

PREDIKSI CUACA PADA DATA *TIME SERIES* MENGGUNAKAN *BACKPROPAGATION NEURAL NETWORK (BPNN)*

Muhammad Azzam Imaduddin
Prodi S1 Sistem Informasi, Fakultas
Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
azzami@student.telkomuniversity.ac.id

Faqih Hamami S. Kom., M.T.
Prodi S1 Sistem Informasi, Fakultas
Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
faqihhamami@telkomuniversity.ac.id

Riska Yanu Fa'Rifah S.Si., M.Si.
Prodi S1 Sistem Informasi, Fakultas
Rekayasa Industri
Universitas Telkom
Bandung, Indonesia
riskyanu@telkomuniversity.ac.id

Perubahan cuaca yang ekstrim dapat menimbulkan bencana. Kerugian akibat bencana alam ini dapat kita minimalisir apabila ada persiapan yang matang dalam menghadapi kemungkinan terjadinya bencana alam. Dan persiapan yang matang dalam menghadapi bencana alam tentunya didasarkan pada pengetahuan tentang prediksi kapan dan dimana bencana alam tersebut akan terjadi. Perubahan cuaca ini dapat diprediksi berdasarkan data cuaca di masa lampau. Data pada penelitian kali ini bersumber dari database Badan Meteorologi, Klimatologi dan Geofisika (BMKG). Kemudian dilakukan preprocessing, berupa cleansing dan penyesuaian data. Backpropagation neural network (BPNN) merupakan algoritma yang dipakai penulis dalam melakukan forecasting terkait perubahan kondisi cuaca. Backpropagation neural network (BPNN) pada penelitian ini dibangun dengan menggunakan library keras dan Tensorflow. Bahasa pemrograman yang digunakan pada penelitian kali ini adalah python dan dengan menggunakan tools jupyter notebook. Model yang digunakan adalah menggunakan 1 input layer, 6 hidden layer dan 1 output layer. Sedangkan untuk epochs yang digunakan berjumlah 10000. Dan model evaluasi dengan Mean Squared Error (MSE) Hasil dari penelitian ini berbentuk grafik per parameter cuaca di wilayah Bandung pada tahun 2021.

Kata kunci— *backpropagation neural network, cuaca, time series, machine learning*

I. PENDAHULUAN

Cuaca merupakan keadaan udara pada satu tempat tertentu dengan jangka waktu yang terbatas. Cuaca memiliki jangka waktu yang terbatas sehingga kondisi cuaca senantiasa berubah dari waktu ke waktu. Keadaan cuaca sendiri dipengaruhi oleh temperatur, cahaya matahari, kelembaban, kecepatan angin, curah hujan dan masih banyak lagi parameter yang mempengaruhi cuaca. Indonesia sebagai salah satu negara beriklim tropis tentunya memiliki

karakteristiknya tersendiri terkait pola cuaca. Perubahan cuaca di Indonesia yang sangat drastis mengakibatkan banyak wilayah di Indonesia yang turut merasakan dampaknya. Dampak dari perubahan cuaca yang ekstrim dapat berupa banjir, tanah longsor, dan timbulnya wabah penyakit. Penelitian-penelitian terkait perubahan cuaca terus berkembang dari waktu ke waktu. Manusia telah mencoba untuk memprediksi cuaca secara informal selama ribuan tahun, dan secara formal setidaknya sejak abad kesembilan belas (Craft, 2003). BPNN merupakan salah satu arsitektur jaringan syaraf tiruan yang memiliki proses pembelajaran maju dan koreksi kesalahan secara mundur. Model jaringan ini banyak digunakan baik itu untuk proses pengenalan, prediksi, dan peramalan dengan tingkat akurasi yang cukup baik (Dewi & Muslikh, 2013). Pada penelitian yang dilakukan oleh penulis ini, akan membahas terkait pengimplementasian dari metodologi BPNN dalam memprediksi cuaca.

II. KAJIAN TEORI

2.1 Data Mining

Data Mining merupakan proses dalam mencari hubungan dan pola yang tersembunyi pada suatu data dalam jumlah yang besar. (Holsheimer & Siebes, 1994). Analisis *Data Mining* dikembangkan dengan orientasi terhadap volume data yang besar, dengan memanfaatkan data sebanyak mungkin sehingga sampai pada kesimpulan dan keputusan yang dapat andal (Nagabhushana, 2006). Sedangkan keenam proses dari *Data Mining* adalah sebagai berikut menurut Fayyad et al (2015) :

1. Seleksi, merupakan pengelompokkan data berdasarkan beberapa kriteria

2. Preprocessing adalah tahap pembersihan data di mana informasi tertentu dihapus yang dianggap tidak perlu dan dapat memperlambat proses. Selain itu data dikonfigurasi ulang untuk memastikan format yang konsisten karena ada kemungkinan inkonsisten pada format data karena data diambil dari beberapa sumber.
3. Transformasi, data di transformasi dan ditambahkan *overlays* apabila dibutuhkan.
4. *Data Mining*, tahapan ini berkaitan dengan ekstraksi pola dari data.
5. Interpretasi dan evaluasi, pola yang telah didapat dari proses *Data Mining* diidentifikasi oleh sistem ditafsirkan menjadi pengetahuan yang kemudian dapat digunakan untuk mendukung pengambilan keputusan manusia.

