JTIM : Jurnal Teknologi Informasi dan Multimedia

p-ISSN : 2715-2529
e-ISSN : 2684-9151
https://journal.sekawan-org.id/index.php/jtim

Penerapan Metode K-Nearest Neighbor Untuk Prediksi Jumlah Kasus HIV di Provinsi Jawa Barat

Muhammad Adam Rizky Habibi [1], Shofa Shofia Hilabi [1], Bayu Priyatna [1] dan Elfina Novalia [1]

[1] Program Studi Sistem Informasi, Universitas Buana Perjuangan Karawang, Indonesia.

* Korespondensi: si21.muhammadhabib@mhs.ubpkarawang.ac.id

Sitasi: Habibi, M. A. R.; Hilabi, S. S.; Priyatna, B.; dan Novalia, E. (**2025**). Penerapan Metode K-Nearest Neighbor Untuk Prediksi Jumlah Kasus HIV di Provinsi Jawa Barat. JTIM: Jurnal Teknologi Informasi Dan Multimedia, 7(2), 372-386. https://doi.org/10.35746/jtim.v7i2.721

Diterima: 14-03-2025
Direvisi: 24-04-2025
Disetujui: 29-04-2025

Copyright: © 2025 oleh para penulis. Karya ini dilisensikan di bawah Creative Commons Attribution-ShareAlike 4.0 International License. (https://creativecommons.org/licenses/by-sa/4.0/).

Abstract: The high number of HIV/AIDS cases in Indonesia, especially in West Java Province, is a serious challenge in the field of public health. Limitations in understanding the pattern of spread and predicting the trend of HIV cases cause countermeasures to be less than optimal. To overcome this, this study was conducted with the aim of predicting the number of HIV cases in West Java using the K-Nearest Neighbor (KNN) algorithm, based on historical data from Open Data Jabar from 2019 to 2023 which includes 1,617 data from various districts / cities. The research stages include data collection, preprocessing, feature selection, normalization, division of training and test data, and model evaluation using regression metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R^2). The evaluation results show that the KNN model with an optimal K value of 19 produces an MAE of 142.31, MSE of 40,442.92, RMSE of 201.10, and R^2 value of 0.2397. Predictions for 2024 show that areas with the highest number of HIV cases are in Bandung City, Bogor Regency, Bekasi City, Bekasi Regency, and Indramayu Regency.

Keywords: *HIV Prediction, K-Nearest Neighbor, Data Mining, West Java*

Abstrak. Tingginya jumlah kasus HIV/AIDS di Indonesia, khususnya di Provinsi Jawa Barat, menjadi tantangan serius dalam bidang kesehatan masyarakat. Keterbatasan dalam memahami pola penyebaran dan memprediksi tren kasus HIV menyebabkan upaya penanggulangan menjadi kurang optimal. Untuk mengatasi hal tersebut, penelitian ini dilakukan dengan tujuan mem-prediksi jumlah kasus HIV di Jawa Barat menggunakan algoritma K-Nearest Neighbor (KNN), berdasarkan data historis dari Open Data Jabar tahun 2019 hingga 2023 yang mencakup 1.617 data dari berbagai kabupaten/kota. Tahapan penelitian mencakup pengumpulan data, prepro-cessing, pemilihan fitur, normalisasi, pembagian data latih dan uji, serta evaluasi model menggunakan metrik regresi: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan R-squared (R^2). Hasil evaluasi menunjukkan bahwa model KNN dengan nilai K optimal sebesar 19 menghasilkan MAE sebesar 142,31, MSE sebesar 40.442,92, RMSE sebesar 201,10, dan nilai R^2 sebesar 0,2397. Prediksi tahun 2024 menunjukkan bahwa daerah dengan jumlah kasus HIV tertinggi berada di Kota Bandung, Kabupaten Bogor, Kota Bekasi, Kabupaten Bekasi, dan Kabupaten Indramayu.

Kata kunci: Prediksi HIV, K-Nearest Neighbor, Data Mining, Jawa Barat

1. Pendahuluan

HIV/AIDS merupakan salah satu tantangan kesehatan masyarakat yang masih menjadi perhatian serius di berbagai negara, termasuk Indonesia[1]. Penyakit ini tidak hanya berdampak pada kesehatan individu tetapi juga memberikan beban sosial dan ekonomi yang signifikan[2]. Angka kasus HIV di Indonesia masih sangat signifikan. Menurut Kementerian Kesehatan Republik Indonesia, jumlah pengidap HIV di Indonesia diperkirakan mencapai 543.100 orang pada tahun 2020[3]. Data juga menunjukkan bahwa 29.557 di antaranya adalah infeksi baru, menunjukkan bahwa HIV masih menyebar di masyarakat. Selain itu, jumlah kematian akibat HIV yang mencapai 30.137 orang pada tahun 2020 menimbulkan kekhawatiran mengenai kualitas hidup serta kesejahteraan individu yang terdampak penyakit ini[4]. Sementara itu, Jumlah kasus HIV di Jawa Barat mengalami peningkatan yang cukup signifikan. Berdasarkan data dari Dinas Kesehatan Jawa Barat tahun 2022, terjadi kenaikan kasus HIV positif dibandingkan dengan tahun sebelumnya, yang tercatat sebanyak 5.444 kasus. Peningkatan jumlah kasus HIV di Provinsi Jawa Barat menandakan perlunya pendekatan berbasis data untuk memahami pola penyebarannya dan mengidentifikasi wilayah dengan tingkat infeksi tertinggi[3].

