

PENERAPAN DATA MINING UNTUK CLUSTERING PENDUDUK MISKIN DI KOTA TANJUNGBALAI MENGGUNAKAN METODE ALGORITMA K-MEANS

Zunaida Sitorus¹, Suhartika²

Universitas Asahan

e-mail: suhartika82@gmail.com

Abstract: *In the Tanjungbalai region, the problem of poverty is a complex matter. A family is called a poor family based on several aspects such as food, clothing and shelter. Based on data quoted from <https://tanjungbalaikota.bps.go.id/> on Tuesday 25 July 2023 at 18.00, it shows that the poverty rate is still high in Tanjungbalai with the number of poor people in 2018 being 14.64%, in 2019 being 14.04%, in 2020 it was 13.33%, in 2021 it was 13.40% and in 2022 it was 12.45%. The aim of this research is to implement Data Mining at the Tanjungbalai Social Service, especially in the field of Social Empowerment & Handling the Poverty, which uses the K-Means algorithm whose visualization will be displayed on the web. The data used in this research is data from residents of Tanjung Balai City. The results of new information from existing data by applying data mining provide information on the clustering of poor residents in Tanjungbalai City so that related parties can use these results for consideration in making decisions.*

Keywords: *Data Mining, Clustering of Poor Population, Tanjungbalai City, K-Means Algorithm Method, Web Based*

Abstrak: Di Wilayah Tanjungbalai masalah kemiskinan merupakan hal yang kompleks. Seorang keluarga disebut sebagai keluarga miskin berdasarkan beberapa aspek seperti aspek pangan, sandang dan papan. Berdasarkan data yang dikutip dari <https://tanjungbalaikota.bps.go.id/> pada hari Selasa tanggal 25 Juli 2023 pukul 18.00 menunjukkan masih tingginya angka kemiskinan di Tanjungbalai dengan jumlah penduduk miskin pada tahun 2018 sebanyak 14,64%, tahun 2019 sebanyak 14,04%, tahun 2020 sebanyak 13,33%, tahun 2021 sebanyak 13,40% dan pada tahun 2022 sebanyak 12,45%. Tujuan penelitian ini adalah untuk mengimplementasikan Data Mining pada Dinas Sosial Tanjungbalai khususnya pada Bidang Pemberdayaan Sosial & Penanganan Fakir Miskin, yang menggunakan algoritma K-Means yang visualisasinya akan ditampilkan di web. Data yang digunakan dalam penelitian ini merupakan data penduduk Kota Tanjung Balai. Hasil informasi baru dari data yang ada dengan melakukan penerapan data mining memberikan sebuah informasi clustering penduduk miskin di Kota Tanjungbalai sehingga pihak terkait bisa menggunakan hasil ini untuk pertimbangan dalam mengambil keputusan.

Kata kunci: Data Mining, Clustering Penduduk Miskin, Kota Tanjungbalai, Metode Algoritma K-Means, Berbasis Web

PENDAHULUAN

Di Wilayah Tanjungbalai masalah kemiskinan merupakan hal yang kompleks. Seorang keluarga disebut sebagai keluarga miskin berdasarkan beberapa aspek seperti aspek pangan,

sandang dan papan. Berdasarkan data yang dikutip dari <https://tanjungbalaikota.bps.go.id/> pada hari Selasa tanggal 25 Juli 2023 pukul 18.00 menunjukkan masih tingginya angka kemiskinan di Tanjungbalai dengan jumlah penduduk miskin pada tahun 2018

sebanyak 14,64%, tahun 2019 sebanyak 14,04%, tahun 2020 sebanyak 13,33%, tahun 2021 sebanyak 13,40% dan pada tahun 2022 sebanyak 12,45%.

Berdasarkan angka tersebut di atas terlihat masih tingginya angka kemiskinan yang ada di wilayah Kota Tanjungbalai. BKKBN (Badan Kependudukan dan Keluarga Berencana Nasional) sebagai badan yang bertugas menghimpun data statistik kemiskinan di Tanjungbalai merasa kesulitan dalam pendistribusian berbagai macam bantuan. Berdasarkan pada kondisi tersebut, maka dalam penelitian ini dibuat model clustering untuk mendapatkan klaster-klaster kemiskinan dengan menganalisa atribut yang berpengaruh maupun tidak. Upaya tersebut dilakukan melalui pembuatan suatu alat bantu berupa aplikasi dengan menggunakan metode K-Means untuk mengetahui pola penduduk miskin. Metode K-Means clustering sebagai salah satu metode data clustering non-hirarki dengan mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain. Kelompok atau cluster yang didapat merupakan pengetahuan atau informasi yang bermanfaat bagi pengguna kebijakan dalam proses pengambilan keputusan, dalam permasalahan ini penulis tertarik membuat penelitian yang berjudul “Penerapan Data Mining Untuk Clustering Penduduk Miskin Di Kota Tanjungbalai Menggunakan Metode Algoritma K-Means”.

