

KLASIFIKASI KUALITAS UDARA DI PROVINSI DKI JAKARTA MENGUNAKAN METODE BACKPROPAGATION

Taufik Satria Nugraha¹, Hazrat Zun Asha²

^{1,2}Informatika, Fakultas Teknik, Universitas Jenderal Soedirman, Indonesia

NIM : ¹H1D020028, ²H1D020041

Email: ¹taufik.nugraha@mhs.unsoed.ac.id, ²hazrat.asha@mhs.unsoed.ac.id

(Artikel dikirimkan tanggal : dd mmm yyyy)

Abstrak

Dalam dunia kesehatan, lingkungan yang baik menjadi kebutuhan paling mendasar bagi seluruh umat manusia untuk mempertahankan kehidupan yang layak, oleh karena itu manusia sangat membutuhkan kualitas lingkungan hidup yang baik salah satunya adalah kualitas udara, hal ini karena proses metabolisme yang terjadi pada tubuh makhluk hidup tidak dapat berlangsung tanpa adanya oksigen yang berasal dari udara. Selain oksigen, udara juga memiliki beberapa kandungan zat-zat lain seperti karbon monoksida, karbon dioksida, sulfur dioksida, nitrogen oksida, ozon dan zat lain sebagainya yang apabila sudah melebihi batas wajar dapat menyebabkan masalah serius pada kesehatan manusia. Di kota besar seperti DKI Jakarta, pencemaran udara merupakan salah satu masalah yang cukup serius yang perlu dihadapi oleh masyarakat DKI Jakarta. Berdasarkan situs pengamat kualitas udara aqicn.org tanggal 23 Juli 2019 pukul 17.00, DKI Jakarta menempati urutan 16 dari 100 negara teratas dengan polusi udara terburuk. Oleh karena itu, agar dapat memberikan informasi yang akurat dan cepat perlu adanya sistem untuk mengklasifikasikan kualitas udara, salah satu metode yang dapat digunakan adalah *Backpropagation*. Pada penelitian ini klasifikasi kualitas udara di Provinsi DKI Jakarta menggunakan metode *Backpropagation* berhasil diterapkan dengan hasil pengujian klasifikasi yang sangat baik berdasarkan nilai *loss function* sebesar 0.1658 dan nilai *accuracy* sebesar 0.9614.

Kata kunci: Adam, Backpropagation, Jakarta, Klasifikasi, Pencemaran Udara

AIR QUALITY CLASSIFICATION IN DKI JAKARTA PROVINCE USING BACKPROPAGATION METHOD

Abstract

In the world of health, a good environment is the most basic need for all human beings to maintain a decent life, therefore humans really need a good quality environment, one of which is air quality, this is because the metabolic processes that occur in living bodies are not can take place in the absence of oxygen from the air. Apart from oxygen, air also contains several other substances such as carbon monoxide, carbon dioxide, sulfur dioxide, nitrogen oxides, ozone and other substances which, when exceeding normal limits, can cause serious problems to human health. In big cities like DKI Jakarta, air pollution is a serious problem that needs to be faced by the people of DKI Jakarta. Based on the air quality monitoring website aqicn.org dated July 23 2019 at 17.00, DKI Jakarta ranks 16th out of the top 100 countries with the worst air pollution. Therefore, in order to provide accurate and fast information, it is necessary to have a system to classify air quality, one of the methods that can be used is backpropagation. In this study, the classification of air quality in DKI Jakarta Province using the Backpropagation method was successfully applied with very good classification test results based on a loss function value of 0.1658 and an accuracy value of 0.9614.

Keywords: Adam, Backpropagation, Jakarta, classification, air pollution

1. PENDAHULUAN

Dengan adanya perkembangan teknologi membuat manusia tertarik untuk menciptakan hal-hal baru yang lebih berguna di masa yang akan datang [1]. Salah satu bidang yang sangat penting dalam kehidupan manusia adalah bidang kesehatan [2].

Dalam dunia kesehatan, lingkungan yang baik menjadi kebutuhan paling mendasar bagi seluruh umat manusia untuk mempertahankan kehidupan yang layak, oleh karena itu manusia sangat membutuhkan kualitas lingkungan hidup yang baik salah satunya adalah kualitas udara. Tidak hanya berlaku untuk manusia, makhluk hidup lainnya

seperti hewan dan tumbuhan juga sangat membutuhkan kualitas udara yang baik, hal ini karena proses metabolisme yang terjadi pada tubuh makhluk hidup tidak dapat berlangsung tanpa adanya oksigen yang berasal dari udara [3]. Selain oksigen, udara juga memiliki beberapa kandungan zat-zat lain seperti karbon monoksida, karbon dioksida, sulfur dioksida, nitrogen oksida, ozon dan zat lain sebagainya. Beberapa zat tersebut masih dapat ditoleransi oleh tubuh dalam konsentrasi yang masih di bawah batas wajar, namun apabila sudah lebih dari batas wajar maka dapat menyebabkan masalah serius pada kesehatan seperti penyakit jantung, infeksi pernapasan, stroke, asma hingga penyakit paru-paru.

