MATHunesa
Jurnal Ilmiah Matematika *Volume 14 No 02*
e-ISSN : 2716-506X | p-ISSN : 2301-9115 *Tahun 2026*

KLASIFIKASI LAJU PERTUMBUHAN PENDUDUK ANTAR PROVINSI DI INDONESIA MENGGUNAKAN METODE K-NEAREST NEIGHBOR

Agung Atra Perkasa*
Program Studi S1 Statistika, FMIPA, Universitas Negeri Medan
email: agungatraperkasa1408@gmail.com

Emon Sejahtera Zendrato
Program Studi S1 Statistika, FMIPA, Universitas Negeri Medan
email: yuhuesz@gmail.com

Ilham Pratama
Program Studi S1 Statistika, FMIPA, Universitas Negeri Medan
email: ilhampratamaaaa595@gmail.com

Hizkia Simamora
Program Studi S1 Statistika, FMIPA, Universitas Negeri Medan
email: hizkiakhairossimamora@gmail.com

Abstrak

Data kependudukan memiliki peranan penting dalam perencanaan pembangunan, namun pemanfaatannya masih belum optimal untuk menghasilkan informasi yang benar-benar informatif. Penelitian ini bertujuan mengklasifikasikan laju pertumbuhan penduduk antar provinsi di Indonesia ke dalam kategori rendah, sedang, dan tinggi menggunakan algoritma K-Nearest Neighbor (KNN). Proses penelitian meliputi pengumpulan dan praproses data, pembentukan label kategori menggunakan metode kuantil, normalisasi data dengan Min-Max Scaling, serta pembagian data menjadi data latih dan data uji dengan rasio 70:30. Model KNN dibangun menggunakan nilai parameter k, yaitu 3, 5, dan 7, kemudian dipilih nilai k terbaik berdasarkan hasil evaluasi model. Evaluasi dilakukan menggunakan confusion matrix dengan menghitung nilai akurasi. Hasil pengujian menunjukkan bahwa model terbaik diperoleh pada nilai k = 5 dengan akurasi sebesar 75%. Temuan ini menunjukkan bahwa KNN dapat mengidentifikasi kemiripan karakteristik demografis antar provinsi secara cukup baik, meskipun masih terdapat kesalahan klasifikasi pada kelas dengan karakteristik yang saling berdekatan. Oleh karena itu, metode KNN dapat dijadikan pendekatan yang sederhana dan efektif dalam analisis data kependudukan serta berpotensi mendukung pengambilan keputusan berbasis data.
Kata Kunci: data mining, K-Nearest Neighbor, klasifikasi, kependudukan, pertumbuhan penduduk.

Abstract

Population data plays a crucial role in development planning, yet its utilization has not yet been optimized to generate truly informative insights. This study aims to classify population growth rates across provinces in Indonesia into low, medium, and high categories using the K-Nearest Neighbor (KNN) algorithm. The research process includes data collection and preprocessing, category labeling using the quantile method, data normalization via Min-Max Scaling, and splitting the data into training and test sets with a 70:30 ratio. The KNN model was built using parameter values of k, namely 3, 5, and 7, and the best value of k was selected based on the model evaluation results. Evaluation was performed using a confusion matrix by calculating the accuracy value. The test results showed that the best model was obtained at k = 5 with an accuracy of 75%. These findings indicate that KNN can identify similarities in demographic characteristics across provinces quite well, although there are still classification errors in classes with closely related characteristics. Therefore, the KNN method can serve as a simple and effective approach for population data analysis and has the potential to support data-driven decision-making.
Keywords: *data mining, K-Nearest Neighbor, classification, population, population growth.*

PENDAHULUAN

Data kependudukan di Indonesia terus mengalami perubahan dari waktu ke waktu, baik dari sisi jumlah penduduk maupun laju pertumbuhannya. Perubahan ini menjadi hal yang penting untuk diperhatikan karena berpengaruh terhadap berbagai aspek pembangunan, seperti penyediaan infrastruktur, pelayanan publik, serta pemerataan kesejahteraan masyarakat di setiap provinsi (Lee, 2011). Informasi mengenai jumlah dan pertumbuhan penduduk dapat memberikan gambaran mengenai kondisi suatu wilayah, sehingga dapat digunakan sebagai dasar dalam perencanaan yang lebih tepat. Namun, data yang tersedia dalam jumlah besar seringkali hanya disajikan dalam bentuk statistik tanpa dilakukan analisis lebih lanjut, sehingga informasi yang terkandung di dalamnya belum dimanfaatkan secara optimal (Kitchin, 2014).

