Metode Imputasi pada Data Debit Daerah Aliran Sungai Opak, Provinsi DI Yogyakarta

(*Imputation Method on Opak Watershed Data, Special Region of Yogyakarta*)

Fahmi Dhimas Irnawan[1], Indriana Hidayah[2], Lukito Edi Nugroho[3]

***Abstract*—The data availability of water resources in Indonesia has several complex problems related to the perfection of data. The problems taking place when collecting data in several Indonesian agencies are the accuracy and completeness of the data. There are several methods that can be used to handle missing value imputation, such as k-Nearest Neighbors Imputation (k-NNi) and Multivariate Imputation by Chained Equation (MICE). This study seeks to compare and find the most appropriate method using the Opak watershed dataset in Special Region of Yogyakarta. The characteristics of the Opak watershed lies in its fan shape that provides a lower concentration-time and produces a higher flow. The results of the statistical validation comparison showed that the most consistent average value of RMSE and MAE was the k-NNi method with a value of k = 28. As for the comparison of R-Squared values, the k-NNi method with a value of k = 28 obtained the best average value with 80%, followed by the k-NNi method of k = 7 as the default k value with a percentage of 73%. Among the applied methods, the MICE comparison method obtained the lowest average percentage value with 63%.**

***Intisari*—Ketersediaan data sumber daya air di Indonesia memiliki beberapa permasalahan yang kompleks terkait dengan kesempurnaan data. Permasalahan yang terjadi saat pendataan di beberapa instansi di Indonesia adalah kurangnya keakuratan dan kelengkapan data. Terdapat beberapa metode yang dapat digunakan untuk imputasi nilai hilang, misalnya *k-Nearest Neighbors Imputation* (k-NNi) dan *Multivariate Imputation by Chained Equation* (MICE). Tujuan makalah ini adalah membandingkan dan menemukan metode yang paling tepat dalam menggunakan dataset DAS Opak di Provinsi DIY. Karakteristik DAS Opak adalah berbentuk kipas atau melebar sehingga memberikan waktu konsentrasi yang lebih rendah dan menghasilkan aliran yang lebih tinggi. Hasil perbandingan validasi statistik, nilai rata-rata RMSE dan MAE, yang paling konsisten adalah metode k-NNi dengan nilai k = 28, sedangkan untuk perbandingan nilai *R-Squared*, metode k-NNi dengan nilai k = 28 mendapatkan nilai rata-rata terbaik sebesar 80%, disusul metode k-NNi sebesar k = 7 sebagai nilai k *default* dengan persentase 73%. Metode perbandingan MICE mendapatkan nilai persentase rata-rata terendah dari metode lainnya, yaitu sebesar 63%.**

***Kata Kunci*—DAS, Debit, k-NNi, MICE, Imputasi, *Missing Value*, *Statistical Validation*.**

[1,2,3] *Departemen Teknik Elektro dan Teknologi Informasi, Fakultas Teknik, Universitas Gadjah Mada, Jln. Grafika No.2, Kampus UGM, Yogyakarta, 55281, INDONESIA (Tel. +62-274-552305, email: [1]fahmi.dhimas.irnawan@mail.ugm.ac.id, [2]indriana.h@ugm.ac.id, [3]lukito@ugm.ac.id)*

I. PENDAHULUAN

Air adalah sumber daya alam yang hadir dalam berbagai bentuk, seperti sungai, sumur, danau dan waduk. Pengembangan sumber daya untuk berbagai keperluan, termasuk konsep hidrologi, sumber daya air, dan penanggulangan banjir, merupakan dasar bagi pembangunan sosial ekonomi masyarakat [1]. Dampak perubahan iklim juga menjadi faktor pendukung perubahan dinamis air dalam wujud benda, mulai dari bentuk, warna, debit, dan bau. Daerah Aliran Sungai (DAS) berperan penting dalam perubahan tersebut, seperti kegiatan sosial yang meningkat dan perkembangan tutupan lahan, yang menyebabkan limpasan air mengalir dari hulu ke hilir. Hal ini menyebabkan kenaikan muka air di DAS yang sangat signifikan di atas normal, sehingga berakibat pada meluapnya air sungai atau dikenal dengan banjir [2].

Masalah yang sering terjadi pada saat pengambilan data yang ada di beberapa lembaga di Indonesia adalah keakuratan data dan kelengkapan data yang kurang. Secara tradisional, sistem pemodelan simulasi hidrologi diklasifikasikan menjadi tiga kelompok utama, yaitu *black box* empiris, konseptual terpusat, dan model berbasis fisik terdistribusi. Secara umum, konsep tradisional kurang memiliki akurasi dalam sampel data yang akan dianalisis [1]. Masalah paling umum dalam *dataset* nyata dan analisis statistik adalah nilai yang hilang, dengan persentase nilai yang hilang bervariasi dari satu kumpulan data ke kumpulan data lainnya. Umumnya, kumpulan data berisi persentase berbeda dari nilai yang hilang di setiap kolom [3], [4]. Rasio nilai hilang yang kurang dari 1% disebut *trivial* dan rasio nilai yang hilang dengan kisaran 1-5% bersifat *flexible*. Metode lanjutan yang diterapkan untuk menangani nilai yang hilang pada kisaran 5-15% dan < 15% berdampak sangat besar pada analisis [5].

Proses untuk memperbaiki data dengan melakukan imputasi perlu dilakukan dengan estimasi untuk mengisi kekosongan data agar kesimpulan statistik menjadi lengkap dan efisien [6]. Sangat penting untuk diperhatikan bahwa terdapat perbedaan antara nilai kosong dan nilai yang hilang. Nilai kosong berarti tidak ada nilai yang dapat diberikan, sedangkan nilai yang hilang berarti nilai aktual untuk variabel itu ada, tetapi tidak tersedia atau ditangkap dalam kumpulan data karena beberapa alasan [7]. Pentingnya pendekatan yang tepat terhadap penanganan nilai hilang merupakan syarat mutlak terhadap kondisi sebuah data, sehingga dibutuhkan beberapa teknik untuk dapat mengakomodasi nilai hilang dan meminimalkan efek negatif terhadap sebuah data [8]. Tiga

pendekatan yang paling banyak digunakan yang diidentifikasi adalah memeriksa kasus yang tidak lengkap, mengganti nilai untuk nilai yang hilang, dan menyediakan statistik bobot untuk menyelesaikan kasus [9].

Sebelum menggunakan metode apa pun untuk menangani nilai yang hilang, sebab nilai hilang perlu dipahami terlebih dahulu [7]. Terdapat tiga konsep kemungkinan mekanisme nilai yang hilang, yaitu *Missing Completely at Random* (MCAR), *Missing at Random* (MAR), dan *Missing Not at Random* (MNAR) [10], [11]. MNAR sering dianggap sebagai tipe hilang yang paling buruk karena dapat menyebabkan hasil yang menyimpang (terjadi bias), sedangkan MCAR dan MAR dapat menyebabkan hilangnya kekuatan statistik [12]. Berbagai jenis nilai yang hilang ini penting diketahui karena menentukan perlakuan statistik dari nilai hilang yang dapat digunakan secara efektif [7].

Imputasi data merupakan proses memperkirakan nilai yang hilang dari suatu pengamatan berdasarkan nilai-nilai valid dari variabel lain [10]. Salah satu hal yang perlu diperhatikan dalam imputasi adalah memastikan metode yang digunakan untuk imputasi nilai hilang. Secara spesifik, imputasi data diklasifikasikan menjadi dua jenis, yaitu *single imputation* dan *multiple imputation* [7].