Tujuan dari dilakukannya *Data Mining* adalah untuk menghasilkan prediksi dan deskripsi. Prediksi melibatkan penggunaan beberapa variabel atau bidang dalam database untuk memprediksi nilai di masa depan dari variabel yang diteliti.

2.2 Machine Learning

Machine Learning merupakan automasi dalam proses pembelajaran. Pembelajaran disini berupa konstruksi aturan berdasarkan pengamatan dan transisi. Algoritma pada *Machine Learning* menggunakan set data dan informasi sebagai masukan. Kemudian *Machine Learning* akan memeriksa berdasarkan kasus-kasus yang telah diolah sebelumnya sehingga akan mereproduksi pengetahuan terkait kasus-kasus baru. Pada umumnya sistem *Machine Learning* tidak menggunakan data tunggal dalam proses pembelajarannya, namun menggunakan beberapa set data yang disebut *training set*. (Zaccone et al., 2017)

Pada penelitian ini, metode yang digunakan adalah *supervised learning* dikarenakan algoritma yang digunakan adalah *backpropagation neural network*. Algoritma *backpropagation neural network* sendiri merupakan salah satu dari pembelajaran *neural network* yang merupakan salah satu algoritma klasifikasi. Algoritma klasifikasi ini tergolong pada metode *supervised learning*, dikarenakan terdapat variabel target pada model data yang dibuat pada penelitian kali ini.

2.3 Supervised Learning

Supervised Learning merupakan metode pembelajaran bagi *machine* untuk memproses *input* menjadi *output* berdasarkan kelas-kelas atau kasus-kasus yang telah ditentukan sebelumnya (Friedman, 2008). Sistem pembelajaran ini seperti kegiatan belajar antara guru murid, dimana seorang guru membantu sistem membangun model dengan mendefinisikan kelas dan menyediakan contoh masing-masing kelas. Sistem harus menemukan deskripsi dari setiap kelas yaitu properti umum dicontoh (Nagabhushana, 2006).

Dalam *supervised learning* sendiri terdapat banyak algoritma yang dapat digunakan dalam memprediksi sesuatu, diantaranya adalah *decision trees*, *decision rules*, *neural network* dan *bayesian network* (Zaccone et al., 2017).

2.4 Metode Forecasting Time Series

Forecasting merupakan kegiatan melakukan peramalan terkait apa yang akan terjadi di masa yang akan datang dengan menggunakan data-data yang telah ada sebelumnya. *Forecasting* memiliki banyak manfaat bagi semua aspek yang terlibat didalamnya. Pada bidang jual-beli, *Forecasting* digunakan untuk memprediksi penjualan, penawaran dan permintaan. Hal ini dapat digunakan oleh pelaku usaha dalam mengambil keputusan terkait bisnis mereka. berdasarkan (Heizer & Render, 2009),

Pada penelitian ini, proses prediksi akan dilakukan untuk memprediksi kumpulan model data *time series*. *Time series* adalah rangkaian data yang berupa nilai pengamatan yang diukur selama kurun waktu tertentu, berdasarkan waktu dengan interval yang sama (Murray R. Spiegel, 1975). Sedangkan dataset disini dapat dihitung berdasarkan satuan waktu, jumlah pengamatan hasil penelitian kimia, dan sebagainya. penerapan *Time Series* juga sangat bermanfaat dalam berbagai bidang, baik itu ekonomi, sains, bisnis, Teknik. *Time Series* digunakan untuk melakukan analisis yang memungkinkan pengguna untuk mengetahui dan memahami apa yang telah terjadi di masa lampau serta memprediksi apa yang akan terjadi di masa yang akan datang (Hornik, 2017).

2.5 Backpropagation Neural Network (BPNN)

Backpropagation adalah salah satu algoritma pada jaringan syaraf tiruan yang sering digunakan dalam mencari bobot optimal. Pada jaringan *backpropagation* terdapat pola *input* dan pola *output* yang diinginkan. Pada saat jaringan diberikan suatu pola, nilai bobot- bobot diubah agar dapat memperkecil perbedaan antara pola *output* dari jaringan dan pola *output* yang diinginkan. Pelatihan jaringan dilakukan berulang-ulang sampai semua pola *output* dari jaringan dapat mengenali pola *output* yang diinginkan.

Menurut Hansun (2013) *Backpropagation* merupakan salah satu metode pelatihan Jaringan Syaraf Tiruan (JST) terawasi (Supervised Learning) yang terdiri atas 3 langkah utama, yakni:

1. Data dimasukkan dalam *input* jaringan (*feed forward*).
2. Perhitungan dan propagasi balik dari *error* yang ditemukan (*backpropagation*).
3. Pembaruan bobot dan bias.

