

Prediksi Hujan Jangka Pendek Berbasis Data Spasial Menggunakan Metode Extreme Learning Machine

Surya Aditama¹⁾, Indrabayu²⁾, Ais Prayogi³⁾

Departemen Teknik Informatika, Fakultas Teknik, Universitas Hasanuddin
Jl. Poros Malino, Gowa, Sulawesi Selatan, Indonesia
suryaadi4848@gmail.com, indrabayu16@gmail.com, ais.prayogi@gmail.com

Abstrak

Tingkat Curah hujan merupakan salah satu parameter yang berkaitan dengan dampak perubahan iklim. Oleh karena itu, dalam penelitian klimatologi dan meteorologi, parameter tersebut sering dilibatkan. Beberapa penelitian yang telah dilakukan belum mampu memberikan hasil yang cukup memadai. Kelemahannya, penelitian tersebut hanya dapat dipergunakan pada pendugaan curah hujan berdasarkan data curah hujan yang ada di stasiun penakar hujan (temporal) tanpa mempertimbangkan pengaruh parameter dari lokasi stasiun lain (spasial) dengan data resolusi perhari. Beberapa penelitian sudah membuktikan adanya pengaruh parameter dari lokasi lain, yakni dibuktikan dengan korelasi antar lokasi, salah satunya korelasi temperatur dan curah hujan daerah Ambon dan Makassar dengan korelasi 0,995, Palembang dan Makassar dengan 0,991, informasi tentang prakiraan cuaca yang cepat dan tepat menjadi suatu hal yang penting karena cuaca menjadi bagian yang tidak dapat dipisahkan dari aktifitas manusia dan mempengaruhi berbagai bidang kehidupan, seperti penentuan masa tanam (pertanian) dan kelayakan keberangkatan pesawat udara (transportasi). Menyikapi hal tersebut maka diperlukan kemampuan dan metode yang efektif dalam prakiraan cuaca, khususnya prakiraan cuaca jangka pendek.

Dalam penelitian ini, dikembangkan sistem yang dapat memprediksi hujan dengan sampling resolusi per 30 menit. Metode yang digunakan adalah *Extreme Learning Machine* dengan memanfaatkan aplikasi Matlab 2016a. Data hujan yang diolah berasal dari data yang di unduh dari website wunderground.com dimana data yang dimiliki berasal dari 250.000 lebih anggota yang mengirim data real-time dari stasiun pribadi mereka, 26.000 stasiun dari National Oceanic and Atmospheric Administration (NOAA) dan 6000 stasiun cuaca di setiap bandara. Data yang di gunakan adalah data Makassar, Palembang dan Ambon. Daerah spasial ini dipilih berdasarkan penelitian terkait dan korelasi parameter hujan.

Hasil penelitian dengan menggunakan metode extreme learning machine menghasilkan rata-rata sensitivity sebesar 83,99%, precision sebesar 94,68% dan accuracy sebesar 98,37% untuk tahun 2015 dan memiliki rata-rata sensitivity sebesar 89,22%, precision sebesar 93,62% dan accuracy sebesar 98,62% pada tahun 2016 untuk prediksi 1 jam kedepan dan untuk prediksi 3 jam kedepan sistem menghasilkan rata-rata sensitivity sebesar 85,08%, precision sebesar 84,39% dan accuracy sebesar 94,06% untuk tahun 2015 dan rata-rata sensitivity sebesar 79,56%, precision sebesar 85,86% dan accuracy sebesar 79,78% untuk tahun 2016.

Kata Kunci : Prediksi hujan; *Extreme Learning Machine*; Spasial; Korelasi; *Confusion Matrix*.

1. Pendahuluan

Indonesia merupakan negara yang mempunyai kompleksitas dalam fenomena cuaca dan iklim disebabkan letak geografis negara yang dilewati oleh garis khatulistiwa serta dikelilingi oleh dua samudra dan dua benua. Posisi ini menjadikan Indonesia sebagai daerah pertemuan sirkulasi meridional (Utara – Selatan) dikenal sebagai Sirkulasi Hadley dan sirkulasi zonal (Timur – Barat) dikenal sebagai Sirkulasi Walker, dua sirkulasi yang sangat mempengaruhi keragaman iklim di Indonesia. Pergerakan matahari yang berpindah dari 23.5° Lintang Utara ke 23.5° Lintang Selatan sepanjang tahun mengakibatkan timbulnya aktivitas monsoon yang juga berperan dalam mempengaruhi keragaman iklim [1].

Beberapa penelitian yang telah dilakukan belum mampu memberikan hasil yang cukup memadai. Kelemahannya, penelitian tersebut hanya dapat dipergunakan pada pendugaan curah hujan berdasarkan data curah hujan yang ada di stasiun penakar hujan (temporal) tanpa mempertimbangkan pengaruh parameter dari lokasi stasiun lain (spatial) dengan data resolusi perhari. Beberapa penelitian sudah membuktikan adanya pengaruh parameter dari lokasi lain [2].

Metode *Extreme Learning Machine* (ELM) merupakan salah satu metode pembelajaran baru dari

Neural network. Metode ini memiliki tingkat pembelajaran yang lebih baik dan cepat dengan beragam data di dibandingkan dengan metode konvensional lainnya [3], oleh sebab itu pada penelitian kali ini digunakan metode ELM dengan data spasial iklim per 30 menit dari www.wunderground.com untuk berbagai data parameter curah hujan di beberapa lokasi bukan hanya Makassar dengan harapan dapat memberikan hasil prediksi curah hujan yang lebih baik dan akurat.

Pada tugas akhir ini, metode yang digunakan yaitu metode *Extreme Learning Machine*. Metode ini akan diimplementasikan menggunakan matlab 2016a.

2. Landasan Teori

2.1 Cuaca dan Iklim

Cuaca adalah keadaan udara pada saat tertentu dan di wilayah tertentu yang relatif sempit dan pada jangka waktu yang singkat. Di Indonesia keadaan cuaca diumumkan untuk jangka waktu sekitar 24 jam melalui prakiraan cuaca yang dikembangkan oleh Badan Meteorologi Klimatologi dan Geofisika (BMKG), Departemen Perhubungan [4].

