

## MULTIVARIATE TIME SERIES FORECASTING USING RECURRENT NEURAL NETWORKS FOR METEOROLOGICAL DATA

Victor Hariadi<sup>1</sup>, Ahmad Saikhu<sup>2</sup>, Nurotuz Zakiya<sup>3</sup>,  
Arya Yudhi Wijaya<sup>4</sup>, Fajar Baskoro<sup>5</sup>

Departemen Informatika, Fakultas Teknologi Informasi dan Komunikasi  
Institut Teknologi Sepuluh Nopember (ITS)  
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia  
Email : <sup>1</sup>victor@if.its.ac.id

### *Abstract*

*Rainfall is related to a number of factors that are interdependent and influenced by dynamic global time, region and climate factors. Determination of relevant predictors is important for the efficiency of the rainfall estimator model. Although some climate modeling studies in one region/country have high accuracy, this model is not necessarily suitable for other regions. Determination of predictor variables by considering spatio-temporal factors and local / global features results in a very large number of inputs. Feature selection produces minimal input so that it gets relevant predictor variables and minimizes variable redundancy. Recurrent Neural Networks is one of the artificial neural networks that can be used to predict time series data. This study aims to predict rainfall by combining the SVM classification method and the RNN method. Tests on the Perak 1 daily and monthly weather data (WMO ID: 96933) and Perak 2 Station daily and monthly data(WMO ID: 96937), showed high accuracy results with an  $R^2$  are 92.1%; 94.1%; 90.9% and 89.6%.*

**Keywords :** *Feature Selection, minimum Redundancy Maximal Relevance, Support Vector Machine, Recurrent Neural Network.*

### **1. Pendahuluan**

Cuaca, iklim dan musim senantiasa akrab dalam kehidupan manusia sehari-hari. Gejala alam tersebut berpengaruh kepada banyak sektor kehidupan, diantaranya pertanian, kehutanan, kemaritiman, perikanan, kesehatan, transportasi, pariwisata, pertambangan, energi dan infrastruktur. Dalam kondisi yang ekstrim, cuaca dan iklim bahkan dapat menyebabkan terjadinya bencana, diantaranya wabah penyakit, banjir, angin kencang, tanah longsor, kekeringan dan kebakaran hutan/lahan. Oleh karenanya, informasi, analisis dan pengetahuan terhadap kondisi cuaca dan iklim serta pola perubahan iklim sangat penting untuk dipelajari agar manusia dapat beradaptasi dengan perubahan yang dinamis, menghindarkan atau mengurangi terjadinya resiko akibat bencana dan melakukan tindakan mitigasi [1].

Pemodelan curah hujan di suatu wilayah cukup penting untuk perencanaan sumber daya air secara optimal, pertanian, infrastruktur, pertambangan dan merupakan penelitian penting dalam bidang hidrologi. Berbagai penelitian telah dilakukan untuk pemodelan curah hujan, baik pendekatan model teoritis/konseptual maupun pendekatan data empiris/histori. Pemodelan konseptual bertumpu pada interaksi dan proses antara dinamika atmosfer, suhu permukaan laut, dan posisi matahari. Model ini memiliki kompleksitas tinggi dan memerlukan komputasi yang mahal. Model empiris dipilih untuk menemukan hubungan antara data input dan output ketika pengguna tidak memahami secara lengkap hubungan faktor-faktor fisik yang terjadi dan lebih berfokus pada akurasi.

Berdasarkan sejumlah studi, curah hujan di Indonesia dipengaruhi oleh berbagai faktor lokal (cuaca) dan iklim regional (global). Oleh karenanya penentuan prediktor yang relevan adalah penting untuk efisiensi model penduga curah hujan. Meskipun beberapa penelitian pemodelan iklim di suatu wilayah/negara memiliki akurasi tinggi, belum tentu model tersebut sesuai untuk wilayah lainnya dikarenakan berbagai faktor. Hal ini menyebabkan bahwa pemodelan untuk iklim dan prediksi hujan menjadi khas untuk suatu wilayah. Model iklim dan curah hujan wilayah tropis yang mampu mensimulasikan dan memprediksi curah hujan di Indonesia secara efektif belum ditemukan [2].

Penentuan prediktor yang relevan adalah penting untuk efisiensi model penduga curah hujan. Meskipun beberapa penelitian pemodelan iklim di suatu wilayah/negara memiliki akurasi tinggi, belum tentu model tersebut sesuai untuk wilayah lainnya dikarenakan berbagai faktor. Untuk mendapatkan prediktor yang relevan adalah dengan mengidentifikasi sebanyak mungkin variabel prediktor dengan mempertimbangkan faktor waktu dan lokasi stasiun cuaca (spatio-temporal) dan fitur lokal/global.