Backpropagation merupakan pengembangan dari algoritma *single layer network* yang memiliki 2 *layer* saja, yaitu *input layer* dan *output layer*. Sedangkan untuk *backpropagation* sendiri disusun berdasarkan 3 *layer*, yaitu *input*, *hidden* dan *output layer*. Dengan adanya *hidden layer* meningkatkan akurasi dari model yang disusun dengan *backpropagation* (Kholis & Rofii, 2017).

2.6 Mean Squared Error (MSE)

Mean squared error (MSE) merupakan salah satu metode yang digunakan untuk mengetahui besarnya kesalahan rata-rata dari hasil peramalan (Hansun, 2013), dengan rumus perhitungan.

$$MSE = \frac{\sum_{t=1}^n (A_t - F_t)^2}{n}$$

Dimana:

MSE = Mean Squared Error

A_t = nilai aktual

F_t = nilai hasil prediksi

n = banyaknya data

Dimana menyatakan nilai dari *Mean squared error* (MSE) dihasilkan dari deret nilai n aktual yang dikurangi nilai n hasil prediksi. Dari operasi pengurangan ini akan dikuadratkan dan hasilnya dibagi dengan banyaknya data.

III. METODE

Terdapat tiga tahapan utama dalam menyelesaikan penelitian ini. Yaitu pendahuluan, *preprocessing* dan *processing data*. Pendahuluan merupakan tahap awal yang melatar belakangi terbentuknya penelitian ini. Dalam tahapan pendahuluan ini berisi empat proses, diawali dengan menentukan latar belakang penelitian ini. Kemudian dilanjutkan dengan perumusan masalah, menentukan tujuan penelitian dan menetapkan batas-batas dalam penelitian yang akan dikerjakan. Tahapan selanjutnya adalah *preprocessing* yang berisi empat proses, yaitu proses mencari dan mengambil data dari sumber data kemudian menyesuaikan data agar data dapat di *running* dalam *code*, mengecek kolom yang kosong dan mengecek nilai pencilan atau *outliers*. Langkah terakhir pada *preprocessing* adalah mengeliminasi *missing value* dan *outliers*. Tahapan terakhir dari penelitian ini adalah *processing data*. Pada tahapan ini berisi lima proses yang harus dikerjakan. Yaitu membagi data menjadi data *train* dan data *testing*, melakukan pelatihan model dengan data *training*, dilanjutkan dengan menerapkan model pada data *testing*, menghitung *mean squared error*, dan yang terakhir menampilkan data hasil dari *forecast*.

IV. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Data yang digunakan pada penulisan penelitian ini berasal dari data pengamatan harian Badan Meteorologi, Klimatologi dan Geofisika yang terletak di Stasiun Geofisika Bandung. Data yang digunakan adalah data pada sepanjang tahun 2018 dan 2021. Data yang didapat berupa data berformat *xlsx*. Data yang diambil adalah data yang merepresentasikan parameter cuaca yaitu temperatur, udara, angin, kelembapan, penyinaran. Data yang diunduh berupa *raw* data yang nantinya akan disesuaikan kolom dan tabelnya sehingga menghasilkan data seperti berikut.

Tabel 1 Data Kecepatan Angin

| Hari | Jul | Agu | Sep | Okt |
|------|-----|------|------|------|
| 1 | 23 | 24,3 | 24,5 | 25,2 |
| 2 | 23 | 24,1 | 24,1 | 24,8 |
| 3 | 23 | 23 | 24,4 | 24 |
| 4 | 23 | 22,9 | 23,5 | 24,2 |
| 5 | 24 | 22,8 | 24,1 | 24,4 |
| 6 | 23 | 22,9 | 24,2 | 24,5 |
| 7 | 22 | 22,6 | 23,9 | 24,7 |
| 8 | 22 | 23,4 | 23,9 | 24 |
| 9 | 23 | 23,8 | 23,5 | 24,6 |
| 10 | 23 | 24,6 | 23,3 | 24,5 |

Tabel 1 berisi tentang sampel data kecepatan angin rata-rata pada tahun 2018 dan 2021. Total data yang diambil sebanyak 1460 data. Kolom paling kiri merupakan yang menunjukkan hari dalam bulan yang direpresantasikan pada

kolom-kolom berikutnya. Satuan yang digunakan pada kecepatan angin rata-rata merupakan meter per detik(m/s). Format data yang sama digunakan untuk semua parameter cuaca. Untuk kelembapan pada tahun 2018 dan 2021, total data yang diambil sebanyak 1460 data. Satuan yang digunakan pada kelembapan merupakan persen(%). Untuk data lamanya penyinaran matahari pada tahun 2018 dan 2021, total data yang diambil sebanyak 1460 data. Satuan yang digunakan pada lamanya penyinaran matahari merupakan jam. Untuk data perubahan suhu pada tahun 2018 dan 2021. Total data yang diambil sebanyak 1460 data. Satuan yang digunakan pada perubahan suhu merupakan derajat Celsius ($^{\circ}\text{C}$).