Iklim adalah keadaan cuaca rata-rata dalam waktu satu tahun yang penyelidikannya dilakukan dalam waktu yang lama (\pm minimal 30 tahun) dan meliputi

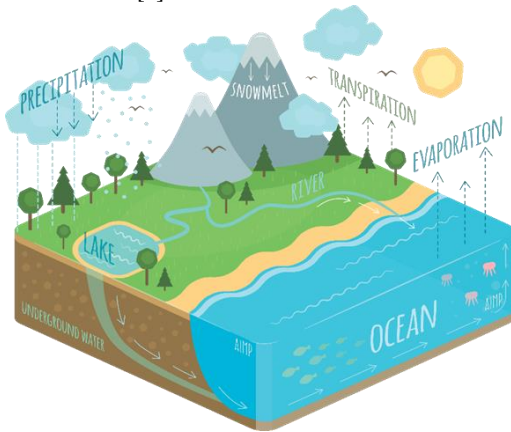
wilayah yang luas. Menurut [1] Iklim dapat terbentuk karena adanya faktor berikut:

- Rotasi dan revolusi bumi sehingga terjadi pergeseran semu harian matahari dan tahunan
- Perbedaan lintang geografi dan lingkungan fisis. Perbedaan ini menyebabkan timbulnya penyerapan panas matahari oleh bumi sehingga besar pengaruhnya terhadap kehidupan di bumi.

2.2 Proses Presipitasi

1.2.1 Siklus Hidrologi

Siklus hidrologi, yaitu proses presipitasi dimana terjadinya sirkulasi air dari bumi ke atmosfer dan kembali lagi ke bumi, siklus hidrologi terjadi mulai dari evaporasi, transpirasi, evapotranspirasi, sublimasi, kondensasi, adveksi, presipitasi, run-off dan infiltrasi [5].



Gambar 1. Siklus Hidrologi

A. Evapotranspirasi

Evapotranspirasi adalah penguapan air keseluruhan yang terjadi di seluruh permukaan bumi, baik yang terjadi pada badan air dan tanah, maupun pada jaringan makhluk hidup. Evapotranspirasi merupakan gabungan antara evaporasi dan transpirasi. Dalam siklus hidrologi, laju evapotranspirasi ini sangat mempengaruhi jumlah uap air yang terangkat ke atas permukaan atmosfer [5].

B. Sublimasi

Sublimasi adalah proses perubahan es di kutub atau di puncak gunung menjadi uap air tanpa melalui fase cair terlebih dahulu. Meski sedikit, sublimasi juga tetap berkontribusi terhadap jumlah uap air yang terangkat ke atas atmosfer bumi melalui siklus hidrologi panjang. Akan tetapi, dibanding melalui proses penguapan, proses sublimasi dikatakan berjalan sangat lambat [5].

C. Kondensasi

Ketika uap air yang dihasilkan melalui proses evaporasi, transpirasi, evapotranspirasi, dan proses sublimasi naik hingga mencapai suatu titik ketinggian tertentu, uap air tersebut akan berubah menjadi partikel-partikel es berukuran sangat kecil melalui proses kondensasi. Perubahan wujud uap air menjadi es tersebut terjadi karena pengaruh suhu udara yang sangat rendah di titik ketinggian tersebut, semakin banyak partikel es yang bergabung, awan yang terbentuk juga akan semakin tebal dan hitam [5].

D. Adveksi

Awan yang terbentuk dari proses kondensasi selanjutnya akan mengalami adveksi. Adveksi adalah proses perpindahan awan dari satu titik ke titik lain dalam satu horizontal akibat arus angin atau perbedaan tekanan udara. Adveksi memungkinkan awan akan menyebar dan berpindah dari atmosfer lautan menuju atmosfer daratan. Perlu diketahui bahwa, tahapan adveksi tidak terjadi pada siklus hidrologi pendek [5].

E. Presipitasi

Awan yang mengalami adveksi selanjutnya akan mengalami proses presipitasi. Proses presipitasi adalah proses mencairnya awan akibat pengaruh suhu udara yang tinggi. Pada proses inilah hujan terjadi. Butiran-butiran air jatuh dan membasahi permukaan bumi [5].

F. Run Off

Setelah presipitasi terjadi sehingga air hujan jatuh ke permukaan bumi, proses run off pun terjadi. Run off atau limpasan adalah suatu proses pergerakan air dari tempat yang tinggi ke tempat yang rendah di permukaan bumi. Pergerakan air tersebut misalnya terjadi melalui saluran-saluran seperti saluran got, sungai, danau, muara, laut, hingga samudra. Dalam proses ini, air yang telah melalui siklus hidrologi akan kembali menuju lapisan hidrosfer [5].

G. Infiltrasi

Tidak semua air hujan yang terbentuk setelah proses presipitasi akan mengalir di permukaan bumi melalui proses run off. Sebagian kecil di antaranya akan bergerak ke dalam pori-pori tanah, merembes, dan terakumulasi menjadi air tanah [5].

2.3 Klasifikasi Hujan

Mekanisme dimana udara diangkat merupakan cara yang digunakan untuk mengklasifikasi tipe hujan [6].

Ada 3 mekanisme pengangkatan udara, yaitu:

- Orographic Lifting**, yaitu proses dimana udara dipaksa untuk naik ke pegunungan atau ketempat yang lebih tinggi. Udara akan mengembang dan mendingin hingga akhirnya proses presipitasi dalam bentuk hujan pun terjadi [6].
- Frontal Surface Lifting**, yaitu proses saat udara yang lebih panas dipaksa ke atas udara yang lebih dingin pada suatu equilibrium dengan permukaan yang lebih dingin. Proses ini juga dapat diakibatkan oleh pemuatan udara pada pendakian disepanjang atau dekat permukaan frontal [6].
- Convective Lifting** merupakan proses dimana udara panas naik dari permukaan yang panas dan semakin mendingin. Proses ini paling sering terjadi pada daerah tropis. Gambar berikut mengilustrasikan proses ini [7].

2.4 Curah Hujan

Curah hujan adalah jumlah air hujan yang turun pada suatu daerah dalam waktu tertentu. Sementara intensitas curah hujan merupakan perbandingan derivasi jumlah hujan terhadap interval waktu tertentu dalam satuan m/s, mm/menit atau inci/jam. Dalam menentukan intensitas atau laju hujan, ada beberapa faktor yang perlu diperhatikan [8].

- Durasi hujan dalam detik, menit atau jam.