Dalam penelitian pada pemanfaatan data UCI *machine learning repository***,** *single imputation* dengan metode *k-Nearest Neighbors Imputation* (k-NNi) memiliki kinerja yang cukup signifikan dibandingkan dengan *multiple imputation* [7]. Sementara itu, pada penelitian analisis perbandingan imputasi dengan memanfaatkan data *Integrated City Sustainability Database* (ICSD), *multiple imputation* menggunakan metode *Multivariate Imputation by Chained Equation* (MICE) yang memungkinkan dilakukannya analisis dengan ukuran sampel yang lebih besar, bias yang lebih sedikit, dan kemampuan untuk menginterpretasikan data seolah-olah data tersebut tidak hilang [8].

Nilai *Root Mean Squared Error* (RMSE) dari MICE didasarkan pada algoritme yang jauh lebih kompleks dan perilakunya tampaknya terkait dengan ukuran kumpulan data, yaitu cepat dan efisien pada kumpulan data kecil, tetapi kinerja sedikit menurun dan waktu eksekusi meningkat ketika diterapkan pada kumpulan data besar. Sementara itu, metode k-NNi menghasilkan nilai yang stabil meningkat berdasarkan ukuran data yang terkecil hingga ukuran data besar [13].

Makalah ini secara khusus menggunakan *package* dan *library* yang tersedia pada pemrograman R. *Package* MICE sangat kompatibel dengan pemrograman R [14], sedangkan k-NNi dapat dilakukan dengan menggunakan paket VIM di R [15].

Makalah ini diharapkan mampu memberikan pengetahuan tentang pengisian nilai yang hilang sebagai bahan untuk menyusun prosedur *pre-processing* yang menangani anomali beberapa data dalam suatu *dataset*. Hal ini dilakukan karena pendekatan konvensional memperlakukan setiap anomali data sebagai kasus yang terisolasi, tetapi anomali data juga terjadi pada data. Metode MICE dan k-NNi menjadi fokus makalah, dengan data DAS Sungai Opak dengan kategori data numerik dan *time series* dengan ukuran data sedang.

Karena pengaruh pengolahan data terhadap hasil analisis statistik belum banyak diteliti, maka perlu dilakukan penelitian di bidang ini untuk menghasilkan hasil yang akurat, terutama dalam pengisian nilai yang hilang. Akibatnya cakupan yang lebih luas dari penggabungan *imputation gap* dan data nilai yang hilang berdampak pada tingkat akurasi dalam penerapan prosedur *pre-processing*. Untuk itu, diperlukan penerapan prosedur *imputation missing data* agar data anomali, bias, dan *noise* dapat diminimalkan dan dapat memberikan perbandingan untuk menemukan metode yang paling tepat pada kasus DAS Opak di Provinsi DIY. Hasil perbandingan tersebut dapat menjadi acuan untuk melengkapi nilai yang hilang pada beberapa stasiun *Automatic Water Level Recorder* (AWLR).

II. Imputasi Nilai Hilang

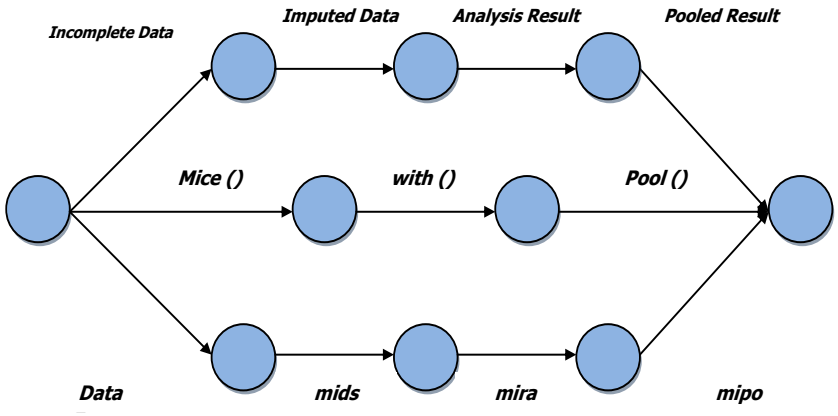
Nilai yang hilang dapat sangat mengganggu kualitas dan utilitas data. Masalah kesenjangan dalam *data series* dapat diselesaikan secara teoretis dengan melengkapi catatan aliran harian dari data yang ada di stasiun pengukuran terdekat, baik hulu atau hilir aliran air yang sama (misalnya, teknik interpolasi), meskipun pemilihan stasiun donor mungkin sangat penting dan menjadi faktor yang memengaruhi hasil akhir [16].

Beberapa ahli menggunakan berbagai teknik alternatif untuk mengakomodasi nilai yang hilang dan meminimalkan efek negatifnya. Imputasi data adalah salah satu teknik penanganan nilai yang hilang untuk membuat data lengkap dan siap dianalisis dengan mengganti nilai yang hilang dengan nilai yang paling masuk akal [7].

Dalam penelitian pada DAS di Little Ruaha, Tanzania, ditunjukkan bahwa terdapat pendekatan umum yang sering diadaptasi dalam menangani kesenjangan nilai yang hilang. Cara pertama hanya menggunakan catatan data secara kontinu dan mengabaikan peristiwa sebelumnya. Hal ini berdasarkan sebuah peristiwa atau kejadian sebelum *dataset* tersedia dengan melakukan asumsi bahwa data adalah satu seri catatan yang berkelanjutan. Cara selanjutnya adalah menghapus waktu hilangnya data dan menganggap data yang tersisa sebagai kumpulan data berkelanjutan [1].

Referensi [17] memberikan beberapa jenis teknik *mean imputation techniques* yang diperkenalkan pada metode imputasi nilai hilang. Kemudian nilai yang hilang dilakukan perhitungan menggunakan setiap metode dan dibandingkan dengan nilai yang diamati [17].

Salah satu metode *machine learning* yang dikembangkan untuk menangani nilai yang hilang adalah k-NN [18]. Metode ini menggunakan jarak antar data latih sebagai klasifikasi untuk melakukan pengujian [19]. Metode k-NN merupakan metode yang fleksibel, baik dalam data kontinu maupun data diskret [20]. Metode ini dapat digunakan karena pada beberapa data *filler* yang hilang [21], tidak diperlukan model prediksi untuk setiap atribut [22]. Salah satu kelemahan metode k-NN adalah kebutuhan waktu yang sangat tinggi dalam menganalisis *dataset* besar karena metode ini mencari data yang serupa di semua *dataset*. Selain itu, akurasi k-NN dapat sangat terdegradasi dengan data berdimensi tinggi



Gbr. 1 Langkah dalam konsep MICE.

karena terdapat sedikit perbedaan antara tetangga terdekat dan terjauh [3].

III. METODOLOGI

A. Metode k-Nearest Neighbors Imputation (k-NNi)

Dalam k-NN, untuk melakukan prediksi nilai yang hilang, digunakan rata-rata dari contoh data yang memiliki kemiripan [20]. Kemiripan yang terjadi diambil dengan rata-rata fungsi jarak antara stasiun AWLR hulu dan hilir. Dengan pertimbangan tersebut, dilakukan perbandingan nilai yang hilang dengan stasiun sebelumnya.

Metode k-NN melakukan pendekatan yang berbeda dengan melakukan estimasi, dengan asumsi bahwa harus dimiliki hasil keluaran yang serupa dengan data dengan fungsi jarak sebelumnya. Dengan demikian, tingginya resistansi terhadap mekanisme dan model nilai yang hilang menjadikan metode k-NNi sebagai salah satu pendekatan penanganan nilai yang hilang [23]. Hambatan ini juga dilihat dengan imputasi menggunakan parameter *k* yang berbeda. *Neighbors* dari titik data yang ditanyakan harus digunakan untuk menentukan nilainya dan bukan titik yang jauh. Dalam kasus regresi, nilai *k-Neighbors* dipilih dan rata-ratanya dianggap mengatur nilai titik yang ditanyakan [24]. Secara *default*, p = 2, sehingga menjadi *Euclidean distance*. Berikut ini adalah rumus k-NNi [18].

$$d_{(x,y)} = \sqrt{\sum_{j=1}^{s}(x_j - y_j)^2} \quad (1)$$

dengan

$d_{(x,y)}$ = *Euclidean distance*
j = atribut data, dengan j = 1,2,3, ..., s
s = dimensi data
x_{aj} = nilai dari atribut ke-j berisi nilai yang hilang
y_{bj} = nilai selain atribut ke-j berisi data yang lengkap.