4.2 Cleansing Data

Proses *cleansing data* merupakan proses membersihkan data agar data dapat diproses tanpa terdapat *error* dan agar hasil prediksi lebih akurat. Terdapat dua tahap pada *cleansing data* yaitu menghapus kolom yang kosong dan mengecek dan menghapus nilai pencilan. Tahap pertama dilakukan proses *cleansing* untuk menghilangkan nilai kosong pada data tersebut dan dapat digunakan dalam analisis data. Hasil dari proses ini merupakan dataset tanpa adanya baris yang kosong. Hasil dari penghapusan *missing value* menghasilkan data yang bersih dari nilai kosong. Kemudian akan dilakukan proses mengecek nilai pencilan atau *outliers*. Pengecekan nilai pencilan ini melihat apakah terdapat nilai yang ekstrim dari sebuah kumpulan data. Nilai yang ekstrim ini nantinya akan membuat hasil prediksi menjadi tidak akurat. Pembersihan nilai outlier ini akan dilakukan dengan metode IQR. Metode IQR merupakan singkatan dari *Interquartile Range*. Metode ini akan menghapus nilai pencilan dengan nilai yang sudah ditentukan oleh rumus IQR. nilai IQR didapat dengan mengurangi nilai kuartil 3 atau Q3 dengan nilai kuartil 1 atau Q1. Kemudian nilai pencilan akan dieliminasi dengan menggunakan rumus IQR tersebut. Rumus IQR ini akan diterapkan di setiap tahun data yang digunakan. Sehingga data yang dihasilkan bersih dari *missing value* untuk setiap tahunnya. Langkah ini merupakan tahapan terakhir dari pembersihan data. Sebelum dilakukan jumlah data untuk setiap parameter berjumlah sama yaitu 1447 data. Namun dalam tersebut masih terdapat *missing value* dan nilai pencilan Namun setelah dilakukan *cleansing*, jumlah data menjadi bervariasi yang dikelompokkan berdasarkan parameternya. Berikut adalah tabel jumlah data setelah di *cleansing*.

Tabel 2 Jumlah data setelah *cleansing*

| Parameter | Jumlah data setelah <i>cleansing</i> |
|-----------------------------|--------------------------------------|
| Temperatur rata-rata | 1404 |
| Kecepatan angin | 1260 |
| Lamanya penyinaran matahari | 1404 |
| kelembapan | 1344 |

Tabel 2 merupakan tabel yang merepresentasikan jumlah data setelah mengalami proses *cleansing* untuk keempat parameter yang digunakan.

4.3 split Data

Data yang sudah di *cleansing* selanjutnya dibagi menjadi data *training* dan data *testing*. Berikut adalah sampel data yang telah dilakukan pembagian.

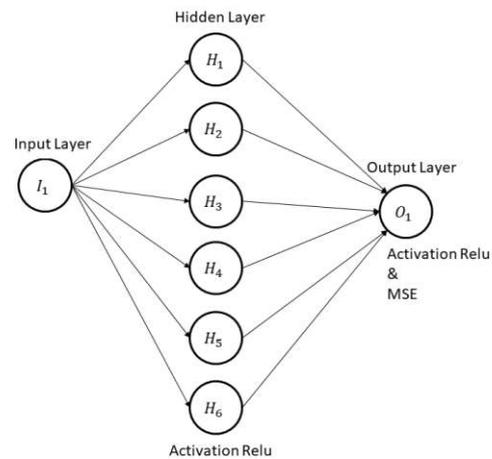
Tabel 3 Split Data

| suhu 2018 | Suhu 2019 | suhu 2020 | suhu 2021 |
|-----------|-----------|-----------|-----------|
| 23,5 | 23 | 22,3 | 23,3 |
| 23,1 | 24 | 23,6 | 24,9 |
| 23,6 | 24,1 | 24,8 | 24,1 |
| 22,9 | 24,2 | 24,4 | 23 |
| 23,5 | 24,8 | 24,1 | 23,4 |

Tabel 3 merupakan sampel data yang telah dibagi menjadi 4 kategori data yaitu data berdasarkan tahun. Dari keempat kategori tersebut nantinya akan dibagi lagi menjadi 2 yaitu variabel *x* dan *y*. Variabel *x* digunakan sebagai *data source*. Sedangkan variabel *y* digunakan untuk membuat label atau data target. Hal ini dilakukan untuk membuat model data *time series*. Variabel *x* berisi data dari 1 Januari 2018 sampai dengan 30 Desember 2021. Sedangkan untuk variabel *y* berisi data dari 2 Januari 2018 sampai dengan 31 Desember 2021. Selanjutnya variabel *x* dan *y* ini akan di split lagi dengan komposisi 75% data *training* dan 25% data *testing*. Komposisi pembagian data *split* ini dibuat berdasarkan pada tahun. Dikarenakan data *time series* terkait cuaca bergantung dari pola musiman. Sehingga *cut-off* antara *training* dan *testing* didasarkan pada tahun, yaitu untuk *training* adalah data 2018 sampai 2020 dan *testing* adalah data tahun 2021. Dari proses pembagian data ini akan menghasilkan 4 variabel yaitu *x_train*, *y_train*, *x_test* dan *y_test*. Parameter *x_train* dan *y_train* nantinya akan digunakan dalam membuat model algoritma. Sedangkan *x_test* akan digunakan untuk menguji apakah model algoritma yang telah dibuat akurat dan sesuai. Hasil dari *x_test* berupa variabel *ypredik* akan dibandingkan dengan *y_test* untuk mengetahui apakah hasil prediksi sesuai dengan data aktualnya.

