PERBANDINGAN IMPUTASI DAN PARAMETER SUPPORT VECTOR REGRESSION UNTUK PERAMALAN CUACA

Arif Mudi Priyatno
Fakultas Teknik Informatika, Program Studi Teknik Informatika
Institut Teknologi Sepuluh November
Email: arif.18051@mhs.its.ac.id

Agung Wiratmo
Fakultas Teknik Informatika, Program Studi Teknik Informatika
Institut Teknologi Sepuluh November
Email: agung.18051@mhs.its.ac.id

Fahmi Syuhada
Fakultas Teknik Informatika, Program Studi Teknik Informatika
Institut Teknologi Sepuluh November
Email: fahmi.18051@mhs.its.ac.id

Putri Cholidhazia
Fakultas Matematika Dan Ilmu Pengetahuan Alam Departement Ilmu Komputer
Bogor Agricultural University
Email: Putri_cho@apps.ipb.ac.id

ABSTRAK

Curah hujan adalah informasi penting di bidang transportasi, pertanian, industri dll. Dengan mengetahui informasi curah hujan, tindakan dapat diambil secara tepat di beberapa bidang tersebut. sehingga tidak ada kerugian karena kesalahan dalam informasi curah hujan. Makalah ini bertujuan untuk menemukan metode yang sesuai dalam peramalan curah hujan yang terkait dengan metode pemrosesan data imputasi dan nilai parameter dalam *Support Vector Regression* (SVR). Hasil percobaan menunjukkan bahwa metode preprocessing data imputasi terbaik diperoleh untuk digunakan ke dalam SVR berdasarkan nilai *Mean Squared Error* (MSE) dan *Mean Absolute Error* (MAE). Berdasarkan hasil MSE, k-nearest neighbor adalah metode terbaik yang digunakan untuk preprocessing data imputasi. Data preprocessing menghasilkan eksperimen pada SVR Polinomial dengan parameter C 1000, toleransi 0,001, epsilon 0,01 dan iterasi tak terbatas. Di sisi lain, hasil MAE menunjukkan bahwa *Artificial Neural Network* (ANN) adalah metode terbaik dalam imputasi data preprocessing. ANN dengan *radial basis function* kernel, gamma 0,001, C 1000, toleransi 0,001 dan iterasi tanpa batas. JST diuji pada RBF SVR dengan gamma 0,001, C 1000, toleransi 0,001 dan iterasi tak terbatas.

Kata kunci: curah hujan; imputasi data; *support vector regression*; *k-nearest neighbor; radial basis function kernel; artificial neural network.*

ABSTRACT

Rainfall is important information in the fields of transportation, agriculture, industry etc. By knowing rainfall information, action can be taken precisely in some of these fields so that there is no loss due to errors in rainfall information. This paper aims to find a suitable method in forecasting rainfall related to the data imputation preprocessing method and parameter values in the Support Vector Regression (SVR). The experiment results showed that the best imputation data preprocessing method obtained for use into SVR based on Mean Squared Error (MSE) and Mean Absolute Error (MAE) values. Based on the results of MSE, k-nearest neighbor is the best method used for preprocessing imputation data. The preprocessing data results experiment on the Polynomial SVR with parameters C 1000, tolerance 0.001, epsilon 0.01 and infinite iterations. On the other hand, the MAE results show that Artificial Neural Network (ANN) is the best method in imputation preprocessing data. ANN with radial basis function kernel, gamma 0.001, C 1000, tolerance 0.001 and infinite iterations. ANN was tested on RBF SVR with gamma 0.001, C 1000, tolerance 0.001 and infinite iterations.

Keywords: *rainfall; data imputation; support vector regression; k-nearest neighbor; radial basis function kernel; artificial neural network.*

1 PENDAHULUAN

Ketidakpastian adalah masalah yang sering terjadi dalam banyak ilmu alam seperti meteorologi dan hidrologi. ini membuat prediksi keadaan alami seperti prediksi curah hujan menjadi layak untuk dipelajari oleh masing-masing ilmu alam [1]. Curah hujan adalah hal terbesar yang berkaitan dengan proses siklus air. Karakteristik curah hujan penting untuk memodelkan akurasi dan nilai kebetulan hidrologi lainnya seperti limpasan dan evapotranspirasi. Ini karena karakteristik curah hujan mengontrol perilaku dan evaluasi pemodelan [2] [3] [4].

Cuaca adalah kondisi udara di suatu tempat dalam waktu yang relatif singkat. Parameter yang mempengaruhi kondisi cuaca di suatu daerah adalah suhu, tekanan udara, kecepatan angin, kelembaban udara, dan berbagai fenomena atmosfer lainnya [5] [6] [7]. Pilihan metode yang tepat untuk menentukan kondisi cuaca adalah kegiatan yang sering dilakukan oleh beberapa peneliti cuaca atau atmosfer akhir-akhir ini. Ini karena ada banyak tuntutan dari berbagai pihak yang menginginkan informasi tentang kondisi atmosfer yang lebih cepat, lebih akurat, dan terperinci [8] [9].

