



# Klasifikasi Kemiskinan Di Jawa Timur Menggunakan Data Mining Algoritma K-Nearest Neighbor (KNN)

Yua Isman Islam<sup>1\*</sup>, Zaehol Fatah<sup>2</sup>

Teknologi Informasi, Sains dan Teknologi, Universitas Ibrahimy  
Sistem Informasi, Universitas Ibrahimy  
[yuaismanislam000@email.com](mailto:yuaismanislam000@email.com), [zaeholfatah@email.com](mailto:zaeholfatah@email.com),

## Abstrak

Kemiskinan merupakan masalah sosial-ekonomi yang serius di Jawa Timur, yang memerlukan analisis mendalam untuk memahami faktor-faktor yang memengaruhi kondisi tersebut. Algoritma *K-Nearest Neighbor* (KNN) dalam *data mining* digunakan untuk mengklasifikasikan tingkat kemiskinan berdasarkan data sosial ekonomi di Jawa Timur. Data yang digunakan mencakup beberapa atribut, seperti persentase penduduk miskin, pengeluaran per kapita, indeks pembangunan manusia, dan akses sanitasi. Dengan mengeksplorasi berbagai nilai K, akurasi KNN dievaluasi dalam memprediksi kategori kemiskinan (tinggi atau rendah). Hasil analisis menunjukkan bahwa model KNN mencapai akurasi 92,11%, memberikan kontribusi dalam memetakan kemiskinan yang dapat mendukung kebijakan pengentasan kemiskinan di tingkat daerah.

**Kata Kunci:** Kemiskinan, K-Nearest Neighbor, Data Mining, Klasifikasi

## PENDAHULUAN

Kemiskinan merupakan permasalahan global yang berdampak luas pada kondisi sosial dan ekonomi masyarakat. Di Indonesia, kemiskinan tetap menjadi tantangan besar meskipun telah banyak upaya yang dilakukan oleh pemerintah untuk mengatasinya (DURAESA, 2021). Menurut Badan Pusat Statistik, Provinsi Jawa Timur memiliki jumlah penduduk miskin yang cukup signifikan dibandingkan dengan provinsi lainnya di Indonesia. Kondisi ini menuntut adanya kebijakan yang efektif dalam menanggulangi kemiskinan melalui pemahaman yang lebih mendalam tentang karakteristik penduduk miskin dan faktor-faktor yang memengaruhi kondisi tersebut (TIMUR, 2024).

Data mining adalah salah satu metode yang efektif untuk menganalisis data dalam jumlah besar guna menemukan pola-pola tersembunyi yang dapat digunakan untuk pengambilan keputusan (Muhammad Arhami, S.Si., M.Kom., Muhammad Nasir, S.T., 2020). Dalam konteks kemiskinan, data mining dapat membantu mengidentifikasi karakteristik sosial dan ekonomi yang berhubungan erat dengan kondisi kemiskinan sehingga kebijakan yang diambil dapat lebih tepat sasaran. Di antara berbagai algoritma dalam data mining, algoritma K-Nearest Neighbor (KNN) sering digunakan karena kesederhanaan dan keakuratannya dalam klasifikasi data.

Algoritma KNN adalah metode non-parametrik yang bekerja dengan mengidentifikasi sejumlah K tetangga terdekat untuk menentukan kelas dari suatu data baru (Suntoro, 2019). Dengan menggunakan KNN, data kemiskinan dapat diklasifikasikan ke dalam beberapa kategori berdasarkan karakteristik seperti presentase, pengeluaran, indeks, kesehatan, kualitas air minum, status pekerjaan, dan pdrb. Implementasi KNN dalam klasifikasi kemiskinan dapat membantu pemerintah daerah memahami lebih baik kondisi penduduk miskin di wilayah tertentu, termasuk Jawa Timur, serta mendukung perencanaan program pengentasan kemiskinan yang lebih efektif.