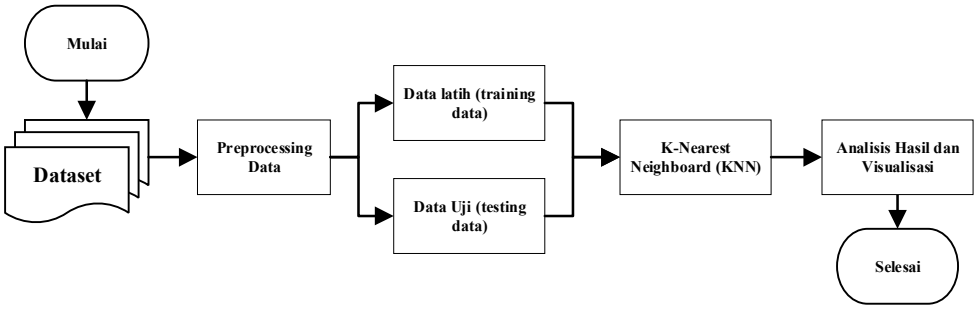
Pengambilan data dalam penelitian ini bersumber dari Open Data Jabar, dengan total 1.617 data yang mencakup jumlah kasus HIV sebanyak 31.986 kasus yang tersebar di berbagai kota dan kabupaten di Jawa Barat. Dalam analisis epidemiologi, *data mining* merupakan metode yang efektif untuk mengungkap pola penyebaran penyakit secara akurat, termasuk dalam kasus HIV/AIDS[5]. Metode prediksi berbasis data memungkinkan strategi pencegahan dan penanganan yang lebih tepat sasaran dan efisien[6]. Penelitian ini menggunakan algoritma *K-Nearest Neighbor* (KNN) untuk memprediksi jumlah kasus HIV di berbagai wilayah di Jawa Barat. Pemilihan metode KNN didasarkan pada kemampuannya dalam memprediksi nilai berdasarkan pola kedekatan data historis, sehingga cocok untuk kasus regresi dengan data yang bersifat spasial dan temporal. Beberapa penelitian sebelumnya menunjukkan efektivitas KNN dalam konteks serupa. Misalnya, Yunus & Pratiwi (2023) menerapkan KNN untuk prediksi status gizi balita dan mendapatkan hasil prediksi yang cukup akurat[7]. Demikian pula, Syafana et al. (2024) berhasil menggunakan KNN untuk memprediksi angka kelahiran berdasarkan kelompok umur ibu. Keberhasilan ini menunjukkan bahwa KNN mampu mengidentifikasi tren dari data historis dan menghasilkan prediksi yang bermanfaat dalam konteks kesehatan masyarakat.[8]

Penelitian ini bertujuan untuk prediksi jumlah kasus HIV pada 27 kota dan kabupaten di Jawa Barat berdasarkan data historis menggunakan metode KNN[9]. Dengan hasil analisis yang diperoleh, diharapkan pemerintah dan lembaga kesehatan dapat memiliki acuan yang lebih akurat dalam menyusun kebijakan pencegahan, meningkatkan intervensi medis, serta merancang program edukasi masyarakat guna menekan angka penyebaran HIV di wilayah Jawa Barat.

2. Bahan dan Metode

2.1. Perancangan Penelitian

Penelitian ini bertujuan untuk prediksi jumlah kasus HIV pada 27 kota dan kabupaten di Jawa Barat berdasarkan data historis menggunakan algoritma *K-Nearest Neighbor* (KNN). Diagram alur penelitian ditunjukkan pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

Proses penelitian ini terdiri dari beberapa tahap utama, dimulai dengan pengumpulan dataset, yaitu data kasus HIV yang mencakup informasi kabupaten/kota, tahun, dan jumlah kasus. Selanjutnya dilakukan preprocessing data yang meliputi pembersihan nilai kosong, pengkodean fitur kategorikal seperti nama kabupaten/kota, serta normalisasi agar seluruh fitur memiliki skala yang seragam [9]. Setelah data siap, model K-Nearest Neighbor (KNN) diterapkan untuk melakukan pelatihan dan prediksi jumlah kasus HIV berdasarkan pola data historis. Evaluasi model dilakukan menggunakan *Mean Absolute Error* (MAE), yaitu metrik yang mengukur rata-rata selisih absolut antara nilai aktual dan hasil prediksi. MAE dipilih karena memberikan gambaran langsung tentang tingkat kesalahan dalam satuan jumlah kasus HIV, sehingga hasilnya lebih mudah dipahami dan relevan untuk konteks epidemiologi.

2.2. Dataset

Sumber data pada penelitian ini berasal dari situs resmi Open Data Jabar, yang dapat diakses melalui tautan https://opendata.jabarprov.go.id/id/dataset/jumlah-kasus-hiv-berdasarkan-kabupatenkota-di-jawa-barat[10], yang menyediakan data jumlah kasus HIV berdasarkan kabupaten/kota di Provinsi Jawa Barat. Dataset terdiri dari total 1.617 entri data yang dikumpulkan selama periode tahun 2019 hingga 2023. Pemilihan rentang waktu 2019–2023 dilakukan karena pada periode tersebut terjadi peningkatan perhatian terhadap isu kesehatan masyarakat pasca pandemi COVID-19. Selain itu, data pada periode ini dinilai paling mutakhir dan representatif untuk memetakan tren penyebaran kasus HIV terkini di Jawa Barat. Dataset ini memiliki atribut-atribut utama sebagai berikut: nama_kabupaten_kota: lokasi kejadian kasus HIV, kelompok_umur: klasifikasi usia penderita, jenis_kelamin: kategori LAKI-LAKI atau PEREMPUAN, jumlah_kasus: total kasus yang tercatat (dalam satuan ORANG), tahun: periode pencatatan.

Tabel 1. Atribut Dataset Kasus HIV

nama_kabupaten_kota	kelompok_umur	jenis_kelamin	jumlah_kasus	tahun
Kabupaten Bogor	15-19	LAKI-LAKI	8	2019
Kabupaten Bogor	25-49	PEREMPUAN	133	2019
Kabupaten Sukabumi	20-24	LAKI-LAKI	20	2019
Kabupaten Purwakarta	25-49	LAKI-LAKI	61	2021
Kabupaten Kuningan	25-49	LAKI-LAKI	86	2022
Kota Bogor	25-49	LAKI-LAKI	230	2022
Kota Bandung	15-24	LAKI-LAKI	150	2023
Total Data: 1617				

Dataset ini disimpan dalam format Comma Separated Values (.csv), yang memudahkan dalam proses pemrosesan dan analisis data. Atribut-atribut dalam dataset memiliki peran penting dalam melatih model pembelajaran mesin, terutama dalam memahami tren jumlah kasus HIV di berbagai kabupaten/kota di Jawa Barat.

Dari keseluruhan data, tercatat sebanyak 1.210 kasus penderita berjenis kelamin laki-laki dan 407 kasus penderita berjenis kelamin perempuan. Dalam proses pengolahan, dataset ini dibagi menjadi dua bagian, yakni data latih dan data uji, dengan rasio 80% untuk data latih dan 20% untuk data uji. Dari total 1.617 data, sebanyak 1.294 data digunakan sebagai data latih, sedangkan 323 data sisanya digunakan sebagai data uji. Pembagian data ini dilakukan secara acak untuk memastikan distribusi data yang merata, sehingga dapat meningkatkan kinerja model dalam proses pelatihan dan pengujian.

2.3. Prepocessing

Gambar 2. Flowchart Tahapan Preprocessing

Tahap preprocessing data dilakukan untuk menjamin bahwa data yang diterapkan dalam model memiliki mutu optimal dan siap diolah dengan efisien[12]. Berikut adalah langkah-langkah preprocessing yang dilakukan:

1. Menghapus Data Kosong (Missing Values)

   Data yang memiliki nilai kosong terlebih dahulu diidentifikasi. Karena persentase data kosong sangat kecil, maka metode yang digunakan adalah penghapusan baris (*row deletion*) yang mengandung nilai kosong. Pendekatan ini dipilih agar tidak menimbulkan bias pada data dan tidak mempengaruhi distribusi nilai yang telah ada[13].