Dalam hal perkiraan curah hujan, data diperoleh dari agensi yang menangani masalah cuaca. Badan meteorologi dan geofisika umumnya menyediakan data yang dibutuhkan dan dapat diakses melalui portal mereka. data menyajikan nilai parameter yang mempengaruhi curah hujan. Data yang diambil dari proses peristiwa langsung disebut sebagai data nyata. Curah hujan termasuk dalam data nyata yang memiliki ukuran besar. Namun, penggunaan data nyata akan mendapatkan anomali data. Data anomali biasanya disebabkan oleh berbagai faktor seperti faktor pendeteksi sensor hingga kesalahan manusia itu sendiri. Data anomali tentu akan mempengaruhi hasil peramalan curah hujan. Oleh karena itu, perbaikan pada data sebelum memasuki proses peramalan perlu dilakukan.

Makalah ini bertujuan untuk menganalisis perbandingan model imputasi dan pengaruh perubahan nilai parameter pada metode Support Vector Regression (SRV) untuk peramalan curah hujan. Data imputasi diperoleh dari proses preprocessing data menggunakan 3 buah algoritma yang diuji. Algoritma yang digunakan untuk menghasilkan data imputasi adalah algoritma K-Nearest Neighbors (KNN), Jaringan Syaraf Tiruan (JST), dan Binning. Sedangkan parameter SVR yang dianalisis adalah gamma, parameter C, toleransi, epsilon, dan jumlah iterasi. Model data imputasi dengan perubahan nilai parameter metode SVR dikombinasikan untuk mendapatkan algoritma preprocessing terbaik dan nilai parameter sebagai prakiraan curah hujan. Algoritma dan nilai parameter terbaik ditentukan berdasarkan nilai MSE dan MAE dari masing-masing teknik percobaan. Pengetahuan tentang algoritma preprocessing untuk mengatasi data anomali diperoleh dalam penelitian ini. Selain itu, pengetahuan diperoleh tentang pengaruh perubahan nilai parameter pada metode SVR dalam proses klasifikasi.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan data primer yang berasal dari Stasiun Meteorologi Sultan Syarif Kasim II di Kota Riau Pekanbaru dari 1 Januari 2000 hingga 13 Oktober 2018 [10]. Data memiliki 10 fitur, yaitu suhu minimum (° C), suhu maksimum (° C), suhu rata-rata (° C), kelembaban rata-rata (%), curah hujan (mm), waktu iradiasi (jam), kecepatan angin rata-rata (knot), arah angin tertinggi (deg), kecepatan angin terbesar (knot), arah angin pada kecepatan maksimum (deg). Data berisi nilai anomali karena ada data yang bernilai "8888" dan "9999". Nilai data "8888" berarti bahwa nilai parameter cuaca tidak diukur pada hari tertentu. sedangkan nilai "9999" berarti nilai data yang tidak direkam satu hari. Jumlah data yang diperoleh adalah 6851 data.

Penelitian ini mengimplementasikan beberapa metode untuk meningkatkan data anomali. proses ini merupakan preprocessing dari data yang digunakan. Langkah preprocessing menghasilkan 3 buah data imputasi berdasarkan algoritma yang digunakan. Algoritma yang diterapkan adalah Algoritma *k-nearest neighbor*, *Artificial Neural Network*, dan *Binning*. Setelah *preprocessing* selesai, setiap data imputasi dihasilkan untuk memasuki proses outlier. dalam proses ini, data dikoreksi dengan metode deviasi standar untuk menghilangkan data outlier. Tiga potong data imputasi yang telah diperbaiki kemudian dimasukkan ke dalam proses peramalan menggunakan metode Support Vector Regression.

2.1 Outlier Process (Imputation)

Outliers adalah pengamatan yang jauh (ekstrem) dari pengamatan lain. Secara umum, *outlier* dibagi menjadi dua, yaitu outlier dalam model observasi dan outlier dalam model linier. Berdasarkan jumlah variabel yang dipertimbangkan, Outliers dibagi menjadi Outliers dalam observasi univariat atau multivariat. Outliers dalam model linear multivariat dapat dibagi menjadi tiga kategori, yaitu *Outliers* untuk *leverage* dan *residual* atau keduanya [11] [12].

Deteksi *outlier* adalah proses umum dalam aplikasi penambangan data. Deteksi dilakukan untuk menghindari distribusi antara setiap data yang memiliki nilai yang sangat berbeda. Deteksi outlier dari data mining sering didasarkan pada pengukuran jarak, cluster, dan metode spasial [13].

Dalam deteksi outlier, data dalam pemrosesan data mining dapat dikategorikan menjadi data univariat dan multivariat [14] [15]. Data univariat adalah data yang memiliki satu jenis variabel.

Sedangkan data multivarian lebih dari satu tipe data. Dalam penelitian ini, data yang digunakan adalah data dengan model multivariat.