Proses klasifikasi tingkat kemiskinan di Provinsi Jawa Timur menggunakan algoritma KNN. Dataset yang digunakan meliputi data sosial ekonomi yang diperoleh dari sumber-sumber pemerintah dan lembaga statistik. Dengan mengeksplorasi beberapa nilai K dan menganalisis akurasi yang dihasilkan, penelitian ini akan menilai efektivitas KNN dalam klasifikasi data kemiskinan. Hasil penelitian ini diharapkan dapat memberikan kontribusi nyata dalam upaya perencanaan kebijakan berbasis data yang lebih tepat sasaran untuk pengentasan kemiskinan di Jawa Timur.

## METODE

### Tahapan Penelitian

Bagian ini menjelaskan metode yang digunakan untuk mengumpulkan data dalam studi yang sedang dilakukan. Data dikumpulkan melalui studi literatur menggunakan data uji dan data latih berupa data yang diambil dari situs Kaggle.com. Setelah disaring, data ini akan diproses menggunakan aplikasi RapidMiner untuk memprediksi tingkat kemiskinan (rendah atau tinggi) di Jawa Timur. Beberapa atribut yang dianalisis meliputi persentase Penduduk Miskin,

Pengeluaran per Kapita, IPM, Akses Sanitasi, Akses Air Minum, Tingkat Pengangguran, TPAK, PDRB, dan Klasifikasi. Dataset ini ditampilkan pada gambar di bawah ini:

Tabel 1. Data Training

Tahun	Kota/ Kab	Penduduk Miskin (%)	Pengeluaran per Kapita (Ribu)	IPM	Akses Sanitasi (%)	Akses Air Minum (%)	Tingkat Pengangguran (%)	TPAK	PDRB (Rupiah)	Klasifikasi
2020	Pacitan	15,11	8887	68,57	71,7	76,71	2,04	80,57	11107402	Rendah
2020	Ponorogo	10,26	9851	71,06	87,49	96,3	4,38	72,63	14619969	Rendah
2020	Trenggalek	12,14	9743	70,06	75,48	75,78	3,53	72,36	12959018	Rendah
2020	Tulungagung	7,51	10807	73,15	85,37	97,4	4,91	72,26	27390424	Rendah
2020	Blitar	9,65	10757	71,05	75,67	93,33	3,66	70,44	25700019	Rendah
2020	Kediri	11,64	11127	72,56	86,92	96,35	5,15	69,34	29361672	Rendah
2020	Malang	10,5	10163	70,6	80,79	94,14	5,4	68,49	68619103	Rendah
2020	Lumajang	10,05	9203	66,07	83,24	95,73	3,51	66,19	22623402	Rendah
2020	Jember	10,41	9410	67,32	64,17	94,35	5,44	68,97	54688719	Rendah
2020	Banyuwangi	8,07	12217	71,38	81,15	94,16	5,42	72,32	55471065	Rendah
2020	Bondowoso	14,73	10690	66,59	52,82	93,44	4,46	73,89	13921654	Rendah

Tabel 2. Data Testing

Tahun	Kota/Kab	Penduduk Miskin (%)	Pengeluaran per Kapita (Ribu)	IPM	Akses Sanitasi (%)	Akses Air Minum (%)	Tingkat Pengangguran (%)	TPAK	PDRB (Rupiah)	Klasifikasi
2023	Gresik	12,42	13280	76,5	91,56	88,99	8	69,43	10131868	6
2023	Bangkalan	21,57	8673	64,36	39,44	97,39	8,07	68,66	17152779	
2023	Sampang	23,76	8790	62,8	76,22	89,34	3,45	70,19	13984568	
2023	Pamekasan	15,3	8804	66,4	70,85	89,15	3,1	65,88	11496236	
2023	Sumenep	20,51	9000	67,04	65,66	98,29	2,31	75,63	24161351	
2023	Kota Kediri	7,75	12359	78,6	95,75	98,95	6,37	67,35	86485594	
2023	Kota Blitar	7,89	13816	78,98	96,77	97,08	6,61	69,96	4924572	
2023	Kota Malang	4,62	16663	82,04	87,08	99,93	9,65	67,59	53309702	
2023	Kota Probolinggo	7,44	12245	73,66	89,77	99,66	6,55	69,71	8361142	
2023	Kota Pasuruan	6,88	13354	75,62	92,1	99,65	6,23	71,66	5914585	
2023	Kota Mojokerto	6,39	13610	78,43	95,49	99,63	6,87	67,09	4976490	
2023	Kota Madiun	5,09	16095	81,25	97,31	99,53	8,15	66,87	10748101	
2023	Kota Surabaya	5,23	17862	82,31	95,2	98,06	9,68	67,3	40772679	9
2023	Kota Batu	4,09	12887	76,28	95,57	100	6,57	73,74	11471435	