- B. Kedalaman hujan yang dinyatakan sebagai ketipisan lapisan air pada permukaan dalam mm atau inci.
- C. Daerah, adalah luas geografis dari curah hujan dalam km².
- D. Frekuensi terjadinya, biasanya dinyatakan sebagai periode kembali, misalnya sekali dalam 10 tahun

2.5 Fungsi Aktivasi

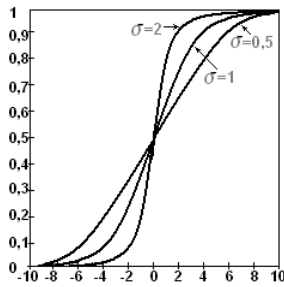
Dalam jaringan saraf tiruan, fungsi aktivasi dipakai untuk menentukan keluaran suatu *neuron*, argumen fungsi aktivasi adalah *net input* (kombinasi linier *input* dan bobotnya) [9]. Beberapa fungsi aktivasi yang dipakai dalam jaringan saraf tiruan salah satunya fungsi *sigmoid biner*.

Fungsi ini digunakan untuk jaringan saraf tiruan yang dilatih dengan menggunakan metode *backpropagation*. Fungsi *sigmoid biner* memiliki nilai pada *range* 0 sampai 1. Oleh karena itu, fungsi ini sering digunakan untuk jaringan saraf tiruan yang membutuhkan nilai *output* yang terletak pada *interval* 0 sampai 1. Namun, fungsi ini juga digunakan oleh jaringan saraf tiruan yang nilai *output*nya 0 atau 1.

Fungsi *Sigmoid Biner* ;

$$y = g(z) = \frac{1}{1 + e^{-\sigma x}}$$

Dengan $g'(z) = \sigma g(z)[1 - g(z)]$



Gambar 2. Fungsi aktivasi *Sigmoid* [10]

2.6 Extreme Learning Machine

Extreme Learning Machine merupakan jaringan saraf tiruan *feedforward* dengan satu *hidden layer* atau biasa disebut dengan istilah *single hidden layer feedforward neural network* (SLFNs) [3]. Metode ELM dibuat untuk mengatasi kelemahan-kelemahan dari jaringan saraf tiruan *feedforward* terutama dalam hal *learning speed*.

Algoritma ELM tidak melatih bobot input ataupun bias, ELM melatih untuk memperoleh bobot keluarannya dengan menggunakan *norm-least-squares solution* dan *moore-penrose inverse* pada sistem linier secara umum. Dengan menemukan *node* yang memberikan nilai output maksimal, pada ELM parameter-parameter seperti *input weight* dan *hidden bias* dipilih secara *random*, sehingga ELM memiliki *learning speed* yang cepat dan mampu menghasilkan *good generalization performance*. Metode ELM mempunyai model matematis yang berbeda dari jaringan syaraf tiruan *feedforward*. Model matematis dari ELM lebih sederhana dan efektif.

a. Single hidden layer feedforward neural network

Standard SLFNs dengan unit hidden sebanyak *m* dan fungsi aktivasi dapat di gambarkan secara matematis [11] sebagai berikut:

$$\sum_{j=1}^m \beta_j g(z_{netj}) = t_i \quad (1)$$

$$z_{netj} = w_{ji} * x_i + b_j \quad (2)$$

Dimana;

$i = 1, 2, \dots, n$ (input).

$j = 1, 2, \dots, m$ (output).

β_j = Vektor bobot menghubungkan antara lapisan *hidden* ke-*i* dan lapisan *output*.

w_{ji} = Vektor bobot yang menghubungkan *input* ke-*i* kelapisan *hidden* ke-*j*.

x_i = Lapisan input yang terdiri dari *i* jumlah *input*.

b_j = Vektor bias pada lapisan *hidden* ke-*j*.

y = Nilai *output*.

Untuk jumlah unit *input* sebanyak *n* dan jumlah unit *hidden* sebanyak *m*, dapat disusun sebuah matriks *H* yang berisi hasil keluaran dari lapisan *hidden* yang berukuran $m \times n$. Persamaan yang terbentuk sebagai berikut:

$$H = \begin{bmatrix} g(b_1 + w_{11} \cdot x_1) & \dots & g(b_m + w_{1m} \cdot x_1) \\ \vdots & \ddots & \vdots \\ g(b_1 + w_{n1} \cdot x_n) & \dots & g(b_m + w_{nm} \cdot x_n) \end{bmatrix} \quad (3)$$

Sehingga persamaan (4) dapat dituliskan sebagai berikut;

$$y = \beta H \quad (4)$$

Pada metode *extreme learning machine*, bobot *input* dan bias ditentukan secara *random*, dan bobot *output* diperoleh dari persamaan sebagai berikut:

$$\beta = H^+ y \quad (5)$$

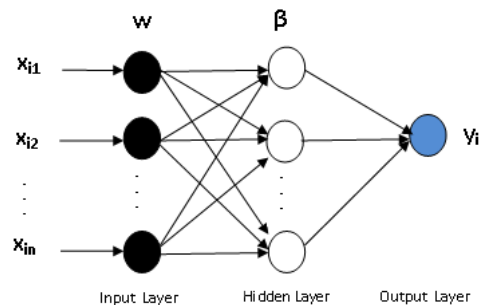
Dengan H^+ adalah matrix *pseudoinverse* dari matrix *H*.

b. Algoritma Extreme Learning Machine

[11] mengatakan bahwa ada tiga tahapan dalam metode ELM, yaitu jika diberikan data training, fungsi aktivasi, dan *m* unit hidden, maka:

1. Menentukan vektor bobot *input* w_j dan bias b_j faktor pengaruh unit *hidden* ke-*j*, $b_j, j = 1, \dots, m$.
2. Menghitung matriks keluaran pada *hidden layer* $H_{n \times m}$.
3. Menghitung bobot keluaran β .

c. Struktur Jaringan Syaraf Tiruan ELM



Gambar 3. Model Jaringan Saraf Tiruan Extreme Learning Machine.