Asumsi inti yang dibuat dengan metode k-NNi adalah bahwa contoh dengan vektor fitur serupa harus memiliki keluaran yang serupa. Tetangga dari titik data yang ditanyakan harus digunakan untuk menentukan nilainya dan bukan titik yang jauh. Dalam kasus regresi, nilai *k* tetangga dipilih dan nilai rata-rata dianggap menetapkan nilai titik yang ditanyakan [24].

B. Multivariate Imputation by Chained Equation (MICE)

MICE dikenal juga dengan "*Fully Conditional Spesification*" atau "*Sequential Regression Multiple Imputation*", yang digunakan dalam acuan statistika sebagai salah satu metode untuk menangani *missing data* [25]. Pada Gbr. 1 [26], ditunjukkan mekanisme *multiple imputation* dengan metode MICE, yang terbagi menjadi tiga tahap, yaitu imputasi data, analisis data, dan *pooling* [27].

Tahap pertama, yaitu imputasi data, merupakan tahap *dataset* dikenai proses imputasi dari distribusi yang menghasilkan *dataset* lengkap. Tahap analisis data adalah tahap ketika nilai hilang telah terisi dengan nilai yang mendekati nilai asli. Pada tahap selanjutnya, yaitu *pooling*, keluaranyang diperoleh setelah analisis data dikumpulkan untuk mendapatkan hasil akhir menggunakan aturan sederhana terhadap mekanisme *multiple imputation* [28].

Metode *multiple imputation* digunakan untuk mengganti nilai yang hilang dengan probabilitas nilai yang tepat. Kumpulan data yang tidak lengkap kemudian diubah menjadi kumpulan data yang lengkap dengan menggunakan metode imputasi, yang kemudian dapat dianalisis menggunakan metode analisis standar apa pun [28].

MICE dapat digunakan untuk berbagai model data, seperti data kontinu, data biner (regresi logistik), data kontinu 2-level, regresi logistik *polycotomus*, dan *odds proportional* [25]. Prosedur MICE mengikuti serangkaian model regresi, yaitu masing-masing variabel dari nilai yang hilang dimodelkan bersyarat pada variabel lain dalam data tersebut, sehingga setiap variabel dapat dimodelkan menurut distribusinya [26].

C. Statistical Validation

Tantangan terbesar pada proses *machine learning* adalah membuatnya berfungsi secara akurat pada data yang tidak terlihat. Untuk mengetahui model yang dirancang berfungsi dengan baik atau tidak, harus dilakukan pengujian terhadap titik-titik data yang tidak ada selama pelatihan model [26]. Salah satu teknik terbaik untuk memeriksa efektivitas model pembelajaran mesin adalah teknik validasi statistik yang dapat dengan mudah diterapkan menggunakan bahasa pemrograman

R. *Statistical validation* nantinya akan memberikan hasil validasi dengan metode akurasi RMSE, *R-squared* (R^2), dan *Mean Absolute Error* (MAE).

1) R-Squared (R^2): *R-Squared* adalah ukuran persentase variasi total dalam variabel dependen yang diperhitungkan oleh variabel independen [29]. *R-Squared* sebesar 1,0 menunjukkan bahwa data sangat cocok dengan model linier. Apabila nilai *R-Squared* kurang dari 1,0, setidaknya beberapa variabilitas dalam data tidak dapat diperhitungkan oleh model. Sebagai contoh, *R-Squared* sebesar 0,5 menunjukkan bahwa 50% variabilitas dalam data hasil tidak dapat dijelaskan oleh model atau metode tertentu. Nilai *R-Squared* dapat diturunkan secara matematis menggunakan (2) [29].

$$R^2 = 1 - \frac{SSE}{SS_{yy}} \quad (2)$$

dengan

R^2 = *R-Squared coefficient of determination*
SSE = *sum of squares of residuals*
SS_{yy} = *total sum of squares*.

2) Root Mean Square Error (RMSE): RMSE mengukur perbedaan antara nilai yang diperhitungkan dengan nilai yang sebenarnya. Pada dasarnya, hal ini mewakili standar deviasi sampel dari perbedaan antara data sebelum dan sesudah dilakukan imputasi [13]. Persamaan (3) merupakan rumus dari RMSE [13].

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i^{obs} - x_i^{imputed}\right)^2}{n}} \quad (3)$$

dengan

$RMSE$ = *root mean square error*
X_i^{obs} = *corresponding prediction*
$X_i^{imputed}$ = pengukuran ke-*i*
n = jumlah data poin.

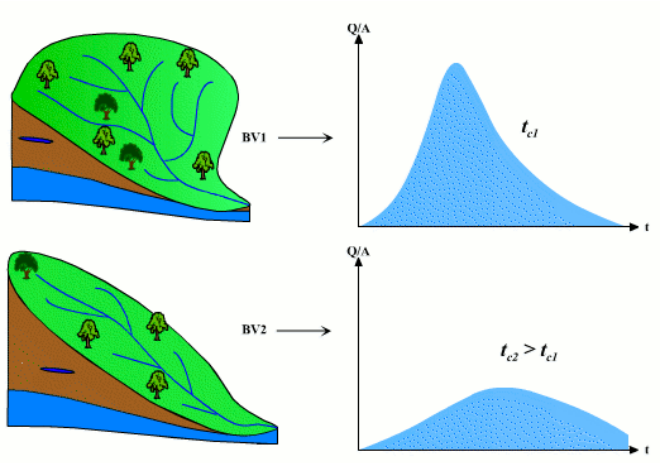
3) Mean Absolute Error (MAE): MAE adalah nilai mutlak dari selisih antara nilai prakiraan dengan nilai sebenarnya [30]. Untuk evaluasi model peramalan, MAE lebih intuitif dalam memberikan rata-rata *error* dari keseluruhan data [31]. Rumus matematis dari MAE ditunjukkan pada (4) [32].

$$MAE = \frac{\sum_{n=1}^{N} |\check{r}_n - r_n|}{N} \quad (4)$$

dengan

MAE = *mean absolute error*
$\check{r}_n$ = *prediction rating*
r_n = *true rating in testing data set*.

Statistical validation digunakan sebagai validasi model untuk menilai hasil statistik analisis imputasi dengan menggeneralisasi kumpulan data independen [33], [34]. *Statistical validation* dapat memberikan wawasan teoretis dan teori *asymptotic* pemilihan model dengan sejumlah variabel tetap sebagai model linier [35], [36].



Gbr. 2 Karakteristik DAS.

IV. Persiapan Eksperimen

A. Studi Area

Alur penelitian diawali dengan tinjauan pustaka dan pengambilan data dari pihak terkait, dalam hal ini dari SISDA Balai Besar Sumber Daya Air Progo-Serayu-Opak Kementrian Pekerjaan Umum dan Perumahan Rakyat. Makalah ini berfokus pada imputasi nilai hilang dari data debit DAS Opak di Provinsi DIY dengan area cakupan seluas 2,9 km^2 yang melintasi dua Kabupaten dan satu kota, dengan panjang aliran sejauh 62,83 km [37].