Kondisi tersebut menunjukkan bahwa masih terdapat keterbatasan dalam pemanfaatan data kependudukan untuk memahami pola pertumbuhan penduduk secara lebih mendalam. Padahal, analisis terhadap pola tersebut sangat penting untuk mengidentifikasi perbedaan karakteristik antar wilayah, seperti daerah dengan pertumbuhan rendah, sedang, maupun tinggi. Tanpa adanya pengolahan data yang tepat, sulit untuk memperoleh informasi yang dapat mendukung pengambilan keputusan secara akurat. Oleh karena itu, diperlukan suatu pendekatan yang mampu mengolah data secara sistematis agar dapat menghasilkan informasi yang lebih bermakna.

Salah satu pendekatan yang dapat digunakan adalah data mining, yaitu proses untuk menemukan pola atau hubungan tertentu dari sekumpulan data dalam jumlah besar (Kitchin, 2014). Dalam penelitian ini, teknik klasifikasi digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan karakteristik yang dimiliki. Melalui pendekatan ini, data yang sebelumnya hanya berupa angka dapat diolah menjadi informasi yang lebih mudah dipahami dan digunakan dalam proses analisis.

Algoritma yang digunakan dalam penelitian ini adalah K-Nearest Neighbor (KNN), yang bekerja dengan menentukan kedekatan antar data berdasarkan perhitungan jarak tertentu. Metode ini mengelompokkan suatu data ke dalam kategori berdasarkan mayoritas kelas dari tetangga terdekatnya (Feng et al., 2018; Zhou & Wang, 2015). Penelitian ini bertujuan untuk mengklasifikasikan laju pertumbuhan penduduk antar provinsi ke dalam kategori rendah, sedang, dan tinggi secara objektif berdasarkan data jumlah penduduk dan laju pertumbuhan penduduk. Proses penelitian dilakukan melalui beberapa tahapan, meliputi preprocessing, pelabelan data, normalisasi, serta pembagian data untuk keperluan pelatihan dan pengujian model. Dengan demikian, penelitian ini diharapkan dapat memberikan gambaran mengenai pola pertumbuhan penduduk serta menjadi referensi dalam penerapan metode data mining pada analisis data kependudukan.

KAJIAN TEORI

Klasifikasi merupakan salah satu metode dalam machine learning yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan karakteristik atau fitur yang dimiliki. Metode ini termasuk ke dalam supervised learning, karena model dibangun menggunakan data berlabel untuk memprediksi kelas pada data baru. Secara umum, klasifikasi bertujuan membentuk fungsi yang mampu memetakan atribut data ke dalam kelas diskrit berdasarkan pola yang telah dipelajari. Dalam penerapannya, teknik ini banyak dimanfaatkan di berbagai bidang, termasuk demografi, karena mampu menghasilkan prediksi yang cukup akurat serta mendukung pengambilan keputusan berbasis data (Hendriyanto & Sari, 2022; Insan & Kusrini, 2021).

K-Nearest Neighbors (KNN) dikenal sebagai algoritma klasifikasi non-parametrik yang mengelompokkan data berdasarkan kedekatan jarak antar objek dalam ruang fitur. Kelas suatu data ditentukan melalui mayoritas dari k tetangga terdekatnya. Untuk mengukur kedekatan tersebut, umumnya digunakan Euclidean distance, sehingga objek dengan jarak paling kecil dianggap memiliki tingkat kemiripan yang lebih tinggi (Kotjek et al., 2025; Ujianto et al., 2025).

Adapun langkah-langkah dalam algoritma KNN adalah sebagai berikut:

1. Menentukan nilai k sebagai jumlah tetangga terdekat yang akan digunakan.

2. Menghitung jarak antara data uji dengan seluruh data training menggunakan metode jarak (misalnya Euclidean distance).
3. Mengurutkan hasil perhitungan jarak dari yang terkecil hingga terbesar.
4. Memilih k data dengan jarak terdekat sebagai tetangga.
5. Melakukan proses voting terhadap kelas dari tetangga tersebut.
6. Menentukan kelas mayoritas sebagai hasil klasifikasi data uji.