Diberikan sebuah model jaringan saraf tiruan ELM dengan n input, m neuron hidden layer dan fungsi aktivasi. Misalkan $x=[x_1, x_2, x_3, \dots, x_n]$ dengan merupakan nilai input pada jaringan tersebut, H merupakan matriks bobot penghubung input layer dan hidden layer maka matriks H mempunyai ukuran $n \times m$. Penentuan nilai elemen-elemen matriks tersebut dilakukan secara random. Kemudian setiap nilai input tersebut diproses pada hidden layer menggunakan fungsi aktivasi sigmoid, dan nilai tersebut dihimpun dalam sebuah matriks H dengan ordo $n \times m$.

d. Proses Belajar dengan pengawasan

Jaringan memberikan tanggapan dengan mendapatkan target tertentu. Sebelum jaringan mengubah sendiri bobotnya untuk mencapai target, bobot interkoneksi diinisialisasi. Proses belajar jaringan saraf tiruan dengan pengawasan adalah proses belajar dengan memberikan latihan untuk mencapai suatu target keluaran yang ditentukan. Jaringan saraf tiruan mendapatkan pelatihan untuk mengenal pola-pola tertentu, dengan memberikan target keluaran, perubahan masukan akan diadaptasi oleh keluaran dengan mengubah bobot interkoneksinya mengikuti algoritma belajar yang ditentukan. Set pelatihan dipilih dari fungsi keluaran maksimal setiap keadaan parameter yang diubah. Dengan menginisialisasi bobot tiap set, jaringan saraf tiruan akan mencari error terkecil, sehingga bentuk fungsi keluaran mendekati target yang diinginkan.

Berdasarkan proses belajar yang dilakukan, [12] mengatakan dalam bukunya bahwa perlu memperhatikan beberapa hal dalam menyusun set pelatihan, yaitu:

1. Pemberian urutan pola yang akan diajarkan
2. Kriteria perhitungan error
3. Kriteria proses belajar
4. Jumlah iterasi yang harus dilalui
5. Inisialisasi bobot, bias, dan parameter awal

Pelatihan dilakukan dengan memberikan pasangan pola-pola masukan dan keluaran. Untuk keperluan pengendalian, pasangan pola tidak mengikuti rumusan tertentu. Jaringan saraf tiruan harus dapat mengadaptasi masukan yang acak supaya keluaran tetap mengikuti target. Lebih lanjut, proses pelatihan dilakukan dengan memberikan pola yang menggunakan masukan acak dan bobot inerkoneksi yang besar. Dengan pemberian bobot yang besar, perbedaan target dan keluaran berkurang lebih cepat, sehingga proses adaptasi akan lebih cepat pula.

e. Inisialisasi Bobot dan Bias

Prosedur yang umum dilakukan pada inisialisasi acak adalah menginisialisasi bias dan bobot, baik dari unit input ke unit hidden maupun dari unit hidden ke unit output ke dalam sebuah interval tertentu ($-\gamma$ dan γ). Misalnya antara -0.4 sampai 0.4, -0.5 sampai 0.5, dan -1 sampai 1 [13]

1.7 Confusion Matrix

Pengukuran terhadap kinerja suatu sistem merupakan hal yang penting. Kinerja sistem klasifikasi menggambarkan seberapa baik sistem dalam mengklasifikasikan data. Confusion matrix merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya confusion matrix mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya [14].

Pada pengukuran kinerja menggunakan confusion matrix, terdapat 4 (empat) istilah sebagai representasi hasil proses. Keempat istilah tersebut adalah True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Nilai TN merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan FP merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, TP merupakan data positif yang terdeteksi benar. FN merupakan kebalikan dari TP, sehingga data positif, namun terdeteksi sebagai data negatif.

Tabel 1. Confusion Matrix

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positif	TP (True Positive)	FN (False Negative)
Negatif	FP (False Positive)	TN (True Negative)

Berdasarkan nilai True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP) dapat diperoleh nilai akurasi, presisi dan recall. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh dengan Persamaan 6.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan Persamaan 7.

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Sementara itu, recall (Sensitivity) menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai recall diperoleh dengan Persamaan 8.

$$recall = \frac{TP}{TP+FN} \quad (8)$$

3. Metodologi Penelitian

3.1. Instrumen Penelitian

Instrumen yang digunakan dalam penelitian ini meliputi :

1. Software MATLAB 2016a.
2. Perangkat Personal Computer CPU Intel Core i7 4700MQ, RAM 16GB, GPU GTX 755, Windows 10 Pro x64

3.2. Teknik Pengambilan Data

Data yang akan dikumpulkan didasarkan pada variabel yang mempengaruhi curah hujan, antara lain:

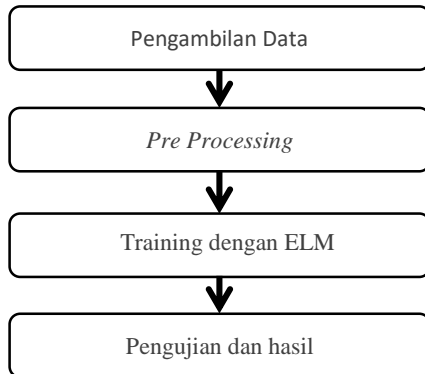
1. Curah hujan
2. Temperatur
3. Titik embun
4. Tekanan permukaan laut
5. Kelembaban
6. Jarak Pandang
7. Kecepatan dan arah angin

Data dari berbagai variabel tersebut diperoleh dari situs <https://www.wunderground.com> yang memiliki 250.000+ anggota mengirim data real-time dari stasiun cuaca pribadi mereka, 26.000

stasiun dari *National Oceanic and Atmospheric Administration* (NOAA) dan 6000 stasiun cuaca di setiap bandara [15].

3.3. Perancangan Implementasi Sistem

Dalam perancangan sistem untuk prediksi hujan terdapat beberapa proses sebagaimana ditunjukkan sebagai berikut.



Gambar 4. Blok diagram implementasi sistem

a. Pengambilan Data

Pada tahap ini data yang digunakan di ambil dari situs <https://www.wunderground.com> dimana memiliki 250.000+ anggota yang mengirim data real-time dari stasiun cuaca pribadi mereka, 26.000 stasiun dari *National Oceanic and Atmospheric Administration* (NOAA) dan 6000 stasiun cuaca di setiap bandara. Data yang di gunakan adalah data makassar, Palembang dan ambon. Daerah spasial ini dipilih berdasarkan penelitian terkait dan korelasi parameter hujan [15].

b. Pre Processing

Pada tahap *pre-processing*, data inputan sebelum di training data melalui beberapa tahap yaitu interpolasi, korelasi dan normalisasi.