Bentuk DAS memengaruhi bentuk hidrografi karakteristiknya. Pada Gbr. 2 [38], terlihat dua bentuk DAS dengan curah hujan yang sama. Pada DAS berbentuk memanjang, aliran yang keluar lebih rendah karena waktu konsentrasi lebih tinggi. Sementara itu, DAS yang berbentuk kipas memberikan waktu konsentrasi yang lebih rendah, sehingga menghasilkan aliran yang lebih tinggi [38]. DAS Opak merupakan DAS dengan karakteristik berbentuk kipas atau melebar, seperti yang ditunjukkan pada Gbr. 3.

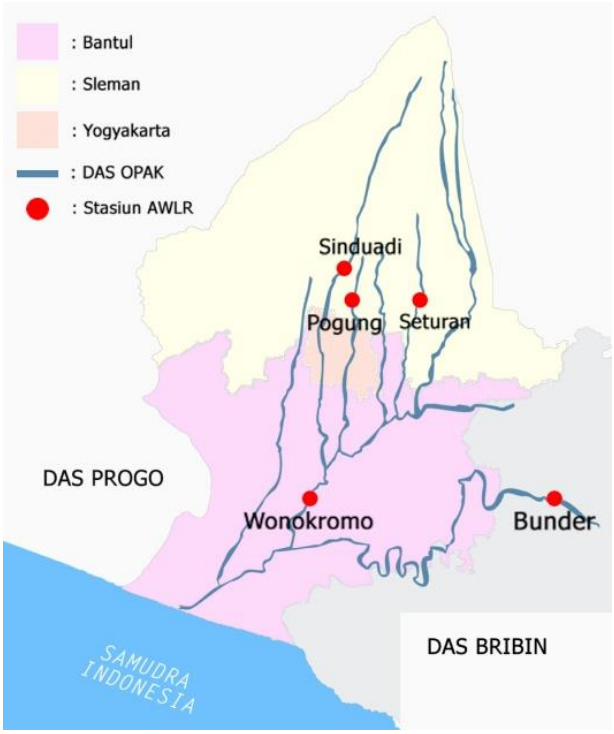
Makalah ini terfokus pada aliran stasiun AWLR dengan area cakupan DAS Opak sebanyak lima stasiun dari hulu hingga hilir. Tabel I menunjukkan letak koordinat AWLR dan sungai yang dilewati. Pengambilan data didasarkan pada titik pos stasiun AWLR yang tersebar di beberapa anak sungai. Kelima stasiun AWLR tersebut adalah stasiun Bunder, stasiun Pogung, stasiun Sinduadi, stasiun Seturan, dan stasiun Wonokromo.

B. Dataset Deskripsi

Dataset tersebut merupakan data yang diambil langsung dari Sistem Informasi Unit Balai Besar Wilayah Sumber Daya Air (BBWSDA) Provinsi DIY. Data debit DAS Opak di Provinsi DIY merupakan *dataset time series* yang diambil dari lima stasiun AWLR DAS Opak dengan durasi waktu *dataset* dimulai dari Januari 2007 hingga Desember 2017. Dari kelima stasiun, terdapat satu stasiun AWLR yang memiliki rasio 0% nilai hilang dan keempat stasiun lainya berada pada rasio 3%-19% nilai hilang. Tabel II menunjukkan atribut dari *dataset* lima stasiun AWLR.

TABEL I
DESKRIPSI AWLR

Stasiun AWLR	Sungai	Daerah Adm.	Koordinat	Hulu Sungai	Data Hilang
Stasiun Bunder	Sungai Oyo	Gunungkidul	-7,896006, 110,513925	Gajah Mungkur	0
Stasiun Pogung	Sungai Code	Sleman	-7,759429, 110,370078	Boyong	4
Stasiun Sinduadi	Sungai Winongo	Sleman	-7,748396, 110,357493	Denggung, Doso	5
Stasiun Seturan	Sungai Tambakbayan	Sleman	-7,747839, 110,357912	Opak	17
Stasiun Wonokromo	Sungai Gajahwong	Bantul	-7,866613, 110,394481	Lereng Merapi	25



Gbr. 3 Peta DAS Opak.

Kolom parameter *Rain* merupakan curah hujan stasiun penangkar hujan otomatis atau *Automatic Rain Recorder* (ARR). *Dataset* DAS Opak menggunakan konsep metode rasional USSCS 1973 yang terbatas untuk DAS dengan ukuran kecil, yaitu kurang dari 300 ha [39]. Secara keseluruhan, luas DAS Opak adalah 2,9 km^2 atau 290 ha, lebih kecil dari 300 ha.

Metode rasional berhubungan dengan koefisien aliran permukaan, intensitas hujan, dan waktu konsentrasi, sehingga parameter *dataset* DAS Opak adalah debit puncak bulanan, curah hujan, dan temperatur udara. Curah hujan dan temperatur merupakan variabel yang memengaruhi debit bulanan. Tabel II menjelaskan *dataset* DAS Opak.

Dataset DAS Opak masih berupa data kasar dan perlu dinormalisasi agar dapat diproses pada sistem *machine learning*. Proses normalisasi dilakukan pada kolom Bulan dan Tahun, yaitu menjadi kolom terpisah, yang menyatakan bulan dan tahun dalam kolom tersendiri. Tabel. III menyajikan hasil *dataset* DAS Opak yang telah melalui proses normalisasi.

Data stasiun AWLR yang memiliki 0% rasio nilai hilang, yaitu stasiun Bunder, menjadi *dataset* yang dikenai proses imputasi nilai hilang sebagai bahan validasi metode yang

TABEL II
ATRIBUT *DATASET*

Atribut	Tipe Data	*Range of Value*
Monthyear	*Date*	(Jan-07 – Des-17)
StationName_waterflow	*Real*	(0.07-131.42)
Rain	*Real*	(0,07 - 19.95)
Temp	*Real*	(0,01 - 4,16)

TABEL III
NORMALISASI *DATASET*

Attribut	Tipe Data
X1	*Numeric*
Bunder Station	*Numeric*
Rain	*Numeric*
Temp	*Numeric*
Year	*Numeric*
Month	*Numeric*

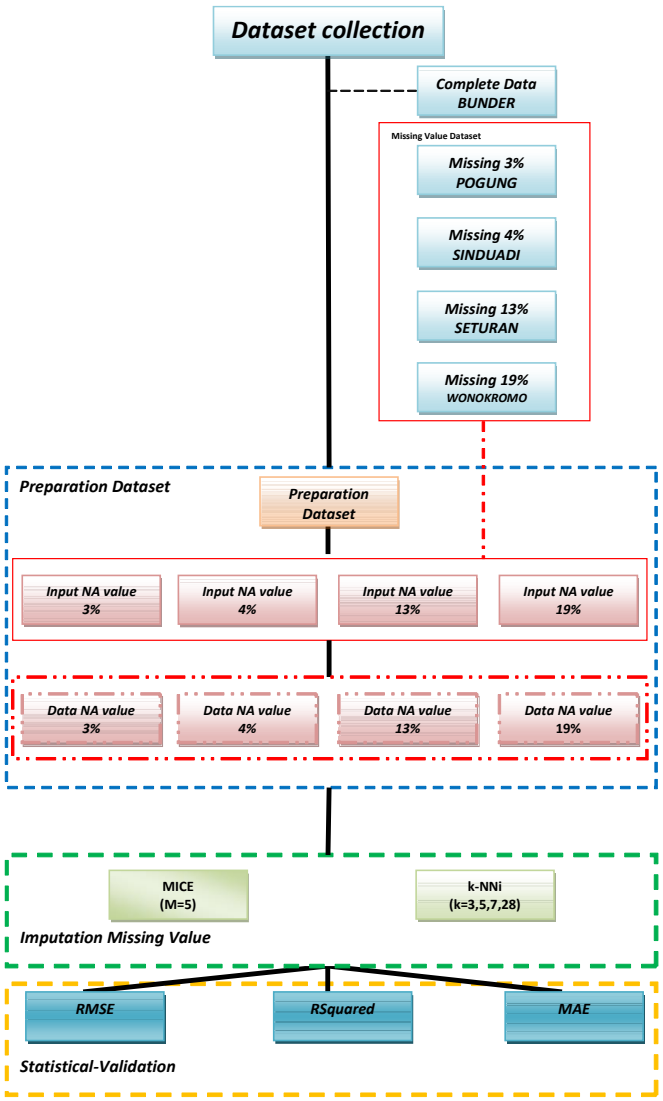
dilakukan. Kemudian, empat stasiun lain dihitung jumlah nilai kosongnya dan digunakan sebagai acuan dengan melakukan imputasi nilai kosong ke stasiun Bunder.