Laporan dari State of Global Air (SOGA) 2019 yang diterbitkan oleh Health Effect Institute di Amerika dan Universitas British Columbia di Kanada menyebutkan bahwa pencemaran udara merupakan faktor penyebab ke-5 paling besar yang memperpendek usia manusia [4]. Peraturan pemerintah (PP) Tahun 1999 menjelaskan bahwa pencemaran udara merupakan masuknya maupun dimasukkannya zat, energi, serta komponen lainnya ke udara ambien karena aktivitas manusia yang pada akhirnya kualitas udara ambien turun hingga level tertentu sehingga fungsi udara ambien tidak terpenuhi [5]. Beberapa hal yang dapat menyebabkan pencemaran udara diantaranya adalah kendaraan bermotor, pabrik-pabrik industri, pabrik zat kimia, hingga limbah rumah tangga [6].

Pencemaran udara tentu menjadi masalah utama bagi setiap daerah baik perkotaan di negara maju maupun di negara berkembang. Dengan meningkatnya pencemaran udara, maka suatu negara menerapkan indeks standar akan pencemaran udara, salah satunya ISPU [7]. Indeks Standar Pencemaran Udara (ISPU) adalah angka yang tidak mempunyai satuan yang menggambarkan kondisi kualitas udara ambien di lokasi dan waktu tertentu yang dikeluarkan oleh kementerian lingkungan hidup dan kehutanan [8]. ISPU memberikan laporan kualitas udara kepada masyarakat untuk menerangkan seberapa bersih atau tercemarnya kualitas udara dan bagaimana dampaknya terhadap kesehatan setelah menghirup udara tersebut setelah beberapa jam/hari/bulan. Penetapan ISPU ini mempertimbangkan tingkat mutu udara terhadap kesehatan manusia, hewan, tumbuhan, bangunan, dan nilai estetika. Parameter ISPU meliputi Sulfur dioksida (SO₂), Partikulat (PM₁₀), Karbon monoksida (CO), Ozon (O₃), dan Nitrogen dioksida (NO₂) [9].

Di kota besar seperti DKI Jakarta, pencemaran udara merupakan salah satu masalah yang cukup serius yang perlu dihadapi oleh masyarakat DKI Jakarta [10]. Berdasarkan situs pengamat kualitas udara aqicn.org tanggal 23 Juli 2019 pukul 17.00, DKI Jakarta menempati urutan 16 dari 100 negara teratas dengan polusi udara terburuk [11]. Kualitas udara di DKI Jakarta mencapai level 155 US *Air Index Quality* (AIQ) yang mengindikasikan udara di

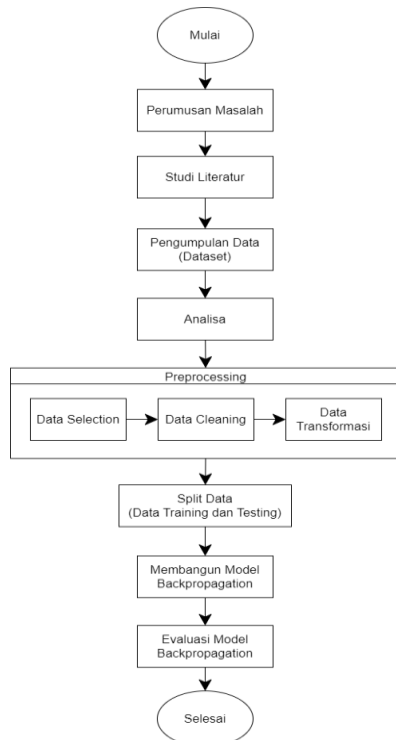
DKI Jakarta tidak sehat. Udara dengan kategori tidak sehat dapat menyebabkan banyak masalah bagi kesehatan manusia. Efek yang ditimbulkan bergantung pada berapa lama seorang individu terpapar polusi dan kerentanan individu tersebut. Contoh yang ditimbulkan seperti mengalami gangguan pernapasan, mengalami vertigo, kejang mendadak, hilang kesadaran, mengaktifkan sel kanker dalam tubuh, bahkan dapat mengakibatkan kematian [12]. Pencemaran udara bila tidak segera teratasi akan berdampak buruk bagi kehidupan makhluk hidup. Sehingga perlu adanya upaya tingkat lanjut guna mengatasi masalah tersebut. Oleh karena itu, Dinas Lingkungan Hidup (DLH) DKI Jakarta membuat stasiun pemantauan kualitas udara di lima kota besar di DKI Jakarta untuk memantau polusi udara seperti Sulfur dioksida (SO₂), Partikulat (PM₁₀), Karbon monoksida (CO), Ozon (O₃), dan Nitrogen dioksida (NO₂) [13].

Pada penelitian yang dilakukan oleh Agus Nurkhozin, Mohammad Isa Irawan, dan Imam Mukhlas dengan judul “Komparasi Hasil Klasifikasi Penyakit Diabetes Mellitus Menggunakan Jaringan Syaraf Tiruan *Backpropagation* dan *Learning Vector Quantization*” menghasilkan kesimpulan bahwa dalam klasifikasi penyakit diabetes mellitus metode *Backpropagation* memberikan kinerja lebih baik dibandingkan LVQ [14]. Kemudian pada penelitian yang dilakukan oleh Aldillah Nurul Hidayah yang berjudul “Penerapan Jaringan Syaraf Tiruan *Backpropagation* Klasifikasi Penyakit ISPA” dengan perbandingan 90% data latih dan 10% data training mendapatkan hasil akurasi tertinggi mencapai 93,33% pada *learning rate* 0,6 [15]. Ini menunjukkan bahwa algoritma *Backpropagation* memiliki akurasi yang cukup tinggi dalam melakukan pengklasifikasian.