Penentuan variabel prediktor dengan mempertimbangkan faktor spatio-temporal dan fitur lokal/global mengakibatkan jumlah input yang sangat banyak. Seleksi fitur dapat dilakukan untuk memilih sejumlah variabel input seminimal mungkin sehingga mendapatkan prediktor yang relevan dengan meminimalkan redundansi [3].

Recurrent Neural Networks (RNN) merupakan salah satu metode yang sesuai untuk memprediksi data time series. Untuk memperoleh variabel prediktor yang efisien, dalam penelitian ini digunakan metode Support Vector Machine (SVM). Studi kasus untuk implementasi seleksi fitur dan prediksi adalah data meteorologi pada stasiun cuaca Perak 1 (WMO ID: 96933) dan stasiun cuaca Perak 2 (WMO ID: 96937).

Tujuan dari penelitian ini adalah mengembangkan model prediksi curah hujan menggunakan metode RNN melalui proses seleksi fitur input (prediktor) variabel cuaca dan iklim dengan menggunakan SVM. Manfaat dari penelitian ini adalah tersedianya metode seleksi fitur input untuk data *time-series* multivariat dan tersedianya model penduga curah hujan yang berbasis spatio-temporal.

## 2. Dasar Teori

### 2.1 Penelitian Terkait

Beberapa penelitian yang telah dilakukan untuk model penduga data menggunakan spatio-temporal. Penelitian di bidang meteorologi dilakukan pada data curah hujan mingguan di wilayah Brasil. Dengan menggunakan pendekatan Spatio Temporal Data Mining menghasilkan MSE lebih baik [4]. Penentuan prediktor yang relevan adalah penting untuk efisiensi model penduga curah hujan pada data spatio-temporal. Berdasarkan penelitian yang dilakukan oleh Nicolas De Jay dkk (2013), seleksi fitur menggunakan *mRMRe* bekerja secara efisien dalam memilih prediktor yang relevan dan tidak redundan [5].

Dalam membangun model penduga curah hujan, beberapa penelitian telah dilakukan menggunakan Support Vector Machine (SVM). Klasifikasi menggunakan SVM untuk prediksi curah hujan harian lebih baik dibandingkan dengan metode decision trees dan K-nearest neighbor [6]. Penelitian lain yang dilakukan untuk prediksi data time-series membandingkan metode Recurrent Neural Network (RNN) dengan metode ARIMA dimana metode RNN lebih baik digunakan pada data non-linier [7].

Beberapa penelitian terkait prediksi curah hujan adalah:

- a. Metode Local Regression Smoothing (LRS) dan Fuzzy Grammatical Evolution (GE) digunakan sebagai model penduga curah hujan bulanan di wilayah Bandung dengan nilai  $R^2=84.62\%$ . Apabila dibandingkan dengan ANFIS,  $R^2=80\%$  dan GE mempunyai  $R^2=74,35\%$  [8], [9].

- b. Metode Seemingly Unrelated Regression (SUR) untuk model GSTARIMA sebagai penduga curah hujan per bulan di wilayah Malang, dengan  $R^2=58\%$  [10].
- c. Metode Adaptif Jaringan Syaraf Tiruan untuk memprediksi curah hujan per bulan di daerah Bongan, Bali dengan prediktor ENSO, SOI, OLR dan SST mempunyai  $R^2 = 74.6\%$  [11].

## 2.2 Data Time Series Multivariat

Deret Waktu (*time series*) adalah data yang memiliki serangkaian nilai-nilai variabel yang terurut berdasarkan waktu kejadian di mana antar nilai saling berhubungan [12]. Data time series dapat dinotasikan sebagai berikut

$$Xi(t); [i = 1, 2, 3, \dots, n, t = 1, 2, 3, \dots, m] \quad (1)$$

$i$  merupakan jumlah variabel dan  $t$  merupakan deret waktu. Jika nilai  $n > 2$ , disebut sebagai *multivariate time series (MTS)*, sedangkan  $n=1$  dinamakan *univariate time series (UTS)*. UTS mengacu pada deret waktu yang terdiri dari pengamatan tunggal yang direkam secara berurutan atas kenaikan waktu yang sama.