1. Korelasi

Korelasi dilakukan Untuk mengukur kuat tidaknya hubungan antara variabel input terhadap variabel output yang diharapkan, Semakin kuat hubungan/korelasi antara variabel input terhadap variabel output, maka hasil prediksi yang didapatkan akan semakin akurat, korelasi antara variabel input dan output disajikan dalam bentuk koefisien korelasi yang diberikan pada persamaan (9) yang terdapat pada Microsoft Excel:

$$\text{Correlation} = \text{Correl}(\text{Array1}; \text{Array2}) \quad (9)$$

Dimana :

Array1 = Sampel data variabel 1

Array2 = Sampel data variabel 2

2. Interpolasi

Interpolasi adalah suatu cara untuk mencari nilai di antara beberapa titik data yang telah

diketahui. Di dunia nyata, *interpolasi* dapat digunakan untuk memperkirakan suatu fungsi, yang mana fungsi tersebut tidak terdefinisi dengan suatu formula, tetapi didefinisikan hanya dengan data-data atau tabel, misalnya tabel dari hasil percobaan, ada berbagai macam interpolasi berdasarkan fungsinya, di antaranya adalah interpolasi linier, interpolasi kuadrat, dan interpolasi polinomial, *interpolasi* yang dilakukan untuk mengisi data yang kosong disajikan pada persamaan (10) yang terdapat pada Microsoft Excel :

$$\text{Interpolasi} = (\text{end_x} - \text{start_x}) / (\text{row}(\text{end_x}) - \text{row}(\text{start_x})) \quad (10)$$

Dimana :

end_x = Batas akhir alamat sel untuk data yang ingin dicari

start_y = Batas Awal alamat sel untuk data yang ingin dicari

row = fungsi untuk letak sel

3. Normalisasi

Proses *preprocessing* yang dilakukan dalam penelitian ini menggunakan transformasi minimum-maksimum, sehingga setiap data input ditransformasikan menjadi dalam *range* -0 sampai dengan +1. Menurut Iskandar (2010), *Algoritma* minimum-maksimum adalah sebagai berikut :

$$x_i = \frac{(x_i - x_{\min_i})}{(x_{\max_i} - x_{\min_i})} \quad (11)$$

Dimana ;

$i = 1, 2, \dots, n$

x_i = Lapisan input yang terdiri dari i jumlah input.

x_i' = Data yang sudah ditransformasi

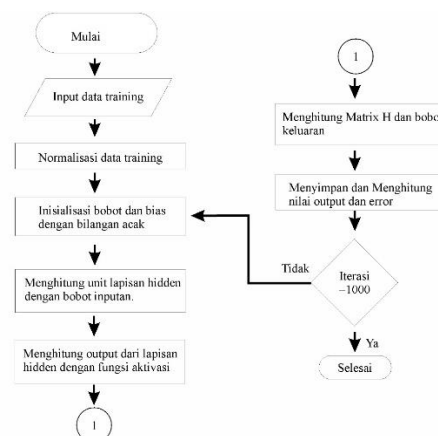
x_{\min_i} = Data minimum dari input ke- i

x_{\max_i} = Data maximum dari input ke- i

c. Training dan Testing dengan *Extreme Learning Machine*

a. Tahapan Training

Langkah-langkah yang dilakukan pada proses *training* ini diilustrasikan pada gambar 5.



Gambar 5. Flowchart Pelatihan

1. Pembagian Data *Training* dan *Testing*

Data akan dibagi dua bagian, bagian pertama digunakan untuk sebagai data *training* sebanyak 80% dari total data dan yang lainnya untuk data *testing* sebesar 20%.

2. Desain Arsitektur Jaringan

Arsitektur yang akan digunakan dalam jaringan ini terdiri dari 3 *layer* yaitu *input layer*, *hidden layer*, dan *output layer*. Data yang akan dimasukkan dalam *input layer* akan ditransformasi terlebih dahulu.

3. Pelatihan (*Training*) Jaringan

Pelatihan dilakukan guna mencari bobot dan bias optimal atau sesuai untuk digunakan pada proses *testing*.

Langkah-langkah *training* yang akan diproses adalah sebagai berikut:

Langkah 1: Inisialisasi semua bobot dan bias dengan bilangan acak.

Langkah 2: Jika kondisi penghentian belum terpenuhi, maka lakukan langkah sampai langkah 7.

Langkah 3: Tiap unit masukan $x_i (i = 1, 2, \dots, n)$ menerima sinyal dan meneruskan sinyal tersebut ke semua unit pada lapisan *hidden*.

Langkah 4: Menghitung tiap-tiap unit lapisan *hidden* $z_j (j = 1, 2, \dots, m)$ dengan menjumlahkan sinyal-sinyal *input* terbobot:

$$z_{net_j} = b_j + \sum_{i=1}^n x_j w_{ji} \quad (12)$$

selanjutnya hitung *output* dari lapisan *hidden* dengan fungsi aktivasi;

$$g(z_{net_j}) = \frac{1}{1 + e^{-z_{net_j}}} \quad (13)$$

Setelah mendapatkan *output* dari lapisan *hidden*, maka langkah selanjutnya:

Langkah 5: Menghitung matriks H dengan ukuran $n \times m$.

$$H = \begin{pmatrix} g(w_{11} \cdot x_1 + b_{10}) & g(w_{12} \cdot x_1 + b_{20}) & \dots & g(w_{1n} \cdot x_1 + b_{n0}) \\ g(w_{21} \cdot x_2 + b_{10}) & g(w_{22} \cdot x_1 + b_{20}) & \dots & g(w_{2n} \cdot x_2 + b_{n0}) \\ g(w_{31} \cdot x_3 + b_{10}) & g(w_{32} \cdot x_1 + b_{20}) & \dots & g(w_{3n} \cdot x_3 + b_{n0}) \end{pmatrix}$$

setelah mendapatkan matriks H dengan ukuran $n \times m$, selanjutnya hitung H^+ yang merupakan matriks *pseudoinvers* dari matriks H yang akan digunakan pada pencarian nilai bobot antara *hidden layer* dan *output layer*, persamaan H^+ sebagai berikut:

$$H^+ = (H^T H)^{-1} H^T \quad (14)$$

kemudian mencari bobot ke *output layer* (β).

$$\beta = H^+ t_i \quad (15)$$

t_i adalah target dari proses *training*.

Langkah 6: Menghitung nilai *output* dengan menggunakan persamaan:

$$\sum_{j=1}^m \beta_j g(z_{net_j}) = y \quad (16)$$

Langkah 7: Menghitung nilai *error* pada unit *output*.