2.1.1 K-Nearest Neighbors Algorithm (KNN)

KNN termasuk dalam kategori *Lazy Learning*. Pendekatan metode KNN dilakukan dengan menyimpan semua data pelatihan dalam ruang n-dimensi. Saat menguji suatu data, itu akan menentukan tetangga K terdekat yang akan menjadi nilai dari data pengujian. Metode KNN digunakan untuk melakukan perbaikan pada data anomali. Algoritma menghasilkan data hasil imputasi yang siap untuk memasuki proses peramalan. Data curah hujan yang digunakan dalam penelitian ini dikarakterisasi menjadi dua berdasarkan pada kelengkapan nilai dari setiap parameter. Pertama, data untuk setiap baris memiliki nilai parameter lengkap. Tidak ada nilai yang hilang dari setiap parameter. Kedua, data untuk setiap baris memiliki nilai yang hilang pada suatu parameter. Dalam data yang digunakan, nilai yang hilang ditandai dengan nilai "8888" dan "9999". Data dengan karakteristik pertama menjadi data clean atau data pelatihan. Sedangkan data dengan karakteristik kedua menjadi data unclean atau data uji yang akan diperbaiki. Algoritma 1 adalah algoritma untuk mengatasi data yang tidak bersih.

Algoritma 1: Penggantian data mengandung missing value dengan metode KNN

***Input* :** data clean, data unclean
***Output* :** Data matrik hasil perbaikan data *missing value* dengan metode KNN
Step:

a. Pengecekan per baris data *unclean*.
b. Pengecekan jumlah missing value per baris.
   Jika jumlah missing value per baris lebih dari lima missing value maka baris tersebut dihapus dan diulang ke langkah ke 1. Jika tidak lanjut ke langkah berikutnya
c. Pengecekan data dalam fitur yang mengandung *missing value*. data dalam fitur yang tidak mengandung *missing value* akan menjadi dataset dan data dalam fitur yang mengandung *missing value* dijadikan sebagai data target.
d. Pelatihan KNN menggunakan data clean dengan ketentuan dataset dan data target sesuai dengan yang didapatkan pada langkah ke 3
e. Dilakukan proses prediksi terhadap baris data yang diujikan..
f. Hasil dari prediksi yang tidak mengandung missing value ditambahkan kedalam data clean dan dijadikan data baru.
g. Data baru disimpan pada matrik dan dilakukan kembali langkah 1 sampai semua data *unclean* diproses menjadi data *clean*.

2.1.2 Artificial Neural Network Algorithm (ANN)

Secara umum, proses perbaikan data menggunakan metode JST sama dengan metode KNN. tetapi dalam proses memperbaiki setiap baris data, metode ini membuat parameter network yang digunakan untuk menentukan nilai yang hilang dari data yang dikoreksi. Langkah 4 dan 5 menjelaskan bagaimana proses perbaikan data dengan metode JST menghasilkan data imputasi.

Algoritma 2: Penggantian data mengandung missing value dengan metode *Artificial neural network*

Input : data clean, data unclean
Output : Data matrik hasil perbaikan data *missing value* dengan metode *Artificial neural network*
Step:

a. Pengecekan jumlah *missing value* per baris.
b. Jika jumlah missing value per baris lebih dari lima missing value maka baris tersebut dihapus dan diulang ke langkah ke 1. Jika tidak lanjut ke langkah berikutnya
c. Pengecekan data dalam fitur yang mengandung missing value. data dalam fitur yang tidak mengandung missing value akan menjadi dataset dan data dalam fitur yang mengandung missing value dijadikan sebagai data target.
d. Pelatihan ANN menggunakan data clean dengan ketentuan dataset dan data target sesuai dengan yang didapatkan pada langkah ke 3
e. Dilakukan proses prediksi terhadap baris data yang diujikan..
f. Hasil dari prediksi yang tidak mengandung missing value ditambahkan kedalam data clean dan

dijadikan data baru.

g. Data baru disimpan pada matrik dan dilakukan kembali langkah 1 sampai semua data unclean diproses menjadi data clean.

2.1.3 Binning Algorithm

Metode binning diusulkan untuk memperbaiki data dengan mengurutkan nilai tetangga dari nilai itu. Penyortiran data didistribusikan ke beberapa 'bin'. Karena distribusi data dalam metode binning mempertimbangkan nilai tetangga, nilai yang dihasilkan dalam metode binning ini adalah nilai rata-rata bin. Perhitungan dalam metode binning yang biasa digunakan adalah untuk menghitung rata-rata di setiap "bin" atau menghitung median di setiap "bin".

Metode binning dalam penelitian ini digunakan sebagai metode ketiga untuk meningkatkan parameter nilai yang hilang. Perhitungan dalam metode binning digunakan dengan menghitung nilai tengah. Langkah 2 adalah algoritma untuk meningkatkan data dengan metode ini.

