

Klasifikasi Kesehatan Koperasi Di Sumatera Barat Tahun 2023 Menggunakan Metode K-Nearest Neighbour

Ayunda Setiani¹, Meira Parma Dewi²

^{1,2} Program Studi Matematika, Fakultas Matematika Ilmu Pengetahuan dan Alam, Universitas Negeri Padang (UNP)

Article Info

Article history:

Received May 12, 2025
Revised May 26, 2025
Accepted June 02, 2025

Keywords:

Cooperation
Classification
K-Nearest Neighbour
Elbow

Kata Kunci:

Koperasi
Klasifikasi
K-Nearest Neighbour
Elbow

ABSTRACT

Cooperation are economic institutions that have a role in supporting the welfare of society collectively. However to ensure the sustainability of cooperation, it is necessary to conduct a cooperation health check. Cooperation health assessment requires a data classification system to help the decision making process of cooperation health assessment in West Sumatera. The KNN method was chosen because of its high ability to classify data based on certain patterns and provide results that are in accordance with actual conditions. This study aims to analyze the classification results and the the optimal k selection with the elbow method. From 84 testing data, 80 data were correctly classified and there were 4 data that are predicted wrong. The classification results of the 80 data obtained 62 cooperations with healthy status, then 1 cooperation is quite healthy, then 11 cooperations are less healthy, and 6 cooperations are unhealthy with an accuracy rate of 95%. As for the selection of the optimal k value using the Elbow method, the Sum of Square Error value is 563,7 with a value k = 3.

ABSTRAK

Koperasi merupakan institusi ekonomi yang memiliki peran dalam mendukung kesejahteraan masyarakat secara kolektif. Namun untuk memastikan keberlanjutan koperasi maka perlu diadakan pemeriksaan kesehatan koperasi. Penilaian kesehatan koperasi membutuhkan sistem pengklasifikasian data guna membantu proses pengambilan keputusan kesehatan koperasi di Sumatera Barat. Metode KNN dipilih karena kemampuan yang tinggi dalam mengklasifikasikan data berdasarkan pola tertentu dan memberikan hasil yang sesuai dengan kondisi sebenarnya. Penelitian ini bertujuan menganalisis hasil klasifikasi serta pemilihan k optimal dengan metode *Elbow*. Dari 84 data *testing* diperoleh 80 data tepat klasifikasi dan terdapat 4 data diprediksi salah. Hasil klasifikasi 80 data tersebut diperoleh 62 koperasi berstatus sehat, lalu 1 koperasi cukup sehat, kemudian 11 koperasi kurang sehat, dan 6 koperasi tidak sehat dengan tingkat akurasi sebesar 95%. Sedangkan untuk pemilihan nilai *k* optimal menggunakan metode *elbow* diperoleh nilai *Sum of Squared Error* sebesar 563,7 dengan nilai *k* = 3

This is an open access article under the CC BY-SA license.



Ayunda Setiani

Program Studi Matematika, Departemen Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Universitas Negeri Padang, Jl. Prof. Dr. Hamka, Air Tawar barat, Padang Utara, Padang, Indonesia. Kode Pos: 25131
Email: ayundasetiani20@gmail.com

1. PENDAHULUAN

Koperasi merupakan sebuah organisasi ekonomi yang dikelola dan dimiliki secara bersama-sama oleh anggotanya, dengan tujuan memenuhi kebutuhan serta meningkatkan kesejahteraan masyarakat. Di Sumatera Barat yang dikenal dengan kekayaan budaya serta sumber daya alamnya, koperasi memiliki peran strategis dalam mendukung pembangunan ekonomi lokal. Namun untuk memastikan efektivitas koperasi tersebut, penilaian kesehatan koperasi menjadi hal yang sangat penting. Penilaian ini bertujuan untuk mengevaluasi sejauh mana koperasi mampu mencapai tujuan sosial dan ekonominya. Pada tahun 2022 pelaksanaan pemkab ditargetkan sebanyak 649 unit koperasi dengan realisasi sebanyak 543 unit koperasi (capaian 83,66%) [1]. Melalui evaluasi yang komprehensif terhadap berbagai aspek koperasi, seperti keuangan, tata kelola, profil resiko dan modal, pihak terkait dapat memahami kondisi kesehatan koperasi secara mendetail serta menentukan langkah-langkah perbaikan yang diperlukan.