METODE

Kerangka Kerja Penelitian

Kerangka kerja penelitian ini penjabaran dari langkah-langkah yang akan dilakukan dalam penyelesaian masalah yang dibahas. Adapun kerangka

kerja penelitian ini dapat dilihat pada Gambar 1 berikut:



HASIL DAN PEMBAHASAN

Dari pengolahan dan pengujian data yang telah dilakukan dengan menggunakan metode *Algoritma K-Means*. Hasil dari *clustering* tersebut akan menghasilkan tingkat persentase (%) keakuratannya dalam penentuan penduduk miskin yang tepat sehingga akan berdampak pada peningkatan tarap hidup yang akan di berikan oleh pemerintah.

Objek Penelitian

Objek penelitian merupakan tempat dilakukanya penelitian ini, objek ini sangat mempengaruhi dalam pelaksanaan suatu penelitian, karena dengan adanya objek tersebut penelitian dapat memperoleh bahan yang dibutuhkan dalam penelitian. Dalam hal ini peneliti mengambil objek penelitian penduduk miskin di Kota Tanjungbalai.

Operasional Variabel

Operasional variabel berisikan komponen-komponen dari suatu variabel yang mungkin peneliti mengumpulkan data yang relevan untuk variabel tersebut. Variabel penelitian merupakan objek

penelitian atau apa yang menjadi titik perhatian suatu penelitian data penduduk miskin yang telah dikumpulkan. Dalam proses ini diambil 6 sampel data khusus untuk proses pengujian clustering data yaitu data yang diambil atas nama eria doni siregar, Abdurrahman, agus salim, faisal adha, ardi surya dan palit sitorus. Kemudian ke 6 data sampel ini diubah menjadi data berikut dalam table dibawah untuk proses clusteringnya.

Sample Data	X	Y
1	100	50
2	40	60
3	30	70
4	90	10
5	65	40
6	25	35

Keterangan :

-Sampel data 1-6 adalah urutan nama warga,

-X adalah Jumlah bantuan tahun lalu dan Y

adalah jumlah bantuan tahun ini.

Kita akan mengaplikasikan K-Means Clustering

untuk data diatas menjadi 2 cluster. Pertama kita akan hitung Centroid.

Cluster	X	Y
K1	100	50
K2	40	60

Perhitungan menggunakan persamaan Euclidean Distance

$$[(x,y) , (a,b)] = \sqrt{(x - a)^2 + (y - b)^2}$$

Perhitungan Pertama ;

$$\text{Cluster 1 (100, 50)} = \sqrt{(100 - 100)^2 + (50 - 50)^2} = 0$$

(jarak cluster 1 ke cluster 1)

$$\text{Jarak dari Cluster 2 ke cluster 1 (40 , 60)} \leftrightarrow (100,50) =$$

$$\sqrt{(40 - 100)^2 + (60 - 50)^2} = \sqrt{(60)^2 + (10)^2} = \sqrt{3600 + 100} = \sqrt{3700} = 60.83$$

$$\text{Jarak dari cluster 1 ke cluster 2 (100, 50)} \leftrightarrow (40, 60) =$$

$$\sqrt{(100 - 40)^2 + (50 - 60)^2} = \sqrt{(60)^2 + (-10)^2} = \sqrt{3600 + 100} = \sqrt{3700} = 60.83$$

$$\text{Jarak cluster 2 ke cluster 2 (40, 60)} =$$

$$\sqrt{(40 - 40)^2 + (60 - 60)^2} = 0$$

Sehingga ;

Cluster	Centroid		Kelompok Cluster
	X	Y	
K1 (100, 50)	0	60.83	1
K2 (40, 60)	60.83	0	2

Perhitungan kedua.

Langkah selanjutnya kita beralih ke data 3 yaitu (30, 70).