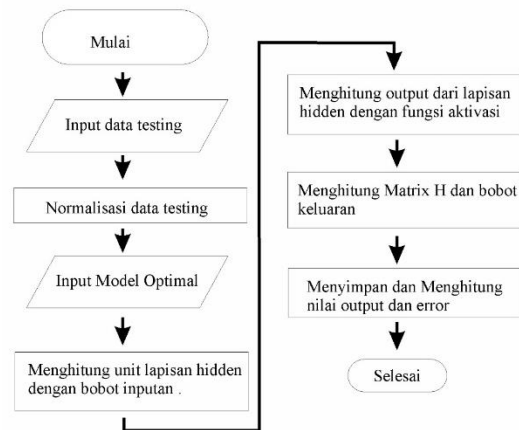
$$E = ||y - t|| \quad (17)$$

Selanjutnya tahap akhir untuk *training*, hasil di kalkulasi dengan bobot keluaran dan nilai *root mean*

square error (RMSE) terbaik. Setelah didapatkan nilai RMSE terbaik dari 1000 iterasi maka keluaran untuk *training* adalah model *training* optimal yang akan dipakai untuk pengujian sistem, dimana terdapat *Input Weight*, *Bias of Hidden Neurons*, *Output Weight*, *Training Accuracy*, actual.

b. Tahapan *Testing*

Pada bagian pengujian, model optimal dari hasil pelatihan digunakan untuk pengujian, selanjutnya dilakukan langkah-langkah yang sama dengan tahap *training* dari langkah 3 sampai 7, Tahapan pengujian di ilustrasikan pada gambar 6.



Gambar 61. Flowchart pengujian sistem

4. Hasil dan Pembahasan

Pada bab ini disajikan hasil kinerja sistem Short-Term Rain Prediction Berbasis Data Spasial Online Menggunakan Metode Extreme Learning Machine.

4.1 Analisis Korelasi Variabel input

Untuk mengukur kuat tidaknya hubungan antara variabel input terhadap variabel output yang diharapkan, maka pengujian korelasi penting dilakukan. Semakin kuat hubungan/korelasi antara variabel input terhadap variabel output, maka hasil prediksi yang didapatkan akan semakin akurat.

Telah dilakukan korelasi terhadap 8 parameter masing masing daerah spasial, yaitu curah hujan, temperatur, kelembaban, jarak pandang, titik embun, tekanan permukaan laut, arah angin dan kecepatan angin. Hasil pengujian ini selanjutnya digunakan dalam pemilihan variabel-variabel yang paling kuat korelasinya atau yang memiliki koefisien korelasi yang paling mendekati nilai 1 atau -1, sebagai variabel input untuk beberapa metode prediksi yang akan digunakan.

Tabel 2 – Korelasi Parameter hujan

PrecipitationMa kasar	T	SLP	H	DP	V	WS-WD
Palembang	0,99 54	0,49 25	0,94 31	0,99 30	0,87 05	0,05 91
Lagging	Lag -1	Lag -5	Lag 0	Lag 0	Lag 0	Lag 0
Ambon	0,99 09	0,61 41	0,96 19	0,91 60	0,77 82	0,02 11
Lagging	Lag -1	Lag -8	Lag 0	Lag -4	Lag 0	Lag 0

Dimana :

CH	: Curah hujan
M-P	: Makassar - Palembang
M-A	: Makassar – Ambon
T	: Temperatur
SLP	: Tekanan permukaan laut
.H	: Kelembaban
DP	: Titik embun
WS	: Kecepatan angin
WD	: Arah angin
V	: Jarak pandang

Berdasarkan tabel hasil uji korelasi antara variabel-variabel yang mempengaruhi curah hujan, terlihat bahwa variabel yang memiliki koefisien tertinggi adalah temperatur, kelembaban, titik embun, jarak pandang, yang artinya variabel tersebut telah memenuhi kriteria korelasi antara 2 variabel seperti yang ditunjukkan pada tabel 3

Tabel 3 Koefisien korelasi [16].

Koefisien korelasi	Tingkat korelasi
$\rho = 1$	korelasi positif sempurna
$0,80 \leq \rho < 1$	korelasi tinggi sekali
$0,60 \leq \rho < 0,80$	korelasi tinggi
$0,40 \leq \rho < 0,60$	korelasi sedang
$0,20 \leq \rho < 0,40$	Korelasi rendah
$0,00 < \rho < 0,20$	korelasi rendah sekali
$\rho = 0$	tidak mempunyai korelasi linier
$\rho = -1$	korelasi negatif sempurna
$-1 < \rho \leq -0,80$	korelasi negatif tinggi sekali
$-0,80 < \rho \leq -0,60$	korelasi negatif tinggi
$-0,60 < \rho \leq -0,40$	korelasi negatif sedang
$-0,40 < \rho \leq -0,20$	korelasi negative rendah
$-0,20 < \rho < 0$	korelasi negatif rendah sekali
Makin jauh ρ dari 0	korelasi akan semakin tinggi

Dari uji simulasi juga ditemukan bahwa dengan hanya memasukkan kelima parameter meteorology temperatur, kelembaban, titik embun dan jarak pandang maka tingkat prediksi dapat meningkat . Data yang digunakan adalah data tahun 2009–2016.

4.2 Hasil Analisa Algoritma Extreme learning machine

Untuk mengetahui hasil prediksi cuaca menggunakan metode ELM ini yaitu bisa di lihat dari hasil proses training dan testing jaringan. Dalam pencarian bobot dengan error yang minimum pada proses training. Proses training dilakukan dengan 1000 iterasi, Kemudian

dilakukan proses trial and error untuk jumlah hidden node hingga dapat diperoleh jumlah hidden node yang optimal, lama pelatihan dan nilai mean square error yang minimum. Kemudian dari hasil pelatihan yang optimal maka akan digunakan untuk melatih data pengujian. Dari hasil analisis data pelatihan yang telah dilatih seperti pada tabel 3 di bawah ini:

Tabel 3 Hasil Analisis Jaringan Saraf Tiruan dengan Perubahan Iterasi dan Hidden Node untuk tahun 2015

No	Hidden Node	Iterasi	RMSE
1	2	1000	0,0008724
2	5	1000	0,0008628
3	10	1000	0,0008618
4	20	1000	0,0008617
5	30	1000	0,0008608
6	40	1000	0,0008609

Pencarian MSE pada proses training di atas menggunakan bantuan program matlab 2016. Hasil uji coba pada proses training yang telah dilakukan menggunakan kombinasi unit hidden dengan menggunakan 1000 iterasi. Hasil akurasi paling optimal dapat dilihat tabel 4 pada no 5 dengan 30 unit hidden menghasilkan RMSE paling kecil yaitu 0,0008608. Dari proses trial and error pada data pelatihan, maka yang akan digunakan pada pelatihan data pengujian adalah 15 unit input dengan unit hidden 30, berikut hasil confusion matrix untuk data 2015.

