujuan untuk mencari referensi ilmiah yang terkait dengan permasalahan penelitian guna mendukung penelitian. Prosesnya tentu melalui pencarian bahan dari buku, jurnal ilmiah, dan hasil penelitian terkait penerapan algoritma k-means clustering dan algoritma k-nearest neighbor. Tahap analisis menggunakan tahapan pertama, kedua dan ketiga dalam pengimplementasian RAD, yaitu pemodelan bisnis, pemodelan data dan pemodelan proses. Dalam tahap pemodelan bisnis, penulis mendapatkan data yang dibutuhkan untuk dilakukan perhitungan dalam aplikasi yang dipakai. Teknik pengumpulan data yang dilakukan penulis

adalah wawancara. Penulis melakukan wawancara kepada Ketua Seksi Kesejahteraan Sosial di Kecamatan Gedebage dengan memberikan beberapa pertanyaan yang menghasilkan sekumpulan data penduduk seperti jumlah penduduk, tingkat pendidikan, tingkat usia, data mata pencaharian. Tahap pemodelan data adalah menganalisis setiap kebutuhan yang sebelumnya dikumpulkan guna memahami kebutuhan yang diperlukan dalam perhitungan yang dilakukan (Bhagya & Prakarsa, 2020). Tahap selanjutnya yaitu pemodelan proses, merupakan tahapan ketiga dalam pengimplementasian RAD. Tujuannya untuk mengidentifikasi ataupun mencari pola pada data, dimana pola yang ditemukan bisa berguna dan bisa dimengerti adalah proses yang dilakukan oleh Knowledge Discovery in Database (KDD). Untuk penerapan KDD terdapat proses data mining, di mana cara ini melibatkan satu atau lebih metode pembelajaran mesin untuk mempelajari dan mendapatkan informasi secara otomatis, atau serangkaian langkah untuk menemukan nilai baru dari kumpulan data yang sebelumnya tidak diketahui dengan cara manual. Setelah menganalisis kebutuhan, penulis menggunakan algoritma K-Means Clustering dan algoritma K-Nearest Neighbor sebagai dasar dalam pengambilan keputusan yang tepat. Pola yang dibuat kemudian diuji pada aplikasi yang telah ada untuk menghasilkan kebijakan dalam pengambilan keputusan. Setelah melakukan tahap pemodelan proses, penulis melakukan implementasi dari dua algoritma tersebut. Setelah melakukan perhitungan algoritma k-means clustering dan algoritma k-nearest neighbor, hasil yang telah diperoleh lalu penulis melakukan perbandingan dengan tujuan untuk mengetahui akurasi hasil penentuan penerima bantuan diantara algoritma K-Means dan algoritma K-Nearest Neighbor.

a. K-Means Clustering

Algoritma K-Means Clustering adalah algoritma yang berprinsip partitional, dikarenakan K-Means Clustering didasarkan pada awalnya dengan pendefinisian nilai centroid dengan penentuan jumlah awal kelompok (Prasetyo, 2014). Algoritma K-Means Clustering bekerja dengan cara yang diulang-ulang untuk mendapatkan kelompok data. Kita perlu memberi tahu berapa banyak kelompok yang kita inginkan sebagai input, dan hasilnya akan menunjukkan jumlah kelompok akhir. Jika algoritma ini digunakan untuk membuat K kelompok, maka kita akan mulai dengan K awal dan berakhir dengan K akhir. Algoritma K-Means akan secara acak memilih titik awal centroid sebagai pola. Berapa kali kita harus ulang untuk menemukan posisi centroid kelompok akan ditentukan oleh titik awal centroid yang dipilih secara acak, terutama jika posisi centroid baru tidak berubah. Nilai K yang dipilih sebagai pusat awal dihitung menggunakan rumus Jarak Euclidean, yang mencari jarak terdekat antara titik centroid dengan data atau objek. Data yang memiliki jarak terdekat ke centroid akan membentuk sebuah kelompok. Adapun langkah-langkah berdasarkan algoritma K-Means adalah: Tentukan nilai K sebagai jumlah *cluster* yang akan dibentuk. Lalu selanjutnya inisialisasi K pusat *cluster* dengan cara acak/random. Setelah itu, ukur seberapa jauh setiap data dari setiap pusat kelompok menggunakan rumus Euclidean Distance. Tujuannya adalah untuk mencari jarak terpendek antara setiap data dengan pusat kelompok. Berikut adalah persamaan *Euclidean Distance*:

$$d(x_i, c_i) = \sqrt{\sum_{i=1}^r (x_i - c_i)^2} \quad (1)$$

Di mana penjelasan persamaan diatas sebagai berikut: d merupakan titik dokumen, x_i merupakan obyek x ke- i , y_i merupakan data y ke- i , dan n merupakan banyaknya obyek. Setelah menghitung jarak menggunakan persamaan (1), selanjutnya menggolongkan setiap data berdasarkan kedekatannya menggunakan centroid (jarak terkecil). Lakukan pengulangan ke step-3, jika terdapat data yang berpindah cluster atau jika perubahan centroid. Pembaharuan titik *centroid* bisa dilakukan melalui rumus berikut ini:

$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (2)$$

Di mana penjelasan persamaan diatas sebagai berikut: μ_k merupakan titik *centroid* dari *cluster* ke-K, N_k merupakan jumlah data yang tergabung dalam sebuah *cluster*, x_q merupakan data ke-q pada *cluster* ke-K.

b. K-Nearest Neighbor

K-Nearest Neighbor adalah salah satu algoritma berbasis NN yang cukup populer. Nilai K yang digunakan menunjukkan berapa banyak tetangga terdekat yang kita pilih, kemudian kita memberikan suara untuk menentukan kelas berdasarkan tetangga terdekat tersebut. Kelas yang dengan suara terbanyak dari tetangga itu akan menjadi label kelas untuk hasil prediksi pada data yang sedang diuji. Algoritma NN murni termasuk dalam kategori lazy learner karena ia menunda proses pelatihan sampai ada data uji yang kita mau tahu label kelasnya, baru kemudian metode ini akan menjalankan algoritmanya. Algoritma Nearest Neighbor melakukan pengklasifikasian berdasarkan seberapa mirip suatu data dengan data lainnya (Adinugroho & Sari, 2018). Langkah-langkah yang mengikuti aturan K-NN adalah: Tentukan nilai K. Hitung jarak data uji dari data latih. Jika datanya adalah angka, maka gunakan *Euclidean Distance*. Berikut adalah persamaan *euclidean distance* seperti pada persamaan 3.

$$d(x_i, c_i) = \sqrt{\sum_{i=1}^r (x_i - c_i)^2} \quad (3)$$

Setelah kita menghitung jarak, kita urutkan jarak-jarak itu dari yang terbesar sampai yang terkecil. Kemudian kita ambil jarak-jarak yang paling dekat sampai mencapai jumlah k yang sudah ditentukan. Dari situ, kita pilih kelas yang muncul paling sering dan kita beri label sesuai klasifikasi tersebut.

c. WEKA (Waikato Environment for Knowledge Analysis)

Waikato Environment for Knowledge Analysis (WEKA) adalah aplikasi pembelajaran mesin yang terkenal ini ditulis menggunakan bahasa pemrograman Java. WEKA dibuat di Universitas Waikato yang ada di Selandia Baru. WEKA memiliki berbagai langkah pemecahan yang juga disertai dengan tampilan visual untuk membantu dalam analisis data dan pembuatan model prediksi. Algoritma dan langkah pemecahan dalam WEKA bisa digunakan untuk menyelesaikan masalah dalam bidang penggalian data. (Suara, 2020).