4.4 Pembuatan Model *Backpropagation Neural Network* (BPNN)

Algoritma *backpropagation neural network* dibangun berdasarkan konsep *neural network* yang mengalami dua tahap proses. Yaitu alur maju dan alur mundur. Kedua tahapan ini digunakan dalam proses melatih data. Dari proses *training* terhadap jaringan dilakukan, akan diperoleh nilai bobot dan bias terbaik yang membentuk suatu model. Dan nantinya model ini akan diterapkan pada data *testing* untuk menghasilkan suatu prediksi. Berikut merupakan struktur model algoritma yang digunakan untuk membuat model.



Gambar 1 Model *Backpropagation*

Gambar 1 merupakan model *backpropagation neural network* (BPNN) yang digunakan pada penelitian ini. Konsep utama pada algoritma *Backpropagation Neural Network* (BPNN) adalah terdapat 3 layer yang membentuk struktur. Yaitu *input layer*, *hidden layer* dan *output layer* (Kholis & Rofii, 2017). Pada BPNN data akan melalui alur maju (*feed forward*) dan alur mundur (*backpropagation*). Pada alur mundur ini, akan dilakukan pengecekan *error* dengan metode *Mean Square Error* (MSE). Pada model di atas arsitektur BPNN dibangun berdasarkan 1 *node input layer*, 6 *node hidden layer*, dan 1 *node output layer*. Hal ini dibuat berdasarkan banyaknya atribut yang digunakan pada penelitian yang mana dilakukan per parameter cuaca. Jumlah *hidden layer* yang ideal adalah maksimal dua kali lipat lebih besar dari *input layer*. Namun disini penulis menguji menggunakan 5 variabel *node*. Yaitu 2 sampai dengan 6 *node*. Berikut adalah tabel perbandingan jumlah *node* terhadap tingkat akurasi model algoritma dengan menggunakan MSE.

Tabel 4 Perbandingan jumlah *node* pada *hidden layer*

| Jumlah <i>node</i> pada <i>hidden layer</i> | MSE |
|---|---------------------|
| 2 | 0,47302626184063445 |
| 3 | 0,46991285451914394 |
| 4 | 0,47045214754523146 |
| 5 | 0,47914185198715065 |
| 6 | 0,46968934902058357 |

Pada Tabel 4 menjelaskan tentang pengaruh banyaknya jumlah *node* pada *hidden layer* terhadap tingkat akurasi MSE. terlihat pada tabel tersebut adanya hubungan linear. Semakin banyak jumlah *node* menentukan akurasi MSE yang dihasilkan. Namun dengan bertambahnya jumlah *node* akan membuat waktu *running* semakin lebih lama serta menggunakan *resource* komputasi yang lebih besar. Sehingga dipilihlah arsitektur dengan 6 *hidden node*. Proses pelatihan menggunakan 10000 *epochs* dan dengan menggunakan *activation* RELU. *Output* yang diharapkan pada proses ini adalah hasil numerik terkait prediksi cuaca berdasarkan parameter yang dihitung. Selain itu pada proses ini juga menghasilkan nilai MSE. nilai MSE ini menunjukkan seberapa efektif model algoritma yang sudah dibangun.

Pemilihan metode MSE ini didasarkan pada kesesuaian metode MSE dengan penelitian ini. MSE cocok digunakan untuk mengevaluasi model prediksi dan peramalan. Semakin nilai MSE mendekati nol, menunjukkan seberapa baik dan efektif sebuah model algoritma. Proses *training data* ini menggunakan *library* dari Tensorflow. *Library* ini digunakan untuk melatih dan menjalankan *machine learning* utamanya adalah *neural network*. Model yang digunakan pada proses adalah TF.keras yaitu model pada Tensorflow untuk membuat dan melatih *machine learning*.

4.5 Menghitung Loss dengan Mean Square Error (MSE)

Tahap terakhir dalam memproses data adalah menghitung Mean Squared Error (MSE). Mean square error menghitung rata-rata kesalahan kuadrat pada nilai aktual dan nilai forecast. Proses menghitung MSE ini masih menggunakan library dari Scikit-learn. Dengan menghitung kesalahan pada model algoritma memungkinkan penulis untuk menghitung seberapa efektif model algoritma *backpropagation neural network* ini dalam membuat prediksi terhadap cuaca. Nilai mean squared error (MSE) dihasilkan dari perbandingan antara nilai aktual (y_{test}) dan hasil prediksi (y_{predik}). Berikut adalah hasil nilai MSE untuk setiap parameter.