C. Skenario Eksperimen

Tahap awal proses penelitian ini adalah persiapan *dataset* untuk dinormalisasi agar dapat dilakukan proses *machine learning* menggunakan pemrograman R. Pada pemrograman R, data yang diproses menggunakan data *numeric*, sehingga poin data berupa bulan dan tahun diubah menjadi angka dengan menambahkan kolom Tahun dan Bulan ke dalam *dataset*, kemudian menghapus kolom *Monthyear*.

Tahap selanjutnya adalah memberikan nilai *Not Available* (NA) dari *dataset* dengan 0% nilai hilang, yaitu *Bunder_waterflow*. Pemberian nilai NA dibagi menjadi empat *dataset*, dengan masing-masing persentase berbeda, sesuai dengan empat *dataset* stasiun AWLR lain sesuai nilai hilang pada Tabel I. Pada tahap ini, nilai NA akan diberikan ke dalam *dataset* stasiun Bunder dengan masing masing nilai hilang sebesar 3% NA, 4% NA, 13% NA, dan 19% NA.

Dataset tersebut dikenai proses imputasi nilai hilang pada *software* Rstudio dengan metode MICE dan k-NNi. Setelah proses imputasi nilai hilang dilakukan, dilanjutkan dengan proses validasi menggunakan *statistical validation* untuk mengetahui nilai R^2, MAE, dan RMSE. Hasil perhitungan validasi akan membandingkan metode imputasi menggunakan MICE m = 5, k-NNi k = 3, k-NNi k = 5, k-NNi k = 7, dan k-NNi, k = 28. Gbr. 4 memperlihatkan diagram proses yang dilakukan.



Gbr. 4 Skenario eksperimen.

V. Hasil dan Pembahasan

A. Persiapan Dataset

Pada tahap awal persiapan *dataset* akan dimasukkan nilai NA pada stasiun Bunder dengan persentase 3%, 4%, 13%, dan 19%, sehingga menjadi empat *dataset* stasiun bunder seperti ditunjukkan pada Tabel IV. Proses persiapan *dataset* dilanjutkan untuk mencari Nilai *k* pada merode k-NNi. Gbr. 5 memperlihatkan hasil proses *machine learning* pemrograman R dengan asumsi nilai yang hilang adalah nilai titik yang paling dekat dengan variabel lain. Dalam hal ini, variabel *Rain* dan *Temp* merupakan variabel yang paling dekat dengan variabel *Bunder_waterflow*.

Range nilai 25-30 memiliki titik biru yang merupakan nilai NA dengan asumsi bahwa nilai *Rain* dengan *Temp* merupakan variabel tetangga yang berpengaruh terhadap debit atau variabel *Bunder_waterflow*, yaitu hujan berpengaruh terhadap debit sungai dari sisi curah hujan. Sementara itu, temperatur berpengaruh terhadap potensi curah hujan dan proses hidrologi sungai. Nilai *k* = 28 diambil dari titik biru dengan

Tabel IV
Masukan NA

Bunder_waterflow **(Dataset Asli)**	***Bunder_waterflow*** **(Masukan NA)**			
	3%	**4%**	**13%**	**25%**
5,83	5,83	5,83	5,83	5,83
18,54	18,54	18,54	18,54	18,54
9,55	9,55	9,55	9,55	9,55
26,07	26,07	26,07	26,07	26,07
3,00	3,00	3,00	3,00	3,00
2,77	2,77	2,77	2,77	2,77
1,64	1,64	1,64	NA	NA
0,34	0,34	0,34	NA	NA
0,17	0,17	0,17	0,17	0,17
...	...	...	...	...
5,93	5,93	5,93	5,93	5,93
9,01	9,01	9,01	NA	NA

intensitas terbanyak pada *range* 27-30, sehingga titik tengah, yaitu nilai 28, dipilih sebagi nilai *k* untuk proses imputasi. Nilai *k* = 5 merupakan nilai *default* metode k-NNi dan *k* = 3 adalah nilai *k* terendah menggunakan rumus *Euclidean distance* dengan satu dimensi variabel. Nilai parameter ketiga adalah *k* = 7, yang merupakan nilai hasil perhitungan rata-rata *Euclidean distance* pada variable *Bunder_waterflow*. Nilai *k* = 28 merupakan nilai yang diambil dari GGPLOT.

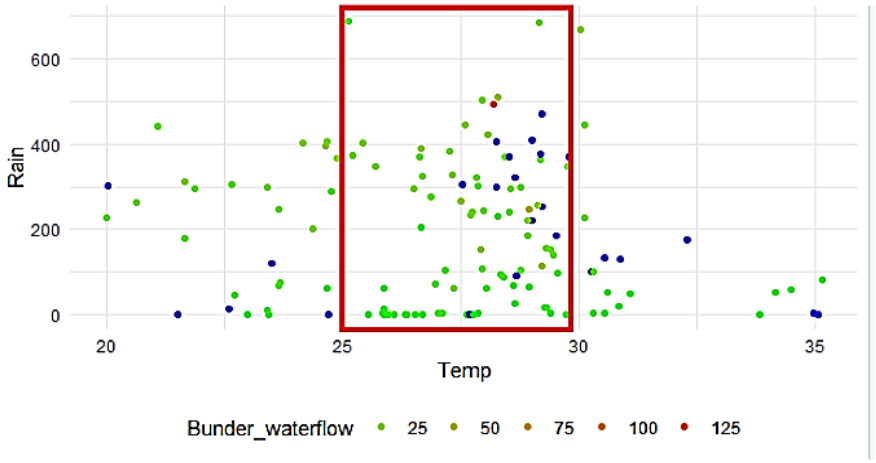
B. Implementasi Imputasi Nilai Hilang

1) Imputasi MICE: Proses pengisian nilai hilang menggunakan metode MICE dilakukan menggunakan software Rstudio dengan *package library* MICE. Terdapat empat langkah dalam implementasi pengisian nilai hilang menggunakan MICE, yaitu persiapan *dataset*, pengisian nilai hilang, *pooling*, dan validasi.

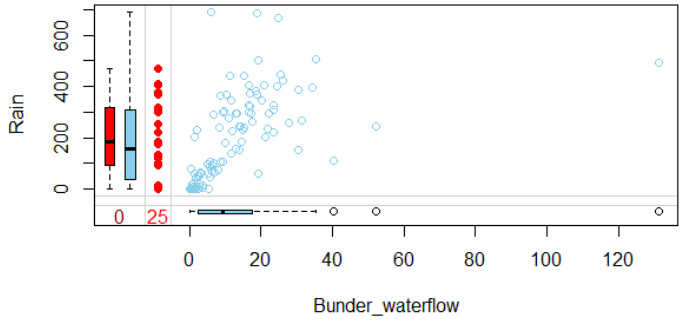
Persiapan *dataset* diawali dengan impor *dataset Bunder_waterflow* yang telah dibagi menjadi empat *dataset* yang telah diisi dengan nilai NA pada proses persiapan *dataset*. Proses selanjutnya adalah melihat pola (*pattern*) dari nilai hilang menggunakan *library(mice)* dengan *package margin plot*. Gbr. 6 menunjukkan sebaran nilai yang hilang dari dua variable *Bunder_waterflow* dan variabel *Rain*, dengan asumsi nilai pada *plot* kotak merah di sebelah kiri menunjukkan distribusi variabel *Rain* dengan variabel *Bunder_waterflow* yang hilang. Sementara itu, plot kotak biru menunjukkan distribusi titik data yang tersisa antara dua variabel. Asumsi *plot* ini berdasarkan model MCAR.