### Klasifikasi

. Pada dasarnya klasifikasi merupakan proses membagi sekumpulan data sehingga setiap data menjadi anggota suatu kategori atau kelas (Adinugroho, 2018). Tujuan utama dari klasifikasi adalah memahami pola atau hubungan tersembunyi dalam data dan memanfaatkan informasi ini untuk memprediksi kategori dari data baru yang belum diketahui. (Sari, 2024).

## Data Mining

Data mining merupakan proses penggalian informasi dan pola yang bermanfaat dari data yang sangat besar. Data mining mencakup pengumpulan data, ekstraksi data, analisis data, dan statistik data (Muhammad Arhami, S.Si., 2020). Secara sederhana, data mining adalah proses menemukan informasi baru melalui pencarian pola atau aturan tertentu dalam kumpulan data yang sangat besar. Tujuan utama dari data mining adalah memeriksa database berukuran besar untuk menemukan pola atau bentuk baru yang bermanfaat dalam pengambilan keputusan (Mustika, 2021)

## K-Nearest Neighbor (KNN)

Algoritma k-Nearest Neighbor adalah metode pembelajaran terawasi yang menggunakan data pelatihan, di mana hasil klasifikasi dari instance baru ditentukan berdasarkan kategori dengan kedekatan jarak terbanyak dari objek-objek sekitarnya (Setiawan, 2023). Metode ini termasuk dalam algoritma supervised learning, yang berarti dataset harus memiliki target. Dalam penelitian ini, penentuan nilai K dihitung menggunakan rumus Euclidean Distance (Yogianto et al., 2024).

Kemiripan data dengan label yang digunakan dihitung menggunakan rumus Euclidean distance sebagai berikut:

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

Keterangan:

$x_i$  = data latih (data training)

$y_i$  = data uji (data testing)

$i$  = variabel data

$euc = d(x, y)$  = jarak

$n$  = jumlah data

Euclidean Distance sering digunakan untuk menghitung jarak dan berfungsi sebagai ukuran yang dapat menginterpretasikan kedekatan antara dua objek.

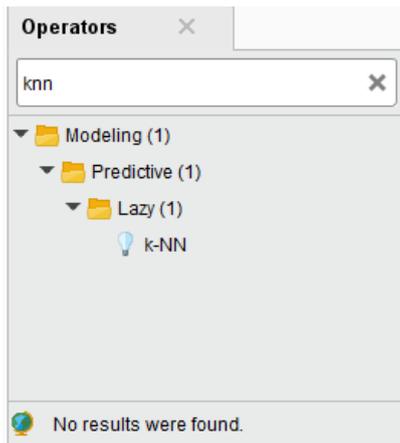
## HASIL DAN PEMBAHASAN

Model klasifikasi yang diterapkan adalah K-Nearest Neighbors (K-NN) menggunakan algoritma Machine Learning pada RapidMiner versi 10.4. Dataset yang digunakan berasal dari situs Kaggle.com, mencakup 9 atribut, termasuk klasifikasi tentang rendah atau tingginya kemiskinan di Jawa Timur.