Tabel 5 Hasil dari MSE untuk setiap parameter cuaca

| Parameter cuaca | Nilai MSE |
|-----------------------------|---------------------|
| Suhu | 0,46968934902058357 |
| Kelembapan | 25,83351323331317 |
| Kecepatan angin | 0,43786243260248814 |
| Lamanya penyinaran matahari | 6.4057194502727315 |

Tabel 5 merupakan hasil nilai MSE yang didapatkan dari menjalankan model yang telah dibuat di setiap parameter. Untuk setiap parameter menghasilkan nilai yang berbeda-beda. Hal ini disebabkan oleh banyak faktor, diantaranya adalah range nilai yang tinggi, simpangan yang besar dan flutuatifnya suatu data.

4.6 Hasil Prediksi

4.6.1 Hasil prediksi kelembapan

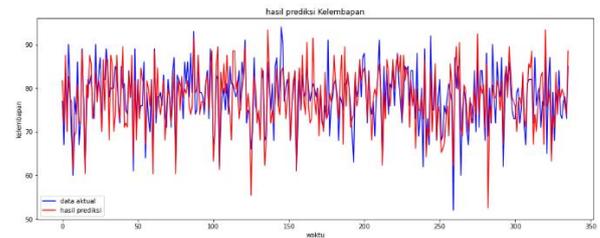
Data yang dihasilkan pada *running* model pada penelitian ini adalah data excel yang berisi data prediksi terkait kelembapan dalam satuan persen. Berikut adalah sampel data hasil prediksi pada kelembapan.

Tabel 6 Sampel data hasil prediksi kelembapan

| Data Aktual(%) | Hasil prediksi(%) |
|----------------|-------------------|
| 77 | 81,71463776 |
| 67 | 71,98764038 |
| 84 | 87,55082703 |
| 74 | 70,04223633 |
| 90 | 82,68733215 |

Tabel 6 merupakan sampel data hasil prediksi kelembapan. Nantinya data hasil prediksi ini akan dibandingkan dengan data aktual pada parameter kelembapan. Membandingkan kedua data ini akan dilakukan dengan visualisasi menggunakan line chart. Dengan sumbu x merupakan *time series* dalam satuan hari. Sedangkan sumbu

y merupakan tingkat kelembapan dalam satuan persen. Berikut hasil dari visualisasi data kelembapan.



Gambar 2 Data hasil prediksi pada parameter kelembapan

Gambar 2 menjabarkan gambaran umum data perbandingan antara data aktual dan data hasil prediksi. Dari data prediksi tersebut menghasilkan angka rata-rata kelembapan di Bandung untuk periode tahun 2021 berkisar di angka 77,77 persen. Hasil ini berbanding dengan data aktual yang memiliki rata-rata kelembapan di angka 77,94 persen. Sedangkan untuk standar deviasi data hasil prediksi berkisar di angka 6,98. Standar deviasi pada data aktual yang berada di angka 6,84. Tingginya standar deviasi menggambarkan perubahan kelembapan harian yang cukup tinggi di Bandung pada setiap harinya.

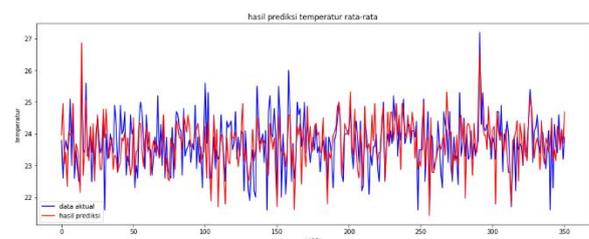
4.6.2 Hasil Prediksi Parameter Suhu

Semakin tinggi suhu menandakan semakin panas cuaca di suatu wilayah. Pada parameter suhu atau temperatur ini yang digunakan merupakan temperatur rata-rata harian di Bandung. Data hasil prediksi pada parameter suhu adalah sebagai berikut.

Tabel 7 Sampel data hasil prediksi suhu

| Data Aktual(°C) | Hasil prediksi(°C) |
|-----------------|--------------------|
| 23,8 | 23,96328163 |
| 22,6 | 24,95898819 |
| 23,5 | 23,05809021 |
| 23,8 | 23,42016602 |
| 23,5 | 22,3339386 |

Tabel 7 merupakan sampel data hasil prediksi pada parameter suhu. Data yang dihasilkan menggunakan satuan Celsius. Nantinya data tersebut akan dibandingkan dengan data aktual yang terjadi. Data prediksi tersebut akan dibandingkan untuk setiap parameter masing-masing. Untuk visualisasi sendiri menggunakan library matplotlib. Berikut adalah hasil dari visualisasi prediksi dengan *backpropagation neural network* untuk parameter suhu.