Algoritma KNN memiliki beberapa keunggulan, seperti kemudahan dalam implementasi serta tidak memerlukan proses pelatihan yang kompleks. Selain itu, metode ini mampu menangani pola data yang tidak linear dan dapat diterapkan pada berbagai jenis data numerik. Namun demikian, terdapat keterbatasan, terutama pada tingginya beban komputasi karena perhitungan jarak dilakukan terhadap seluruh data training, khususnya pada dataset berukuran besar. Di samping itu, KNN cukup sensitif terhadap skala data sehingga diperlukan normalisasi agar setiap fitur memiliki kontribusi yang seimbang. Akurasi model juga dapat dipengaruhi oleh pemilihan nilai k serta keberadaan noise atau outlier dalam data (Bakri & Harahap, 2025; Harahap & Harahap, 2025).

Pada konteks data kependudukan, metode klasifikasi seperti KNN dapat digunakan untuk mengelompokkan wilayah berdasarkan karakteristik tertentu, seperti jumlah penduduk, laju pertumbuhan, persentase penduduk, dan kepadatan penduduk. Data kependudukan yang bersifat multidimensi serta memiliki variasi antarwilayah memerlukan metode yang mampu mengenali pola kemiripan secara efektif. Oleh karena itu, KNN dapat dimanfaatkan untuk mengelompokkan provinsi ke dalam kategori tertentu, misalnya tingkat pertumbuhan penduduk (rendah, sedang, tinggi), sehingga memberikan gambaran kondisi demografi secara lebih sistematis (Sari et al., 2024).

Berbagai penelitian menunjukkan bahwa KNN memiliki performa yang cukup baik dalam beragam kasus klasifikasi berbasis data sosial, termasuk data kependudukan. Tingkat akurasi yang dihasilkan umumnya dipengaruhi oleh tahap preprocessing serta ketepatan dalam pemilihan parameter (Hendriyanto & Sari, 2022; Ujianto et al., 2025). Hal ini menunjukkan bahwa KNN cukup fleksibel dan dapat diterapkan pada berbagai jenis data dengan karakteristik yang berbeda.

Penerapan algoritma KNN dalam penelitian dilakukan melalui beberapa tahapan sebagai berikut:

1. Pengumpulan data, yaitu mengumpulkan data kependudukan yang relevan.
2. Preprocessing data, meliputi pembersihan data dan normalisasi untuk meningkatkan akurasi perhitungan jarak
3. Pembagian data, menjadi data training dan data testing.
4. Pemilihan fitur, untuk menentukan variabel yang paling berpengaruh dalam klasifikasi.
5. Penentuan nilai k, yang dapat dilakukan menggunakan metode validasi seperti cross-validation.
6. Proses klasifikasi, dengan menghitung jarak dan menentukan kelas berdasarkan tetangga terdekat.
7. Evaluasi model, menggunakan metrik seperti akurasi, presisi, dan recall.

Secara umum, tahapan tersebut merupakan prosedur standar dalam penerapan algoritma klasifikasi berbasis instance seperti KNN. Oleh karena itu, keberhasilan metode ini sangat dipengaruhi oleh kualitas data, tahap preprocessing, serta ketepatan dalam menentukan parameter, sehingga model yang dibangun dapat menghasilkan klasifikasi yang lebih optimal (Kotjek et al., 2025; Insan & Kusrini, 2021).

Berbagai penelitian menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) memiliki performa yang baik dalam berbagai permasalahan klasifikasi berbasis data sosial maupun numerik. Penerapan KNN pada klasifikasi judul berita hoaks berbahasa Indonesia menunjukkan bahwa metode ini mampu mengenali pola kemiripan data secara efektif dan menghasilkan tingkat akurasi yang cukup tinggi (Hendriyanto & Sari, 2022). Selain itu, perbandingan algoritma ID3 dan KNN pada klasifikasi emosi teks berita menunjukkan bahwa KNN memiliki kinerja yang kompetitif dalam memprediksi kelas berdasarkan kedekatan karakteristik data (Insan & Kusrini, 2021).