## 2.3 Seleksi Fitur

Seleksi fitur adalah proses memilih sejumlah fitur yang merupakan subset dari fitur awal sehingga diperoleh fitur-fitur yang relevan terhadap akurasi klasifikasi [13]. Seleksi fitur dapat meningkatkan akurasi klasifikasi karena dapat menghapus fitur yang tidak relevan [14], [15]. Proses seleksi fitur yang ideal dalam menghasilkan subset fitur yang terbaik adalah dengan mencoba semua kemungkinan kombinasi elemen vektor fitur yang ada untuk diproses dalam proses data mining. Kombinasi elemen vektor menghasilkan subset fitur. Subset fitur terbaik didapatkan dari nilai akurasi yang tertinggi dari proses klasifikasi.

## 2.4 Minimum Redundancy Maximal Relevance (mrMR)

*mrMR* adalah metode pemilihan fitur yang cepat untuk menemukan serangkaian fitur yang relevan dan tidak redundan dengan menggunakan *Mutual Information (MI)* [5] sesuai persamaan 2.

$$I(x,y) = -\frac{1}{2} \ln(1-\rho(x,y)^2), \quad (2)$$

dimana  $I$  dan  $\rho$  masing-masing mewakili MI dan koefisien korelasi antara variabel  $x$  dan  $y$ . *mrMR* memungkinkan pemilihan efisien fitur yang relevan dan tidak redundan [15]. Variabel  $y$  menjadi variabel output dan  $X=\{x_1, \dots, x_n\}$  menjadi set fitur input  $n$ . Metode ini memaksimalkan MI dari  $x$  dengan  $y$  (relevansi maksimum) dan meminimalkan MI rata-rata dengan semua variabel yang dipilih sebelumnya (redundansi minimum).  $x_i$  menjadi fitur dengan MI tertinggi terhadap variabel respon dan dengan demikian akan dipilih terlebih dahulu sesuai persamaan 3.

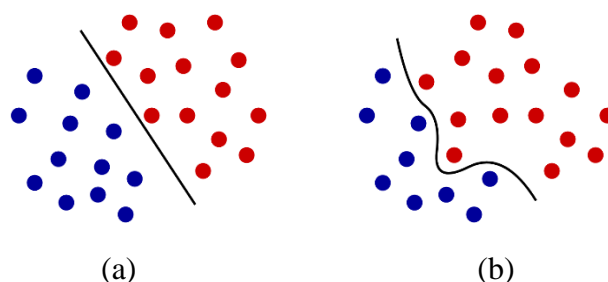
$$Xi = \arg \max_{x_i \in X} I(x_i, y) \quad (3)$$

Selanjutnya, fitur lain ditambahkan ke  $S$  dengan memilih fitur yang memiliki relevansi tertinggi dengan variabel output dan redundansi terendah dengan fitur yang dipilih sebelumnya, sehingga memaksimalkan skor  $q$  pada langkah  $j$ , sesuai persamaan 4.

$$q_j = I(x_j, y) - \frac{1}{|S|} \sum_{x_k \in S} I(x_j, x_k) \quad (4)$$

## 2.5 Support Vector Machine

SVM didasarkan pada teori pada suatu bidang keputusan yang menggambarkan batas-batas keputusan [7]. Metode klasifikasi ini membagi antara satu set objek yang terdiri dari kelas yang berbeda. Gambar dibawah ini adalah contoh objek yang kelasnya terpisah secara linier yaitu kelas merah dan kelas biru. Garis antara kedua objek mengklasifikasikan batas dimana objek kelas biru berada di satu sisi dan kelas merah berada di sisi lain.



Gambar 1. Klasifikasi Linier (a) dan klasifikasi non-linier (b)

Solusi untuk masalah klasifikasi dapat dengan mudah diperoleh ketika data terpisah secara linier sesuai Gambar 1(a). Namun, komplikasi muncul ketika data tidak dapat dipisahkan secara linier. Jika dibandingkan dengan model sebelumnya, pembagian objek merah dan biru membutuhkan kurva yang lebih mudah daripada pembagian objek pada data non linier (gambar 1b).

### 2.6 Fungsi Kernel

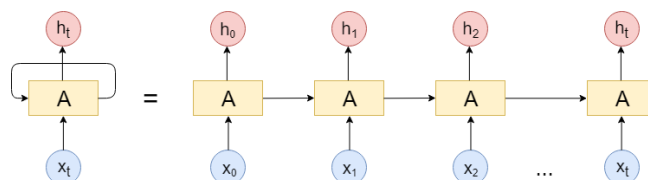
Fungsi *kernel* merupakan fungsi yang dapat mengatasi masalah non-linier yang umumnya terjadi pada dunia nyata. Fungsi ini diaplikasikan pada setiap data untuk memetakan data asli non-linier ke dalam ruang dimensi yang lebih tinggi (*higher-dimensional space*) [2]. Terdapat beberapa fungsi *kernel* yang umum digunakan yaitu linier, *polynomial* dan *Radial Basis Function* (RBF) [4] yang ditunjukkan oleh **Tabel 1**.