Gambar 3 Data hasil prediksi parameter suhu

Gambar 3 merupakan visualisasi dari parameter suhu di Bandung pada tahun 2021 berdasarkan pembuatan model di

tahun 2018 sampai dengan 2020. Garis merah menggambarkan data hasil prediksi dan garis biru menggambarkan data aktual. Hasil visualisasi ini menggambarkan data keseluruhan terkait temperatur rata-rata di Bandung dalam satuan Celsius. Dari data tersebut dapat diambil informasi bahwa temperatur rata-rata hasil prediksi berada di angka 23,70°C. Sedangkan pada data aktual temperatur rata-rata berada di angka 23,69°C. Terlihat pada visualisasi terdapat nilai-nilai yang selalu berubah secara ekstrim pada sumbu waktu. Namun pada nilai rata-rata hanya terdapat perbedaan 0,01°C. Ini menandakan bahwa perubahan suhu harian di Bandung cukup ekstrim. Walaupun terlihat pada grafik terdapat perubahan nilai yang sangat fluktuatif, Namun, nilai prediksi ini sudah mendekati nilai aktual.

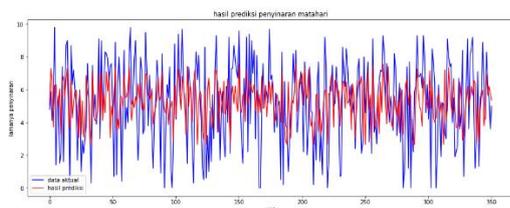
4.6.3 Hasil Prediksi Parameter Penyinaran Matahari

Hasil pada parameter penyinaran matahari adalah data dalam bentuk tabel. Tabel ini hanya beratribut tunggal yang berisi data tentang data rata-rata lamanya penyinaran matahari di Bandung. Lamanya penyinaran matahari menggambarkan durasi suatu wilayah mendapatkan paparan sinar matahari. Hal ini dapat dipengaruhi oleh keberadaan awan dan musim di wilayah tersebut. Penyinaran matahari juga akan mempengaruhi parameter-parameter cuaca lainnya seperti suhu dan kelembapan di suatu wilayah. Lamanya penyinaran dihitung dalam satuan jam. Dan lamanya penyinaran ini juga dihitung per harinya dalam mengumpulkan data. Berikut adalah sampel data yang dihasilkan dari *running* model untuk parameter lamanya penyinaran matahari.

Tabel 8 Data terkait lamanya penyinaran matahari

| Data Aktual(jam) | Hasil prediksi(jam) |
|------------------|---------------------|
| 4,8 | 5,062355518 |
| 5,9 | 7,313557148 |
| 4,1 | 6,513129711 |
| 5,4 | 3,711634398 |
| 9,8 | 6,262996197 |

Pada Tabel 8 menjelaskan tentang sampel data harian pada parameter lamanya penyinaran matahari. Lamanya penyinaran ini dihitung berdasarkan jam per harinya di Bandung. Hasil data prediksi ini nanti akan dibandingkan dengan data aktual yang sebenarnya terjadi untuk melihat seberapa akurat hasil prediksi yang dihasilkan pada model yang telah dibuat dengan menggunakan *backpropagation neural network*. Grafik perbandingan akan ditampilkan menggunakan *line chart* dengan menggunakan *library matplotlib*. Berikut adalah grafik perbandingan pada parameter lamanya penyinaran matahari.



Gambar 4 Hasil prediksi lamanya penyinaran matahari

Gambar 4 merupakan hasil dari perbandingan antara data prediksi dan data aktual yang sebenarnya terjadi pada parameter lamanya penyinaran matahari. Pada data prediksi, rata-rata lamanya penyinaran matahari di Bandung berada di angka 5,26 jam perhari. Sedangkan untuk data aktual menunjukkan lamanya penyinaran matahari di angka 5,19 jam perharinya. Ini berarti selisih rata-rata dari data prediksi dan data aktual sebesar 0,07 jam. Namun pada standar deviasi menunjukkan perbedaan yang cukup signifikan. Standar deviasi merupakan simpangan data dari rata-rata yang ada. Pada standar deviasi untuk data prediksi berada di angka 1,24 jam. Sedangkan pada data aktual menunjukkan standar deviasi yang lebih tinggi di angka 2,55 jam. Selain itu terlihat pada grafik menunjukkan data aktual yang memiliki simpangan yang hampir selalu lebih ekstrim dari data hasil prediksi. Hal ini menandakan bahwa, lamanya penyinaran matahari di Bandung mengalami perubahan yang signifikan pada setiap harinya.