Pada bidang sosial dan ekonomi, metode KNN juga telah digunakan untuk mengklasifikasikan tingkat kemiskinan sebagai pendukung pengambilan keputusan sosial. Hasil penelitian tersebut menunjukkan bahwa KNN mampu memberikan

klasifikasi yang cukup akurat berdasarkan kemiripan indikator antar wilayah (Juanuari et al., 2026). Penelitian lain menunjukkan bahwa KNN efektif digunakan untuk memprediksi ketidaklulusan mahasiswa berdasarkan data akademik, sehingga menegaskan fleksibilitas algoritma ini dalam menangani data multidimensi (Nur et al., 2025).

Meskipun berbagai penelitian telah menunjukkan keberhasilan penerapan KNN pada beragam bidang, penerapannya pada klasifikasi laju pertumbuhan penduduk antar provinsi di Indonesia masih relatif terbatas, khususnya menggunakan data kependudukan terbaru. Oleh karena itu, penelitian ini dilakukan untuk mengisi celah tersebut dengan menerapkan metode KNN dalam mengklasifikasikan laju pertumbuhan penduduk berdasarkan karakteristik demografis provinsi (Sari et al., 2024). Penelitian ini diharapkan dapat memperluas penerapan metode KNN pada analisis data kependudukan serta memberikan gambaran pola pertumbuhan penduduk secara lebih sistematis.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi berbasis algoritma K-Nearest Neighbor (K-NN) untuk mengelompokkan provinsi di Indonesia ke dalam kategori laju pertumbuhan penduduk, yaitu rendah, sedang, dan tinggi. Klasifikasi dilakukan berdasarkan kedekatan jarak antar data, di mana suatu data ditentukan kelasnya melalui mayoritas tetangga terdekat dalam ruang fitur. Algoritma K-NN dikenal sebagai metode yang sederhana dan efektif dalam berbagai kasus klasifikasi (Hakim et al., 2026). Perhitungan jarak antar data menggunakan Euclidean Distance sebagai berikut:

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

Populasi penelitian mencakup seluruh data provinsi di Indonesia yang memiliki informasi jumlah penduduk, laju pertumbuhan penduduk, persentase penduduk, dan kepadatan penduduk. Teknik sampling yang diterapkan adalah sampling jenuh, sehingga seluruh populasi digunakan sebagai sampel. Data yang dianalisis merupakan data sekunder yang diperoleh dari publikasi resmi statistik kependudukan tahun 2024 dengan jumlah total 38 provinsi.

Pengumpulan data dilakukan melalui studi dokumentasi dengan mengunduh dataset dari sumber terkait, kemudian diseleksi berdasarkan variabel yang relevan. Variabel dalam penelitian ini dibedakan menjadi variabel fitur dan variabel target. Variabel fitur meliputi jumlah penduduk, persentase penduduk, dan kepadatan penduduk, sedangkan variabel target berupa kategori laju pertumbuhan penduduk. Variabel laju pertumbuhan penduduk tidak digunakan sebagai fitur, melainkan hanya sebagai dasar pembentukan label kelas. Hal ini dilakukan untuk menghindari terjadinya data leakage, yaitu kondisi di mana informasi target digunakan dalam proses pelatihan model yang dapat menyebabkan hasil klasifikasi menjadi bias.

Pembentukan label kelas dilakukan menggunakan metode kuantil (tertile), yaitu dengan membagi data laju pertumbuhan penduduk ke dalam tiga kelompok berdasarkan distribusinya. Kategori kelas ditentukan sebagai berikut: rendah untuk data pada rentang kuantil 0%–33%, sedang untuk rentang 33%–66%, dan tinggi untuk rentang 66%–100%. Metode ini dipilih karena mampu menghasilkan pembagian kelas yang proporsional dan objektif tanpa bergantung pada asumsi distribusi tertentu.

Tahap analisis diawali dengan preprocessing data, meliputi pengecekan missing value, duplikasi data, serta penyesuaian format data numerik. Data kemudian dibagi menjadi data latih dan data uji dengan rasio 70:30, sehingga diperoleh sekitar 26 data latih dan 12 data uji. Sebelum pemodelan, dilakukan normalisasi menggunakan metode Min-Max Scaling untuk menyamakan rentang nilai antar fitur. Proses normalisasi dinyatakan dengan rumus sebagai berikut:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Nilai minimum (X_{min}) dan maksimum (X_{max}) diperoleh dari data latih, kemudian digunakan untuk mentransformasikan data uji guna mencegah kebocoran informasi dari data uji ke dalam proses pelatihan model.