Tabel 1. Kernel pada Support Vector Machine

Kernel	Fungsi
<i>Linier</i>	$k(x_i . x) = x_i . x$
<i>Polynomial</i>	$k(x_i . x) = (x_i . x + 1)^d$
<i>RBF</i>	$k(x_i . x) = \exp\left[-\gamma \ x_i - x\ ^2\right]$

### 2.7 Recurrent Neural Network

RNN adalah jaringan yang mengakomodasi output jaringan untuk menjadi input pada jaringan tersebut yang kemudian digunakan untuk menghasilkan output yang baru. RNN merupakan jaringan dengan kemampuan dinamis karena perilaku jaringan tidak hanya bergantung pada input saat ini saja melainkan pada operasi sebelum jaringan [16].



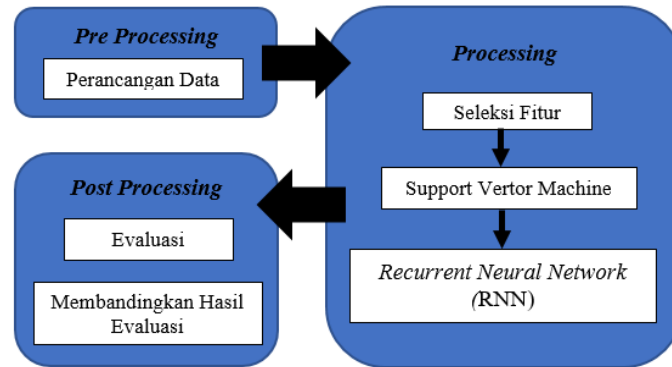
Gambar 2. Recurrent Neural Network

## 3. Analisis Dan Perancangan

### 3.1 Analisis Metode Secara Umum

Penelitian ini mengembangkan model penduga curah hujan untuk data *spatio-temporal* meteorologi (cuaca dan iklim) dengan studi kasus 2 stasiun cuaca di Surabaya dengan periode bulan Januari 2010 hingga bulan Desember 2016. Tahap implementasi dari pengembangan model meliputi pra-proses, tahap pemodelan, dan tahap evaluasi. Pada pra-

proses, dilakukan seleksi fitur menggunakan *mRMR* untuk memperoleh himpunan variabel *time-lags*. Selanjutnya, dilakukan klasifikasi dengan SVM menggunakan variasi kernel untuk mengetahui subset variabel terbaik dari seleksi fitur. Kemudian dilakukan *forecasting* menggunakan *RNN* untuk memprediksi curah hujan pada 2 lokasi secara serentak, yaitu Perak 1 dan Perak 2. Diagram alir dari keseluruhan proses ditunjukkan oleh Gambar 3.



Gambar 3. Blok Proses Prediksi Curah Hujan

Evaluasi dilakukan terhadap hasil uji coba bertujuan untuk mengetahui metode terbaik berdasarkan nilai *Root Mean Square Error (RMSE)* dan Koefisien Determinasi ( $R^2$ ).

### 3.2 Data Studi Kasus

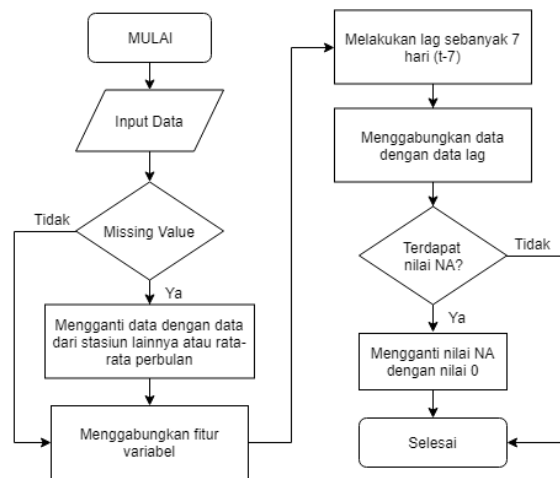
Data input merupakan data *time series spatio temporal* yang terdiri dari variabel lokal/global, waktu dan wilayah. Fitur input dari Stasiun Perak 1 akan digunakan sebagai prediktor untuk Stasiun Perak 2 dan juga sebaliknya. Hal ini bertujuan untuk mengetahui pengaruh wilayah terhadap nilai variabel respon. Tabel 2 adalah daftar variabel yang digunakan yang pada pemodelan.