Algoritma 3: Penggantian data mengandung missing value dengan metode binning

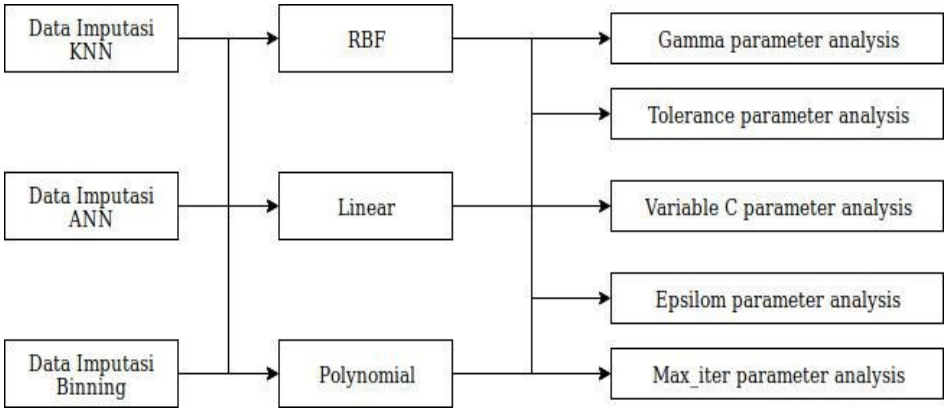
Input : Seluruh Data
Output : Data hasil perbaikan dengan metode binning
Step:

a. Pengecekan jumlah missing value per baris. Jika jumlah missing value per baris lebih dari lima missing value maka baris tersebut dihapus dan diulang ke langkah ke 1. Jika tidak lanjut ke langkah berikutnya
b. Pengecekan data dalam fitur yang mengandung missing value setiap baris akan dilakukan perhitungan nilai tengah dalam fitur tersebut dapat mengganti missing value.
c. Langkah kedua diulangi sebanyak baris data lolos pada langkah 1 dan setiap baris diulangi sebanyak fitur.
d. Data perbaikan dari missing value disimpan dalam data clean data

2.2 Support Vector Regression

Support Vector Regression (SVR) adalah salah satu metode untuk menyelesaikan masalah dalam regresi. SVR digunakan untuk menemukan fungsi g(x) yang cocok untuk seluruh data di mana nilai e mendekati nilai yang disepakati atau mencapai iterasi yang ditentukan. Nilai e menunjukkan keakuratan perhitungan SVR.

SVR terdiri dari sejumlah parameter, yaitu kernel, derajat, gamma, coef0, toleransi, parameter, epsilon, menyusut, cache_size, verbose, dan iterasi maksimum. Namun, pengujian model data imputasi dalam penelitian ini tidak menggunakan semua parameter SVR. Parameter yang digunakan adalah kernel, gamma, toleransi, epsilon, variabel C, dan iterasi maksimum. Kernel SVR yang digunakan untuk pengujian adalah kernel RBF, Linear, dan Polinomial. sedangkan untuk nilai parameter lainnya, yaitu dengan nilai yang ditentukan. Tabel 1 adalah deskripsi parameter beserta nilainya untuk menguji data imputasi.



Gambar 1. Model Pengujian Data Imputasi Menggunakan Metode SVR

Tabel 1. Parameter uji SVR dan nilainya

Kernel	***RBF, Linear, Polynomial***				
Gamma	1	10	100	1000	10000
tolerance	0.1	0.01	0.001	0.0001	0.00001
Variable C	1	1.00E+02	1.00E+03	1.00E+04	1.00E+05
Epsilon	0.1	0.01	0.001	0.0001	0.00001
Maximum Iteration	50	100	500	1000	-1

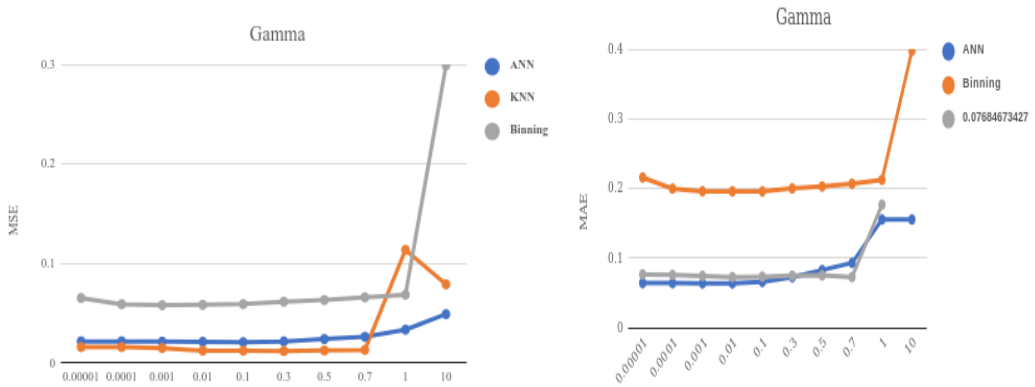
3. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil penelitian komparatif pada model data imputasi dan perubahan nilai parameter untuk prakiraan curah hujan. Hasil pengujian data imputasi yang diklasifikasikan oleh metode SVR ditampilkan dalam bentuk tabel hasil pengujian. Tingkat Optimalitas rekayasa pengujian ditampilkan dalam bentuk dua parameter nilai, yaitu parameter MSE dan MAE.

Gambar 1 menjelaskan bagaimana proses pengujian data imputasi diperoleh dari 3 metode preprocessing. Hasil pengujian akan dijelaskan dalam sub-bagian analisis parameter. Ada lima analisis sub-bagian parameter seperti yang ditunjukkan pada Gambar 1. Analisis ini mendapatkan kesimpulan bahwa model peningkatan data yang baik digunakan, nilai kernel, dan parameter optimal diberikan.

