

## 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 memprediksi 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



KARAWANG

## 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.*

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