Tahun 2023 menjadi periode krusial bagi klasifikasi kesehatan koperasi di Sumatera Barat. Di lapangan terdapat sejumlah tantangan utama yang memengaruhi kesehatan koperasi. Pertama keterbatasan modal yang disebabkan rendahnya kemampuan koperasi dalam menarik investasi. Kedua, tata kelola yang lemah menjadi persoalan yang serius. Ketiga, koperasi sering kali memiliki profil resiko yang tinggi karena minimnya strategi untuk mengidentifikasi dan mengelola resiko sehingga menyebabkan koperasi kurang siap menghadapi tantangan. Keempat aspek keuangan merupakan indikator penting dalam mengevaluasi pendapatan, pengeluaran dan profitabilitas koperasi, sehingga penting untuk melakukan evaluasi keuangan secara menyeluruh [2].

Klasifikasi merupakan teknik untuk mengklasifikasikan suatu data tertentu ke dalam kelas yang sebelumnya telah ditentukan kelasnya pada pola atau atribut yang terdapat pada data tersebut [3]. Algoritma KNN adalah sebuah metode untuk melakukan klasifikasi objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut [4]. Keunggulan KNN dibandingkan metode lain seperti K-Means adalah ketahanan terhadap data latih yang banyak noise (gangguan) dan efektif terhadap data latih yang berukuran besar [5]. Berbeda dengan K-Means, KNN memiliki algoritma yang dapat mengurangi waktu pemrosesan dan memberikan akurasi yang baik dalam sampel pengujian [6]. Dengan kemampuan ini KNN dapat digunakan untuk mengklasifikasikan kasus berdasarkan kesamaan data dalam sampel pengujian [7].

Metode KNN dipilih karena kemampuannya yang tinggi dalam mengklasifikasikan data berdasarkan pola tertentu dan memberikan hasil yang akurat meskipun pada dataset dengan jumlah terbatas [8]. KNN bekerja dengan membandingkan karakteristik koperasi sehingga dapat mengidentifikasi kategori kesehatan koperasi dengan cepat dan efisien [9]. Kelebihan ini membuat KNN menjadi metode yang efektif dalam mengklasifikasikan kesehatan koperasi di Sumatera Barat. Penelitian ini menggunakan metode *Machine Learning* dengan menggunakan metode KNN. Machine Learning adalah bagian dari kecerdasan buatan, metode ini membangun model matematika berdasarkan data sampel, disebut “data latih”, yang digunakan untuk membuat prediksi atau keputusan [10].

Penelitian ini bertujuan untuk menganalisis klasifikasi kesehatan koperasi di Provinsi Sumatera Barat tahun 2023 dengan menggunakan metode KNN. Dengan memanfaatkan data historis, diharapkan metode KNN ini dapat menjadi alat bantu yang andal dalam pengambilan keputusan. Tujuan akhir dari penelitian ini adalah memberikan dukungan kepada Dinas Koperasi dan UMKM Provinsi Sumatera Barat agar dapat menentukan status kesehatan koperasi secara lebih akurat dan efisien. Hasil dari penelitian diharapkan memberikan wawasan terhadap pemerintah daerah dan pengelola koperasi dalam meningkatkan kualitas pengelolaan koperasi kedepanya.

2. METODE

Penelitian ini adalah penelitian terapan. Penelitian ini dimulai dengan tinjauan literatur yang kemudian dilanjutkan dengan pengumpulan data sekunder. Data sekunder merujuk pada informasi yang telah tersedia dan diperoleh dari organisasi atau perusahaan. Penelitian ini pengolahan data

dibantu dengan excell dan *google colabs*. Langkah-langkah yang dilakukan dalam penelitian ini [11] : a) Melakukan analisis deskriptif data, b) Menentukan nilai k optimal dengan metode *Elbow*, c) Menghitung jarak menggunakan metode pengukuran jarak *Euclidean*, d) Mengurutkan hasil perhitungan jarak yang terbentuk dari yang terkecil ke yang terbesar berdasarkan nilai inputan k , e) Analisis hasil prediksi klasifikasi yang telah dihasilkan dan evaluasi akurasi model menggunakan *Confusion Matriks*.