HASIL PENELITIAN DAN PEMBAHASAN

Pada penelitian ini dilakukan perbandingan implementasi antara algoritma *k-means* dan *k-nearest neighbor*. Kemudian setelah hasil implementasi data dari masing-masing algoritma dilakukan proses perbandingan antara algoritma *k-means* dan *k-nearest neighbor*.

a. Implementasi K-Means Clustering

Atribut yang digunakan dalam perhitungan K-Means Clustering adalah 3 atribut, yaitu atribut pekerjaan, jumlah penghasilan dan jumlah tanggungan. Ketiga atribut ini diambil berdasarkan ketersediaan data, wawancara dengan pakar serta kecocokan dengan algoritma data mining yang digunakan.

Tabel 1 Atribut

No	Nama Atribut	Keterangan
1.	Pekerjaan	a. Bekerja b. Tidak Bekerja
2.	Penghasilan	Gaji kepala keluarga perbulan
3.	Jumlah Tanggungan	Jumlah anak yang menjadi tanggungan oleh kepala keluarga

Setelah variabel data terdefiniskan, untuk menentukan penerima bantuan berdasarkan

data kependudukan, ukuran tingkat kelayakan pendapatan harus didefinisikan terlebih dahulu. Kelayakan pendapatan bulanan kepala keluarga dapat dilihat pada Tabel 2.

Tabel 2 Tingkat Kelayakan Pendapatan

Rentang Nilai	Tingkat Kelayakan
< 1.000.000	Sangat Layak
1.000.000 – 2.000.000	Cukup Layak
2.000.0000 – 3.000.000	Kurang Layak
>3.000.000	Tidak Layak

Tingkat kelayakan pendapatan dipakai untuk membagi kelayakan menurut pendapatan bulanan kepala keluarga. Selain itu, ada juga tingkat jumlah tanggungan. Tingkat jumlah tanggungan bisa dilihat pada Tabel 3.

Tabel 3 Tingkat Kelayakan Jumlah Tanggungan

Jumlah Tanggungan	Keterangan Kelayakan
2	Layak
> 2	Tidak Layak

Tingkat jumlah tanggungan dalam keluarga disarankan untuk 2 anak, karena dikatakan oleh Kepala BKKBN Hasto Wardoyo dalam penelitian ilmiah menyebutkan keluarga yang mempunyai 2 anak mampu menurunkan jumlah kematian ibu dengan anak. Ini merupakan hasil kajian yang dikerjakan peneliti secara global yang menerangkan bahwa paritas atau jumlah anak sangat erat kaitannya menggunakan kejadian mortalitas ibu dengan bayi [18].

Tahap pertama yang dilakukan dalam perhitungan K-Means Clustering dengan menentukan jumlah *cluster*. Dalam menentukan jumlah *cluster* (*cluster* yang digunakan adalah 5, karena mengacu pada metode *Latent Class Cluster Analysis* yang menghasilkan 5 *cluster* untuk memperoleh banyak variasi). Setelah menentukan jumlah *cluster*, selanjutnya menghitung jarak antara objek dengan *centroid*. Selanjutnya mengelompokkan berdasarkan jarak minimum. Jika sudah sama, posisi jarak perhitungan akan berhenti. Jika belum sama, perhitungan akan mengulang ke tahap perhitungan jarak antara objek dengan *centroid*. Dalam perhitungan algoritma *k-means clustering*, penulis menggunakan aplikasi WEKA. Tujuan dari penggunaan *software* ini adalah untuk memprediksi data penduduk yang layak mendapatkan bantuan. Pada aplikasi WEKA pengambilan pusat *cluster* diambil secara random dan disebutkan nilainya dari 3 atribut, atribut yang dipakai adalah pekerjaan, jumlah tanggungan dan penghasilan. Pada WEKA dijelaskan jumlah iterasi yang dilakukan untuk mendapatkan 5 *cluster* yaitu sebanyak 8 iterasi. Hasil dari perhitungan *k-means clustering* menghasilkan 5 *cluster*.