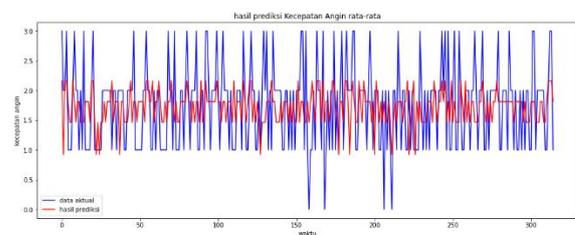
4.6.4 Hasil Prediksi Parameter Kecepatan Angin

Kecepatan angin merupakan salah satu parameter cuaca yang berpengaruh pada jenis aktivitas yang dilakukan manusia di wilayah yang terdampak. Parameter kecepatan angin juga berdampak pada kecepatan pergerakan awan dari satu wilayah ke wilayah lainnya. Kecepatan angin pada penelitian ini dihitung dalam satuan *meter/second* atau disingkat (m/s). satuan ini menggambarkan kecepatan angin dalam meter yang lewat setiap detiknya. Pada pengaplikasian model *backpropagation* ini akan menghasilkan data berupa excel sebagai berikut.

Tabel 9 Data hasil prediksi kecepatan angin

| Data Aktual(m/s) | Data hasil prediksi(m/s) |
|------------------|--------------------------|
| 3 | 2,164396524 |
| 2 | 0,919819593 |
| 2 | 2,164396524 |
| 3 | 2,164396524 |
| 1 | 1,814352632 |

Tabel 9 menjabarkan data prediksi untuk parameter rata-rata kecepatan angin yang berhembus setiap detiknya. Data hasil prediksi yang mendekati data aktual yang sebenarnya terjadi menggambarkan bahwa model algoritma yang dibuat memiliki akurasi yang baik. Berikut adalah visualisasi yang digunakan untuk menggambarkan perbandingan pada data hasil prediksi dan data aktual yang sebenarnya terjadi.



Gambar 5 Hasil prediksi kecepatan angin rata-rata

Gambar 5 menggambarkan *line chart* antara variabel data hasil prediksi kecepatan angin dengan data yang terjadi. Terlihat bentuk garis yang lurus menandakan bahwa nilai

yang sama pada kecepatan angin. Namun sesekali terdapat garis yang tajam menandakan adanya perubahan kecepatan angin yang drastis di hari tersebut. Dari grafik tersebut menunjukkan bahwa kecepatan angin di Bandung cukup stabil walaupun ada hari-hari tertentu yang hembusan angin terasa sangat cepat atau terkadang tidak ada angin yang berhembus. Untuk rata-rata kecepatan angin berhembus di Bandung ada di angka 1,75 m/s. sedangkan hasil prediksi berada di angka 1,72 m/s. sedangkan untuk standar deviasi hasil prediksi berada di angka 0,28 dan untuk data aktual berada di angka 0,71.

V. KESIMPULAN

Pada penelitian ini digunakan arsitektur dengan menggunakan 3 layer. Pada penelitian ini dilakukan percobaan pada jumlah node pada hidden layer menggunakan 5 variabel, yaitu 2 sampai dengan 6 node. Keefektifan jumlah node diukur dengan nilai *mean squared error* (MSE). Dari percobaan yang dilakukan, didapatkan 6 node sebagai jumlah yang efektif pada hidden layer. Sehingga pada penelitian ini arsitektur model algoritma BPNN dibangun dengan menggunakan 1 node pada input layer, 6 node pada hidden layer dan 1 node pada output layer. Untuk evaluasi dari model algoritma menggunakan MSE.

REFERENSI

- [1] Craft, E. D. (2003). An Economic History of Weather Forecasting. *The American Historical Review*, 37(2), 350.
- [2] Dewi, C., & Muslikh, M. (2013). Perbandingan Akurasi Backpropagation Neural Network dan ANFIS Untuk Memprediksi Cuaca. *Journal of Scientific Modeling & Computation*, 1(1), 7–13.
- [3] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (2015). From Data Mining to Knowledge Discovery in Databases. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9078(3), 637–648.
- [4] Hansun, S. (2013). *Peramalan Data IHSG Menggunakan Metode Backpropagation*. IV(1), 26–30.
- [5] Heizer, J., & Render, B. (2009). *Principios de Administración de Operaciones*. PEARSON EDUCACIÓN.
- [6] Holsheimer, M., & Siebes, A. P. J. M. (1994). *Data Mining: the search for knowledge in databases*.
- [7] Hornik, K. (2017). *Introductory Time Series with R* | Paul S.P. Cowpertwait | Springer.
- [8] Kholis, I., & Rofii, A. (2017). ANALISIS VARIASI PARAMETER BACKPROPAGATION ARTIFICIAL NEURAL NETWORK PADA SISTEM PENGENALAN WAJAH BERBASIS PRINCIPAL COMPONENT. 2(1), 1–12.
- [9] Murray R. Spiegel. (1975). *Schaum's Outline of Theory and Problems of Probability and Statistics.pdf*.
- [10] Nagabhushana, S. (2006). Data Warehousing OLAP and Data Mining. In *NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS*.
- [11] Zaccone, G., Karim, R., & Menshaw, A. (2017). *Deep Learning with TensorFlow*.