3. HASIL DAN PEMBAHASAN

3.1. Deskripsi Data

Data yang digunakan dalam penelitian ini adalah data kesehatan koperasi yang diperoleh dari Dinas Koperasi dan UMKM Sumatera Barat. Data yang digunakan untuk menjadi *dataset* adalah data yang dikumpulkan dalam 1 file excel berjudul “Rekap Pemkes Koperasi 2023.xlsx”, yang berjumlah 263 data. Data yang digunakan adalah data *splitting*, yaitu proses pemisahan data *training* sebanyak 179 data dan data *testing* sebanyak 84 data.

3.2. Analisis Data

3.2.1. Statistik Deskriptif

Pada penelitian ini terdapat variabel dependen (terikat) yang merupakan status kesehatan koperasi (0: sehat, 1: cukup sehat, 2: kurang sehat dan 3: tidak sehat) serta variabel independent (bebas) yang terdiri dari tata kelola (X_1), profil resiko (X_2), keuangan (X_3), modal (X_4). Setelah selesai pengumpulan data maka tahap selanjutnya menentukan karakteristik data akan dicari nilai rata-rata, nilai maksimum dan nilai minimum dengan menggunakan Excel dapat dilihat melalui Tabel 1.

Tabel 1. Analisis Deskriptif Variabel *Independen*

Variabel	N	Mean	Std Deviasi	Minimum	Maximum
X_1	263	74,38	20,32	26	97
X_2	263	73,69	21,73	22	97
X_3	263	74,77	20,10	22	97
X_4	263	75,25	20,27	25	97

Berdasarkan Tabel 1, hasil nilai rata-rata X_1 sebesar 74,38 dengan nilai minimum 26 dan maksimum 97. Serta X_4 memiliki hasil nilai rata rata yang besar yaitu 75,25 dengan nilai minimum sebesar 25 dan maximum 97. Secara keseluruhan, analisis deskriptif menunjukkan bahwa keempat variabel memiliki jumlah observasi yang sama besar. Rata-rata nilai relatif berdekatan, begitu pula dengan tingkat penyebaran datanya (standar deviasi).

3.2.2. Menentukan k optimal dengan *Elbow*

Tahap selanjutnya yaitu menetapkan nilai k optimal dengan mengimplementasikan metode elbow. Metode ini dinamakan “elbow” karena grafik yang dihasilkan biasanya berbentuk siku pada titik yang dianggap sebagai jumlah paling optimal [12]. Untuk menentukan nilai k optimal berdasarkan metode elbow dapat dilakukan perhitungan menggunakan nilai SSE. Adapun persamaan yang digunakan untuk mencari nilai sum of square error sebagai berikut [13].

$$SSE = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2$$

k = jumlah klaster

x = titik data (baris data)

μ_i = titik pusat centroid dari klaster ke- i

C_i = cluster ke- i

Tahap selanjutnya yaitu melakukan perhitungan nilai sum of square error elbow pada data training, berikut tampilan 10 data training dapat dilihat pada Tabel 2.

Tabel 2. Data Training k optimal Elbow

No	X_1	X_2	X_3	X_4
1	87	84	91	81
2	93	88	84	92
3	94	82	80	81
4	90	80	82	78
5	92	89	83	79
6	91	94	76	86
7	85	92	94	81
8	77	89	75	97
9	74	89	92	95
10	80	91	89	87