Tabel 4 Hasil Perhitungan K-Means Clustering

Cluster	Hasil	Persentase
Cluster 0	224 Data	12%
Cluster 1	284 Data	16%
Cluster 2	198 Data	11%
Cluster 3	461 Data	26%
Cluster 4	637 Data	35%

Perhitungan *k-means clustering* menghasilkan 5 *cluster*, dimana dalam setiap *cluster*-nya yaitu *cluster* 0 sebanyak 224 data atau sebesar 12%, *cluster* 1 sebanyak 284 data atau sebesar 16%, *cluster* 2 sebanyak 198 data atau sebesar 11%, *cluster* 3 sebanyak 461 data atau sebesar 26%, dan *cluster* 4 sebanyak 637 data atau sebesar 35%.

b. Implementasi *K-Nearest Neighbor*

Atribut yang digunakan dalam perhitungan *K-Nearest Neighbor* adalah 3 atribut, yaitu atribut pekerjaan, jumlah penghasilan dan jumlah tanggungan. Atribut yang digunakan sama seperti pada perhitungan algoritma *K-Means Clustering*.

Tahap pertama yang dilakukan dalam perhitungan *K-Nearest Neighbor* adalah menentukan jumlah *K*. Setelah ditentukan jumlah *K*, selanjutnya menghitung jarak *test* data (data uji) ke data *training* (data latih). Selanjutnya mengurutkan data berdasarkan data yang memiliki jarak terkecil, lalu dikelompokkan kelas yang sesuai. Dan tahap terakhir adalah mencari jumlah kelas dari tetangga yang terdekat. Dalam perhitungan algoritma *k-nearest neighbor*, penulis menggunakan aplikasi WEKA. Tujuan dari penggunaan *software* ini adalah untuk memprediksi data penduduk yang layak mendapatkan bantuan. Pada aplikasi WEKA, penulis memilih atribut yang akan dipakai, atribut yang dipakai adalah pekerjaan, jumlah tanggungan dan penghasilan. Lalu pilih menu *classify* untuk melakukan perhitungan klasifikasi. Hasil perhitungan algoritma *k-nearest neighbor*, dalam perhitungan *k-nearest neighbor* menggunakan *test option-use training set*.

Tabel 5 Hasil Perhitungan K-Nearest Neighbor

Klasifikasi	Hasil	Persentase
Correctly Classified Instance	1752 Data	99.5455%
Incorrectly Classified Instance	8 Data	0.4545%

Perhitungan *k-nearest neighbor* menghasilkan persentase untuk *Correctly Classified Instance* adalah sebesar 99.5455% sementara persentase untuk *Incorrectly Classified Instance* adalah sebesar 0.4545%. Dimana dari 1760 data penduduk, ada 1752 data berhasil diklasifikasikan dengan benar dan 8 data tidak berhasil diklasifikasikan dengan benar.

c. Perbandingan Algoritma *K-Means* dan *K-Nearest Neighbor*

Tahapan akhir dari penelitian ini adalah melakukan perbandingan antara *k-means clustering* dan *k-nearest neighbor*.

Tabel 6 Perbandingan Algoritma

<i>K-MEANS CLUSTERING</i>	<i>K-NEAREST NEIGHBOR</i>
a. Cara pengambilan keputusan menggunakan 5 <i>cluster</i> .	a. Cara pengambilan keputusan ditentukan dengan "Layak" dan "Tidak Layak"
b. Hasil akurasi perhitungan <i>k-means</i> sebesar 65.145%.	b. Hasil akurasi perhitungan <i>k-nearest neighbor</i> sebesar 99.6501%
c. Tingkat <i>error</i> yang dihasilkan dari perhitungan <i>k-means</i> sebesar 34.855%.	c. Tingkat <i>error</i> yang dihasilkan dari perhitungan <i>k-nearest neighbor</i> sebesar 0.3409%.