Pengisian nilai hilang dalam metode MICE menggunakan nilai *m* dengan nilai *default m* = 5, yang mengacu pada jumlah *dataset* yang diperhitungkan. Pertimbangan penggunaan nilai *m* = 5 terdapat pada variabel dan jumlah poin *dataset* yang tidak terlalu banyak, yang hanya berisi 132 objek dari lima variabel. Metode imputasi yang dilakukan mengacu pada *Predictive Mean Matching* (PMM) dari bagian *library* MICE yang tersedia, seperti yang ditunjukkan pada *source code* berikut ini.

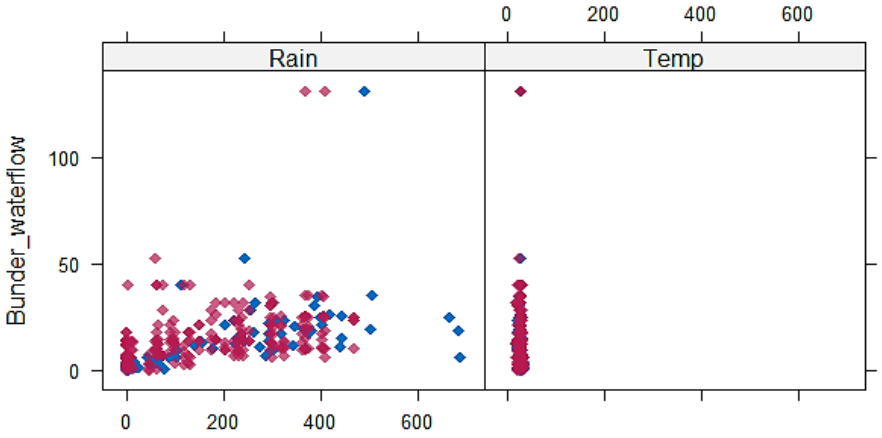
```
tempData < mice(data,m=5,maxit=50,
meth='pmm',seed=500)
summary(tempData)
```



Gbr. 5 GGPLOT k-NNi.



Gbr. 6 Margin *plot* MICE.



Gbr. 7 Hasil *xyplot* sebagai inspeksi distribusi original dan data terimputasi.

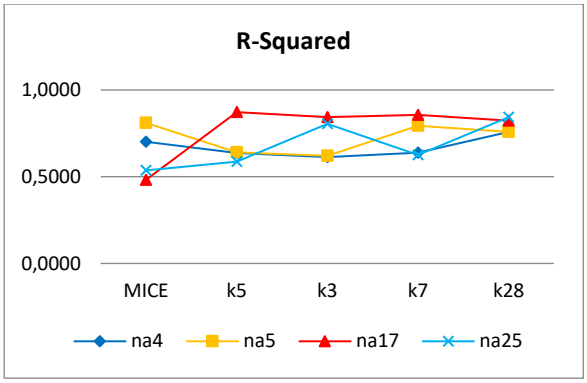
Proses persiapan *dataset* selanjutnya adalah pemeriksaan distribusi data asli dengan data imputasi menggunakan *xyplot*, seperti pada Gbr. 7. Perbandingan ini dapat memberikan gambaran bahwa titik magenta atau yang diperhitungkan memiliki kecocokan dengan titik biru, sehingga dapat dipastikan bahwa nilai yang diperhitungkan adalah nilai yang masuk akal (*plausible value*).

Proses *pooling* merupakan proses terakhir sebelum validasi. Proses ini berfokus pada penyesuaian model linier dengan data yang berisi hasil penyesuaian pada *dataset* yang diperhitungkan dengan fungsi *pool()* untuk mengumpulkan semua penyesuaian data.

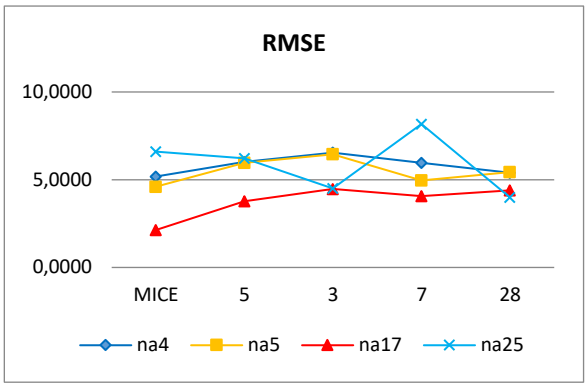
```
modelFit1 <- with(tempData,lm(Bunder_waterflow~
Rain+Temp))
summary(pool(modelFit1))

tempData2 <- mice(data,m=50,seed=245435)
modelFit2<- with(tempData2,lm(Bunder_waterflow~
Rain+Temp))
summary(pool(modelFit2))
```

2) k-Nearest Neighbors Imputation (k-NNi): Pada eksperimen k-NNi, proses pengisian nilai hilang dilakukan menggunakan *dataset* yang telah dikenai pengisian nilai NA kolom *Bunder_waterflow* pada proses persiapan *dataset*.



Gbr. 8 *R-Squared*.



Gbr. 9 RMSE.

TABEL V
R-*SQUARED*

R-Squared					
	MICE	***k* = 5**	***k* = 3**	***k* = 7**	***k* = 28**
NA4	70%	64%	61%	64%	76%
NA5	81%	64%	62%	79%	76%
NA17	48%	87%	84%	86%	82%
NA25	54%	59%	80%	63%	84%
AVG	63%	68%	72%	73%	80%
MIN	48%	59%	61%	63%	76%
MAX	81%	87%	84%	86%	84%

TABEL VI
RMSE

RMSE					
	MICE	***k* = 5**	***k* = 3**	***k* = 7**	***k* = 28**
NA4	5,1821	6,0157	6,5407	5,9661	5,4011
NA5	4,6029	5,9575	6,4457	4,9650	5,4417
NA17	2,1337	3,7764	4,4743	4,0721	4,3934
NA25	6,6053	6,2122	4,5113	8,1720	3,9910
AVG	4,6310	5,4904	5,4930	5,7938	4,8068
MIN	2,1337	3,7764	4,4743	4,0721	3,9910
MAX	6,6053	6,2122	6,5407	8,1720	5,4417

Makalah ini menggunakan parameter k = 3,5,7,28. Tidak ada metode khusus dalam menentukan nilai k untuk metode k-NNi [24]. Dengan demikian, tingginya resistansi terhadap mekanisme dan model nilai yang hilang menjadikan metode k-NNi sebagai salah satu pendekatan penanganan nilai yang hilang.

Hambatan ini juga dilihat dengan imputasi menggunakan parameter k yang berbeda. Jika nilai k terlalu kecil, akan terjadi banyak *noise* dan bias yang mengurangi tingkat akurasi dalam imputasi; dan apabila terlalu besar, dapat terjadi kesalahan dalam membatasi nilai yang diambil dan secara tidak langsung memengaruhi keakuratan [23]. Namun, terdapat metode *machine learning* yang mampu melakukan identifikasi penentuan nilai k dengan salah satu *library* yang ada di pemrograman R.