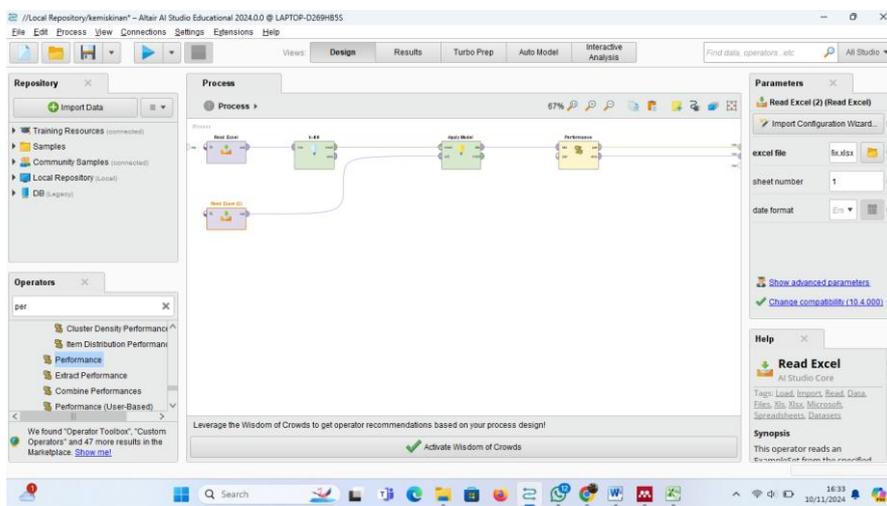
1. **Memasukkan data ke RapidMiner.** Gambar di bawah ini menunjukkan tampilan antarmuka pada tahap pengaturan data di Altair AI Studio, di mana data yang diimpor siap untuk diproses lebih lanjut dalam analisis klasifikasi kemiskinan di Jawa Timur menggunakan algoritma K-Nearest Neighbor (KNN). Klasifikasi, berisi label kategori kemiskinan, yang akan digunakan sebagai target dalam proses klasifikasi. Pengaturan data ini sangat penting dalam penelitian karena atribut-atribut tersebut merupakan variabel penentu untuk memprediksi tingkat kemiskinan menggunakan KNN, di mana nantinya model akan belajar dari pola data ini untuk mengklasifikasikan wilayah berdasarkan tingkat kemiskinan.

	Akses Sanit...	Akses Air M...	Tingkat Pe...	TPAK	PDRB (Rupi...	Klasifikasi
	real	real	real	real	integer	polynomial label
1	71.700	76.710	2.040	80.570	11107402	Rendah
2	87.490	96.300	4.380	72.630	14619969	Rendah
3	75.480	75.780	3.530	72.360	12959018	Rendah
4	85.370	97.400	4.910	72.260	27390424	Rendah
5	75.670	93.330	3.660	70.440	25700019	Rendah
6	86.920	96.350	5.150	69.340	29361672	Rendah
7	80.790	94.140	5.400	68.490	68619103	Rendah
8	83.240	95.730	3.510	66.190	22623402	Rendah
9	64.170	94.350	5.440	68.970	54688719	Rendah
10	81.150	94.160	5.420	72.320	55471065	Rendah
11	52.820	93.440	4.460	73.890	13921654	Rendah
12	60.700	98.880	3.680	71.630	13715831	Rendah

2. **Menambahkan Operator KNN.** Naïve Bayes menggunakan metode probabilistik untuk memprediksi peningkatan kinerja berdasarkan data yang tersedia.



3. **Menjalankan Model.** Setelah semua operator terhubung dengan benar, klik tombol **Run** untuk menjalankan proses KNN.



4. **Hasil Klasifikasi Tabel.** Gambar di bawah menunjukkan hasil klasifikasi kemiskinan di Jawa Timur pada Altair AI Studio dengan asumsi menggunakan algoritma K-Nearest Neighbor (KNN). Data menampilkan prediksi kategori kemiskinan rendah atau tinggi untuk setiap kota/kabupaten pada tahun 2023

