

Penerapan penambangan data dalam mengisi Missing Data Curah Hujan Harian: Studi kasus Pusat Database BMKG = Application of data mining in Daily Rainfall Missing Data Imputation: A Case study of the Pusat Database BMKG

Reza Bayu Perdana, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920567177&lokasi=lokal>

Abstrak

Data curah hujan harian merupakan data yang penting dan paling banyak diakses oleh pengguna portal data terbuka BMKG. Adanya data yang hilang (missing data) pada data curah hujan harian dapat menyebabkan turunnya kualitas dataset curah hujan harian. Penelitian ini bertujuan untuk mengevaluasi metode pengisian missing data curah hujan harian pada dataset dari Pusat Database BMKG menggunakan pendekatan penambangan data. Algoritma pemodelan yang digunakan di antaranya Multiple Imputation by Chained Equations (MICE), Random Forest (RF), Inverse Distance Weighting (IDW), K-Nearest Neighbor (KNN), Feed-forward Neural Network (FNN), Mean Imputation, dan Linear Interpolation (LI). Evaluasi dilakukan dengan menggunakan metrik koefisien determinasi (R^2), Root Mean Square Error (RMSE), dan waktu komputasi. Hasil penelitian menunjukkan RF unggul di wilayah Monsunal pada semua tingkat kehilangan data, sementara MICE-LR lebih direkomendasikan untuk wilayah Ekuatorial dengan performa konsisten hingga kehilangan data 50%. Di wilayah Anti-Monsunal, IDW lebih efektif menangani tingkat kehilangan data tinggi karena pola spasial yang unik, sementara pada tingkat kehilangan data rendah hingga sedang, MICE-LR direkomendasikan. Penelitian ini menyimpulkan bahwa rekomendasi pemodelan untuk mengisi data curah hujan harian yang hilang harus mempertimbangkan karakteristik wilayah dan tingkat kehilangan data untuk menghasilkan data curah hujan yang lebih akurat.

.....Daily rainfall data is crucial and among the most accessed datasets on BMKG's open data portal. Missing data in daily rainfall records can significantly degrade the quality of the dataset. This study aims to evaluate imputation methods for missing daily rainfall data in the BMKG Database Center dataset using data mining approaches. The modeling algorithms used include Multiple Imputation by Chained Equations (MICE), Random Forest (RF), Inverse Distance Weighting (IDW), K-Nearest Neighbor (KNN), Feed-forward Neural Network (FNN), Mean Imputation, and Linear Interpolation (LI). Evaluation metrics include the coefficient of determination (R^2), Root Mean Square Error (RMSE), and computational time. The findings indicate that RF performs best in Monsunal regions across all levels of missing data. Meanwhile, MICE-LR is recommended for Equatorial regions due to its consistent performance, even with up to 50% missing data. In Anti-Monsunal regions, IDW is more effective in handling high levels of missing data due to unique spatial patterns, whereas MICE-LR is recommended for low to moderate levels of missing data. The study concludes that selecting the appropriate imputation method should consider regional characteristics and the extent of missing data to produce more accurate rainfall datasets.