3.1 Gamma Parameter Analysis

Parameter gamma adalah kebalikan dari standar deviasi kernel RBF (fungsi Gaussian), yang digunakan sebagai ukuran kesamaan antara dua titik. Dalam SVM dengan kernel RBF, gamma digunakan untuk mempercepat fungsi-fungsi dalam kernel RBF untuk mendapatkan akurasi model klasifikasi yang optimal. Gambar 2 adalah grafik hasil MSE dan MAE dari pengujian data imputasi. Hasilnya diilustrasikan dalam bentuk grafik karena ada 3 model data imputasi yang diuji. Parameter gamma hanya dimiliki oleh kernel RBF. Oleh karena itu, setiap model data imputasi memiliki garis nilai grafik tunggal. Nilai MSE terkecil 0,012 dalam data imputasi KNN yang terletak parameter gamma dengan nilai 0,3 kernel. sedangkan nilai MAE adalah 0,064 pada data imputasi JST yang dipecah pada nilai parameter gamma 0,001.



Gambar 2. Grafik MSE dan MAE Perubahan Nilai Parameter Gamma

3.2 Tolerance Parameter Analysis

Toleransi adalah untuk melihat nilai toleransi terendah dalam proses SVR yang ditetapkan. Secara umum, semakin kecil toleransi yang diberikan, semakin baik nilai yang diperoleh. Tabel 2 menunjukkan bahwa jika nilai toleransi terkecil telah ditemukan, yang berikutnya hanya akan mengikuti toleransi sebelumnya dan perubahan tidak akan berubah secara signifikan. Nilai MSE terendah dihasilkan dari kernel polinomial dengan metode praproses KNN. Nilai MSE adalah 0,013. sedangkan MAE minimum yang dihasilkan adalah 0,07585 dalam model data imputasi KNN dengan Kernel Polinomial dan toleransi 0,01.

Tabel 2. Hasil MSE dan MAE parameter toleransi

MSE Results Tolerance Parameter Analysis						
Data Imputasi	**0.1**	**0.01**	**0.001**	**0.0001**	**0.00001**	**MIN**
ANN Linear	0.02190	0.02186	0.02177	0.02177	0.02178	0.02177
ANN RBF	0.02029	0.02017	0.02016	0.02016	0.02016	0.02016
ANN Poly	0.02184	0.02122	0.02119	0.02119	0.02119	0.02119
KNN Linear	0.01387	0.01561	0.01566	0.01566	0.01563	0.01387
KNN RBF	0.01973	0.07808	0.01507	0.01507	0.01565	0.01507
KNN Poly	0.01267	0.01279	0.01277	0.01277	0.01277	**0.01267**
Binning Linear	0.05840	0.05841	0.05829	0.05829	0.05829	0.05828
Binning RBF	0.22443	0.23898	0.24025	0.24025	0.24024	0.22443
Binning Poly	0.05901	0.05925	0.05921	0.05921	0.05921	0.05901
MAE Results Tolerance Parameter Analysis						
ANN Linear	0.06477	0.06441	0.06443	0.06442	0.06441	0.06441
ANN RBF	0.09651	0.09271	0.09270	0.09270	0.09270	0.09270
ANN Poly	0.06981	0.06543	0.06511	0.06510	0.06511	0.06510
KNN Linear	0.08358	0.07795	0.07793	0.07790	0.07792	0.07790
KNN RBF	0.10068	0.17573	0.10068	0.08766	0.08957	0.08766
KNN Poly	0.07683	0.07585	0.07597	0.07599	0.07599	**0.07585**
Binning Linear	0.19651	0.19605	0.19577	0.19575	0.19576	0.19575
Binning RBF	0.33112	0.33651	0.33709	0.33713	0.33713	0.33112
Binning Poly	0.19512	0.19600	0.19592	0.19588	0.19589	0.19512

3.3 Variable C Parameters Analysis

Nilai parameter C menunjukkan seberapa besar optimasi SVR dalam menghindari kesalahan peramalan di setiap pelatihan. Semakin besar nilai parameter C akan menghasilkan semakin lebar jarak tepi hyperplane sehingga jika parameter C kecil maka akan mempersempit jarak tepi hyperplane. Ada perubahan yang sangat signifikan dalam perubahan nilai parameter C di Binning, RBF, dan Polinomial. Hasil MSE terkecil berada di metode KNN dengan kernel polinomial pada nilai parameter C1e + 5 dengan hasil MSE 0,0126. Sedangkan untuk hasil MAE terkecil adalah MAE dalam metode JST dengan kernel polinomial pada nilai parameter C 1 dengan hasil MAE 0,0640