Kita mulai menghitung jarak dataset terhadap cluster 1.

$$(100, 50) \leftrightarrow (30, 70) =$$

$$\sqrt{(30 - 100)^2 + (70 - 50)^2} = \sqrt{(-70)^2 + (20)^2} = \sqrt{4900 + 400} = \sqrt{5300} = 72.80$$

Kemudian kita hitung jarak dataset terhadap cluster 2.

$$(40, 60) \leftrightarrow (30, 70) =$$

$$\sqrt{(30 - 40)^2 + (70 - 60)^2} = \sqrt{(-10)^2 + (10)^2} = \sqrt{100 + 100} = \sqrt{200} = 14.14$$

Sehingga didapatkan :

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(30, 70)	72.80	14.14	2

Dataset ke-3 masuk dalam kelompok cluster 2,

karena jarak minimum / terdekat dataset adalah terhadap cluster 2 yaitu 14.14.

Kemudian kita update Centroid

Cluster	X	Y
K1	100	50
K2	$\frac{40 + 30}{2} = 35$	$\frac{60 + 70}{2} = 65$

Jadi Cluster Centroid yang baru adalah sebagai berikut :

Cluster	X	Y
K1	100	50
K2	35	65

Perhitungan ketiga

Kita lanjutkan kembali menghitung dataset ke-4 yaitu (90, 10).

Kita mulai menghitung jarak dataset terhadap cluster 1.

$$(100, 50) \leftrightarrow (90, 10) =$$

$$\sqrt{(90-100)^2 + (10-50)^2} = \sqrt{(-10)^2 + (-40)^2} = \sqrt{100+1600} = \sqrt{1700} = 41.23$$

Kemudian kita hitung jarak dataset terhadap cluster 2.

Sangat perlu diingat, centroid 2 yang kita gunakan adalah

yang sudah di update di perhitunga kedua diatas yaitu (35, 65).

$$(35, 65) \leftrightarrow (90, 10) =$$

$$\sqrt{(90-35)^2 + (10-65)^2} = \sqrt{(55)^2 + (-55)^2} = \sqrt{3025+3025} = \sqrt{6050} = 77.78$$

Sehingga didapatkan :

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(90, 10)	41.23	77.78	1

Dataset ke-3 masuk dalam kelompok cluster 1,

karena jarak minimum / terdekat dataset adalah terhadap cluster 1 yaitu 41.23.

Kemudian kita update Centroid

Cluster	X	Y
K1	$\frac{100 + 90}{2} = 95$	$\frac{50 + 10}{2} = 30$
K2	35	65

Jadi Cluster Centroid yang baru adalah sebagai berikut :

Cluster	X	Y
K1	95	30
K2	35	65

Perhitungan keempat Kita lanjutkan Kembali

menghitung dataset ke-5 yaitu (80, 40).

Kita

mulai menghitung jarak dataset terhadap cluster 1.

Perhatikan kembali, centroid 1 yang digunakan

adalah yang telah terupdate yaitu (95, 30).

$$(95, 30) \leftrightarrow (65, 40) =$$

$$\sqrt{(65-95)^2 + (40-30)^2} = \sqrt{(-30)^2 + (10)^2} = \sqrt{900+100} = \sqrt{1000} = 31.62$$

Kemudian kita hitung jarak dataset terhadap cluster 2.

$$(35, 65) \leftrightarrow (65, 40) =$$

$$\sqrt{(65-35)^2 + (40-65)^2} = \sqrt{(30)^2 + (-15)^2} = \sqrt{900+225} = \sqrt{1125} = 33.54$$

Sehingga didapatkan :

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(80, 40)	31.62	33.54	1

Dataset ke-4 masuk dalam kelompok cluster 1 ,

karena jarak minimum / terdekat dataset adalah terhadap cluster 1 yaitu 31.62.

Kemudian kita update Centroid

Cluster	X	Y
K1	$= \frac{95 + 65}{2} = 80$	$= \frac{30 + 40}{2} = 35$
K2	35	65

Jadi Cluster Centroid yang baru adalah sebagai berikut :

Cluster	X	Y
K1	80	35
K2	35	65

Perhitungan kelima

Kita lanjutkan kembali menghitung dataset ke-5

yaitu (25, 35). Kita mulai menghitung jarak dataset

terhadap cluster 1. Perhatikan kembali, centroid

1 yang digunakan adalah yang telah terupdate yaitu (80, 35).

$$(80, 35) \leftrightarrow (25, 35) =$$

$$\sqrt{(25-80)^2 + (35-35)^2} = \sqrt{(-55)^2 + (0)^2} = \sqrt{3025+0} = \sqrt{3025} = 55$$