2. Pemilihan Fitur

   Untuk membangun model prediksi yang lebih akurat dan efisien, dilakukan proses pemilihan fitur dari keseluruhan atribut dalam dataset. Fitur-fitur yang digunakan

meliputi nama_kabupaten_kota, tahun, dan jumlah_kasus sebagai target prediksi. Selain itu, fitur tambahan yaitu total_penduduk digunakan untuk merepresentasikan densitas populasi di tiap wilayah. Pemilihan fitur ini didasarkan pada analisis korelasi terhadap target prediksi (jumlah_kasus) dan pertimbangan logis atas relevansi masing-masing atribut dalam konteks epidemiologi. Fitur-fitur lain yang memiliki korelasi rendah atau berpotensi menyebabkan noise, seperti kelompok umur dan jenis kelamin, tidak dimasukkan karena model KNN akan menghitung jarak antar data — sehingga keberadaan fitur yang tidak relevan dapat mengurangi performa model[14].

3. Encoding Variabel Kategorikal

   Kolom nama_kabupaten_kota bersifat kategorikal dan tidak bisa langsung digunakan dalam perhitungan numerik. Oleh karena itu, encoding dilakukan dengan mengubah nama kabupaten/kota menjadi angka menggunakan metode Label Encoding[15].

4. Normalisasi Data

   Untuk memastikan setiap fitur memiliki skala yang seragam, dilakukan proses normalisasi menggunakan Min-Max Scaler dengan rentang nilai antara 0 hingga 1. Normalisasi ini diterapkan pada kolom tahun dan jumlah kasus guna meningkatkan akurasi model KNN.

5. Pembagian Data Latih dan Uji

   Dataset dipisahkan menjadi dua bagian, yakni data latih (training set) dan data uji (test set) dengan rasio 80:20. Dari total 1.617 data, sebanyak 1.293 data digunakan sebagai data latih, sementara 324 data digunakan sebagai data uji. Data latih digunakan dalam proses pembentukan model, sedangkan data uji dimanfaatkan untuk mengevaluasi kinerja model sebelum diterapkan dalam prediksi jumlah kasus HIV pada tahun 2024.

2.4. Algoritma KNN

Algoritma *K-Nearest Neighbor* (KNN) merupakan algoritma non-parametrik yang digunakan dalam regresi untuk prediksi nilai berdasarkan kedekatan dengan data historis. KNN bekerja dengan mencari K tetangga terdekat berdasarkan metrik jarak, dalam hal ini menggunakan Euclidean Distance[16].

Rumus Jarak Euclidean:

$$\mathrm{d}(p,q) = \sqrt{\sum_{i=1}^{n}(p_i - q_1)^2} \quad (1)$$

Keterangan:
$d(p,q)$: jarak antara dua titik p dan q.
p_1: nilai atribut ke-i dari titik p.
q_1: nilai atribut ke-i dari titik q.
n: jumlah dimensi atau atribut data.

2.4.1. Arsitektur Model KNN

Arsitektur model KNN dalam penelitian ini mencakup parameter dan komponen berikut:

a. Parameter Input:

   1. Jumlah Tetangga (K): 19 (nilai optimal hasil metode elbow).
   2. Metrik Jarak: Euclidean Distance (Persamaan 1).

3. Bobot Tetangga: Uniform (semua tetangga memiliki kontribusi sama).
4. Fitur Input:
   1) nama_kabupaten_kota (dienkode dengan Label Encoding).
   2) tahun (dinormalisasi ke rentang [0,1]).
   3) total_penduduk (dinormalisasi ke rentang [0,1]).
5. Target: jumlah_kasus (dinormalisasi ke rentang [0,1]).

b. Proses Pembentukan Model:

1. Normalisasi: Menggunakan Min-Max Scaler untuk menghindari dominasi skala fitur.
2. Pembagian Data: Data dibagi dengan rasio 80:20 (training:testing) menggunakan stratified sampling berdasarkan tahun.

2.4.2. Skenario Uji Coba dan Optimasi Parameter

a. Penentuan Nilai K Optimal:

1. Uji coba nilai K = 1 hingga 30 dengan interval 2.
2. Metode elbow digunakan dengan memplot MSE terhadap nilai K (Gambar 3).
3. Hasil: Nilai K optimal ditetapkan pada K=19 (titik penurunan MSE mulai melandai).

b. Kombinasi Parameter yang Diuji:

Parameter	**Opsi yang Diuji**	**Hasil Terbaik**
Metrik Jarak	Euclidean, Manhattan	Euclidean (MSE = 40.442,92)
Bobot Tetangga	Uniform, Distance-based	Uniform (tidak ada peningkatan signifikan)
Normalisasi	Min-Max, Z-Score	Min-Max (stabilitas lebih baik)

c. Validasi Silang:

1. Dilakukan 5-fold cross-validation untuk memastikan konsistensi model.
2. Rata-rata MSE validasi: 41.200 ± 1.150.

2.4.3. Langkah-langkah Algoritm KNN:

1. Menentukan jumlah tetangga K (dalam kode ini, digunakan K = 5).Menghitung jarak Euclidean antara sampel data uji dengan seluruh sampel data latih[11].
2. Memilih K tetangga terdekat berdasarkan jarak terkecil[17].
3. Menghitung rata-rata nilai target dari K tetangga terdekat sebagai hasil prediksi[18].
4. Mengevaluasi model menggunakan metrik regresi seperti *Mean Squared Error* (MSE), *R-squared* (R^2), *Mean Absolute Error* (MAE), dan *Root Mean Squared Error* (RMSE)[19].
   a. *Mean Absolute Error* (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (2)$$

   Keterangan:
   y_i: nilai sebenarnya
   $\hat{y}_i$: nilai prediksi
   n: jumlah data

   b. *Mean Squared Error* (MSE):

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i) \quad (3)$$

   Keterangan:
   y_i: nilai sebenarnya

$\hat{y}_i$: nilai prediksi
n: jumlah data

c. *Root Mean Squared Error* (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)} \quad (4)$$

y_i: nilai sebenarnya
$\hat{y}_i$: nilai prediksi
$\bar{y}_i$: rata – rata nilai aktual
n: jumlah data

d. *R-squared* (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (5)$$

y_i: nilai sebenarnya
$\hat{y}_i$: nilai prediksi
$\bar{y}_i$: rata – rata nilai aktual
n: jumlah data

Dengan menggunakan model KNN, prediksi jumlah kasus HIV pada tahun 2024 dapat dilakukan dengan memasukkan data tahun 2024 dan melihat pola berdasarkan data historis. Model ini dapat membantu dalam memetakan wilayah dengan tingkat jumlah kasus tertinggi, yang dapat menjadi referensi dalam strategi pencegahan dan pengendalian HIV.