Berikut perhitungan SSE untuk $k = 1$ seperti berikut:

a) Hitung *centroid* (rata-rata kolom)

$$\text{Centroid} = \frac{1}{n} \sum_{i=1}^n x_i$$

x_i = titik data dalam klaster

n = jumlah titik dalam klaster

$$C_1 = \frac{1}{10} \sum x_1 = \frac{1}{10} (87 + 93 + 94 + 90 + 92 + 91 + 85 + 77 + 74 + 80) = 86,3$$

$$C_2 = \frac{1}{10} \sum x_2 = \frac{1}{10} (84 + 88 + 82 + 80 + 89 + 94 + 92 + 89 + 89 + 91) = 87,8$$

$$C_3 = \frac{1}{10} \sum x_3 = \frac{1}{10} (91 + 84 + 80 + 82 + 83 + 76 + 94 + 75 + 92 + 89) = 84,6$$

$$C_4 = \frac{1}{10} \sum x_4 = \frac{1}{10} (81 + 92 + 81 + 78 + 79 + 86 + 81 + 97 + 95 + 87) = 85,7$$

b) Hitung kuadrat jarak (error) tiap baris

$$\text{Error} = \sum_{j=1}^n (x_j - c_j)^2$$

$$\text{Error}_i = (x_{i1} - c_1)^2 + (x_{i2} - c_2)^2 + (x_{i3} - c_3)^2 + (x_{i4} - c_4)^2$$

x_{ij} = titik data ke-j dari baris data ke-i

C_j = hasil klaster ke-j dari pusat klaster

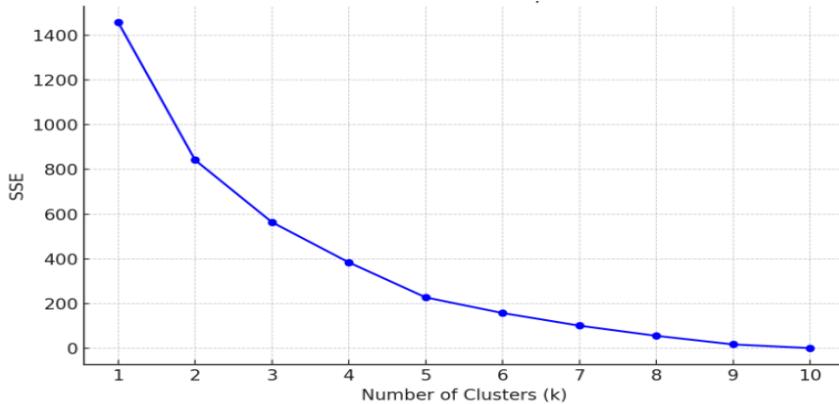
Tabel 3. Hasil Kuadrat Jarak Error Elbow

No	$(x_{i1} - c_1)^2$	$(x_{i2} - c_2)^2$	$(x_{i3} - c_3)^2$	$(x_{i4} - c_4)^2$	Error
1	0,49	14,44	40,96	22,09	77,98
2	44,89	0,04	0,36	39,69	84,98
3	59,29	33,64	21,16	22,09	136,18
4	13,69	60,84	6,76	59,29	140,58
5	32,49	1,44	2,56	44,89	81,38
6	22,09	38,44	73,69	0,09	134,58
7	1,69	17,64	88,36	22,09	129,78
8	86,49	1,44	92,16	127,69	307,8
9	151,29	1,44	54,76	86,49	293,98
10	39,69	10,24	19,36	1,69	79,98

- C) Menentukan nilai SSE (jumlahkan semua error)

$$\begin{aligned} SSE &= 77,98 + 84,98 + 136,18 + 140,58 + 79,98 + 81,38 + 134,58 + 129,78 + 307,8 + 293,98 \\ &+ 79,98 = 1458,2 \end{aligned}$$

Selanjutnya menampilkan representasi dari tabel diatas dapat dilihat pada Gambar 1.



Gambar 1. Grafik Nilai K Optimal Metode *Elbow*

Berdasarkan Tabel 3, diperoleh nilai SSE sebesar 563,7. Hal ini dikarenakan grafik elbow menunjukkan bentuk seperti siku tajam di titik $k = 3$. Dimana k sebelum 3 grafik menurun curam dan k setelah 3 grafik mulai melandai, titik optimal ditunjukkan oleh sudut grafik yang mulai turun setelah penurunan nilai SSE yang signifikan pada sudut-sudut sebelumnya.

3.2.3. Perhitungan Jarak *Euclidean*

Pengukuran dua titik dalam satu dimensi pada jarak euclidean menampilkan hasil yang mirip dengan perhitungan Pythagoras [14].