Selanjutnya, diimpor *dataset* Bunder_waterflow yang telah dibagi menjadi empat *dataset*. Keempat dataset kemudian diisi secara otomatis menggunakan *library* pemrograman R dengan nilai NA pada proses persiapan *dataset*. Proses selanjutnya adalah pengisian nilai hilang menggunakan nilai parameter k = 3, k = 5, k = 7, dan k = 28. Berikut ini merupakan kode pemrograman R dengan metode k-NNi.

```
Bunder3<-k-
NN(s_bunder,variable=c("Bunder_waterflow"),k=3)

Bunder5<-k-
NN(s_bunder,variable=c("Bunder_waterflow"),k=5)

Bunder7<-k-
NN(s_bunder,variable=c("Bunder_waterflow"),k=7)

Bunder28<-k-
NN(s_bunder,variable=c("Bunder_waterflow"),k=28)
```

C. *Statistical Validation*

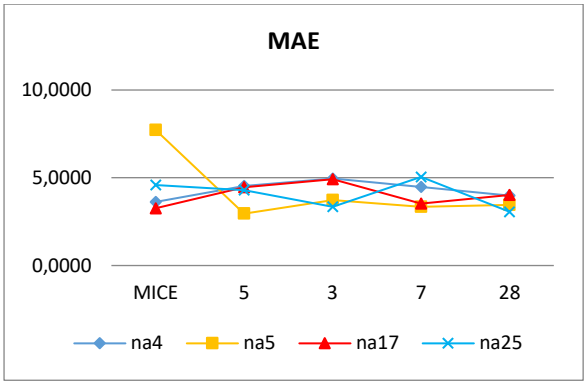
Dalam proses validasi, kelima hasil pada dua metode disajikan dalam bentuk grafik sebagai bahan perbandingan. Hasil *R-Squared* pada Gbr. 8 dan Tabel V menunjukkan bahwa metode k-NNi dengan nilai parameter k = 28 memiliki nilai yang lebih konsisten pada rentang nilai yang berdekatan antara NA4 hingga NA25, yaitu pada 76%-84%, dibanding keempat metode lainnya. Hasil metode imputasi MICE mendapatkan nilai cukup rendah pada rentang nilai hilang 17-25 dengan hasil 48% dan 54%.

Nilai RMSE dengan konsep regresi linier menunjukkan bahwa variasi nilai yang dihasilkan mendekati variasi nilai observasinya. Pada Gbr. 9 dan Tabel VI, nilai RMSE MICE berada pada 6,605, yaitu nilai tertinggi kedua, dan dianggap masih belum memiliki keakuratan dengan variasi nilai observasi. Metode k-NNi memiliki nilai akurasi RMSE yang lebih stabil dibanding metode MICE. Nilai RMSE MICE dengan jumlah tujuh belas nilai hilang mendapatkan nilai terendah, yaitu sebesar 2,134, yang dianggap sangat mendekati nilai sebenarnya.

Nilai MAE dari kelima metode yang dilakukan uji validasi statistik dengan rata-rata nilai *error* terendah diperoleh menggunakan metode k-NNi dengan k = 28. Pada Gbr. 10 dan Tabel VII tampak bahwa nilai k = 28 memiliki kesamaan hasil pada setiap nilai hilang. Hal ini menunjukkan bahwa k = 28 menghasilkan selisih nilai imputasi yang mendekati nilai yang hilang.

VI. KESIMPULAN

Menggunakan metode imputasi dengan MICE dan k-NNi, pengisian nilai hilang dapat dilakukan dengan baik pada setiap



Gbr. 10 MAE.

TABEL VII
MAE

MAE					
	MICE	***k* = 5**	***k* = 3**	***k* = 7**	***k* = 28**
NA4	3,6233	4,5277	4,9585	4,4766	3,9815
NA5	7,7253	2,9549	3,7377	3,3535	3,4615
NA17	3,2636	4,4551	4,9173	3,5324	4,0264
NA25	4,5924	4,2854	3,3388	5,0502	3,0615
AVG	4,8011	4,0558	4,2381	4,1032	3,6327
MIN	3,2636	2,9549	3,3388	3,3535	3,0615
MAX	7,7253	4,5277	4,9585	5,0502	4,0264

metode. Setelah dilakukan validasi, ditemukan bahwa nilai rata-rata RMSE dan MAE yang paling konsisten diperoleh pada metode k-NNi dengan nilai k = 28. Sementara itu, nilai *R-Squared* metode k-NNi dengan nilai k = 28 mendapatkan nilai rata-rata persentase terbaik, yaitu pada nilai 80%, disusul dengan metode k-NNi k = 7 sebagai nilai k *default*, dengan persentase 73%. Metode pembanding MICE mendapatkan rata-rata nilai persentase paling rendah daripada metode lain, dengan hanya mendapatkan nilai 63%. Dari hasil tersebut, metode MICE dianggap kurang sesuai untuk melakukan proses imputasi dengan menggunakan *dataset* DAS Sungai Opak yang berada di Provinsi DIY.

Penentuan nilai parameter k pada metode k-NNi menggunakan pertimbangan variabel tetangga dengan mekanisme GGPLOT pada pemrograman R menjadi tahap yang penting dalam proses imputasi nilai hilang. Penentuan parameter k yang lebih tepat dapat meminimalkan risiko tingginya resistansi terhadap mekanisme dan model nilai yang hilang.

Hasil pengujian yang telah dilakukan pada model imputasi nilai hilang akan sangat membantu penelitian selanjutnya. Hasil imputasi nilai hilang menggunakan metode k-NNi dapat memberikan akurasi yang cukup baik dalam memecahkan masalah nilai hilang. Beberapa keilmuan multidisiplin seperti mekanika fluida dalam keilmuan fisika, limpasan air sungai dalam keilmuan hidrologi geografi, dan dimensional sungai dalam keilmuan teknik sipil dapat dikombinasikan dengan metode *machine learning* seperti k-NNi. Hasil kombinasi tersebut nantinya dapat dilakukan pada penelitian selanjutnya, salah satunya adalah prediksi banjir dan klasifikasi tata kelola ruang dengan gabungan algoritme multidisiplin ilmu.

REFERENSI

[1] S. Kamwaga, D.M.M. Mulungu, dan P. Valimba, "Assessment of Empirical and Regression Methods for Infilling Missing Streamflow Data in Little Ruaha Catchment Tanzania," *Phys. Chem. Earth*, Vol. 106, hal. 17–28, 2018.

[2] R.J. Abrahart, F. Anctil, P. Coulibaly, C.W. Dawson, N.J. Mount, L.M. See, A.Y. Shamseldin, D.P. Solomatine, E. Toth, dan R.L. Wilby, "Two decades of Anarchy? Emerging Themes and Outstanding Challenges for Neural Network River Forecasting," *Prog. Phys. Geogr.*, Vol. 36, No. 4, hal. 480–513, 2012.

[3] E. Acuña dan C. Rodriguez, "The Treatment of Missing Values and its Effect on Classifier Accuracy," dalam *Classification, Clustering, and Data Mining Applications*, D. Banks, F.R. McMorris, P. Arabie, dan W. Gaul, Eds., Berlin, Jerman: Springer, 2004, hal. 639-647.

[4] J. Luengo, S. García, dan F. Herrera, "A Study on the Use of Imputation Methods for Experimentation with Radial Basis Function Network Classifiers Handling Missing Attribute Values: The Good Synergy between RBFNs and EventCovering Method," *Neural Networks*, Vol. 23, No. 3, hal. 406–418, Apr. 2010.

[5] L. Sunitha, M. Balraju, dan J. Sasikiran, "Data Mining: Estimation of Missing Values Using Lagrange Interpolation Technique," *Int. J. Adv. Res. Comput. Eng. Technol.*, Vol. 2, No. 4, hal. 1579–1582, 2013.

[6] W.M. Campion dan D.B. Rubin, "Multiple Imputation for Nonresponse in Surveys," *J. Mark. Res.*, Vol. 26, No. 4, hal. 485-486, 1989.