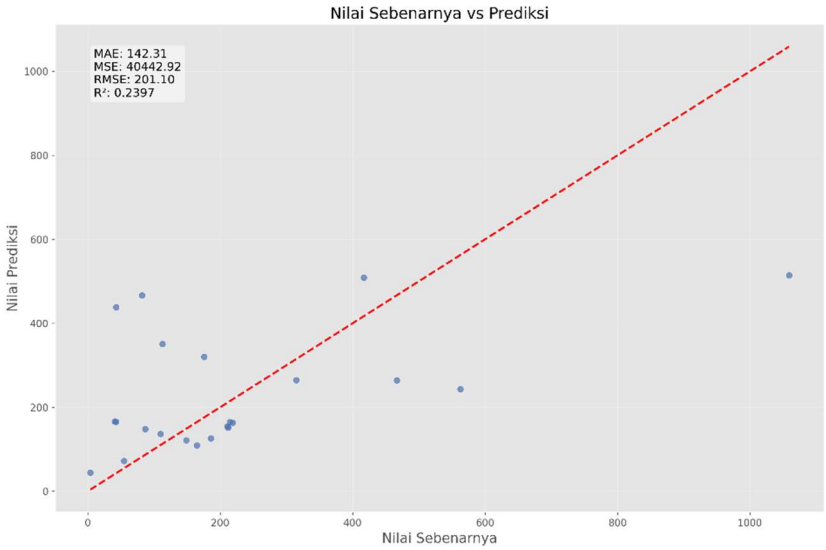
3. Hasil

Evaluasi model *K-Nearest Neighbor* (KNN) dilakukan menggunakan Python melalui Google Colab[9]. Hasil evaluasi ditunjukkan pada Gambar 2, dengan metrik utama yang digunakan meliputi kesalahan absolut rata-rata (MAE), kesalahan kuadrat rata-rata (MSE), akar dari kesalahan kuadrat rata-rata (RMSE), serta koefisien determinasi (R^2).

3.1. Evaluasi Model

Pada bagian ini, model KNN yang digunakan untuk memprediksi jumlah kasus HIV di Jawa Barat akan dievaluasi. Evaluasi dilakukan berdasarkan metrik berikut:

1. Mean Absolute Error (MAE): Hal ini mengindikasikan rata-rata perbedaan absolut antara nilai sebenarnya dan nilai yang diprediksi.
2. Mean Squared Error (MSE): Pengukuran rata-rata kesalahan kuadrat yang memberikan pengaruh yang lebih besar pada kesalahan besar.
3. Root Mean Squared Error (RMSE): Menunjukkan seberapa jauh nilai prediksi menyimpang dari nilai aktual.
4. R-squared (R^2): Menunjukkan sejauh mana model dapat menjelaskan variasi dalam data.



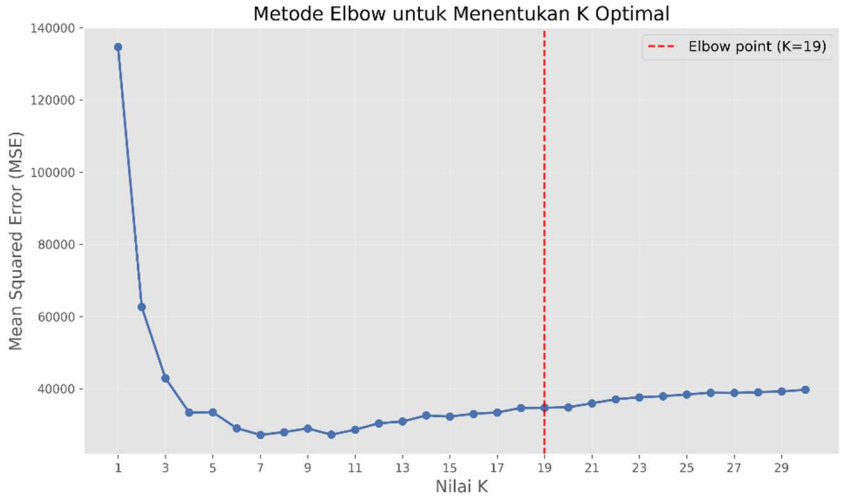
Gambar 3. Nilai Sebenarnya dan Prediksi

Berdasarkan gambar di atas menunjukan hasil bahwa:

1. Nilai MAE = 142.31, menunjukkan bahwa rata-rata selisih absolut antara prediksi dan nilai aktual cukup besar.
2. Nilai MSE = 40442.92 dan RMSE = 201.10, menandakan ada variasi yang cukup tinggi dalam kesalahan prediksi.
3. Nilai R^2 = 0.2397, menunjukkan bahwa model hanya dapat menjelaskan sekitar 23.97% variasi dalam data, yang mengindikasikan bahwa model belum optimal dalam memprediksi jumlah kasus HIV.

3.2. Penentuan Nilai K Optimal

Untuk meningkatkan akurasi model KNN, pemilihan nilai K yang optomal sangat penting. Metode elbow digunakan untuk menentukan nilai K terbaik berdasarkan nilai Mean Squared Error (MSE).

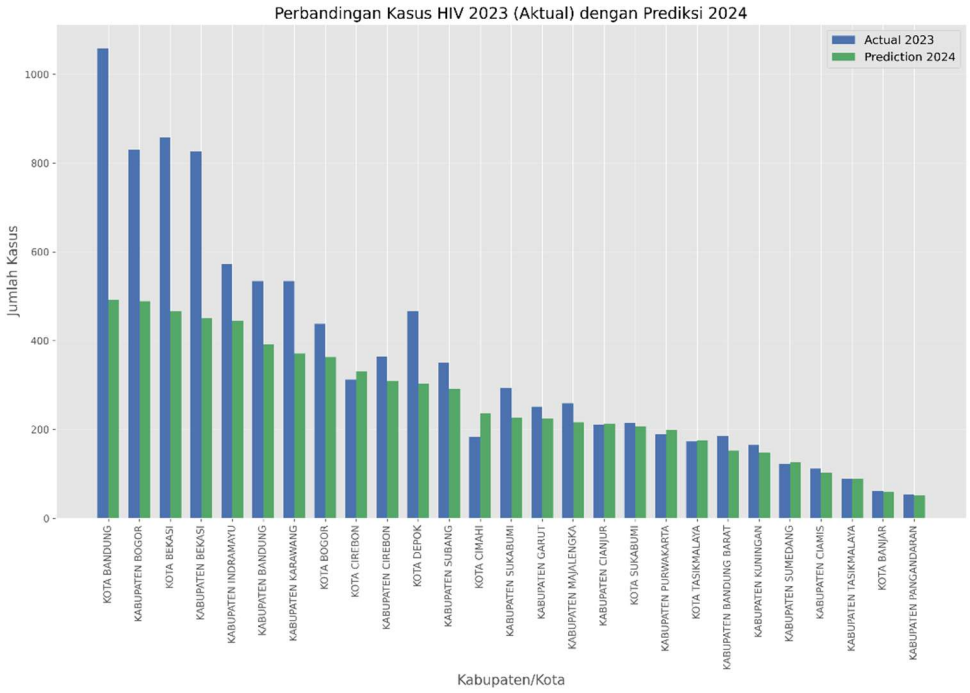


Gambar 4. Hasil Elbow

Dari hasil metode elbow, nilai K optimal ditemukan pada K = 19, karena setelah titik ini, penurunan nilai MSE tidak signifikan.