Model K-NN dibangun dengan menguji beberapa nilai parameter k, yaitu k = 3, 5, dan 7. Pengujian ini dilakukan untuk memperoleh nilai k yang memberikan performa terbaik dalam proses klasifikasi. Klasifikasi dilakukan dengan menghitung jarak antar data yang telah dinormalisasi, kemudian

kelas ditentukan berdasarkan mayoritas dari k tetangga terdekat.

Evaluasi kinerja model dilakukan menggunakan confusion matrix untuk membandingkan hasil prediksi dengan data aktual. Berdasarkan confusion matrix tersebut dihitung nilai akurasi sebagai indikator utama kinerja model. Rumus akurasi dinyatakan sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} x\ 100\%$$

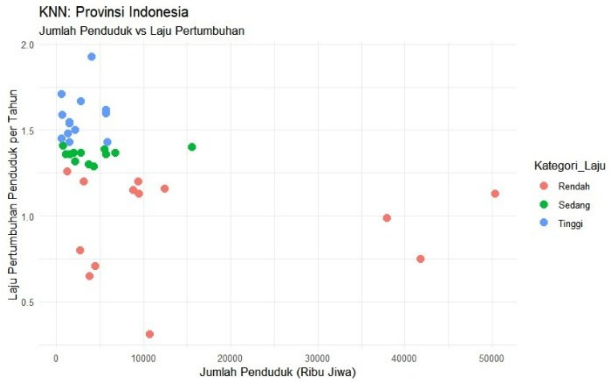
Dimana TP (True Positive), TN (True Negative), FP (False Positive), dan FN (False Negative) merupakan komponen dari confusion matrix.

Penelitian ini menggunakan perangkat lunak R sebagai alat utama dalam pengolahan dan analisis data. Library yang digunakan antara lain package *class* untuk implementasi algoritma K-NN dan *ggplot2* untuk visualisasi data. Data yang digunakan berupa dataset numerik hasil olahan yang berasal dari data kependudukan provinsi di Indonesia, sehingga tidak memerlukan bahan fisik, melainkan berupa data digital yang diolah secara komputasional.

HASIL DAN PEMBAHASAN

Dataset pada penelitian ini berasal dari data kependudukan 38 provinsi di Indonesia tahun 2024 yang mencakup lima variabel utama, yaitu provinsi, jumlah penduduk, laju pertumbuhan penduduk, persentase penduduk, dan kepadatan penduduk. Variabel-variabel tersebut digunakan sebagai dasar klasifikasi laju pertumbuhan penduduk antarprovinsi menggunakan metode K-Nearest Neighbor (KNN), karena mampu merepresentasikan karakteristik demografis secara menyeluruh.

Sebelum klasifikasi, data disusun secara terstruktur. Variabel laju pertumbuhan penduduk dijadikan sebagai dasar pembentukan label kategori menjadi tiga kelas, yaitu rendah, sedang, dan tinggi menggunakan metode kuantil (tertile). Pengelompokan ini bertujuan untuk mempermudah model dalam mengenali pola distribusi data. Penggunaan atribut numerik yang relevan juga mendukung kinerja KNN karena algoritma ini mengandalkan perhitungan jarak antar data dalam ruang multidimensi (Hadi et al., 2024). Dataset yang terdiri dari 38 observasi dinilai cukup merepresentasikan variasi pertumbuhan penduduk secara nasional dan menjadi dasar sebelum tahap preprocessing.



Gambar 1. Visualisasi Data Provinsi Berdasarkan Jumlah Penduduk dan Laju Pertumbuhan.

Pada Gambar 1. menunjukkan visualisasi data berdasarkan jumlah penduduk dan laju pertumbuhan penduduk. Perbedaan warna menunjukkan kategori rendah, sedang, dan tinggi. Sebagian data membentuk kelompok yang cukup jelas, namun terdapat titik yang saling berdekatan. Hal ini menunjukkan adanya kemiripan karakteristik antarprovinsi, terutama pada kategori sedang dan tinggi (Harahap & Harahap, 2025). Kondisi tersebut mengindikasikan adanya overlap antar kelas yang dapat memengaruhi hasil klasifikasi, karena KNN menentukan kelas berdasarkan mayoritas tetangga terdekat (Juanuari et al., 2026).