Tabel 3. Hasil MSE dan MAE parameter C

MSE Results Variable C Parameter Analysis						
Data Imputasi	1	1.00E+02	1.00E+03	1.00E+04	1.00E+05	**MIN**
ANN Linear	0.0219	0.0218	0.0218	0.0201	0.0189	0.0189
ANN RBF	0.0202	0.0202	0.0202	0.0202	0.0202	0.0202
ANN Poly	0.0216	0.0213	0.0212	0.0212	0.0212	0.0212
KNN Linear	0.0162	0.0162	0.0156	0.0137	0.0230	0.0137
KNN RBF	0.0161	0.0197	0.0197	0.0197	0.0197	0.0161
KNN Poly	0.0133	0.0128	0.0128	0.0127	0.0126	**0.0126**
Binning Linear	0.0584	0.0583	0.0583	0.0583	0.0686	0.0583
Binning RBF	0.0822	0.1441	0.2403	0.3444	0.5171	0.0822
Binning Poly	0.0605	0.0587	0.0592	0.0600	0.0605	0.0587
MAE Results Tolerance Parameter Analysis						
ANN Linear	0.0643	0.0644	0.0644	0.0670	0.0760	0.0643
ANN RBF	0.0927	0.0927	0.0927	0.0927	0.0927	0.0927
ANN Poly	0.0639	0.0646	0.0651	0.0653	0.0651	**0.0639**
KNN Linear	0.0768	0.0768	0.0779	0.0856	0.1207	0.0768
KNN RBF	0.0883	0.1007	0.1007	0.1007	0.1007	0.0883
KNN Poly	0.0756	0.0759	0.0760	0.0761	0.0790	0.0756
Binning Linear	0.1965	0.1959	0.1958	0.1963	0.2185	0.1958
Binning RBF	0.2314	0.2893	0.3371	0.3780	0.4163	0.2314
Binning Poly	0.2034	0.1955	0.1959	0.1970	0.1979	0.1955

3.4 Epsilon Parameter Analysis

Epsilon dalam ϵ-SVR adalah parameter yang sangat mudah dipahami. Ini menunjukkan berapa banyak rentang kesalahan yang diizinkan perdata pelatihan acara. Jadi, kisaran biasanya 0 hingga *MAX_ALLOWABLE _ERROR*. Dalam penelitian ini, efek perubahan epsilon pada hasil uji SVM RBF diuji. Rentang epsilon yang diuji adalah dari 0,1, 0,001, 0,0001, 0,00001. Dari hasil tes, semakin kecil rentang epsilon yang diuji, semakin rendah tingkat MSE dan MAE. Tabel 4 menunjukkan bahwa hasil MSE terkecil dalam metode KNN dengan kernel polinomial pada nilai parameter epsilon 0,01 dengan hasil MSE 0,0126. Tabel 4 menunjukkan bahwa hasil MAE terkecil dalam metode JST dengan kernel linier adalah nilai parameter epsilon 0,0001 dengan hasil MAE 0,0064.

Tabel 4. Hasil MSE dan MAE parameter epsilon

MSE Results Tolerance Parameter Analysis						
Data Imputasi	**0.1**	**0.01**	**0.001**	**0.0001**	**0.00001**	**MIN**
ANN Linear	0.0201	0.0213	0.0218	0.0218	0.0218	0.0201
ANN RBF	0.0243	0.0202	0.0202	0.0202	0.0202	0.0202
ANN Poly	0.0202	0.0208	0.0212	0.0212	0.0212	0.0202
KNN Linear	0.0146	0.0138	0.0139	0.0139	0.0141	0.0138
KNN RBF	0.0135	0.0186	0.0197	0.0199	0.0199	0.0135
KNN Poly	0.0133	0.0126	0.0128	0.0128	0.0128	**0.0126**
Binning Linear	0.0581	0.0583	0.0583	0.0583	0.0583	0.0581
Binning RBF	0.1534	0.2313	0.2403	0.2411	0.2411	0.1534
Binning Poly	0.0587	0.0592	0.0592	0.0591	0.0591	0.0587
MAE Results Tolerance Parameter Analysis						
ANN Linear	0.1063	0.0650	0.0644	0.0644	0.0644	**0.0644**
ANN RBF	0.1297	0.0950	0.0927	0.0925	0.0925	0.0925
ANN Poly	0.1069	0.0665	0.0651	0.0651	0.0651	0.0651
KNN Linear	0.0936	0.0843	0.0836	0.0839	0.0837	0.0836
KNN RBF	0.0835	0.0972	0.1007	0.1011	0.1011	0.0835
KNN Poly	0.0871	0.0768	0.0760	0.0759	0.0759	0.0759
Binning Linear	0.1968	0.1957	0.1958	0.1957	0.1957	0.1957
Binning RBF	0.2937	0.3325	0.3371	0.3376	0.3376	0.2937
Binning Poly	0.1957	0.1956	0.1959	0.1958	0.1958	0.1956

3.5 Maximum Iteration Parameter Analysis

Tes iterasi ini adalah tes melihat keseluruhan sistem MSE dan MAE. Semakin besar jumlah iterasi semakin baik akurasi yang diperoleh, ini untuk MSE dan MAE, tetapi ada ketidakstabilan pada tingkat akurasi yaitu pada jumlah iterasi ke 500 dan 1000 yang meningkat dan menurun. Ini karena nilai MSE dan MAE terbaik belum ditemukan. Nilai MSE terbaik adalah iterasi tak terhingga (-1), kernel polinomial, preprocess KNN, nilai MSE adalah 0,0128. Gambar 11 menunjukkan nilai MAE terkecil yang diperoleh dalam kernel Linear, iterasi tidak terbatas (-1), preprocessing JST, nilai MAE adalah 0,0644.