3.3. Perbandingan Prediksi Kasus HIV Tahun 2024 dengan Data 2023

Prediksi jumlah kasus HIV tahun 2024 dibandingkan dengan data aktual tahun 2023 untuk melihat tren perubahan.



Gambar 5. Hasil Perbandingan Kasus 2023 dan 2024

Dari hasil perbandingan:

1. Tren Prediksi

   Hasil prediksi menunjukkan bahwa sebagian besar kabupaten/kota mengalami penurunan jumlah kasus HIV pada tahun 2024 dibandingkan dengan data aktual tahun 2023. Penurunan ini terlihat konsisten di banyak wilayah, namun terdapat beberapa daerah yang memiliki tren berbeda.

2. Wilayah dengan Kasus Tertinggi

   Kota Bandung, Kabupaten Bogor, Kota Bekasi, Kabupaten Bekasi, dan Kabupaten Indramayu merupakan daerah dengan jumlah kasus HIV tertinggi pada tahun 2023. Prediksi model untuk 2024 menunjukkan bahwa daerah-daerah ini masih mendominasi sebagai wilayah dengan kasus tertinggi, walaupun dengan angka yang cenderung lebih rendah.

3. Evaluasi Model dan Akurasi

   Setelah dilakukan penyesuaian nilai K optimal menjadi K = 19 berdasarkan metode elbow, evaluasi performa model dilakukan menggunakan Mean Absolute Error (MAE). Hasil evaluasi menunjukkan nilai MAE sebesar 142,31, yang berarti rata-rata kesalahan absolut dalam prediksi jumlah kasus HIV adalah sekitar 142 kasus per kabupaten/kota. Nilai ini memberikan gambaran bahwa model memiliki akurasi yang cukup moderat, tetapi masih belum optimal.

4. Interpretasi Kesalahan

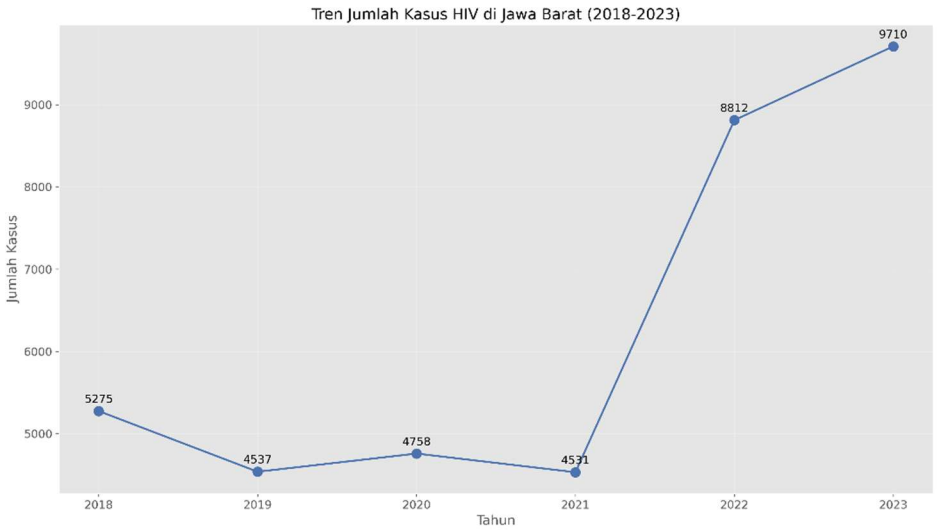
   1) Model KNN cenderung menghaluskan fluktuasi atau ekstrem dalam data historis, sehingga prediksi jumlah kasus di daerah dengan kasus sangat tinggi (seperti Kota Bandung) cenderung lebih rendah dari aktual.

   2) Di sisi lain, untuk kabupaten/kota dengan jumlah kasus menengah hingga rendah, prediksi model terlihat lebih mendekati nilai aktual.

   3) Hal ini menunjukkan bahwa model bekerja lebih baik di rentang tengah data, namun masih kurang akurat dalam menangani outlier atau data ekstrem.

3.4. Visualisasi Tren Kasus HIV di Jawa Barat

Untuk memahami pola penyebaran HIV dari tahun ke tahun, berikut adalah tren jumlah kasus HIV di Jawa Barat dari tahun 2018 hingga 2023.



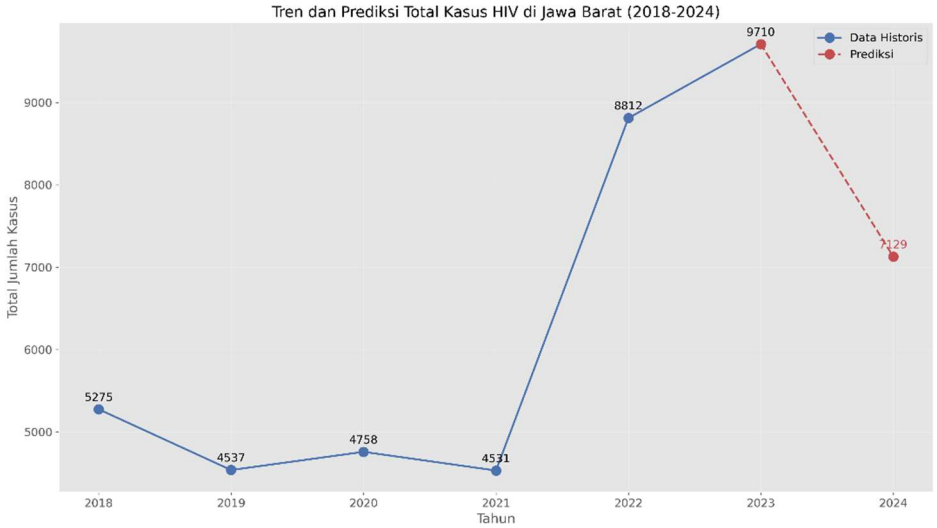
Gambar 6. Hasil Tren Jumlah Kasus

Dari hasil visualisasi:

1. Kasus HIV mengalami peningkatan signifikan pada tahun 2022 dan 2023.
2. Terdapat lonjakan jumlah kasus pada tahun 2022, yang bisa disebabkan oleh peningkatan pelaporan atau peningkatan jumlah infeksi baru.