Berdasarkan uraian diatas, adanya penelitian ini bertujuan untuk mempermudah dalam mengklasifikasi kualitas udara di DKI Jakarta. Metode yang akan digunakan dalam melakukan klasifikasi adalah metode *Backpropagation*, karena berdasarkan penelitian-penelitian sebelumnya menunjukkan bahwa algoritma *Backpropagation* memiliki akurasi yang cukup tinggi dibandingkan algoritma lainnya.

2. METODE

Metode penelitian adalah tahapan sistematis yang akan dikerjakan pada sebuah penelitian untuk menyelesaikan masalah yang terdapat pada penelitian ini. Oleh karena itu diperlukan sebuah *Flowchart* metodologi penelitian yang akan dilakukan pada penelitian kali ini. *Flowchart* tersebut dapat dilihat pada gambar 1 di bawah ini :



Gambar 1. Flowchart Metodologi Penelitian

2.1. Perumusan Masalah

Perumusan masalah merupakan tahapan awal dari metodologi penelitian. Tahapan ini digunakan untuk merumuskan dari atau memberi kesimpulan dari masalah-masalah yang dicantumkan pada latar belakang di bab sebelumnya. Perumusan masalah dibuat agar lebih paham dengan penelitian yang akan dilakukan nantinya, yaitu tentang bagaimana cara membuat sistem Jaringan Syaraf Tiruan untuk klasifikasi kualitas udara di Provinsi DKI Jakarta menggunakan metode *Backpropagation*.

2.3. Studi Literatur

Tinjauan referensi yang dilakukan berupa tinjauan literatur yang berkaitan dengan permasalahan yang diangkat pada penelitian ini. Pencarian informasi dilakukan secara *online* melalui internet. Informasi yang didapat dapat digunakan dalam penyusunan pendahuluan, landasan teori, metodologi penelitian, serta pada pengaplikasiannya.

2.4. Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh data-data maupun informasi yang berhubungan dengan penelitian yang akan dikerjakan. Mencari data yang tersedia, memperoleh data dan mengintegrasikannya ke dalam dataset termasuk variabel yang dibutuhkan dalam pemrosesan data.

Dataset yang digunakan pada penelitian ini berisi mengenai Indeks Standar Pencemaran Udara (ISPU) yang diukur dari 5 stasiun pemantau kualitas udara (SPKU) yang ada di Provinsi DKI Jakarta

selama tahun 2021. Data ini didapatkan dari bank data portal <https://data.jakarta.go.id/> dengan jumlah data sebanyak 1825 data. Dataset tersebut memiliki beberapa variabel yang akan digunakan pada penelitian ini yaitu pm10 (Partikulat), pm25 (Partikulat), so2 (Sulfida), co (Carbon Monoksida), o3 (Ozon), no2 (Nitrogen Dioksida) dan kategori [16].

Tabel 1. Variabel Yang Ada Pada Dataset

| Variabel | Keterangan |
|----------|---|
| tanggal | Tanggal pengukuran kualitas udara |
| stasiun | Lokasi pengukuran di stasiun |
| pm10 | Partikulat salah satu parameter yang diukur |
| pm25 | Partikulat salah satu parameter yang diukur |
| so2 | Sulfida salah satu parameter yang diukur |
| co | Carbon Monoksida salah satu parameter yang diukur |
| o3 | Ozon salah satu parameter yang diukur |
| no2 | Nitrogen Dioksida salah satu parameter yang diukur |
| max | Nilai ukur paling tinggi dari seluruh parameter yang diukur dalam waktu yang sama |
| critical | Parameter yang hasil pengukurannya paling tinggi |
| kategori | Kategori hasil perhitungan indeks standar pencemaran udara |

2.5. Analisa

Analisa merupakan tahap setelah proses pengumpulan data dilakukan. Tahap ini dilakukan agar dalam tahap selanjutnya lebih terarah sehingga dapat menghasilkan hasil yang lebih baik. Tahap analisa terbagi menjadi 2 bagian, yaitu analisa data dan analisa metode.

2.5.1 Analisa Data

Pada tahap ini dilakukan tahap *preprocessing* data. Ini dikarenakan data yang didapat dari tahap sebelumnya belum bisa untuk langsung digunakan. Pada tahap ini menghasilkan data yang awalnya berjumlah 1825 data menjadi 1517 data. *Preprocessing* yang dilakukan adalah sebagai berikut:

1. Data Selection

Data selection merupakan pemilihan data dari sekumpulan data yang akan dilakukan sebelum tahap penggalian informasi dalam data tersebut dimulai. Dalam tahap ini dilakukan pemilihan atribut yang akan dibutuhkan dalam perhitungan pada metode yang digunakan.

2. Data Cleaning

Data yang diperoleh dari asalnya seringkali memiliki isi yang tidak sempurna seperti ada data yang hilang (*missing value*), data yang tidak valid, atribut yang tidak relevan [17]. Data yang tersebut sebaiknya dibuang atau dibersihkan dengan tujuan agar performa dari metode yang digunakan dapat berjalan maksimal.