[7] A. Jadhav, D. Pramod, dan K. Ramanathan, "Comparison of Performance of Data Imputation Methods for Numeric Dataset," *Appl. Artif. Intell.*, Vol. 33, No. 10, hal. 913–933, 2019.

[8] C. Curley, R.M. Krause, R. Feiock, dan C.V. Hawkins, "Dealing with Missing Data: A Comparative Exploration of Approaches Using the Integrated City Sustainability Database," *Urban Aff. Rev.*, Vol. 55, No. 2, hal. 591–615, 2019.

[9] R.J. Little, "Selection Model (Missing Data)," dalam *Wiley StatsRef Stat. Ref. Online*, Hoboken, AS: Wiley, 2016, hal. 1–5.

[10] D.B. Rubin, "Inference and Missing Data," *Biometrika*, Vol. 63, No. 3, hal. 581–592, 1976.

[11] R.J.A. Little, "Missing-Data Adjustments in Large Surveys," *J. Bus. Econ. Stat.*, Vol. 6, No. 3, hal. 287–296, 1988.

[12] J.L. Schafer dan J.W. Graham, "Missing Data: Our View of the State of the Art," *Psychol. Methods*, Vol. 7, No. 2, hal. 147–177, 2002.

[13] P. Schmitt, J. Mandel, dan M. Guedj, "A Comparison of Six Methods for Missing Data Imputation," *J. Biom. Biostat.*, Vol. 6, No. 1, hal. 1–6, 2015.

[14] G. Chhabra, V. Vashisht, dan J. Ranjan, "A Review on Missing Data Value Estimation Using Imputation Algorithm," *J. Adv. Res. Dyn. Control Syst.*, Vol. 11, No. 7-Special Issue, hal. 312–318, 2019.

[15] A. Kowarik dan M. Templ, "Imputation with the R package VIM," *J. Stat. Softw.*, Vol. 74, No. 7, hal. 1-16, 2016.

[16] C. Cortes, L.D., Jackel, dan W-P. Chiang, "Limits in Learning Machine Accuracy Imposed by Data Quality," *Proc. the 1st Int. Conf. Knowl. Discovery Data Mining (KDD-95 Proc.)*, 1994, hal. 57–62.

[17] J.M. Engels dan P. Diehr, "Imputation of Missing Longitudinal Data: A Comparison of Methods," *J. Clin. Epidemiol.*, Vol. 56, No. 10, hal. 968–976, 2003.

[18] D.M.P. Murti, U. Pujianto, A.P. Wibawa, dan M.I. Akbar, "K-Nearest Neighbor (K-NN) based Missing Data Imputation," *Proc. - 2019 5th Int. Conf. Sci. Inf. Technol. Embrac. Ind. 4.0 Towar. Innov. Cyber Phys. Syst. (ICSITech 2019)*, 2019, hal. 83–88.

[19] J. Maillo, S. Ramírez, I. Triguero, dan F. Herrera, "kNN-IS: An Iterative Spark-based Design of the k-Nearest Neighbors Classifier for Big Data," *Knowledge-Based Syst.*, Vol. 117, No. C, hal. 3–15, 2017.

[20] G.E.A.P.A. Batista dan M.C. Monard, "A Study of k-Nearest Neighbour as an Imputation Method," *Conf. Soft Comput. Sys. - Design, Manag.*

Appl. (HIS 2002), 2002, hal. 1–10.

[21] B. Suthar, H. Patel, dan A. Goswami, "A Survey: Classification of Imputation Methods in Data Mining," *Int. J. Emerg. Technol. Adv. Eng.*, Vol. 2, No. 1, hal. 309–312, 2012.

[22] D. Priya, R. dan Sivaraj, R., "A Review of Missing Data Handling Methods," *Int. J. Eng. Technol. Sci.*, Vol. 2, No. 2, hal. 2349–3968, 2015.

[23] Doreswamy, I. Gad, dan B.R. Manjunatha, "Performance Evaluation of Predictive Models for Missing Data Imputation in Weather Data," *2017 Int. Conf. Adv. Comput. Commun. Informatics (ICACCI 2017)*, 2017, hal. 1327–1334.

[24] Y. Sun, A.K.C. Wong, dan M.S. Kamel, "Classification of Imbalanced Data: A Review," *Int. J. Pattern Recognit. Artif. Intell.*, Vol. 23, No. 4, hal. 687–719, 2009.

[25] M.J. Azur, E.A. Stuart, C. Frangakis, dan P.J. Leaf, "Multiple Imputation by Chained Equations: What Is It and How Does It Work?" *Int. J. Methods Psychiatr. Res.*, Vol. 20, No. 1, hal. 40–49, 2011.

[26] S. van Buuren dan K. Groothuis-Oudshoorn, "MICE: Multivariate Imputation by Chained Equations in R," *J. Stat. Softw.*, Vol. 45, No. 3, hal. 1–67, 2011.

[27] P.H. Rezvan, K.J. Lee, dan J.A. Simpson, "The Rise of Multiple Imputation: A Review of the Reporting and Implementation of the Method in Medical Research Data Collection, Quality, and Reporting," *BMC Med. Res. Methodol.*, Vol. 15, No. 1, hal. 1–14, 2015.

[28] G. Chhabra, V. Vashisht, dan J. Ranjan, "A Comparison of Multiple Imputation Methods for Data with Missing Values," *Indian J. Sci. Technol.*, Vol. 10, No. 19, hal. 1–7, 2017.

[29] D.F. Hamilton, M. Ghert, dan A.H.R.W. Simpson, "Interpreting Regression Models in Clinical Outcome Studies," *Bone Jt. Res.*, Vol. 4, No. 9, hal. 152–153, 2015.

[30] T. Chai dan R.R. Draxler, "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? -Arguments Against Avoiding RMSE in the Literature," *Geosci. Model Dev.*, Vol. 7, No. 3, hal. 1247–1250, 2014.

[31] A.A. Suryanto, "Penerapan Metode Mean Absolute Error (MEA) dalam Algoritma Regresi Linear untuk Prediksi Produksi Padi," *Saintekbu*, Vol. 11, No. 1, hal. 78–83, 2019.

[32] W. Wang dan Y. Lu, "Analysis of the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) in Assessing Rounding Model," *IOP Conf. Ser. Mater. Sci. Eng.*, Vol. 324, hal. 1-10, 2018.

[33] M.J. Hartmann dan G. Carleo, "Neural-Network Approach to Dissipative Quantum Many-Body Dynamics," *Phys. Rev. Lett.*, Vol. 122, No. 25, Art. 250502, 2019.

[34] K. Crammer, "On the Algorithmic Implementation of Multiclass Kernel-based Vector Machines," *J. Mach. Learn. Res. (JMLR)*, Vol. 2, No. 2, hal. 265–292, 2002.

[35] X. Wu, V. Kumar, J.R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S. Yu, Z.-H. Zhou, M. Steinbach, D.J. Hand, dan D. Steinberg, "Top 10 Algorithms in Data Mining," *Knowl. Inf. Syst.*, Vol. 14, hal. 1-37, 2008.

[36] J. Shao, "Linear Model Selection by Cross-Validation," *J. Am. Stat. Assoc.*, Vol. 88, No. 422, hal. 486–494, 1993.

[37] Suprapto, dkk., *Katalog Basis Data 2014 Sumber Daya Air*, Jakarta, Indonesia: Pusat Penelitian dan Pengembangan Sumber Daya Air, 2014.

[38] S.L. Dingman, *Physical Hydrology*, 3rd ed., Illinois, AS: Waveland Press, 2015.

[39] S.J. Goldman, T.A. Bursztynsky, and K. Jackson, *Erosion and Sediment Control Handbook*, 1st ed., New York, AS: McGraw-Hill, 1986.