Tabel 2. Perancangan dataset untuk Stasiun Perak 1

Fitur	Nama Variabel
Variabel respon Stasiun Perak 1	Curah hujan
Variabel lokal Stasiun Perak 1	Suhu minimum
	Suhu maksimum
	Suhu rata-rata
	Kelembapan rata-rata
	Lama Penyinaran
	Kecepatan angin rata-rata
	Kecepatan angin terbesar
Variabel lokal Stasiun Perak 2	Suhu minimum
	Suhu maksimum
	Suhu rata-rata
	Kelembapan rata-rata
	Lama Penyinaran
	Kecepatan angin rata-rata
	Kecepatan angin terbesar
Variabel global	NINO12
	NINO3
	NINO34
	NINO4

Setiap wilayah memiliki 19 variabel prediktor yang selanjutnya akan diproses menggunakan fitur waktu. Fitur waktu merupakan *time-lags* dari 18 variabel lokal dan global, yaitu *lags t-1, t-2, t-3, t-4, t-5, t-6* dan *t-7*. Setiap variabel prediktor pada dataset diproses

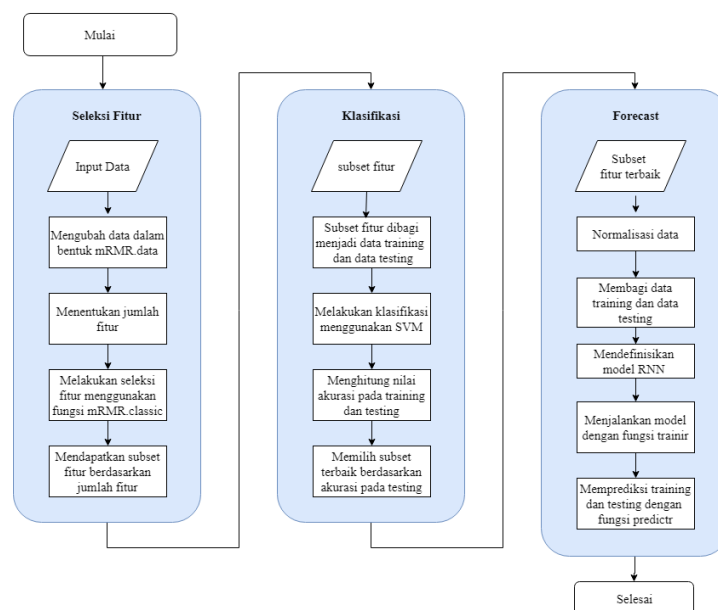
menggunakan fitur waktu sehingga menghasilkan 144 variabel prediktor. Fitur lokasi terdiri dari data pada lokasi stasiun perak 1 dan 2, sehingga didapatkan data dengan fitur variabel lokal/global, waktu dan lokasi. Hasil dari preprocessing menghasilkan dataset dengan 144 fitur variabel. Diagram alir secara detail pada praproses ditunjukkan pada Gambar 4.



Gambar 4. Algoritma *Preprocessing*

### 3.3 Pemodelan

Proses pemodelan terdiri dari seleksi fitur, klasifikasi dan *forecasting*. Hasil dari proses akan dievaluasi untuk memperoleh metode terbaik pada kasus prediksi curah hujan harian pada Stasiun Perak 1 dan Stasiun Perak 2. Diagram alir proses pemodelan ditunjukkan oleh gambar 5.



Gambar 5. Perancangan proses Seleksi fitur, klasifikasi dan forecasting

### 3.4 Seleksi Fitur

*Pseudocode* seleksi fitur menggunakan mRMR dapat dilihat pada Gambar 6.

1. START

2. SET feature set  $F$ , class label  $c$ , the number of selected feature  $m$
3. Rank all feature in  $F$  in descending order based on their relevance
4. WHILE  $|S| < m$  do
5.       Search the top ranked feature score
6.       Move  $f_i$  from  $F$  to  $S$
7. END WHILE
8. END

Gambar 6. Pseudocode Seleksi Fitur

Proses seleksi fitur diawali dengan inisialisasi dataset dengan fitur  $F$ , variabel target atau faktor  $c$  dan jumlah subset fitur  $m$ . Fungsi yang digunakan untuk seleksi fitur pada fitur  $F$  adalah *mRMR.classic*. Skor dari *mRMR.classic* akan diurutkan dari yang terbesar ke yang terkecil. Untuk mendapatkan subset  $|S|$ , dilakukan perulangan sebanyak  $m$  dengan mencari nilai tertinggi pada skor mRMR. Fitur yang memiliki nilai skor tertinggi dimasukkan kedalam subset  $|S|$ .

