

ABSTRACT

The availability of water resources data in Indonesia has several complex problems related to the completeness of the data. Problems that occur during data collection in several agencies at Indonesia are the accuracy of the data and the completeness of the data. The concept of water in the form of objects can be said to be very dynamic starting from the shape, color, discharge, and smell. The impact of climate change and natural disasters is also a contributing factor to the dynamic changes of water. The characteristics of the watershed in DI Yogyakarta Province which is fan-shaped with many downstream branches of the river make the complexity of water resource data more widespread.

One of the problems that arise is the missing value of water resources data which can affect the processing of water resources data. There are several methods that can be used for Missing Value Imputation, one of which is k-Nearest Neighbors Imputation (k-NNi) and Multivariate Imputation by Chained Equation (MICE). The two methods proposed are used to compare and find the most appropriate method using the Opak watershed in Yogyakarta D.I Province.

The results of the statistical validation comparison show that the most consistent average value of RMSE and MAE is the k-NNi method with a value of $k=8$. As for the comparison of R^2 values, the k-NNi method with a value of $k=8$ gets the best average value of 80%, followed by the k-NNi method of $k=20$ as the default k value with a percentage of 78%. The MICE comparison method gets the lowest average percentage value from other methods by only getting a value of 63%.

Keywords: Watershed, Waterflow, k-NNi, MICE, imputation, *missing value*, *statistical validation*

INTISARI

Ketersediaan data sumber daya air di Indonesia memiliki beberapa permasalahan yang kompleks terkait dengan kesempurnaan dan kelengkapan data. Permasalahan yang terjadi pada saat pendataan di beberapa instansi di Indonesia adalah keakuratan data dan kelengkapan data. Konsep air dalam wujud benda dapat dikatakan sangat dinamis mulai dari bentuk, warna, debit, dan bau. Dampak perubahan iklim dan bencana alam juga menjadi faktor pendukung perubahan dinamis dari air. Karakteristik DAS yang ada di Provinsi DI Yogyakarta yang berbentuk kipas dengan hilir cabang sungai yang banyak menjadikan kompleksitas dari data sumber daya air menjadi semakin meluas.

Salah satu permasalahan yang timbul adalah nilai hilang dari data sumber daya air yang dapat berpengaruh terhadap pengolahan data sumber daya air. Ada beberapa metode yang dapat digunakan untuk *imputasi missing value*, salah satunya adalah *k-Nearest Neighbors Imputation* (k-NNi) dan *Multivariate Imputation by Chained Equation* (MICE). Kedua metode akan yang diusulkan digunakan untuk membandingkan dan menemukan metode yang paling tepat menggunakan DAS Opak di Provinsi D.I Yogyakarta.

Hasil perbandingan validasi statistik, nilai rata-rata RMSE dan MAE yang paling konsisten adalah metode k-NNi dengan nilai $k=8$. Sedangkan untuk perbandingan nilai R^2 , metode k-NNi dengan nilai $k=8$ mendapatkan nilai rata-rata terbaik sebesar 80%, disusul dengan metode k-NNi sebesar $k=20$ sebagai nilai k default dengan persentase 78 %. Metode perbandingan MICE mendapatkan nilai persentase rata-rata terendah dari metode lainnya dengan hanya mendapatkan nilai sebesar 63%.

Kata kunci --DAS, Debit, k-NNi, MICE, imputasi, *data hilang*, *statistical validation*