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

$d(x, y)$ = jarak antara objek x ke y

x_i = data training

y_i = data testing

p = banyaknya variabel yang diamati

Salah satu perhitungan jarak *Euclidean* pada data *testing* pertama dengan data *training* ke-1 hingga ke-179 sebagai berikut.

$$d(1,1) = \sqrt{(87 - 96)^2 + (84 - 90)^2 + (91 - 84)^2 + (81 - 92)^2} = 15,71$$

$$d(2,1) = \sqrt{(93 - 96)^2 + (88 - 90)^2 + (84 - 84)^2 + (92 - 92)^2} = 10,63$$

$$d(3,1) = \sqrt{(94 - 96)^2 + (82 - 90)^2 + (80 - 84)^2 + (81 - 92)^2} = 19,62$$

⋮

$$d(179,1) = \sqrt{(29 - 96)^2 + (43 - 90)^2 + (38 - 84)^2 + (46 - 92)^2} = 109,3$$

3.2.4. Mengurutkan Hasil Perhitungan Jarak

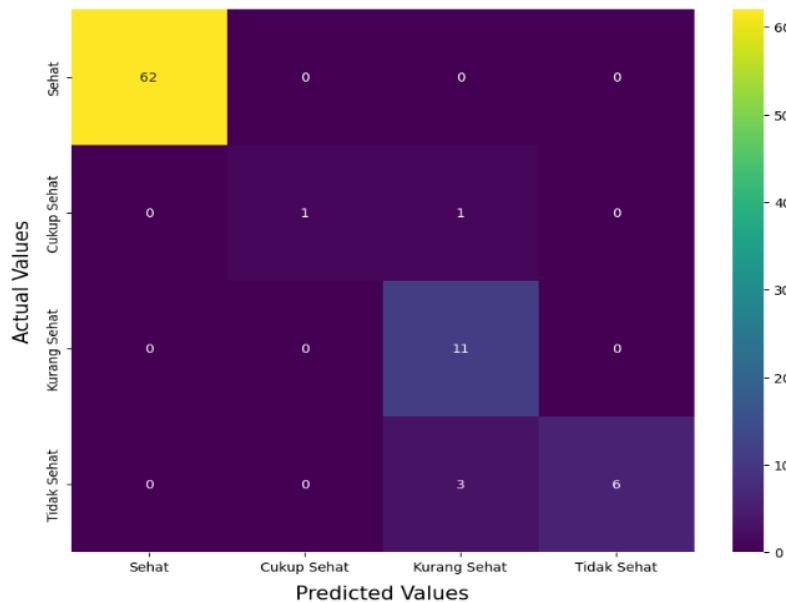
Hasil Perhitungan jarak untuk menentukan urutan terkecil hingga terbesar dari masing-masing data *testing* dan data *training* dapat dilihat pada Tabel 4.

Tabel 4. Ranking Jarak Euclidean antar Data *Training* dan Data *Testing*

Data Testing	Ranking	Data Training	Jarak antar data	Kelas $k = 3$
1	1	49	6,7082	Sehat
	2	11	8,2462	Sehat
	3	46	8,3666	Sehat
2	1	66	4,6904	Sehat
	2	120	5,3851	Sehat
	3	20	5,5677	Sehat
3	1	119	5,0990	Sehat
	2	10	5,7445	Sehat
	3	22	5,9160	Sehat
⋮	⋮	⋮	⋮	⋮
84	1	175	0	Tidak sehat
	2	167	11,916	Kurang sehat
	3	142	13,379	Kurang sehat

3.2.5. Confusion Matrix

Confusion Matrix adalah tabel khusus yang digunakan untuk menunjukkan seberapa baik kinerja algoritma. Setiap baris dalam matriks menggambarkan suatu contoh dalam kelas sebenarnya, sedangkan setiap kolom menggambarkan contoh dalam kelas prediksi, atau sebaliknya [15]. Berikut hasil *Confusion Matrix* yang diperoleh melalui *Google Colabs* dapat dilihat pada Gambar 2.