3. Data Transformation

Agar dapat menghasilkan hasil perhitungan yang optimal menggunakan metode *backpropagation* maka data perlu dilakukan

normalisasi. Normalisasi data adalah proses membuat beberapa variabel memiliki rentang nilai yang sama tidak ada data yang terlalu besar maupun kecil. Proses normalisasi data pada penelitian ini menggunakan normalisasi Min-Max, Min-Max yaitu metode normalisasi data yang mengubah nilai data menjadi skala mulai dari 0 (min) hingga 1 (max).

$$x_s = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

Keterangan :

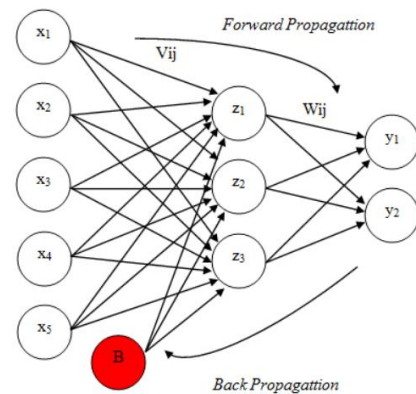
x_s = Data hasil scaling
 x_{\min} = Data minimum
 x_{\max} = Data maksimum

Untuk dapat melakukan pemodelan klasifikasi secara optimal data output perlu dilakukan label encoding, label encoding adalah mengubah setiap nilai dalam kolom menjadi angka yang berurutan.

2.5.1 Analisa Metode

Pada tahap ini merupakan tahap analisa mengenai metode yang akan digunakan yaitu *Backpropagation*. Algoritma *backpropagation* merupakan algoritma untuk melakukan proses pembelajaran terarah pada jaringan syaraf tiruan untuk mencari beban pada setiap neuron yang menghasilkan nilai kesalahan seminimal mungkin melalui data pembelajaran yang diberikan. Berikut adalah tahapan analisa yang akan digunakan, yaitu sebagai berikut :

1. Tahap perambatan maju (*forward propagation*), yaitu pola *input* dihitung secara maju (*feed forward*) mulai dari *input layer* hingga *output layer* dengan menggunakan fungsi aktivasi yang telah ditentukan.
2. Tahap perambatan balik (*backpropagation*) atau disebut proses pengujian, yaitu menghitung selisih antara *output* jaringan dengan target *output* yang diinginkan sebagai *error* yang terjadi, kemudian error tersebut diproporsikan mundur dimulai dari bobot yang berhubungan langsung dengan *node* di *output layer*.
3. Tahap perubahan bobot dan bias yang berfungsi untuk menurunkan *error* yang terjadi.
4. Fungsi aktivasi digunakan untuk menentukan prediksi pada hasil pengujian menggunakan metode *backpropagation*. Dalam penelitian ini, fungsi aktivasi yang digunakan yaitu fungsi aktivasi relu untuk input layer dan hidden layer serta fungsi aktivasi softmax untuk output layer.



Gambar 2. Arsitektur Backpropagation

Keterangan :

$x_1 - x_5$ = Lapisan masukan
 V_{ij} = Bobot ke- pada lapisan tersembunyi
 W_{ij} = Bobot ke- pada lapisan keluaran
 $z_1 - z_3$ = Lapisan tersembunyi
 $y_1 - y_2$ = Lapisan keluaran
 B = Bias

2.6. Implementasi dan Pengujian

Pada tahap implementasi tools yang digunakan adalah Google Collabs dan bahasa yang digunakan adalah Python. Data yang sebelumnya di peroleh kami bagi menjadi 2 untuk dijadikan data training dan data testing dengan perbandingan 80:20.

Pada pemodelan jaringan syaraf tiruan menggunakan model *backpropagation* digunakan fungsi aktivasi ReLU dan Softmax. Rectified Linear Activation Function (ReLU) adalah fungsi linier yang berkisar dari nol hingga tak terbatas, seperti yang ditunjukkan pada Persamaan berikut:

$$R(z) = \max(0, z)$$

di mana z di atas atau sama dengan 0

Fungsi aktivasi Softmax digunakan pada lapisan output untuk menemukan kemungkinan jawaban maksimum untuk masalah klasifikasi, fungsi aktivasi softmax mengacu pada fungsi aktivasi yang menghitung fungsi eksponensial yang dinormalisasi dari setiap unit di lapisan. seperti yang ditunjukkan pada Persamaan berikut:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

di mana σ adalah softmax, Ez adalah vektor, K adalah jumlah kelas.

Pada pemodelan jaringan syaraf tiruan menggunakan model *backpropagation* digunakan juga optimizer *Adam* untuk menghemat memori, mudah diimplementasikan, menangani gradien yang sangat bising atau jarang. Tingkat learning rate dari optimizer *Adam* yang digunakan nilainya 0,001.

Pada pemodelan untuk mengukur tingkat keberhasilan klasifikasi maka perlu dilakukan pengukuran akurasi terhadap metode yang

digunakan. Perhitungan akurasi dilakukan menggunakan *sparse categorical cross-entropy* untuk pengukuran nilai *loss* serta *confusion matrix* untuk menghitung nilai *accuracy*.