Setelah tahap preprocessing dan normalisasi, pemodelan dilakukan menggunakan algoritma KNN dengan menguji beberapa nilai parameter k, yaitu k = 3, 5, dan 7. Pengujian ini bertujuan untuk memperoleh nilai k yang memberikan performa terbaik dalam proses klasifikasi. Hasil evaluasi model berdasarkan nilai akurasi disajikan pada Tabel 1.

Nilai k	Akurasi
3	50%
5	75%
7	75%

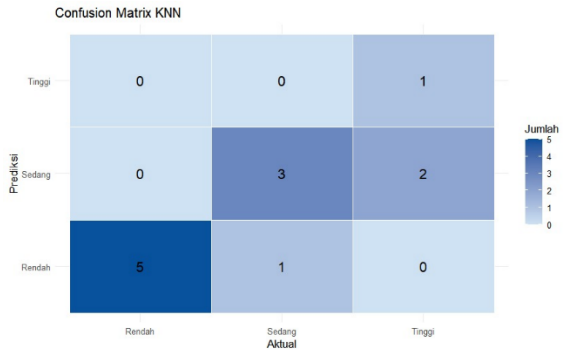
Tabel 1. Perbandingan Akurasi Model KNN Berdasarkan Nilai k.

Berdasarkan hasil pengujian tersebut, diperoleh bahwa nilai k = 5 dan k = 7 menghasilkan tingkat akurasi yang sama, yaitu sebesar 75%, sedangkan k = 3 menghasilkan akurasi yang lebih rendah, yaitu 50%. Nilai k yang terlalu kecil cenderung membuat model sangat sensitif terhadap data terdekat

sehingga rentan terhadap noise dan variasi local (Hadi et al., 2024).Hal ini menyebabkan penurunan akurasi pada k = 3.

Meskipun k = 5 dan k = 7 memiliki akurasi yang sama, nilai k = 5 dipilih sebagai parameter terbaik karena mampu mempertahankan sensitivitas model terhadap pola lokal data. Sementara itu, nilai k yang lebih besar seperti k = 7 cenderung menghasilkan klasifikasi yang lebih umum (over-smoothing), sehingga berpotensi mengurangi kemampuan model dalam menangkap perbedaan karakteristik antar data.

Selanjutnya, dilakukan analisis lebih lanjut menggunakan confusion matrix untuk nilai k terbaik, yaitu k = 5. Hasil klasifikasi menunjukkan bahwa dari 12 data uji, sebanyak 9 data berhasil diklasifikasikan dengan benar, sedangkan 3 data mengalami kesalahan. Pada kategori rendah, sebagian besar data teridentifikasi dengan tepat, sementara pada kategori sedang dan tinggi masih terjadi kekeliruan prediksi. Hal ini menunjukkan bahwa model mengalami kesulitan dalam membedakan kelas dengan karakteristik yang mirip (Juanuari et al., 2026).



Gambar 2. Confusion Matrix Hasil Klasifikasi KNN.

Berdasarkan confusion matrix pada Gambar 2. menunjukkan confusion matrix hasil klasifikasi KNN. Berdasarkan gambar tersebut, kesalahan klasifikasi paling banyak terjadi pada kategori tinggi yang beberapa datanya diprediksi sebagai sedang. Kondisi ini menunjukkan bahwa model mengalami kesulitan dalam membedakan dua kategori tersebut karena memiliki karakteristik nilai yang saling berdekatan. Sementara itu, kategori rendah menunjukkan hasil prediksi yang paling stabil dengan tingkat ketepatan lebih tinggi dibanding kategori lainnya. Pola ini menegaskan bahwa semakin besar kemiripan antar data, semakin tinggi kemungkinan terjadinya salah klasifikasi, karena algoritma KNN menentukan kelas berdasarkan dominasi tetangga terdekat di sekitar data uji. Fenomena tersebut sejalan dengan karakteristik dasar KNN yang sensitif terhadap overlap distribusi data, terutama ketika batas antar kelas tidak terpisah secara jelas (Hadi et al., 2024; Harahap & Harahap, 2025).