Berdasarkan perhitungan yang telah dilakukan, diperoleh hasil akurasi *k-means clustering* adalah sebesar 65.145% dengan tingkat *error* sebesar 34.855%. Dengan menghasilkan 5 *cluster*, *cluster* pertama mendapatkan hasil 138 data dengan persentase 8%, *cluster* kedua mendapatkan hasil 545 data dengan persentase 31%, *cluster* ketiga mendapatkan hasil 393 data dengan persentase 22%, *cluster* keempat mendapatkan hasil 284 data dengan persentase 16%, dan *cluster* terakhir yaitu *cluster* kelima mendapatkan hasil 400 data dengan persentase 23%. Dari 5 *cluster*, yang layak untuk mendapatkan bantuan adalah 3 *cluster* yaitu *cluster* pertama, *cluster* kedua dan *cluster* ketiga. Dari 1760 data yang telah dihitung menggunakan algoritma *k-means clustering*, yang layak untuk mendapatkan bantuan adalah 1076 data. Sedangkan hasil perhitungan yang diperoleh algoritma *k-nearest neighbor* adalah sebesar 99.6501% dengan tingkat *error* sebesar 0.3409%. Dari 1760 data yang telah dihitung menggunakan algoritma *k-*

nearest neighbor; terdapat 1752 data yang benar diklasifikasikan dan 8 data yang salah diklasifikasikan. Dari hasil yang telah diperoleh, algoritma *k-nearest neighbor* lebih unggul dibandingkan dengan algoritma *k-means*.

SIMPULAN

Pola yang digunakan untuk menentukan penerima bantuan social menggunakan algoritma *k-means clustering* dan algoritma *k-nearest neighbor*. Penelitian ini menggunakan data penduduk 2024 dengan jumlah data 1760 untuk menghasilkan data calon penerima bantuan sosial. Hasil perhitungan menggunakan algoritma *K-Means* mendapatkan hasil akurasi dari 5 *cluster*, *cluster* pertama terdapat 138 data dengan hasil persentase 8%, *cluster* kedua terdapat 545 data dengan hasil persentase 31%, *cluster* ketiga terdapat 393 data dengan hasil persentase 22%, *cluster* keempat terdapat 284 data dengan hasil persentase 16%, dan *cluster* kelima terdapat 400 data dengan hasil persentase 23%. Sedangkan hasil perhitungan menggunakan algoritma *K-Nearest Neighbor* menghasilkan 1754 data dengan hasil persentase 99,659%. Berdasarkan perhitungan yang telah dilakukan, diperoleh hasil akurasi *k-means* adalah sebesar 65,145% dengan tingkat *error* sebesar 34,855%. Sedangkan hasil akurasi yang diperoleh algoritma *k-nearest neighbor* adalah sebesar 99,650% dengan tingkat *error* sebesar 0,341%. Berdasarkan hasil yang telah diperoleh, algoritma *k-nearest neighbor* lebih unggul dikarenakan hasil akurasinya lebih tinggi dibandingkan dengan algoritma *k-means*.

Dari hasil penelitian penentuan penerima bantuan, dapat pula dilakukan dengan cara mengkombinasikan kedua algoritma *k-means clustering* dan *k-nearest neighbor*. Dengan cara mengolah data menggunakan algoritma *K-Nearest Neighbor* dengan menggunakan 2 label "Layak" dan "Tidak Layak", lalu hasil dari algoritma *K-Nearest Neighbor* yang "Layak" di olah kembali menggunakan algoritma *K-Means Clustering* untuk menghasilkan beberapa *cluster*.