3.5. Prediksi Total Kasus HIV Tahun 2024

Selain perbandingan per kabupaten/kota, berikut adalah tren total kasus HIV di Jawa Barat dari 2018 hingga 2024 berdasarkan data historis dan prediksi.



Gambar 7. Hasil Tren dan Prediksi Total Kasus

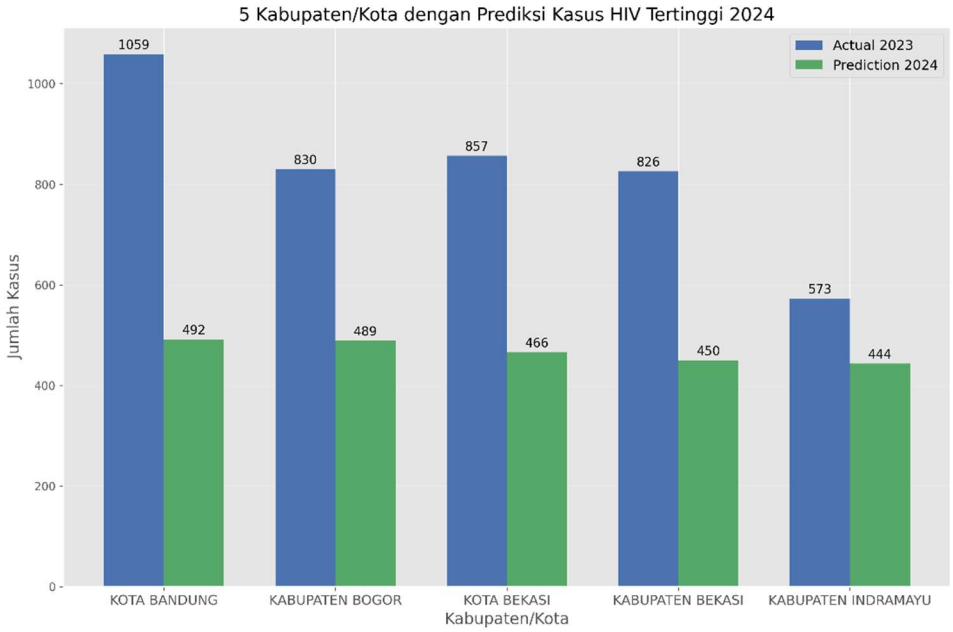
Dari hasil prediksi:

1. Diperkirakan jumlah kasus HIV di Jawa Barat tahun 2024 akan mengalami penurunan dibandingkan 2023.
2. Untuk pengembangan penelitian selanjutnya, penting untuk mengevaluasi hubungan antara faktor eksternal dan jumlah kasus HIV guna memperkaya konteks

prediksi. Faktor-faktor seperti tingkat pendidikan, kesadaran masyarakat terhadap HIV/AIDS, akses terhadap layanan kesehatan, dan kebijakan pencegahan yang diterapkan di tiap daerah memiliki potensi kuat dalam memengaruhi angka kasus HIV. Oleh karena itu, penelitian mendatang dapat mengintegrasikan data eksternal tersebut sebagai variabel tambahan dan melakukan analisis korelasi atau regresi multivariat guna mengetahui sejauh mana faktor-faktor tersebut berkaitan dengan jumlah penderita HIV. Selain itu, penerapan model machine learning yang lebih kompleks seperti Random Forest atau Gradient Boosting juga dapat dipertimbangkan untuk meningkatkan ketepatan prediksi dan kemampuan model dalam menangkap interaksi variabel-variabel kompleks.

3.6. Analisis Kabupaten/Kota dengan Kasus HIV Tertinggi

Untuk mengetahui daerah yang paling terdampak, berikut adalah 5 kabupaten/kota dengan prediksi kasus HIV tertinggi pada tahun 2024.



Gambar 8. Hasil Prediksi 5 Kabupaten HIV Tertinggi 2024

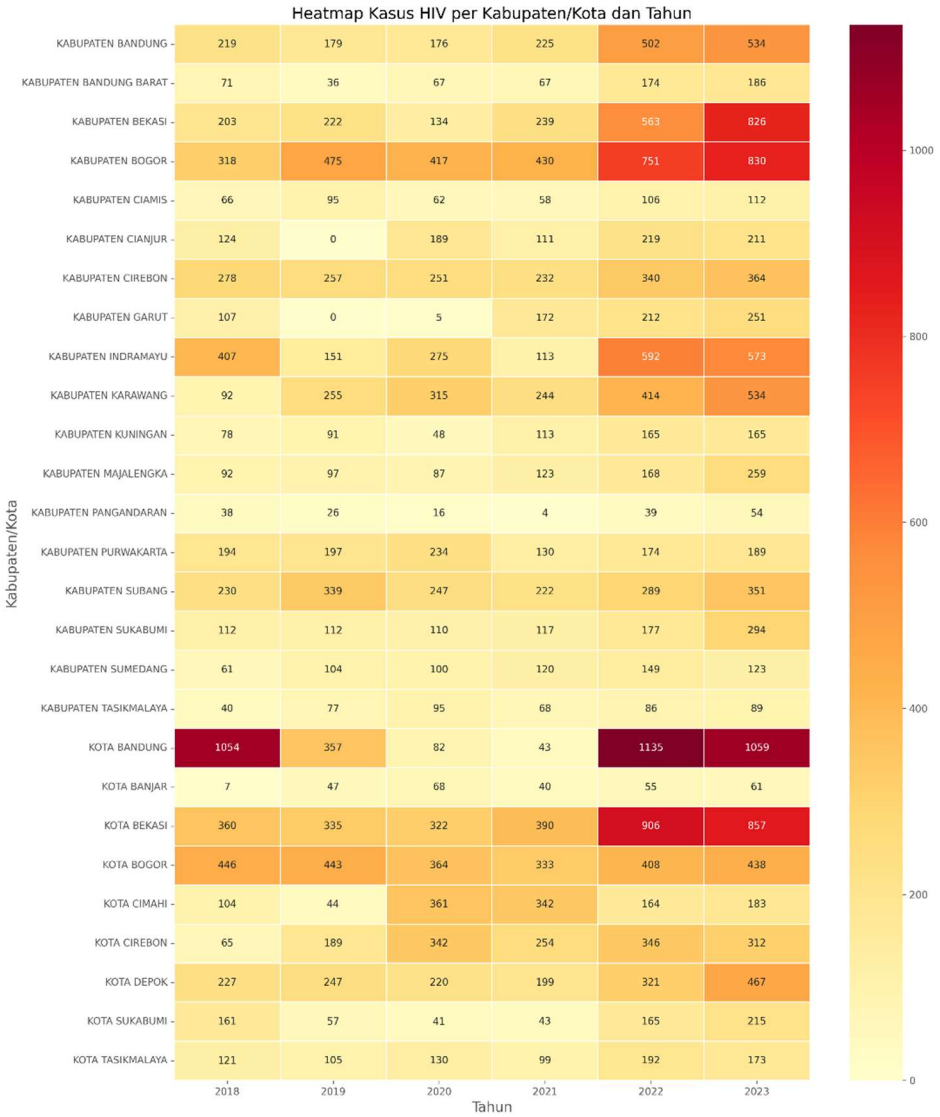
Daerah dengan kasus tertinggi:

1. Kota Bandung: 492 kasus
2. Kabupaten Bogor: 489 kasus
3. Kota Bekasi: 466 kasus
4. Kabupaten Bekasi: 450 kasus
5. Kabupaten Indramayu: 444 kasus

Hasil ini menunjukkan bahwa upaya pencegahan harus difokuskan pada daerah-daerah ini.