Kemudian kita hitung jarak dataset terhadap cluster 2.

$$(35, 65) \leftrightarrow (25, 35) =$$

$$\sqrt{(25-35)^2 + (35-65)^2} = \sqrt{(10)^2 + (-30)^2} = \sqrt{100+900} = \sqrt{1000} = 31.62$$

Sehingga didapatkan :

Dataset	Euclidean Distance		Kelompok Cluster
	Cluster 1	Cluster 2	
(80, 40)	55	31.62	2

Semua data telah kita dapatkan, dan masing-

masing data sudah terkelompokkan

berdasarkan cluster 1 dan cluster 2. Selanjutnya

kita input hasil pengelompokkannya pada table

pertama. jadi hasilnya seperti berikut :

Sample Data	X	Y	Kelompok / Cluster
1	100	50	1
2	40	60	2
3	30	70	2
4	90	10	1
5	65	40	1
6	25	35	2

SIMPULAN

Dari hasil pembahasan pada sebelumnya yang telah diuraikan, maka peneliti mengambil kesimpulan sebagai berikut:

1. Hasil informasi baru dari data yang ada dengan melakukan penerapan data mining memberikan sebuah informasi clustering penduduk miskin di Kota Tanjungbalai sehingga pihak terkait bisa menggunakan hasil ini untuk

- pertimbangan dalam mengambil keputusan.
2. Penerapan Data Mining untuk Clustering data penduduk miskin di Kota Tanjungbalai menggunakan algoritma K-means sehingga hasil clustering data yang di uji sesuai dengan yang diharapkan.
 3. Untuk merancang dan membuat aplikasi data mining untuk clustering penduduk miskin di Kota Tanjungbalai yang mudah di pahami oleh pihak BKKBN peneliti merancanganya dengan menggunakan beberapa langkah mulai dari merancang sistem proses kerja dari tiap-tiap menu dan mendesain aplikasi dengan melakukan rancangan terlebih dahulu sebelum ke tahap implementasi.

DAFTAR PUSTAKA

- Ach, K. (2021). Pemrograman Aplikasi Web Buku Ajar. Media Nusa Creative (MNC Publishing).
- Adawiyah, E. (2020). Kemiskinan Dan Penyebabnya. 1(April), 43–50.
- Afifi, R. M., Putra, E. K., & Pudjiantoro, T. H. (2020). Sistem Electronic Supply Chain Management Menggunakan Metode Just in Time di PT Cemara Agung Mandiri. *Jurnal Media Informatika Budidarma*, 4(4), 970–978.
<https://doi.org/10.30865/mib.v4i4.2338>
- Afrilia, R., Mair, Z. R., & Juansyah, J. (2021). Sistem Informasi Pengelolaan Data Alumni Pada UPT SMK Negeri 1 Musi Banyuasin. *Jurnal Nasional Ilmu Komputer*, 2(2), 112–134.
<https://doi.org/10.47747/jurnalnik.v2i2.527>
- Alda, M. (2021). Rekayasa Perangkat Lunak Berorientasi Objek. Media Sains Indonesia.
- Alkhairi, P., & Windarto, A. P. (2019). Penerapan K-Means Cluster pada Daerah Potensi Pertanian Karet Produktif di Sumatera Utara. *Seminar Nasional Teknologi Komputer & Sains*, 762–767 ISBN 9786025272011.
- Arif, A., Dwi Christyanti, R., & Kaltara, U. (2022). Clustering Calon Penerima Zakat Menggunakan Metode K-Means(Ratna Dwi Christyanti) (Studi Kasus di Provinsi Kalimantan Utara). *73 SMARTICS Journal*, 8(2), 73–79.
- Elgamar, E. (2020). Buku Ajar Konsep Dasar Pemrograman Website dengan PHP. CV Multimedia Publihsers.
- Fathoroni, A., Fathonah, N. S., Andarsyah, R., & Riza, N. (2020). Sistem Pendukung Keputusan Penilaian Kinerja Dosen Menggunakan Metode 360 Degree Feedback. *Kreatif Industri Nusantara*.
- Gusman, A. P. (2019). Analisa Perancangan dan Implementasi Pemesanan Secara Online Berbasis Costumer Relationship Management (CRM). *Majalah Ilmiah UPI YPTK*, 26(1), 7–13.
<https://doi.org/10.35134/jmi.v26i1.17>
- Gusnawan, R., Wijayanto, P., & Rosely, E. (2019). APLIKASI KELOMPOK INFORMASI MASYARAKAT (KIM) BERBASIS WEB MANAGEMENT OF COMMUNITY INFORMATION GROUP APPLICATION (KIM) WEB- BASED. *E-Proceeding of Applied Science*, 5(2), 1123–1132.
- Habibi, R., & Aprilian, R. (2019). Tutorial dan Penjelasan Aplikasi E-Office Berbasis Web Menggunakan Metode RAD. *Kreatif Industri Nusantara*.
- Ismail. (2019). *Evolusi : Jurnal Sains dan Manajemen* Vol 7 No . 2 September 2019 ISSN : 2338-8161 E-ISSN : 2657-0793. *Jurnal Sains Dan Manajemen*, 7(2), 6–14.
- Kaban, R., & Sembiring, D. J. (2021). HTML (Hypertext Markup Language) Pengantar Pemrograman Berbasis Web. *Insan Cendekia Mandiri*.