### 3.5 Klasifikasi

Proses klasifikasi menggunakan berbagai variasi kernel yaitu kernel linier, polinomial dan radial. Sebelum dilakukan proses klasifikasi, variabel faktor pada data akan di diskritisasi menjadi 5 kelas yaitu tidak hujan, hujan ringan, hujan sedang, hujan lebat dan hujan sangat lebat dengan parameter yang ditunjukkan pada Tabel 2.

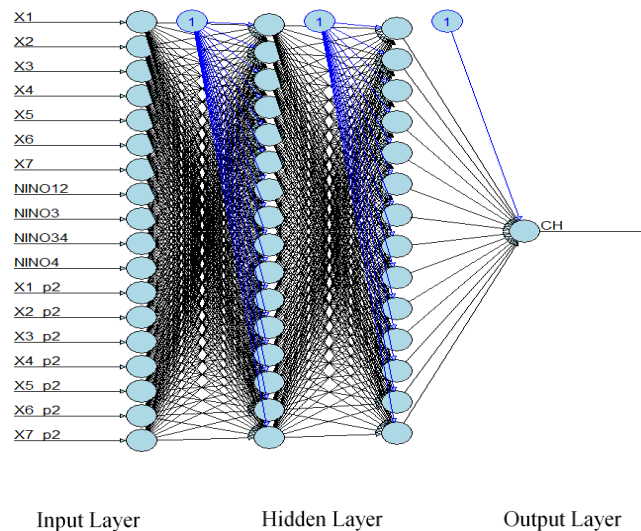
Tabel 1. Kategori hujan harian dan bulanan

No	Kategori	Parameter	
		Harian (mm/jam)	Bulanan (mm/bulan)
1	Tidak hujan	0	0
2	Hujan ringan	1 – 5	1 – 100
3	Hujan sedang	6 – 10	101 – 300
4	Hujan lebat	11 – 20	301 – 400
5	Hujan sangat lebat	> 20	> 400

Variabel curah hujan berfungsi sebagai variabel tujuan dari proses klasifikasi. Tiga nilai akurasi terbaik merupakan subset terbaik yang akan di uji coba pada proses *forecasting*.

### 3.6 Forecasting

Fitur yang diperoleh dari subset terbaik dinormalisasi sebelum proses *forecasting* menggunakan RNN. Arsitektur RNN terdiri dari satu *input layer*, dua *hidden layer* dan satu *output layer* sesuai Gambar 7.



Gambar 7. Arsitektur Jaringan RNN

Hidden layer yang digunakan terdiri dari dua neuron dengan jumlah 16 dan 14 node. Evaluasi terhadap hasil prediksi pada *forecasting* diukur menggunakan *Root Mean Square Error* (RMSE) dan Koefisien Determinasi ( $R^2$ ). Semakin kecil nilai RMSE dan semakin tinggi hasil koefisien determinasi maka semakin bagus hasil prediksi pada *forecasting*.

#### 4. Hasil Uji dan Evaluasi

##### 4.1 Hasil Uji Data Harian

Uji coba data harian Stasiun Perak 1 dan Stasiun Perak 2 dengan metode klasifikasi dengan berbagai kernel dan *forecast* menggunakan RNN dengan epoch 10.000 ditunjukkan pada Tabel 4.

Pada skenario data harian, dari metode klasifikasi SVM menggunakan kernel linear, polinomial dan radial, serta diprediksi menggunakan RNN, hasil terbaik pada data harian Stasiun Perak 1 adalah dengan nilai  $R^2=92,1\%$  dengan kernel polinomial dan jumlah prediktor 2. Sedangkan hasil terbaik pada data harian Stasiun Perak 2 didapatkan nilai  $R^2=94,1\%$  dengan kernel polinomial dan jumlah prediktor 3.