Tabel 4. Confusion matrix hasil prediksi tahun 2015

Prediksi 1 jam kedepan				
Confusion Matrix	Cerah	Ringan	Sedang	Lebat
Cerah	18470	55	0	0
Ringan	190	1256	4	0
Sedang	0	53	228	0
Lebat	0	0	28	61
Prediksi 3 jam kedepan				
Confusion Matrix	Cerah	Ringan	Sedang	Lebat
Cerah	17371	1154	0	0
Ringan	0	1449	1	0
Sedang	0	3	278	0
Lebat	0	0	49	40

Berdasarkan confusion matrix pada Tabel 3, diperoleh nilai sensitivity, precision dan accuracy pada Tabel 4.

Tabel 5. Analisa nilai sensitivity, precision dan accuracy sistem

Hasil analisa 1 jam kedepan			
	Precision	Sensitivity	Accuracy
Cerah	98,98%	99,7%	98,37%
Ringan	92,08%	86,62%	
Sedang	87,69%	81,13%	
Lebat	100%	68,53%	

Rata-rata	94,68%	83,99%	
Hasil anasila 3 jam kedepan			
	Precision	Sensitivity	Accuracy
Cerah	98,98%	99,7%	98,37%
Ringan	92,08%	86,62%	
Sedang	87,69%	81,13%	
Lebat	100%	68,53%	
Rata-rata	94,68%	83,99%	

Berikut contoh perhitungan untuk tabel 5;

$$\text{Precision(lebat)} = \frac{TP}{TP+FP} = 61/(61+0) = 1 = 100\%$$

$$\text{Sensitivity(lebat)} = \frac{TP}{TP+FN} = 61/(61+28) = 0,6853 = 68,53\%$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = 20345/20675 = 0,9837 = 98,37\%$$

Dengan mengamati hasil dari Tabel 3 dan Tabel 4, terlihat bahwa model training dengan 15 unit input dan 30 hidden, sistem memiliki rata rata sensitivity sebesar 83,99%, precision sebesar 94,68% dan accuracy sebesar 98,37%. Berdasarkan nilai sensitivity, specificity, precision dan accuracy ini sistem prediksi hujan dapat dikatakan cukup akurat, bisa dilihat pada nilai precision untuk data lebat menunjukkan nilai 100% dikarenakan dalam prediksi data lebat, nilai lebat hanya diprediksi pada lebat saja, tidak terdapat data cerah, hujan ringan ataupun hujan sedang yang di prediksi ke lebat, begitupula pada nilai sensitivity data lebat menunjukkan nilai 68% dikarenakan dari 89 data lebat ada 28 data diprediksi ke sedang yang seharusnya diprediksi lebat, maka dari itu sensitivitas untuk prediksi lebat kurang tinggi.

Untuk hasil analisa prediksi sistem 3 jam kedepan, sistem memiliki rata rata sensitivity sebesar 84,39%, precision sebesar 85,08% dan accuracy sebesar 94,06%. Berdasarkan nilai sensitivity, precision dan accuracy sistem prediksi hujan dapat dikatakan cukup akurat.

Tabel 6. Hasil Analisis Jaringan Saraf Tiruan dengan Perubahan Iterasi dan Hidden Node untuk tahun 2016

No	Hidden Node	Iterasi	RMSE
1	2	1000	0,0008039
2	5	1000	0,0007904
3	10	1000	0,0007894
4	20	1000	0,0007892
5	30	1000	0,0007890
6	40	1000	0,0007887

Hasil akurasi paling optimal dapat dilihat tabel 6 pada no 6 dengan 40 unit hidden menghasilkan RMSE paling kecil yaitu 0,0007887. Dari proses trial and error pada data pelatihan, maka yang akan digunakan pada pelatihan data pengujian adalah 15 unit input dengan unit hidden 40, berikut hasil confusion matrix untuk data 2016.

Tabel 7. Confusion matrix hasil prediksi tahun 2016

Prediksi 1 jam kedepan				
Confusion Matrix	Cerah	Ringan	Sedang	Lebat
Cerah	18484	49	0	0
Ringan	230	1274	19	0
Sedang	0	44	203	3

Lebat	0	0	3	36
Prediksi 3 jam kedepan				
Confusion Matrix	Cerah	Ringan	Sedang	Lebat
Cerah	13391	5142	0	0
Ringan	0	1522	1	0
Sedang	0	1	249	3
Lebat	0	0	11	28

Berdasarkan confusion matrix pada Tabel 7, diperoleh nilai sensitivity, precision dan accuracy pada Table 8.

Tabel 8. Analisa nilai sensitivity, precision dan accuracy sistem

Hasil analisa 1 jam kedepan			
	Precision	Sensitivity	Accuracy
Cerah	98,77%	99,73%	98,62%
Ringan	93,19%	83,65%	
Sedang	90,22%	81,2%	
Lebat	91,66	91,66%	
Rata-rata	93,62%	89,22%	
Hasil anasila 3 jam kedepan			
	Precision	Sensitivity	Accuracy
Cerah	100%	72,25%	79,78%
Ringan	22,83%	99,93%	
Sedang	95,40%	99,6%	
Lebat	100%	71,66%	
Rata-rata	79,56%	85,86%	

Dengan mengamati hasil dari tabel 7 dan Tabel 8, terlihat bahwa model training dengan 15 unit input dan 40 hidden, sistem memiliki rata rata sensitivity sebesar 89,22%, precision sebesar 93,62% dan accuracy sebesar 98,62%. Berdasarkan nilai sensitivity, specificity, precision dan accuracy ini sistem prediksi hujan dapat dikatakan akurat, bisa dilihat pada nilai precision untuk data cerah menunjukkan nilai 98,77% dikarenakan dalam prediksi data cerah terdapat 230 data yang seharusnya diprediksi ringan tapi diprediksi cerah, berdasarkan dari hasil prediksi cerah ada 18484 yang diprediksi benar, sehingga hasil menunjukkan 98,77% untuk precision data cerah, begitupula pada nilai sensitivity data cerah menunjukkan nilai yang sangat tinggi yaitu 99,73% dikarenakan dari 18533 data terdapat 18484 yang diprediksi benar atau cerah dan ada 49 data negative yang di prediksi salah, maka dari itu sensitivitas untuk prediksi cerah sangat tinggi.