Pada penelitian kali ini menggunakan *parse categorical cross-entropy* karena pada penelitian kali ini digunakan *multi-class* klasifikasi. Persamaan matematika dari *sparse categorical cross-entropy* yaitu :

$$J(w) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (2)$$

Keterangan :

w = Bobot
 y_i = Label yang benar
 \hat{y}_i = Label yang diprediksi [18]

Confusion matrix terdiri dari *true positive*, *false positive*, *true negative*, dan *false negative* untuk menghitung *precision*, *recall*, dan *accuracy* [19]. *Precision* merupakan tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. *Recall* merupakan tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Sedangkan *accuracy* merupakan tingkat kedekatan antara nilai prediksi dengan nilai aktual [20]. *Confusion matrix* dasar berdimensi 2x2 yang terdiri dari baris horizontal adalah kelas prediksi dan kolom vertikal adalah kelas sebenarnya. *Confusion matrix* dasar dapat dilihat pada tabel 2 :

| Tabel 2. <i>Confusion Matrix</i> | | | |
|----------------------------------|-----|---------------------|----|
| | | <i>Actual Class</i> | |
| | | Yes | No |
| <i>Predicted Class</i> | Yes | TP | FP |
| | No | FN | TN |

Dari *confusion matrix* tersebut, untuk menghitung *precision*, *recall*, dan *accuracy* menggunakan persamaan berikut :

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

Keterangan :

TP = True positive yang didapatkan dari jumlah data positif yang diprediksi benar
 TN = True negative yang didapatkan dari jumlah data negative yang diprediksi benar
 FP = False positive didapatkan dari jumlah data negative namun diprediksi sebagai data positif
 FN = False negative yang didapatkan dari jumlah data positif namun diprediksi sebagai data negatif

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Hasil pembuatan model Jaringan Syaraf Tiruan untuk klasifikasi kualitas udara di Provinsi DKI Jakarta menggunakan metode *Backpropagation* dengan tahapan awal pengumpulan data, analisa data dengan tahapan *preprocessing* data dan pembangunan model *backpropagation*, terakhir dilakukan pengujian menggunakan *loss function*, *metrics* dan *confusion matrix* dengan hasil pengukuran berupa *accuracy*.

3.1.1 Pengumpulan Data

Dataset yang digunakan untuk klasifikasi kualitas udara di Provinsi DKI Jakarta berjumlah 1825 dengan 11 variabel terdiri dari tanggal, stasiun, pm10, pm25, so2, co2, co, no2, max, critical, dan kategori, Sample data dapat dilihat pada tabel 3.

| Tabel 3. 10 Data Indeks Standar Pencemaran Udara (ISPU) pada Bulan Januari Tahun 2021 | | | | | | | | | | |
|---|-----------------------|------|------|-----|----|----|-----|-----|----------|----------|
| tanggal | stasiun | pm10 | pm25 | so2 | co | o3 | no2 | max | critical | kategori |
| 2021-01-01 | DKI1 (Bunderan HI) | 38 | 53 | 29 | 6 | 31 | 13 | 53 | pm25 | SEDANG |
| 2021-01-02 | DKI1 (Bunderan HI) | 27 | 46 | 27 | 7 | 47 | 7 | 47 | o3 | BAIK |
| 2021-01-03 | DKI1 (Bunderan HI) | 44 | 58 | 25 | 7 | 40 | 13 | 58 | pm25 | SEDANG |
| 2021-01-04 | DKI1 (Bunderan HI) | 30 | 48 | 24 | 4 | 32 | 7 | 48 | pm25 | BAIK |
| 2021-01-05 | DKI1 (Bunderan HI) | 38 | 53 | 24 | 6 | 31 | 9 | 53 | pm25 | SEDANG |
| 2021-01-06 | DKI1 (Bunderan HI) | 41 | 58 | 23 | 13 | 46 | 13 | 58 | pm25 | SEDANG |
| 2021-01-07 | DKI1 (Bunderan HI) | 35 | 47 | 22 | 6 | 39 | 10 | 47 | pm25 | BAIK |
| 2021-01-08 | DKI1 (Bunderan HI) | 37 | 54 | 26 | 16 | 17 | 10 | 54 | pm25 | SEDANG |
| 2021-01-09 | DKI1 (Bunderan HI) | 47 | 61 | 16 | 27 | 22 | 12 | 61 | pm25 | SEDANG |
| 2021-01-10 | DKI1 (Bunderan HI) | 23 | 25 | 16 | 11 | 33 | 8 | 33 | o3 | BAIK |

3.1.2 Preprocessing Data

Tahap *Preprocessing* Data yang pertama adalah dengan *Data Selection* dengan menyeleksi variabel yang dapat digunakan untuk pemodelan, hasil *data selection* dapat dilihat pada Gambar 3.

| | pm10 | pm25 | so2 | co | o3 | no2 | kategori |
|---|------|------|-----|----|----|-----|----------|
| 0 | 63 | 100 | 43 | 13 | 41 | 30 | SEDANG |
| 1 | 35 | 56 | 42 | 7 | 40 | 14 | SEDANG |
| 2 | 54 | 71 | 43 | 9 | 48 | 17 | SEDANG |
| 3 | 50 | 65 | 45 | 13 | 43 | 16 | SEDANG |
| 4 | 53 | 80 | 44 | 17 | 39 | 29 | SEDANG |