Tabel 5. Hasil MSE dan MAE parameter iterasi maksimum

MSE Results Tolerance Parameter Analysis						
Data Imputasi	50	100	500	1000	-1	**MIN**
ANN Linear	1.0104	0.2035	1.1029	0.0670	0.0218	0.0218
ANN RBF	0.1308	0.1019	0.0379	0.0256	0.0202	0.0202
ANN Poly	0.3008	0.3906	0.3318	0.0686	0.0212	0.0212
KNN Linear	0.1772	0.5277	0.3608	1.1266	0.0156	0.0156
KNN RBF	0.0201	0.0171	0.0167	0.0166	0.0197	0.0166
KNN Poly	0.0984	0.3730	0.5140	0.0936	0.0128	**0.0128**
Binning Linear	2.6803	1.7716	0.5838	0.3230	0.0583	0.0583
Binning RBF	0.1588	0.0904	0.0955	0.1004	0.2403	0.0904
Binning Poly	4.3704	0.8324	2.0746	1.7691	0.0592	0.0592
MAE Results Tolerance Parameter Analysis						
ANN Linear	0.7988	0.3574	0.8517	0.2034	0.0644	**0.0644**
ANN RBF	0.3486	0.3064	0.1783	0.1345	0.0927	0.0927
ANN Poly	0.4437	0.5031	0.4808	0.2054	0.0651	0.0651
KNN Linear	0.3438	0.5968	0.4847	0.8563	0.0779	0.0779
KNN RBF	0.1054	0.0938	0.0905	0.0920	0.1007	0.0905
KNN Poly	0.2554	0.5192	0.6004	0.2417	0.0760	0.0760
Binning Linear	1.3238	1.1276	0.6221	0.4692	0.1958	0.1958
Binning RBF	0.3073	0.2482	0.2525	0.2534	0.3371	0.2482
Binning Poly	1.7827	0.7657	1.1732	1.1508	0.1959	0.1959

Dari hasil eksperimen terlihat dari perubahan metode preprocessing dalam metode SVR ke nilai kesalahan MSE dan MAE sebagai berikut:

a. *Artificial Neural Network* (ANN)
   Proses perbaikan nilai dengan menggunakan regresi JST menggunakan binary sigmoid di lapisan tersembunyi dan purelin pada output. konsep yang digunakan adalah regresi setiap fitur yang dimilikinya sehingga butuh waktu yang lama. hasil proses perbaikan data dengan JST menunjukkan hasil yang lebih optimal pada MAE daripada dibandingkan dengan menggunakan KNN dan Binning. MAE terkecil yang diperoleh setelah proses regresi menggunakan SVR adalah 0,0614 dengan kernel RBF, gamma 0,001, C 1000, toleransi 0,001 dan iterasi tak terbatas. sedangkan nilai MSE terbaik setelah proses SVR adalah 0,0189 dengan kernel linear C 100000, iterasi tidak terbatas, toleransi adalah 0,001.
b. K-Nearest Neighbor (KNN)
   Kesimpulan dari efek preprocessing KNN pada perubahan parameter. Dari hasil pengujian perubahan parameter yang dibuat, nilai parameter optimal dihasilkan untuk digunakan dalam data yang diperbaiki dengan KNN sesuai dengan evaluasi MSE dan MAE. MAE terkecil yang diperoleh setelah proses regresi menggunakan SVR adalah 0,0729 dengan kernel RBF, gamma 1, C 1000, epsilon 0,0001, toleransi 0,001 dan iterasi tanpa batas. sedangkan nilai MSE terbaik setelah proses SVR adalah 0,0125 dengan Poly kernel, C 1000, toleransi 0,001, epsilon 0,01 dan iterasi tak terbatas.
c. Binning
   Proses perbaikan nilai dengan menggunakan metode binning dengan mengambil nilai tengah. hasil dari proses peningkatan data dengan Binning menunjukkan bahwa hasilnya tidak sebaik jika menggunakan KNN dan ANN karena penggantian kesalahan yang hilang hanya mempertimbangkan nilai tengah dari setiap fitur. MAE terkecil yang diperoleh setelah proses regresi menggunakan SVR adalah 0,0583 dengan Poly kernel, C 1000, toleransi 0,001, epsilon 0,0001 dan iterasi tak terbatas. sedangkan nilai MSE terbaik setelah proses SVR adalah 0,1957 dengan Poly kernel, C 1000, toleransi 0,001, epsilon 0,0001 dan iterasi tanpa batas.