Tabel 4. Uji Coba Data Harian

Kernel	Stasiun Perak 1			Stasiun Perak 2		
	Prediktor	Rsquare		Prediktor	$R^2$	
		train	test		train	test
Linear	5	0,936	0,914	8	0,000	0,000
	2	0,939	0,920	29	0,946	0,935
	3	0,940	0,915	25	0,000	0,000
Polinomial	3	0,940	0,915	3	0,942	0,941
	4	0,941	0,914	9	0,940	0,938
	2	0,939	0,921	31	0,949	0,932
Radial	80	0,962	0,888	30	0,946	0,924
	84	0,004	0,010	34	0,943	0,930
	86	0,952	0,901	76	0,172	1,43E-01



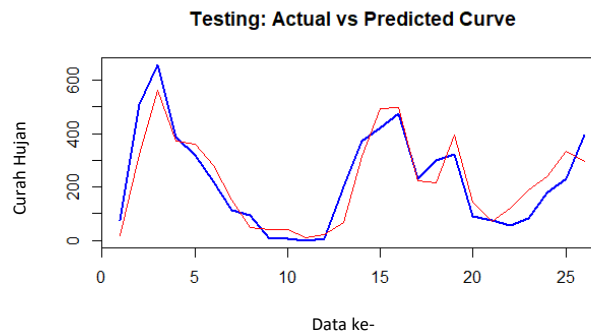
### 4.2 Hasil Uji Data Bulanan

Hasil uji coba menggunakan data bulanan Stasiun Perak 1 dan Stasiun Perak 2 dengan metode klasifikasi dengan berbagai kernel dan *forecasting* menggunakan RNN dengan *epoch* 10.000 dapat dilihat pada Tabel 4.

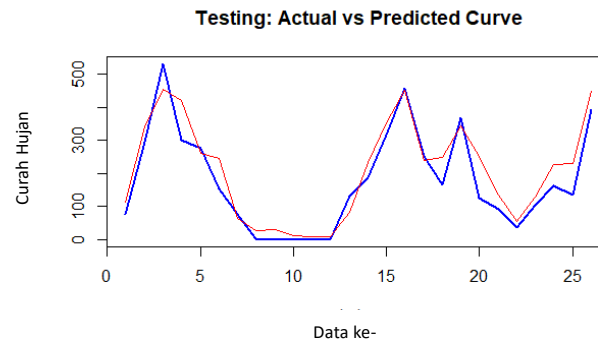
Tabel 5 Uji Coba Data Harian

Kernel	Stasiun Perak 1			Stasiun Perak 2		
	Prediktor	Rsquare		Prediktor	Rsquare	
		train	test		train	test
Linear	31	0,981	0,812	17	0,983	0,803
	4	0,939	0,899	98	0,993	0,562
	23	0,967	0,872	139	0,992	0,644
Poli-nomial	18	0,962	0,827	55	0,986	0,731
	80	0,988	0,780	115	0,995	0,460
	3	0,937	0,909	108	0,991	0,701
Radial	144	0,989	0,453	8	0,978	0,896
	13	0,935	0,900	32	0,988	0,807
	3	0,937	0,908	9	0,980	0,743

Pada data bulanan, dari metode klasifikasi SVM menggunakan kernel linear, polinomial dan radial, serta diprediksi menggunakan RNN, hasil terbaik pada data bulanan Perak 1 didapatkan nilai  $R^2=90,9\%$  dengan kernel polinomial dan jumlah prediktor=3. Sedangkan hasil terbaik pada data bulanan Perak 2 didapatkan nilai  $R^2=89,6\%$  dengan kernel radial dan jumlah prediktor=8.



Gambar 8. Plot aktual dan testing data bulanan Stasiun Perak 1



Gambar 9. Plot aktual dan testing data bulanan Stasiun Perak 2

### 4.3 Evaluasi

Dari hasil uji coba data harian dan data bulanan pada Stasiun Perak 1 dan Stasiun Perak 2, memberikan nilai  $R^2$  yang tinggi. Jumlah prediktor yang mempertimbangkan faktor waktu dan lokasi serta fitur lokal/global menghasilkan 144 prediktor dan dapat direduksi dengan jumlah prediktor=2 pada data harian Stasiun Perak 1, jumlah prediktor=3 pada data harian Stasiun Perak 2, jumlah prediktor=3 pada data bulanan Stasiun Perak 1 dan jumlah prediktor=8 pada data bulanan Stasiun Perak 2.

Metode klasifikasi dengan berbagai kernel dan *forecasting* menggunakan RNN memberikan nilai  $R^2$  yang tinggi dengan nilai 92,1% pada data harian Stasiun Perak 1, 94,1% pada data harian Stasiun Perak 2, 90,9% pada data bulanan Stasiun Perak 1 dan 89,6% pada data bulanan Stasiun Perak 2. Akurasi prediksi dari pengembangan metode pada penelitian ini lebih baik dibandingkan dengan penelitian sebelumnya [8-11].