Secara keseluruhan, model KNN menghasilkan akurasi sebesar 75%, yang menunjukkan bahwa metode ini cukup efektif dalam mengklasifikasikan laju pertumbuhan penduduk antarprovinsi. Namun demikian, masih terdapat keterbatasan dalam membedakan kelas yang memiliki pola serupa. Kondisi ini disebabkan oleh adanya overlap data, sehingga memengaruhi ketepatan prediksi (Hadi et al., 2024).

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Ibu Putri Maulidina Fadillah, M.Si., selaku dosen pengampu mata kuliah Algoritma dan Pemrograman yang telah memberikan bimbingan, arahan, serta masukan yang sangat berharga selama proses penyusunan penelitian ini. Ucapan terima kasih juga disampaikan kepada seluruh anggota kelompok atas kerja sama dan kontribusi dalam pengolahan data serta implementasi metode K-Nearest Neighbor (KNN) menggunakan RStudio.

PENUTUP

SIMPULAN

Berdasarkan hasil penelitian, metode K-Nearest Neighbor (KNN) dapat dimanfaatkan untuk mengelompokkan laju pertumbuhan penduduk antar provinsi di Indonesia ke dalam kategori rendah, sedang, dan tinggi dengan memanfaatkan karakteristik data kependudukan. Proses klasifikasi dilakukan melalui beberapa tahapan, meliputi praproses data, pembentukan label menggunakan metode kuantil, normalisasi data, serta pembagian dataset menjadi data latih dan data uji.

Hasil pengujian menunjukkan bahwa variasi nilai parameter k memengaruhi performa model klasifikasi. Pengujian yang dilakukan pada nilai k = 3, 5, dan 7 memperlihatkan bahwa k = 5 dan k = 7 menghasilkan tingkat akurasi tertinggi sebesar 75%, sedangkan k = 3 memberikan hasil yang lebih rendah, yaitu 50%. Meskipun nilai k = 5 dan k = 7

memiliki akurasi yang sama, k = 5 dipilih sebagai parameter optimal karena mampu menjaga keseimbangan antara kepekaan terhadap pola data dan kestabilan dalam proses klasifikasi.

Secara umum, metode KNN mampu memberikan hasil yang cukup baik dalam mengklasifikasikan data kependudukan. Namun, masih ditemukan kesalahan klasifikasi pada data dengan karakteristik yang mirip, terutama pada kategori sedang dan tinggi. Hal ini mengindikasikan bahwa kemiripan nilai antar data serta adanya overlap antar kelas menjadi faktor yang memengaruhi ketepatan hasil klasifikasi.

SARAN

Berdasarkan hasil penelitian yang telah diperoleh, beberapa hal dapat disarankan untuk penyempurnaan penelitian selanjutnya. Pertama, peningkatan kualitas analisis dapat dilakukan melalui pengolahan data yang lebih optimal, khususnya pada tahap normalisasi serta penanganan data outlier, sehingga tingkat akurasi klasifikasi dapat ditingkatkan. Selain itu, penentuan parameter k pada algoritma K-Nearest Neighbor sebaiknya dilakukan secara lebih sistematis, misalnya melalui teknik cross-validation, agar nilai yang diperoleh benar-benar optimal.

Dari sisi penyajian, penelitian ini masih dapat dikembangkan dengan menambahkan visualisasi data yang lebih variatif dan informatif sehingga pola distribusi data dapat ditampilkan dengan lebih jelas. Penulisan laporan juga perlu diperbaiki dengan memperjelas uraian pada setiap tahapan analisis serta menjaga konsistensi antara sitasi di dalam teks dengan daftar pustaka.

Selain itu, jumlah data dan variasi variabel yang digunakan sebaiknya diperluas agar hasil penelitian menjadi lebih representatif serta mampu memberikan gambaran yang lebih akurat terhadap kondisi yang diteliti.