3.7. Heatmap Penyebaran Kasus HIV per Tahun

Untuk melihat distribusi kasus HIV di berbagai daerah, berikut adalah heatmap jumlah kasus HIV berdasarkan tahun dan kabupaten/kota.



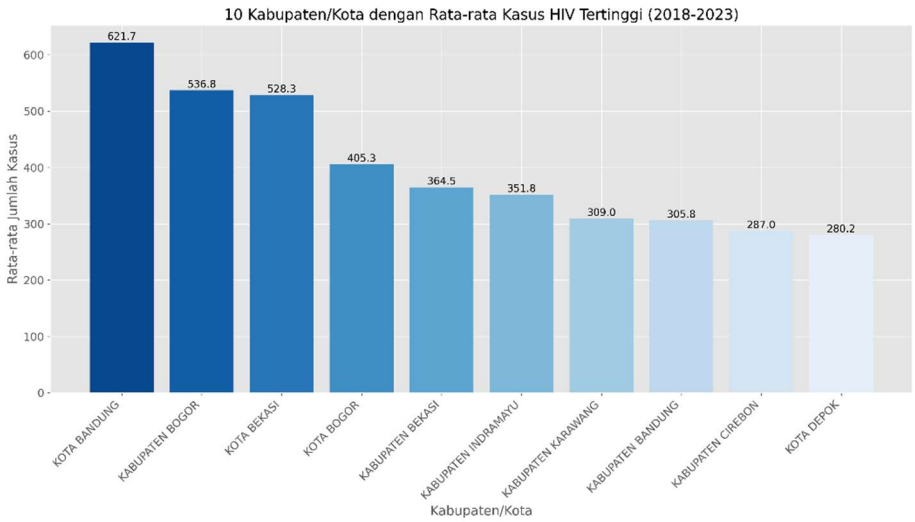
Gambar 9. Hasil Heatmap

Dari hasil heatmap:

1. Kota Bandung, Kota Bekasi, dan Kabupaten Bogor memiliki jumlah kasus tertinggi di sebagian besar tahun.
2. Ada fluktuasi jumlah kasus di beberapa daerah yang mungkin disebabkan oleh perubahan dalam pelaporan atau faktor epidemiologi lainnya.

3.8. Rata-rata Kasus HIV per Kabupaten/Kota

Untuk mengetahui daerah dengan kasus HIV tertinggi secara rata-rata dari 2018 hingga 2023, berikut adalah 10 kabupaten/kota dengan rata-rata kasus tertinggi.



Gambar 10. Hasil 10 Kota dengan Kasus HIV Tertinggi

Dari hasil analisis:

1. Kota Bandung memiliki rata-rata kasus tertinggi (621,7 kasus per tahun).
2. Kabupaten Bogor dan Kota Bekasi juga memiliki angka rata-rata kasus yang cukup tinggi.

4. Pembahasan

4.1. Evaluasi Model KNN

Model KNN yang digunakan menunjukkan performa yang belum optimal, dengan nilai R^2 yang rendah. Beberapa faktor yang mempengaruhi hasil prediksi:

1. Distribusi Data: Penyebaran data tidak merata, dengan banyak daerah memiliki jumlah kasus yang jauh lebih tinggi dari yang lain.
2. Pemilihan Fitur: Model hanya menggunakan data jumlah kasus, tahun, dan lokasi tanpa mempertimbangkan faktor lain seperti tingkat pendidikan, akses layanan kesehatan, atau faktor sosial-ekonomi.
3. Pemilihan Parameter K: Nilai K = 19 ditemukan sebagai optimal, tetapi masih ada kemungkinan peningkatan akurasi dengan parameter tuning lebih lanjut.

4.2. Faktor yang Mempengaruhi Tren Kasus HIV

Beberapa faktor yang mungkin berkontribusi terhadap pola kasus HIV:

1. Akses terhadap layanan kesehatan: Daerah dengan fasilitas kesehatan lebih baik mungkin melaporkan lebih banyak kasus karena peningkatan deteksi.
2. Kesadaran masyarakat: Wilayah dengan program edukasi HIV/AIDS yang kuat mungkin mengalami penurunan kasus baru.
3. Faktor sosial-ekonomi: Kemiskinan, pekerjaan, dan mobilitas penduduk dapat berperan dalam penyebaran HIV.

5. Kesimpulan

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma K-Nearest Neighbor (KNN) dapat digunakan untuk memprediksi jumlah kasus HIV di Provinsi Jawa Barat, meskipun akurasi model masih tergolong rendah dengan nilai R^2 sebesar 0,2397. Hal ini menunjukkan bahwa model hanya mampu menjelaskan sebagian kecil dari variasi data, sehingga perlu adanya peningkatan performa. Arsitektur KNN yang digunakan dalam penelitian ini merupakan versi regresi non-parametrik dengan nilai K optimal sebesar 19, ditentukan melalui metode elbow, yang memberikan keseimbangan antara bias dan

varians. Hasil prediksi menunjukkan bahwa wilayah dengan beban kasus tertinggi tetap terkonsentrasi di Kota Bandung, Kabupaten Bogor, Kota Bekasi, Kabupaten Bekasi, dan Kabupaten Indramayu, yang dapat dijadikan fokus utama dalam intervensi kesehatan. Ke depan, akurasi model dapat ditingkatkan dengan menambahkan fitur-fitur kontekstual seperti kondisi sosial-ekonomi, tingkat edukasi kesehatan masyarakat, serta akses terhadap layanan medis. Selain itu, penerapan model machine learning yang lebih kompleks seperti Random Forest, Gradient Boosting, atau Neural Network juga dapat dieksplorasi untuk menghasilkan prediksi yang lebih akurat dan adaptif terhadap karakteristik data epidemiologi.