- Kadir, A. (2019). Tuntunan Praktis Belajar Database Menggunakan MySQL. ANDI Offset.
- Kusnadi, Y., & Putri, M. S. (2021). Clustering Menggunakan Metode K-Means Untuk Menentukan Prioritas Penerima Bantuan Bedah Rumah (Studi Kasus : Desa Ciomas Bogor). *Jurnal Teknologi Informatika Dan Komputer*, 7(1), 17–24. <https://doi.org/10.37012/jtik.v7i1.498>
- Lestari, R. D. (2021). Analisis Pengaruh AMH, Jumlah Penduduk, Pengangguran, AHH, dan PDB Terhadap Kemiskinan di Indonesia, Malaysia, dan Thailand pada Tahun 2000-2020. *Jurnal Ilmiah*, 10(1).
- Nata, A., Marpaung, N., Yesputra, R., & Syafnur, A. (2021). Pemanfaatan Laporan Keuangan Kas Masjid Berbasis Web. *Community Development Journal : Jurnal Pengabdian Masyarakat*, 2(2), 499–502. <https://doi.org/10.31004/cdj.v2i2.1987>
- Nirsal, Rusmala, & Syafriadi. (2020). Desain Dan Implementasi Sistem Pembelajaran Berbasis E-Learning Pada Sekolah Menengah Pertama Negeri 1 Pakue Tengah. *Journal Ilmiah d'Computare*, 10, 30–37. <http://www.elsevier.com/locate/scp>
- Nugroho, A., Suprihadi, U., & Jaenul, A. (2021). Rancang Bangun Aplikasi Toko Online Berbasis Web Codeigniter 3 untuk Usaha Mikro dan UMKM. *Media Sains Indonesia*.
- Roza, R., Fauzan, M. N., & Rahayu, W. I. (2020). Tutorial Sistem Informasi Prediksi Jumlah Pelanggan Menggunakan Metode Regresi Linier Berganda Berbasis Web Menggunakan Framework Codeigniter. *Kreatif*.
- Sa'ad, M. I. (2020). Otodidak Web Programming: Membuat Website Edutainment. PT Alex Media Komputindo.
- Saputro, H., & Mahendra, D. (2019). Penerapan Aplikasi Penjualan Online Berbasis Customer Relationship Management (CRM) pada Toko Sumber Mulyo di Kabupaten Kudus. *Jurnal Disprotek*, 10(1), 35–42. <https://doi.org/10.34001/jdpt.v10i1.869>
- Suhartini, & Yuliani. (2021). Penerapan Data Mining untuk Mengcluster Data Penduduk Miskin Menggunakan Algoritma K-Means di Dusun Bagik Endep Sukamulia Timur. *Infotek : Jurnal Informatika Dan Teknologi*, 4(1), 39–50. <https://doi.org/10.29408/jit.v4i1.2986>
- Susilowati, Y. (2019). Modul E-Commerce Teaching Factory for Student. Mutiara PUBLISHER.
- Tamtelahitu, T. M. (2020). Komparasi Algoritma Clustering dengan Dataset Penyebaran Covid-19 di Indonesia Periode Maret-Mei 2020. *Jurnal Teknologi Technoscintia*, 13(1), 27–34.
- Wira, B., Budianto, A. E., & Wiguna, A. S. (2019). Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang. *RAINSTEK : Jurnal Terapan Sains & Teknologi*, 1(3), 53–68. <https://doi.org/10.21067/jtst.v1i3.3046>
- Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika*. *Jurnal Edik Informatika*, 2(2), 213–219.