DAFTAR PUSTAKA

Bakri, S. N., & Harahap, L. S. (2025). Analisis klasifikasi Algoritma K-Nearest Neighboar (K-NN) pada struktur Daerah di Kota Medan. *Jurnal Ilmu Komputer Dan Sistem Informasi*, *4*(2), 182–193. https://doi.org/10.70340/jirsi.v4i2.165

Feng, J., Wei, Y., & Zhu, Q. (2018). Natural neighborhood-based classification algorithm without parameter k. *Big Data Mining and Analytics*, *1*(4), 257–265. https://doi.org/10.26599/BDMA.2018.902007

Hadi, R., Luh, N., Suwirmayanti, G. P., Ngurah, G., Kusuma, A., Ayu, G., Saryanti, D., Novayanti, D., Stikom Bali, I., & Raya Puputan, J. (2024). Implementasi Metode Normalisasi dan Seleksi Fitur dalam Optimasi Algoritma K-Nearest Neighbor (KNN) untuk Klasifikasi Data Bank. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, *7*(5). https://doi.org/10.32672/jnkti.v7i5.7989

Hakim, M. I. N., Setyawan, I., Manongga, D., Purnomo, H. D., & Hendry. (2026). Selection of feature data in KNN classification datasets. *IPTEK The Journal for Technology and Science*, 37(1), 8–16. https://doi.org/10.12962/j20882033.v37i1.923

Harahap, M., & Harahap, S. (2025). Klasifikasi kualitas udara menggunakan algoritma K-Nearest Neighbor berdasarkan data PM2.5 di Indonesia. *JIKOMTI: Jurnal Ilmiah Ilmu Komputer dan Teknologi Informasi*, 2(2), 8–12. https://ojs.sains.ac.id/index.php/Jikomti/article/view/137

Hendriyanto, M. D., & Sari, B. N. (2022). Penerapan Algoritma K-Nearest Neighbor Dalam Klasifikasi Judul Berita Hoax. *Jurnal Ilmiah Informatika*, *10*(02), 80–84. https://doi.org/10.33884/jif.v10i02.5477

Indah Sari, S., Adi Suryanto, A., & Haryoko, A. (2024). Klasifikasi Penentuan Penenrima Bantuan Pangan Non Tunai Dengan Menggunakan Metode Knn (K-Nearest Neighbor). *Curtina*, *5*(1), 10–21.

Insan, P., & Kusrini. (2021). Analisis Perbandingan Algoritma ID3 dan KNN Pada Klasifikasi Emosi Teks Berita Berbahasa Indonesia. *METIK JURNAL*, *5*(1), 36–41. https://doi.org/10.47002/metik.v5i1.213

Juanuari, J., Manzis, I., Musyaffa, A., Ngara, S. M., Syahla, H., & Putra, J. L. (2026). Implementasi K-Nearest Neighbor untuk Klasifikasi Tingkat Kemiskinan Sebagai Pendukung Pengambilan Keputusan Sosial di Indonesia. *Infomatek*, *28*(1), 1–10. https://doi.org/10.23969/infomatek.v28i1.266

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data and Society*, *1*(1). https://doi.org/10.1177/2053951714528481

Kotjek, R. A. K., Wijaya, F. S., Wahyudi, M., & Budiman, A. S. (2025). Klasifikasi nilai ujian siswa berdasarkan kebiasaan belajar menggunakan K-Nearest Neighbor dan Support Vector Machine. *Jurnal Sains Informatika Terapan*, 4(2).

Lee, R. (2011). The outlook for population growth. In *Science* (Vol. 333, Number 6042, pp. 569–573). https://doi.org/10.1126/science.1208859

Nur, H. M., Maarif, V., Imaniawan, F. D., Saputro, E., & BSI. (2025). Algoritma K-Nearest Neighbor (K-NN) untuk prediksi ketidaklulusan mahasiswa berdasarkan presensi dan IPK. *Journal Computer Science*, 4(1), 8-12. https://doi.org/10.31294/3smve945

Ujianto, N. T., Gunawan, Fadillah, H., Fanti, A. P., Saputra, A. D., & Ramadhan, I. G. (2025). Penerapan algoritma K-Nearest Neighbors (KNN) untuk klasifikasi citra medis. *IT-Explore: Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, 4(1), 33–43. https://doi.org/10.24246/itexplore.v4i1.2025.pp33-43

Zhou, R. S., & Wang, Z. J. (2015). *A Review of a Text Classification Technique: K-Nearest Neighbor*.