DAFTAR PUSTAKA

- Adinugroho, S., & Sari, Y. A. (2018). *Implementasi Data Mining Menggunakan WEKA* (Edisi Pert). UB Press.
- Bhagya, T. G., & Prakarsa, G. (2020). Model Keputusan Penentuan Jenis Distribusi Dari Kerusakan Bearing pada Mesin TFO di PT XYZ. *SISINFO: Jurnal Sistem Informasi dan Informatika*, 2(2). 134-142.
- Eric Fammaldo, L. H. (2018). *Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar*. V(1), 23-31
- Hafidz, C. A. H. R. A. H. I. (2012). *Clustering Kelompok Swadaya Masyarakat (KSM) dalam Menentukan Kebijakan Bantuan Badan Pemberdayaan Masyarakat di Kota Surabaya dengan*. 1(1).
- Hasanah, R. L., Hasan, M., & Pangesti, W. E. (2019). *Klasifikasi Penerima Dana Bantuan Desa Menggunakan Metode KNN (K-Nearest Neighbor)*. 16(1), 1-6.
- Hermawati, F. A. (2013). *Data Mining* (P. Christian (ed.)). ANDI.
- Isyarah, F., Hasan, A., Wiza, F., Studi, P., Informatika, T., Ilmu, F., Universitas, K., Penduduk, J., Sekolah, L., & Kerja, A. (2020). *Clustering Daerah Miskin Di Provinsi Riau Menggunakan Metode K-Means*. 1(1), 1-12.
- Jajam Haerul Jaman, S. A. F. (2019). Klasifikasi Calon Mahasiswa Bidikmisi dengan Algoritma K-Nearest Neighbor. *Prosiding Annual Research Seminar*, 5(1), 1-5.

- Kusanti, J., & Sutanto, D. (2021). *Journal of Computer Networks , Architecture and High Performance Computing Combination of Decision Tree and K-Means Clustering Methods for Decision Making of BLT Recipients in the Covid-19 Period* *Journal of Computer Networks , Architecture and High Perfor.* 3(1), 80–88.
- Li, Y., & Wu, H. (2012). 2012 International Conference on Solid State Devices and Materials Science A Clustering Method Based on K-Means Algorithm. *Physics Procedia*, 25, 1104–1109. <https://doi.org/10.1016/j.phpro.2012.03.206>
- Noviana, T., Jasmir, J., & Novianto, Y. (2019). Penerapan Data Mining Menentukan Kelompok Prioritas Penerima Bantuan Beras Rastra Dengan Clustering K-Means. *Program Studi Teknik Informatika, Stikom Dinamika Bangsa*, 159–174.
- Permana, A. D., Nasution, V. M., & Prakarsa, G. (2020). *Design and Development of Fuzzy Logic Application Tsukamoto Method in Predicting the Number of Covid-19 Positive Cases in West Java.* 1(2), 85–95.
- Prasetyo, E. (2014). *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab* (A. Sahala (ed.)). ANDI.
- Soeleman, D. M. C. A. S. M. A. (2019). Penentuan Prioritas Penerima Dana Bantuan Operasional Pendidikan Lembaga Pendidikan Anak Usia Dini dengan Metode. *JJurnal Teknologi Informasi*, 15(2), 77–92.
- Suara.com. (2020). *BKKBN Sebut Dua Anak Lebih Sehat untuk Keluarga Indonesia.* <https://www.suara.com/health/2020/09/28/142855/bkkbn-sebut-dua-anak-lebih-sehat-untuk-keluarga-indonesia-apa-sebabnya?page=all>
- Yani, S., Jumeilah, F. S., & Kadafi, M. (2020). Algoritma K-Nearest Neighbor Untuk Menentukan Kelayakan Keluarga Penerima Bantuan Pangan Non Tunai (Studi Kasus: Kelurahan Karya Jaya). *Journal of Information Technology Ampera*, 1(2), 75–87. <https://journal-computing.org/index.php/journal-ita/index>
- Yogachi, E. F., Nasution, V. M., & Prakarsa, G. (2021). Design and Development of Fuzzy Logic Application Mamdani Method in Predicting The Number of Covid-19 Positive Cases in West Java. *IOP Conference Series: Materials Science and Engineering*, 1115(1), 012031. <https://doi.org/10.1088/1757-899x/1115/1/012031>