Gambar 3. Dataset Hasil Data Selection

Data Cleaning dengan menghapus data yang hilang (*missing value*), hasil *data cleaning* yaitu dataset berkurang menjadi 1517 data, dapat dilihat pada Gambar 4.

| Data columns (total 7 columns): | | | |
|---------------------------------|----------|----------------|--|
| # | Column | Non-Null Count | |
| 0 | pm10 | 1517 non-null | |
| 1 | pm25 | 1517 non-null | |
| 2 | so2 | 1517 non-null | |
| 3 | co | 1517 non-null | |
| 4 | o3 | 1517 non-null | |
| 5 | no2 | 1517 non-null | |
| 6 | kategori | 1517 non-null | |

Gambar 4. Dataset Hasil Data Cleaning

Data Transformation menggunakan metode normalisasi Min-Max dan menggunakan Label Encoder untuk memberikan label pada data output, dapat dilihat pada gambar 5.

| | pm10 | pm25 | so2 | co | o3 | no2 | kategori |
|------|----------|----------|--------|----------|----------|----------|----------|
| 0 | 0.140244 | 0.248447 | 0.3375 | 0.097561 | 0.160839 | 0.161290 | 1 |
| 1 | 0.073171 | 0.204969 | 0.3125 | 0.121951 | 0.272727 | 0.064516 | 2 |
| 2 | 0.176829 | 0.279503 | 0.2875 | 0.121951 | 0.223776 | 0.161290 | 1 |
| 3 | 0.091463 | 0.217391 | 0.2750 | 0.048780 | 0.167832 | 0.064516 | 2 |
| 4 | 0.140244 | 0.248447 | 0.2750 | 0.097561 | 0.160839 | 0.096774 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... |
| 1512 | 0.237805 | 0.391304 | 0.4250 | 0.292683 | 0.090909 | 0.709677 | 1 |
| 1513 | 0.176829 | 0.341615 | 0.2250 | 0.219512 | 0.090909 | 0.483871 | 1 |
| 1514 | 0.115854 | 0.254658 | 0.3250 | 0.146341 | 0.118881 | 0.419355 | 1 |
| 1515 | 0.231707 | 0.385093 | 0.2875 | 0.317073 | 0.104895 | 0.661290 | 1 |
| 1516 | 0.274390 | 0.459627 | 0.3250 | 0.414634 | 0.153846 | 0.806452 | 1 |

Gambar 5. Dataset Hasil Data Transformation

3.1.3 Split Data (Data Training dan Data Testing)

Split Data atau membagi dataset menjadi data *training* dan data *testing* diperlukan sebelum melakukan pemodelan. Hasil dari split data dengan perbandingan data training dan testing 80:20 menggunakan 'test_size' sebesar 0.2 serta menggunakan 'random_state' sebesar 42 untuk menginisialisasi data secara acak, menghasilkan data *training* dengan jumlah data 1213 dan data *testing* dengan jumlah data 304.

```
#Split Data (Data Training dan Data Testing)
xtrain, xtest, ytrain, ytest = train_test_split(x,y, test_size=0.2, random_state=42)

xtrain.shape, xtest.shape, ytrain.shape, ytest.shape

((1213, 6), (304, 6), (1213,)), (304,))
```

Gambar 6. Data Training dan Data Testing

3.1.4 Pembangunan Model Backpropagation

Pembangunan model menggunakan model *backpropagation* dengan model arsitektur Sequential, input layer menggunakan fungsi aktivasi relu dengan jumlah units 6, satu hidden layer menggunakan fungsi aktivasi relu dengan jumlah units 28 serta output layer dengan fungsi aktivasi softmax dengan jumlah units 3. Model *backpropagation* dibangun dari beberapa model compile, menggunakan optimizer 'adam' untuk learning rate, menggunakan loss function 'sparse_categorical_crossentropy' untuk klasifikasi multi-class dan menggunakan metrics 'accuracy' untuk menilai kinerja model.

```
#Model Backpropagation
model = tf.keras.models.Sequential() #Membuat model arsitektur
model.add(tf.keras.layers.Dense(units=6, activation='relu')) #Input
model.add(tf.keras.layers.Dense(units=28, activation='relu')) #Hidden Layer
model.add(tf.keras.layers.Dense(units=3, activation='softmax')) #Output

model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])

hist = model.fit(xtrain,ytrain, epochs=100, batch_size=128, validation_data=(xtest,ytest))
```

Gambar 7. Model Backpropagation

3.1.5 Pengujian Model Backpropagation

Pengujian model *backpropagation* terbagi menjadi pengujian untuk *data training* dan *data testing*, pengujian menggunakan 'epoch' sebanyak 100 epochs dengan menerapkan 'batch size' sejumlah 128 batch size serta untuk pengujian data testing menggunakan 'validation data'. Pengujian pada model menggunakan loss function dengan sparse multiclass cross-entropy loss dikarenakan nilai label klasifikasi pada kasus berjumlah tiga kelas dan menggunakan metrics accuracy untuk perhitungan kinerja model serta menggunakan confusion matrix, pengujian model dilakukan sebanyak 4 kali percobaan untuk mendapatkan hasil yang maksimal.