Row No.	Klasifikasi	predictionK...	confidence...	confidence...	Tahun	Kota/Kab	Penduduk M.	Pengusaha...	IPK
1	?	Rendah	1	0	2023	Gresik	12.420	13290	76
2	?	Rendah	0.576	0.424	2023	Bangkalan	21.570	8673	64
3	?	Rendah	0.750	0.250	2023	Sampang	23.760	8790	62
4	?	Rendah	1.000	0	2023	Pamekasan	15.300	8804	66
5	?	Rendah	0.750	0.250	2023	Sumenep	20.510	9000	67
6	?	Rendah	1	0	2023	Kota Kediri	7.750	12359	78
7	?	Rendah	1	0	2023	Kota Blitar	7.890	13816	78
8	?	Rendah	1	0	2023	Kota Malang	4.620	16663	82
9	?	Rendah	1.000	0	2023	Kota Probolin.	7.440	12245	73
10	?	Rendah	1	0	2023	Kota Pasuruan	6.880	13354	75
11	?	Rendah	1	0	2023	Kota Mojokerto	6.390	13610	78
12	?	Rendah	1	0	2023	Kota Madiun	5.090	16095	81
13	?	Rendah	1	0	2023	Kota Surabaya	5.230	17862	82
14	?	Rendah	1	0	2023	Kota Batu	4.090	12887	76

5. **Hasil Akurasi.** Gambar di bawah menampilkan evaluasi model klasifikasi kemiskinan di Jawa Timur menggunakan algoritma K-Nearest Neighbor (KNN) pada Altair AI Studio, dengan akurasi 92,11%. Dari 38

data uji, model berhasil memprediksi 35 data sebagai "Rendah" dengan tepat, sementara 3 data yang sebenarnya "Tinggi" diprediksi sebagai "Rendah". Precision kelas "Rendah" adalah 92,11%, menandakan bahwa sebagian besar prediksi "Rendah" sesuai dengan data sebenarnya, sedangkan recall kelas "Rendah" mencapai 100%, menunjukkan bahwa semua data "Rendah" dengan Nilai  $K = 5$ .

	true Rendah	true Tinggi	class precision
pred. Rendah	35	3	92.11%
pred. Tinggi	0	0	0.00%
class recall	100.00%	0.00%	

## KESIMPULAN

Algoritma K-Nearest Neighbor (KNN) terbukti efektif dalam mengklasifikasikan tingkat kemiskinan di Jawa Timur, dengan mencapai akurasi sebesar 92,11%. Dengan memanfaatkan data sosial ekonomi, termasuk atribut seperti persentase penduduk miskin, pengeluaran per kapita, dan indeks pembangunan manusia, KNN dapat mengidentifikasi pola-pola yang mendukung pemetaan kemiskinan secara tepat. Hasil klasifikasi ini memberikan informasi yang bermanfaat bagi pemerintah daerah untuk memahami distribusi kemiskinan dan mendukung perencanaan kebijakan yang lebih tepat sasaran dalam mengatasi kemiskinan di wilayah tersebut.

## UCAPAN TERIMA KASIH

Terima kasih disampaikan kepada pihak-pihak yang telah mendukung terlaksananya penelitian ini.

## DAFTAR PUSTAKA

- Adinugroho, S. (2018). *Implementasi Data Mining Menggunakan Weka*. Universitas Brawijaya Press.
- DURAESA, M. A. (2021). *kEMISKINAN DI INDONESIA Antara Kajian Empiris dan Teologis* (Ke 1). RAJAWALI PERS.
- Muhammad Arhami, S.Si., M.Kom., Muhammad Nasir, S.T., M. T. (2020). *Data Mining*. Andi Offset.
- Muhammad Arhami, S.Si., M. K. (2020). *Data Mining - Algoritma dan Implementasi*. Andi Offset.
- Mustika. (2021). *DATA MINING DAN APLIKASINYA*. Widina.
- Sari, R. M. (2024). *Klasifikasi Data Mining*. Serasi Media Teknologi.
- Setiawan, Z. (2023). *BUKU AJAR DATA MINING*. PT. Sonpedia Publising Indonesia.
- Suntoro, J. (2019). *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Elex Media Komputindo.
- TIMUR, B. P. S. P. J. (2024). *Profil Kemiskinan Jawa Timur Tahun 2023*. <https://doi.org/3205005.35>
- Yogianto, A., Homaidi, A., & Fatah, Z. (2024). Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung. *G-Tech: Jurnal Teknologi Terapan*, 8(3), 1720–1728. <https://doi.org/10.33379/gtech.v8i3.4495>