Untuk hasil analisa prediksi sistem 3 jam kedepan, sistem memiliki rata rata sensitivity sebesar 79,56%, precision sebesar 85,6% dan accuracy sebesar 74,66%. Berdasarkan nilai sensitivity, precision dan accuracy sistem prediksi hujan dapat dikatakan kurang akurat, dikarenakan hasil akurasi prediksi menurun dari 98,62% menjadi 74,66%, hal ini disebabkan nilai precision untuk hujan ringan sangat kecil yaitu 22,83% artinya dari jumlah prediksi sistem untuk hujan sebanyak 6665 kondisi hujan ringan, ada 5142 yang tidak seharusnya diprediksi hujan ringan tapi di prediksi oleh sistem dimana kondisi itu seharusnya cerah.

5. Kesimpulan

Dari hasil analisis yang telah dilakukan dalam pengujian sistem Prediksi hujan dengan sampling resoulsi

30 menit berbasis data spasial *online* menggunakan metode *extreme learning machine*, maka dapat disimpulkan bahwa:

1. Dari delapan Unsur meteorologi yang diperoleh dari <https://www.wunderground.com>, hanya lima yang digunakan untuk pelatihan yaitu curah hujan, temperatur, kelembaban, titik embun, dan jarak pandang. Keputusan ini diambil setelah uji korelasi terhadap curah hujan yang menunjukkan hanya kelima unsur ini yang memiliki korelasi terbaik.
2. Dari proses trial and error pada data pelatihan nilai *Root means square error* yang minimum dari pelatihan *extreme learning machine* yaitu 0,0008608 untuk pengujian tahun 2015 dan 0,0007887 pada pengujian 2016. *RMSE* tersebut didapatkan dari pelatihan dengan pemberian hidden node 30 untuk tahun 2015, 40 untuk tahun 2016 dan iterasi 1000. Sehingga pada proses testing menggunakan hidden node 30 dan 40 serta menggunakan bobot dan bias optimal. Model arsitektur jaringan saraf tiruan pada kondisi cuaca Makassar, yaitu arsitektur jaringan yang terdiri dari 15 unit input, 30 dan 40 unit hidden pada 1 hidden layer, dan 1 unit output (Cerah, Hujan ringan, sedang dan lebat). Hasil output *extreme learning machine* dengan menggunakan data testing menghasilkan rata-rata *sensitivity* sebesar 83,99%, *precision* sebesar 94,68% dan *accuracy* sebesar 98,79% untuk tahun 2015 dengan prediksi 1 jam kedepan, dan memiliki rata rata *sensitivity* sebesar 89,22%, *precision* sebesar 93,62% dan *accuracy* sebesar 98,62% pada tahun 2016 untuk prediksi 1 jam kedepan, sedangkan hasil untuk prediksi 3 jam kedepan sistem menghasilkan rata-rata *sensitivity* sebesar 85,08%, *precision* sebesar 84,39% dan *accuracy* sebesar 94,06% untuk tahun 2015 dan rata-rata *sensitivity* sebesar 79,56%, *precision* sebesar 85,86% dan *accuracy* sebesar 79,78% untuk tahun 2016%.

6. Daftar Pustaka

- [1] R. Boer, "Penyimpangan Iklim Di Indonesia," 2003.
- [2] M.S. Sreekanth, R. Rajesh, and J. Satheeshku, "Extreme Learning Machine for the Classification of Rainfall and Thunderstorm," *J. Appl. Sci.*, vol. 15, no. 1, pp. 153–156, Jan. 2015.
- [3] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, 2004, vol. 2, pp. 985–990 vol.2.
- [4] S. Wirjomidjojo, Y. Swarinoto, Badan Meteorologi Klimatologi dan Geofisika (BMKG), and Pusat Penelitian dan Pengembangan, *Iklim kawasan Indonesia: dari aspek dinamik-sinoptik*. Jakarta: Pusat Penelitian dan Pengembangan], Badan Meteorologi Klimatologi dan Geofisika, 2010.
- [5] R. H. McCuen, *Hydrologic analysis and design*, 2. ed. Upper Saddle River, NJ: Pearson Education/Prentice Hall, 1998.
- [6] Ragunath, H.M., *Hydrology Principles Analysis Design / Precipitation / Hydrology*. New Delhi: New Age International Publisher, 2007.
- [7] J. Lundquist, "Hydrologic Process," *Mountain Hydrology Research*, 2010. [Online]. Available: http://faculty.washington.edu/jdlund/classes/CEE345/Lundquist_ebook_2010_Hydrology.pdf. [Accessed: 06-Nov-2017].
- [8] B. Tjasyono HK and Harijono, *Meteorologi Indonesia II*, 1st ed., vol. 2, 2 vols. Badan Meteorologi Klimatologi dan Geofisika, 2012.
- [9] joong jek Siang, *Jaringan Syaraf Tiruan & Pemrogramannya Menggunakan MATLAB - Oleh: Jong Jek Siang*. 2006.
- [10] S. Kusumadewi, *Artificial Intelligence (Teknik dan Aplikasinya)*. Yogyakarta: Graha Ilmu, 2003.
- [11] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.
- [12] L. W. Pandjaitan, *Dasar-dasar Komputasi Cerdas*, 1st ed., vol. 182. Yogyakarta: Andi Publisher, 2007.
- [13] D. Puspitaningrum, *Pengantar Jaringan Syaraf Tiruan*. Yogyakarta: Andi Publisher, 2006.
- [14] E. Prasetyo, *Data Mining: Konsep dan Aplikasi menggunakan Matlab*. Yogyakarta: Andi Offset, 2012.
- [15] Weather Underground, "About Our Data," *Weather Underground*, 2017. [Online]. Available: [/about/data](http://about/data). [Accessed: 06-Nov-2017].
- [16] J. . Guildford, *Fundamental Statistics in Psychology and Education*. New York, NY, USA: McGraw Hill, 1956.