Tabel 4. Hasil Pengujian Data Training

| Percobaan Ke- | Loss | Accuracy |
|------------------|--------|----------|
| 1 | 0.1617 | 0.9514 |
| 2 | 0.2293 | 0.9233 |
| 3 | 0.1824 | 0.9349 |
| 4 | 0.1592 | 0.9555 |
| Rata-Rata | 0.1832 | 0.9413 |

Tabel 5. Hasil Pengujian Data Testing

| Percobaan Ke- | Loss | Accuracy |
|------------------|--------|----------|
| 1 | 0.1411 | 0.9671 |
| 2 | 0.2142 | 0.9507 |
| 3 | 0.1701 | 0.9572 |
| 4 | 0.1377 | 0.9704 |
| Rata-Rata | 0.1658 | 0.9614 |

Berdasarkan Tabel 4 dan Tabel 5 terlihat Hasil pengujian data training dan data testing melalui nilai loss pada model didapatkan hasil untuk *data training* mendapatkan nilai loss sebesar 0.1832 sedangkan untuk *data testing* mendapatkan nilai loss sebesar 0.1658 dari rata-rata 4 percobaan.

Pada Tabel 4 dan Tabel 5 terlihat Hasil pengujian data training dan data testing melalui nilai metrics mendapatkan hasil *accuracy* untuk *data training* mendapatkan nilai *accuracy* sebesar 0.9413 sedangkan untuk *data testing* mendapatkan nilai loss sebesar 0.9614 dari rata-rata 4 percobaan.

```
#Pengujian (Confusion Matrix)
confusion_matrix(y_true=ytest, y_pred=yprediksi)
```

```
array([[ 47,  3,  0],
       [ 1, 229,  3],
       [ 0,  3, 18]])
```

```
array([[ 43,  7,  0],
       [ 1, 229,  3],
       [ 0,  4, 17]])
```

```
array([[ 44,  6,  0],
       [ 1, 231,  1],
       [ 0,  5, 16]])
```

```
array([[ 46,  4,  0],
       [ 1, 230,  2],
       [ 0,  2, 19]])
```

Gambar 8. Pengujian Model dengan Confusion Matrix

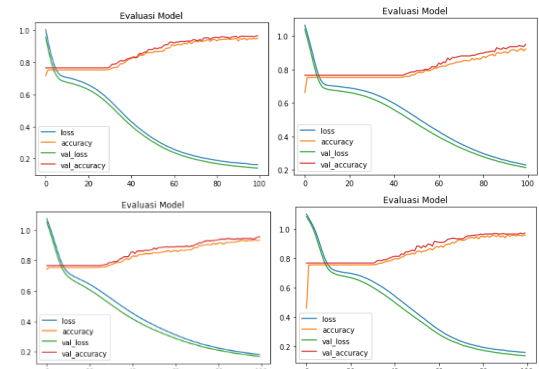
Hasil pengujian model melalui confusion matrix dengan pengukuran *accuracy* dapat dilihat pada gambar 8, terlihat bahwa nilai dari True Positif yang didapat dari 4 percobaan mengalami peningkatan.

3.2 Pembahasan

Berdasarkan hasil pengujian model *backpropagation* didapatkan hasil nilai loss dan accuracy dari pengujian data training serta data testing dijabarkan pada tabel dan grafik berikut.

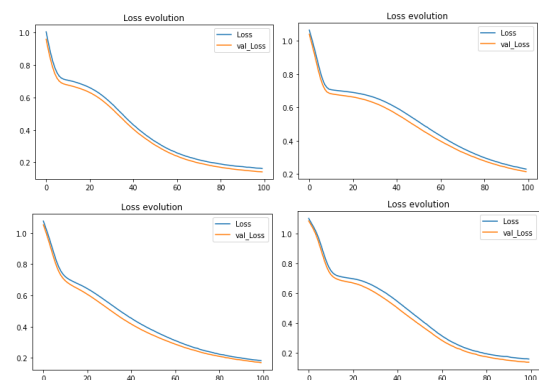
Tabel 6. Hasil Pengujian Model dengan Loss dan Accuracy

| Parameter Pengujian | Data Training | Data Testing |
|---------------------|---------------|--------------|
| Loss | 0.1832 | 0.1658 |
| Accuracy | 0.9413 | 0.9614 |



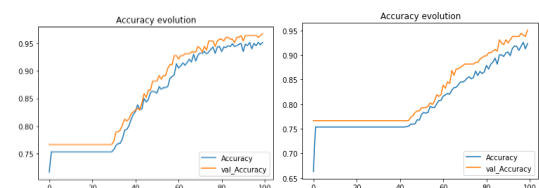
Gambar 9. Hasil Pengujian Model dengan Loss dan Accuracy

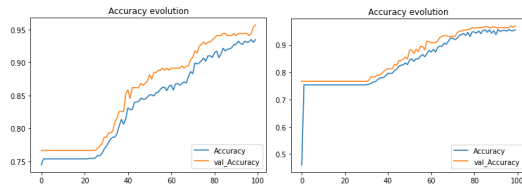
Berdasarkan hasil penerapan model *backpropagation* pada Gambar 10 didapatkan grafik hasil evaluasi nilai loss perbandingan antara nilai loss data training (loss) dan data testing (val_loss) dapat dilihat pada grafik nilai loss dari data testing (val_loss) terlihat lebih rendah dibandingkan dengan nilai loss dari data training (loss). Hasil penerapan model mendapatkan nilai loss 0.1832 dan nilai val_loss adalah 0.1658 semakin rendah nilai loss maka model semakin baik.