Referensi

[1] D. P. S. Dewi and A. Voutama, "Penerapan Algoritma K-means Pada Kasus HIV Di Jawa Barat Untuk Pengelompokan Berdasarkan Tingkat Penyebaran Di Setiap Kabupaten/Kota Menggunakan Rapidminer," *Jurnal Teknologi Sistem Informasi*, vol. 4, no. 2, pp. 289–300, 2023, https://doi.org/10.35957/jtsi.v4i2.4949.

[2] P. Taniwan, A. F. Bilbina, C. R. S. Ganap, and D. Y. Faidah, "Pemodelan Faktor-Faktor yang Memengaruhi Kejadian HIV/AIDS di Provinsi Jawa Barat," *Indo-MathEdu Intellectuals Journal*, vol. 5, no. 3, pp. 3298–3308, 2024, https://doi.org/10.54373/imeij.v5i3.1278.

[3] M. Soni, N. Rahaningsih, and R. Danar Dana, "Komparasi Algoritma K-Means Dan K-Medoids Clustering Pada Data Penyebaran Kasus Hiv Di Provinsi Jawa Barat," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3766–3772, 2024, https://doi.org/10.36040/jati.v7i6.8274.

[4] Dkk. Ismail, I. A., "Hubungan Pengetahuan Dan Sikap Terhadap Pencegahan HIV / AIDS pada Remaja," *Borneo Nursing Journal (BNJ)*, vol. 4, no. 2, pp. 58–63, 2022, https://doi.org/10.61878/bnj.v4i2.39.

[5] N. R. Ihzan Sayid Muallif, Herdi Budiman, "Penerapan Data Mining untuk Prediksi Pergerakan Harga Saham Menggunakan Algoritma K-Nearest Neighbor," *Jurnal Media Informatika Budidarma*, vol. 8, no. 1, p. 497, 2024, https://doi.org/10.30865/mib.v8i1.7279.

[6] V. Syafana, S. S. Hilabi, E. Novalia, and B. Huda, "Prediction of Birth Rates in Different Age Groups of Mothers Using the K-Nearest Neighbor (K-NN) Method Prediksi Angka Kelahiran dalam Berbagai Kelompok Umur Ibu Menggunakan Metode K-Nearest Neighbor," vol. 4, no. July, pp. 1096–1103, 2024, https://www.journal.irpi.or.id/index.php/malcom/article/view/1392.

[7] M. Yunus and N. K. A. Pratiwi, "Prediksi Status Gizi Balita Dengan Algoritma K-Nearest Neighbor (KNN) di Puskemas Cakranegara," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 4, no. 4, pp. 221–231, 2023, https://doi.org/10.35746/jtim.v4i4.328.

[8] V. Syafana, S. S. Hilabi, E. Novalia, and B. Huda, "Prediction of Birth Rates in Different Age Groups of Mothers Using the K-Nearest Neighbor (K-NN) Method Prediksi Angka Kelahiran dalam Berbagai Kelompok Umur Ibu Menggunakan Metode K-Nearest Neighbor," vol. 4, no. July, pp. 1096–1103, 2024, https://www.journal.irpi.or.id/index.php/malcom/article/view/1392.

[9] D. Agung Pratama, E. Novalia, A. Lia Hananto, and U. Buana Perjuangan Karawang, "Implementasi Metode Simple Additive Weighting Menggunakan Python Terhadap Penentuan Menu Minuman Favorit," *INNOVATIVE: Journal Of Social Science Research*, vol. 4, pp. 3395–3407, 2024, http://j-innovative.org/index.php/Innovative/article/view/9594.

[10] O. D. Jabar, "Jumlah Kasus HIV Berdasarkan Kabupaten/Kota di Jawa Barat," Αγαη. Accessed: Sep. 20, 2024, https://opendata.jabarprov.go.id/id/dataset/jumlah-kasus-hiv-berdasarkan-kabupatenkota-di-jawa-barat

[11] Moch. Lutfi, "Prediksi Harga Terendah Dan Harga Tertinggi Dengan Menggunakan Metode Anfis Untuk Analisa Teknikal Pada Forex Market," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 1, no. 3, pp. 261–268, 2019, https://doi.org/10.35746/jtim.v1i3.40.

[12] Moh. A. Azman and N. Nafi'iyah, "Knn for Classification of Fruit Types Based on Fruit Features," *Jurnal Teknologi Informasi Universitas Lambung Mangkurat (JTIULM)*, vol. 7, no. 1, pp. 31–36, 2022, doi: https://doi.org/10.20527/jtiulm.v7i1.100.

[13] H. W. A. Haeruddin, Erick, "Perbandingan Support Vector Machine, Random Forest Classi- fier, dan K-Nearest Neighbour dalam Pendeteksian Anomali pada Jaringan DDos," vol. 7, no. 1, pp. 23–33, 2025, https://doi.org/10.35746/jtim.v7i1.628.

[14] R. Bahtiar, "Implementasi Data Mining Untuk Prediksi Penjualan Kusen Terlaris Menggunakan Metode K-Nearest Neighbor," *Jurnal Informatika MULTI*, vol. 1, no. 3, pp. 203–214, 2023, https://jurnal.publikasitecno.id/index.php/multi/article/view/33.

[15] D. N. Aini, B. Oktavianti, M. J. Husain, D. A. Sabillah, S. T. Rizaldi, and M. Mustakim, "Seleksi Fitur untuk Prediksi Hasil Produksi Agrikultur pada Algoritma K-Nearest Neighbor (KNN)," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 1, p. 140, 2022, https://doi.org/10.30865/json.v4i1.4813.

[16] N. Prasetya and W. Tampubolon, "Penerapan Algoritma K-Nearest Neighbor (KNN) Untuk Prediksi Pengajuan Kredit Perumahan Pada PT. Bopala Kencana Permai," 2022.
[17] R. N. Angraeni, B. Priyatna, A. Hananto, and S. S. Hilabi, "Application of the K-Nearest Neighbor Method to Predict Demand for Goods from Customers at PT Sinergi Prima Enjineering," pp. 99–109, 2024, https://doi.org/10.54209/jurnalinstall.v16i02.200.
[18] U. Nijunnihayah, S. S. Hilabi, F. Nurapriani, and E. Novalia, "Implementasi Algoritma K-Nearest Neighbor untuk Prediksi Penjualan Alat Kesehatan pada Media Alkes," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 695–701, 2024, https://doi.org/10.57152/malcom.v4i2.1326.
[19] D. Puspita Sari, S. Shofia Hilabi, and Agustia Hananto, "Penerapan Data Mining Metode K-Nearest Neighbor Untuk Memprediksi Kelulusan Siswa Sekolah Menengah Pertama," *SMARTICS Journal*, vol. 9, no. 1, pp. 14–19, 2023, https://doi.org/10.21067/smartics.v9i1.8088.