Gambar 2. Representasi Confusion Matrix

Pada gambar 2 menunjukkan bahwa 84 data testing terdapat, 62 koperasi tepat diklasifikasikan menjadi koperasi sehat. Sementara itu 2 koperasi yang cukup sehat, terdapat 1 koperasi yang tepat dan 1 koperasi lainnya tidak tepat diklasifikasikan. Kemudian 11 koperasi tepat diklasifikasikan

menjadi koperasi kurang sehat. Sedangkan 9 koperasi tidak sehat, terdapat 6 koperasi yang tepat dan 3 lainnya tidak tepat diklasifikasikan. Selanjutnya pada Gambar 3 menunjukkan hasil perhitungan akurasi sebesar 0,95. Untuk mencari hasil perhitungan hasil *presisi* dan, *recall* menggunakan software *Google Colabs* dapat dilihat pada Gambar 3.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	62
1	1.00	0.50	0.67	2
2	0.73	1.00	0.85	11
3	1.00	0.67	0.80	9
accuracy				0.95
macro avg		0.93	0.79	0.83
weighted avg		0.97	0.95	0.95

Gambar 3. Hasil Perhitungan Akurasi, *Presisi* dan, *Recall*

Berdasarkan gambar menunjukkan tingkat akurasi sebesar 95%. Sedangkan untuk nilai presisi secara keseluruhan diperoleh sebesar 79,17 % dan Recall sebesar 93,93%.

4. KESIMPULAN

Hasil penelitian, metode K-Nearest Neighbour berhasil diterapkan dengan baik dalam mengklasifikasikan tingkat kesehatan koperasi di Sumatera Barat tahun 2023, dengan akurasi mencapai 95%. Dari total 84 data uji, sebanyak 80 data berhasil diklasifikasikan dengan benar dan 4 data diprediksi salah. Evaluasi performa model menunjukkan presisi 79,17% dan *recall* 93,93% menandakan ketepatan dan keberhasilan model dalam mengenali kategori yang sesuai. Sementara itu nilai *k* optimal diperoleh pada *k* = 3 berdasarkan metode *Elbow* dengan nilai SSE sebesar 563,7.

REFERENSI

- [1] Laporan Keterangan Dan Pertanggung Jawaban (LKPJ), Padang, 2022.
- [2] Indonesia, Juknis No.15 Ttg Pedoman KKP-KK-Pemkes, Indonesia: Yudhi Novianto, 2021.
- [3] C. Vercellis, *Business Intelligence: Data Mining and Optimization for Decision Making*, Italy: A John Wiley and Sons, 2009.
- [4] F. Gorunescu, *Data Mining Concepts, Models and Techniques*, Romania: Springer-Verlag Berlin Heidelberg, 2011.
- [5] A. Indriyalto, "Algoritma K-Nearest Neighbour untuk Prediksi Nasabah Asuransi," 2021.
- [6] K. Khamar, "Short Text Classification Using KNN Based on Distance Function," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 2, pp. 1916-1919, 2013.
- [7] Katarina, A., Singh, M.D, "A review of data classification using K-Nearest Neighbour algorithm," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, pp. 2250-2459, 2013.
- [8] Sivanandam, *Introduction to Data Mining and its Applications*, Springer, 2006.
- [9] Suyanto, *Machine Learning : Teori dan Aplikasi*, Informatika, 2018.
- [10] X. Y. a. Y. Liu, *Machine Learning*, Springer, 2014, pp. 477-517.
- [11] A. Huda, *Data Mining (Teori dan Aplikasinya)*, 2014.
- [12] Purnima. &. Arvind, *A Clustering Technique based on Elbow Method and K-Means in WSN* *International Journal of Computer Applications*, vol. 105(9), pp. 17-24, 2014.
- [13] J. Gentle, *Finding Groups in Data: An Introduction to Cluster Analysis Biometrics*, 47(2), 788, 1991.
- [14] Suasana, "Algoritma Clustering K-Medoids Pada E-Government Bidang Information and Communication Technology dalam Penentuan Status EDGI," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, pp. 1-10, 2018.
- [15] Baskoro, "Prediksi Penerima Beasiswa dengan Menggunakan Teknik Data mining di Universitas Muhammadiyah Pringsewu, " *Prosiding Seminar Nasional Hasil Penelitian Dan Pengabdian Masyarakat 2021*, pp. 87-94, 2021.