### 5. Kesimpulan

Dari hasil pengamatan selama proses perancangan, implementasi, dan pengujian yang dilakukan, dapat diambil kesimpulan sebagai berikut: (i) Metode seleksi fitur yang digunakan pada penelitian ini dapat mereduksi jumlah prediktor dengan jumlah prediktor awal 144 prediktor dengan mempertimbangkan faktor waktu dan lokasi serta fitur lokal/global dan hasil dari seleksi fitur didapatkan jumlah prediktor dengan rentan antara 2 hingga 8 prediktor. (ii) Penelitian ini dapat membangun model penduga curah hujan untuk wilayah surabaya pada Stasiun Perak 1 dan Stasiun Perak 2 baik harian maupun bulanan dengan nilai  $R^2$  yang tinggi.

### Daftar Pustaka

- [1] Aldrian, E., & Budiman, M. K. (2011). Adaptasi dan mitigasi perubahan iklim di Indonesia. Pusat Perubahan Iklim dan Kualitas Udara, Kedeputan Bidang Klimatologi, Badan Meteorologi, Klimatologi, dan Geofisika.
- [2] Ettema, J., & Aldrian, E. (2012). Spatiotemporal Characteristics Of Extreme Rainfall Events Over Java Island, Indonesia. *Indonesian Journal of Geography*, 44(1), 62-86.
- [3] Humphrey, G. B., Galelli, S., Castelletti, A., Maier, H. R., Dandy, G. C., & Gibbs, M. S. (2014). A new evaluation framework for input variable selection algorithms used in environmental modelling.
- [4] Fernandes, M. V., Schmidt, A. M., & Migon, H. S. (2009). Modelling zero-inflated spatio-temporal processes. *Statistical Modelling*, 9(1), 3-25.

- [5] De Jay, N., Papillon-Cavanagh, S., Olsen, C., El-Hachem, N., Bontempi, G., & Haibe-Kains, B. (2013). mRMRe: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics*, 29(18), 2365-2368.
- [6] Ortiz-García, E. G., Salcedo-Sanz, S., & Casanova-Mateo, C. (2014). Accurate precipitation prediction with support vector classifiers: A study including novel predictive variables and observational data. *Atmospheric Research*, 139, 128-136.
- [7] Madan, R., & SarathiMangipudi, P. (2018, August). Predicting Computer Network Traffic: A Time Series Forecasting Approach Using DWT, ARIMA and RNN. In 2018 Eleventh International Conference on Contemporary Computing (IC3) (pp. 1-5). IEEE.
- [8] Pratama, S. W., & Nhita, F. (2016, May). Implementation of local regression smoothing and fuzzy-grammatical evolution on rainfall forecasting for rice planting calendar. In 2016 4th International Conference on Information and Communication Technology (ICoICT) (pp. 1-5). IEEE.
- [9] Nhita, F., Saepudin, D., & Wisesty, U. N. (2015, December). Comparative Study of Moving Average on Rainfall Time Series Data for Rainfall Forecasting Based on Evolving Neural Network Classifier. In 2015 3rd International Symposium on Computational and Business Intelligence (ISCBI) (pp. 112-116). IEEE.
- [10] Nisak, S. C. (2016). Seemingly Unrelated Regression Approach for GSTARIMA Model to Forecast Rain Fall Data in Malang Southern Region Districts. *CAUCHY*, 4(2), 57-64.
- [11] Kanigoro, B., & Salman, A. G. (2016, August). Recurrent gradient descent adaptive learning rate and momentum neural network for rainfall forecasting. In 2016 International Seminar on Application for Technology of Information and Communication (ISemantic) (pp. 23-26). IEEE.
- [12] Vasimalla, K. (2014). A survey on time series data mining. *Intern. J. of Innovative Research in Computer and Communication Engineering*, 2, 170-179.
- [13] Putri, L. A. A. R. (2017). Seleksi Fitur Dalam Klasifikasi Genre Musik. *Jurnal Ilmu Komputer*, 10(1), 19-26.
- [14] Saputro, D. R. S. (2009). Memprediksi Curah Hujan (Data Spatio-Temporal) dengan Metode Bayesian Network. In *Proceeding of National Seminar on Research, Teaching, and Application of Mathematics and Science* (pp. 37-42).
- [15] Ding, C., & Peng, H. (2005). Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology*, 3(02), 185-205.
- [16] Dong, D., Sheng, Z., & Yang, T. (2018, November). Wind Power Prediction Based on Recurrent Neural Network with Long Short-Term Memory Units. In 2018 International Conference on Renewable Energy and Power Engineering (REPE) (pp. 34-38). IEEE.

Victor Hariadi, Ahmad Saikhu, Nurotuz Zakiya, Arya Yudhi Wijaya, Fajar Baskoro