

PENGELOMPOKAN RISIKO DAN ANALISIS TREN KASUS TUBERKULOSIS DENGAN METODE CLUSTERING DAN SARIMA

ABSTRAK

Tuberkulosis (TBC) masih menjadi masalah kesehatan masyarakat yang serius di Kabupaten Aceh Utara, di mana fluktuasi jumlah kasus yang tajam dan perbedaan karakteristik antarwilayah menyebabkan penyebaran penyakit tidak merata serta sulit diprediksi. Penelitian ini bertujuan untuk mengelompokkan risiko wilayah di Kabupaten Aceh Utara berdasarkan karakteristik spesifik kasus TBC, membangun model SARIMA yang optimal pada setiap kelompok risiko untuk memprediksi tren perkembangan kasus di masa depan, serta mengevaluasi akurasi dari model yang diterapkan. Menggunakan data sekunder kejadian TBC bulanan periode 2020-2025 dari Dinas Kesehatan Kabupaten Aceh Utara, tahap pertama membandingkan tiga metode *clustering* (*Gaussian Mixture Model*, *K-Means*, dan *Agglomerative Hierarchical Clustering*) menggunakan metrik *Silhouette Score* dan *Davies-Bouldin Index*. Tahap kedua menggunakan metode SARIMA pada setiap klaster untuk meramal tren kasus 12 bulan ke depan, di mana model terbaik ditentukan berdasarkan nilai AIC terkecil dan ketepatan prediksi diukur menggunakan MAPE dan RMSE. Hasil analisis menunjukkan bahwa meskipun AHC memperoleh hasil evaluasi klaster yang sedikit lebih baik, *K-Means* dipilih sebagai metode pengelompokan karena lebih mudah diinterpretasikan dan sesuai untuk pembentukan deret waktu pada tahap *forecasting*. *K-Means* menghasilkan *Silhouette Score* sebesar 0,3181 dan DBI sebesar 1,1111, yang berhasil membagi wilayah Kabupaten Aceh Utara menjadi tiga klaster tingkat risiko. Pemodelan deret waktu menghasilkan model SARIMA terbaik untuk masing-masing kelompok, yaitu $SARIMA(3,1,0)(0,1,0)_{12}$ untuk Klaster 1 (AIC: -75,692; MAPE: 3,29%; RMSE: 5,12), $SARIMA(0,1,0)(0,1,1)_{12}$ untuk Klaster 2 (AIC: -64,826; MAPE: 4,68%; RMSE: 1,52), dan $SARIMA(0,1,0)(0,1,1)_{12}$ untuk Klaster 3 (AIC: -156,859; MAPE: 2,56%; RMSE: 3,80). Hasil penelitian menunjukkan bahwa integrasi metode *K-Means* dan SARIMA dapat digunakan sebagai pendekatan untuk mendukung pemetaan wilayah prioritas serta perencanaan strategi pengendalian TBC di Kabupaten Aceh Utara.

Kata Kunci: Tuberkulosis, Clustering, K-Means, SARIMA, Peramalan, Kabupaten Aceh Utara.

RISK GROUPING AND TREND ANALYSIS OF TUBERCULOSIS CASES USING CLUSTERING AND SARIMA METHODS

ABSTRACT

Tuberculosis (TB) remains a serious public health problem in North Aceh Regency, where sharp fluctuations in the number of cases and differences in characteristics between regions cause the spread of the disease to be uneven and difficult to predict. This study aims to group regional risks in North Aceh Regency based on the specific characteristics of TB cases, build an optimal SARIMA model for each risk group to predict future case development trends, and evaluate the accuracy of the applied model. Using secondary data on monthly TB incidence for the period 2020-2025 from the North Aceh Regency Health Office, the first stage compares three clustering methods (Gaussian Mixture Model, K-Means, and Agglomerative Hierarchical Clustering) using the Silhouette Score and Davies-Bouldin Index metrics. The second stage uses the SARIMA method on each cluster to forecast the case trends for the next 12 months, where the best model is determined based on the smallest AIC value and prediction accuracy is measured using the MAPE and RMSE. The analysis results show that although AHC obtained slightly better cluster evaluation results, K-Means was chosen as the clustering method because it is easier to interpret and suitable for forming time series in the forecasting stage. K-Means produced a Silhouette Score of 0.3181 and a DBI of 1.1111, which successfully divided the North Aceh Regency area into three risk level clusters. Time series modeling produced the best SARIMA model for each group, namely SARIMA(3,1,0)(0,1,0)₁₂ for Cluster 1 (AIC: -75,692; MAPE: 3,29%; RMSE: 5,12), SARIMA(0,1,0)(0,1,1)₁₂ for Cluster 2 (AIC: -64,826; MAPE: 4,68%; RMSE: 1,52), and SARIMA(0,1,0)(0,1,1)₁₂ for Cluster 3 (AIC: -156,859; MAPE: 2,56%; RMSE: 3,80). The results showed that the integration of the K-Means and SARIMA methods can be used as an approach to support priority area mapping and TB control strategy planning in North Aceh Regency.

Keywords: Tuberculosis, Clustering, K-Means, SARIMA, Forecasting, North Aceh Regency.