Hasil penelitian terlihat dari perubahan parameter seperti gamma, toleransi, C, epsilon, dan iterasi maksimum. Terlihat bahwa nilai MSE terbaik menggunakan metode KNN. Metode KNN dikatakan sebagai yang terbaik dalam menangani MSE karena lima fitur terakhir memiliki nilai penggantian yang ditemukan sama. Tetapi pengujian pada MAE paling baik menggunakan ANN. MAE adalah pengukuran kesalahan dengan menyerap nilai akurasi, ANN memiliki nilai terbaik karena penyebaran nilai prediktif dengan nilai aktual tidak berjauhan sehingga nilai JST lebih dari KNN.

4. KESIMPULAN

Data curah hujan diperoleh dari Stasiun Meteorologi Sultan Syarif Kasim II di Kota Riau Pekanbaru dari 1 Januari 2000 hingga 13 Oktober 2018. Data terdiri dari 10 fitur. Data diproses tanpa pemisahan antara data yang diambil dari musim hujan dan musim kemarau. Hasil pengolahan data menggunakan beberapa metode preprocessing dan parameter metode SVR. Hasil MSE terbaik untuk melengkapi data yang digunakan dalam penelitian ini menggunakan parameter preprocessing KNN dan Poly kernel, C 1000, toleransi 0,001, epsilon 0,01 dan iterasi tak terbatas. Dan hasil MAE terbaik menggunakan preprocessing JST dengan kernel RBF, gamma 0,001, C 1000, toleransi 0,001 dan iterasi tak terbatas.

UCAPAN TERIMA KASIH

Penulis mengucapkan terimaksih kepada Pusat database BMKG yang telah menyediakan data curah hujan wilayah Pekanbaru Riau.

DAFTAR PUSTAKA

[1] J. Liu, C. Li, J. Tian, F. Yu, and Y. Wang, “Self-correcting multi-model numerical rainfall ensemble forecasting method,” 2017.
[2] J. Diez-Sierra and M. del Jesus, “A rainfall analysis and forecasting tool,” *Environ. Model. Softw.*, vol. 97, no. November 2017, pp. 243–258, Nov. 2017.
[3] J. Diez-Sierra and M. del Jesus, “Subdaily Rainfall Estimation through Daily Rainfall Downscaling Using Random Forests in Spain,” *Water*, vol. 11, no. 1, p. 125, Jan. 2019.
[4] R. Mohd, M. A. Butt, and M. Zaman Baba, “SALM-NARX: Self Adaptive LM- based NARX model for the prediction of rainfall,” in *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on*, 2018, pp. 580–585.
[5] C. Dewi and M. Muslikh, “Perbandingan Akurasi Backpropagation Neural Network dan ANFIS Untuk Memprediksi Cuaca,” *J. Sci. Model. Comput.*, vol. 1, no. 1, pp. 7–13, 2013.
[6] N. Ulinnuha and Y. Farida, “Prediksi Cuaca Kota Surabaya Menggunakan Autoregressive Integrated Moving Average (Arima) Box Jenkins dan Kalman Filter,” *J. Mat. “MANTIK,”* vol. 4, no. 1, pp. 59–67, May 2018.
[7] S. Adhy, A. Prasetio, B. Noranita, and R. Saputra, “Usability Testing of Weather Monitoring on Android Application,” in *2018 2nd International Conference on Informatics and Computational Sciences (ICICoS)*, 2018, pp. 1–6.
[8] C. Dewi, D. P. Kartikasari, and Y. T. Mursityo, “Prediksi Cuaca Pada Data Time Series Menggunakan Adaptive Neuro Fuzzy Inference System (Anfis),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 1, pp. 18–24, 2014.
[9] R. F. Rahmat, F. R. Nasution, Seniman, M. F. Syahputra, and O. S. Sitompul, “Implementation of bayesian model averaging on the weather data forecasting applications utilizing open weather map,” in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 309, pp. 1–10.
[10] BMKG, “Data Online - Pusat Database - BMKG,” *dataonline.bmkg.go.id*. [Online]. Available: https://dataonline.bmkg.go.id/home. [Accessed: 11-Dec-2018].
[11] Makkulau Makkulau, Susanti Linuwih, Purhadi Purhadi, and Muhammad Mashuri, “Pendeteksian Outlier dan Penentuan Faktor-Faktor yang Mempengaruhi Produksi Gula dan Tetes Tebu dengan Metode Likelihood Displacement Statistic-Lagrange,” *J. Tek. Ind.*, vol. 12, no. 2, pp. 95–100, 2010.
[12] A. Zimek and P. Filzmoser, “There and back again: Outlier detection between statistical reasoning and data mining algorithms,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 8, no. 6, p. e1280, Nov. 2018.
[13] I. Ben-Gal, “Outlier Detection,” in *Data Mining and Knowledge Discovery Handbook*, New York: Springer-Verlag, 1993, pp. 131–146.
[14] W. J. Faithfull, J. J. Rodríguez, and L. I. Kuncheva, “Combining univariate approaches for ensemble change detection in multivariate data,” *Inf. Fusion*, vol. 45, no. January 2019, pp. 202–214, Jan. 2019.

[15] E. Vilenski, P. Bak, and J. D. Rosenblatt, "Multivariate anomaly detection for ensuring data quality of dendrometer sensor networks," *Comput. Electron. Agric.*, vol. 162, no. April, pp. 412–421, Jul. 2019.