Gambar 10. Hasil Pengujian Model dengan Loss

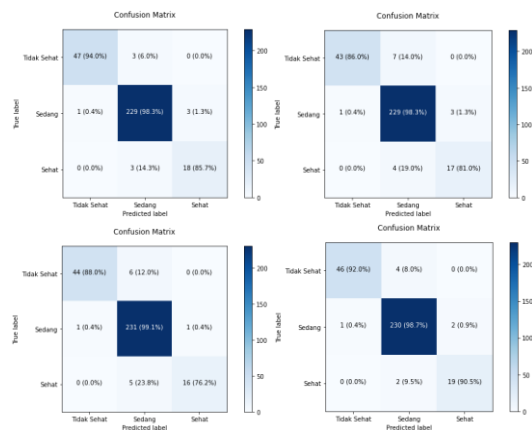
Berdasarkan hasil penerapan model *backpropagation* pada Gambar 11 didapatkan grafik hasil *accuracy* perbandingan antara nilai *accuracy* dari *data training* (accuracy) dan *data testing* (val_accuracy) dapat dilihat pada grafik nilai *accuracy* dari *data testing* (val_accuracy) terlihat lebih rendah dibandingkan dengan nilai *accuracy* dari *data training* (accuracy). Hasil pengujian klasifikasi menggunakan model *backpropagation* mendapatkan nilai *accuracy* 0.9413 dan nilai val_accuracy adalah 0.9614, hasil *accuracy* yang didapat sangat tinggi dan sangat baik..





Gambar 11. Hasil Pengujian Model dengan Accuracy

Pengujian model menggunakan confusion matrix tidak berbeda dengan pengujian menggunakan metrics, pengujian confusion matrix dan metrics sama sama menggunakan indikator pengukuran accuracy tetapi pada confusion matrix dijabarkan hasil dari evaluasi klasifikasi model, dapat dilihat pada gambar 12 hasil pengujian model dengan confusion matrix. pengujian dengan confusion matrix didapatkan hasil accuracy yang meningkat dari percobaan 1 hingga 4 dengan rata rata nilai accuracy sebesar 96.14%.



Gambar 12. Hasil Pengujian Model dengan Confusion Matrix

4. KESIMPULAN

Pemodelan menggunakan metode Backpropagation dengan kasus klasifikasi seperti pada klasifikasi data udara yang memiliki kelas label lebih dari dua atau dapat dikatakan multi class dapat menggunakan sparse categorical cross-entropy untuk pengukuran nilai loss function dan menggunakan optimizer adam untuk mendapatkan hasil pengujian model yang optimal.

Nilai loss function yang didapat dari pemodelan menggunakan metode Backpropagation sebesar 0.1832 untuk data training dan 0.1658 untuk data testing, dengan nilai loss function yang nilainya sudah rendah maka dapat dikatakan model klasifikasi menghasilkan error yang rendah.

Nilai accuracy yang didapat dari pemodelan menggunakan metode Backpropagation sebesar 0.9413 untuk data training dan 0.9614 untuk data testing, dengan hasil accuracy yang sangat baik ditambah hasil data testing sudah lebih besar dari data training maka dapat dipastikan model klasifikasi menggunakan metode Backpropagation sangat baik .

Klasifikasi kualitas udara di Provinsi DKI Jakarta menggunakan metode *Backpropagation* berhasil diterapkan dengan hasil klasifikasi yang sangat baik berdasarkan nilai loss function sebesar 0.1658 dan nilai accuracy sebesar 0.9614.

5. DAFTAR PUSTAKA

- [1] Jamaludin, C. Rozikin, dan A. S. Y. Irawan, "Klasifikasi Jenis Buah Mangga dengan Metode Backpropagation," *Techné: Jurnal Ilmiah Elektroteknika*, vol. 20, no. 1, hlm. 1–12, 2021, doi: <https://doi.org/10.31358/techn.v20i1.231>.
- [2] F. Y. al Irsyadi dan H. Fathina, "Klasifikasi Status Gizi Balita Jenis Kelamin Laki-laki Menggunakan Jaringan Syaraf Tiruan," *Khazanah Informatika*, vol. 1, no. 1, hlm. 16–22, 2015, doi: <https://doi.org/10.23917/khif.v1i1.1177>.
- [3] S. S. A. Umri, M. S. Firdaus, dan A. Primajaya, "Analisis dan Komparasi Algoritma Klasifikasi dalam Indeks Pencemaran Udara di DKI Jakarta," *JIKO (Jurnal Informatika dan Komputer)*, vol. 4, no. 2, hlm. 98–104, 2021, doi: [10.33387/jiko](https://doi.org/10.33387/jiko).
- [4] N. Koesoemaningroem, Endroyono, dan S. M. S. Nugroho, "Peramalan Pencemaran Udara di Kota Surabaya Menggunakan Metode Dsarima dengan Pendekatan Percentile Error Bootstrap (PEB)," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 8, no. 5, hlm. 987–994, 2021, doi: [10.25126/jtik.202185216](https://doi.org/10.25126/jtik.202185216).
- [5] R. Puspitarini dan R. Ismawati, "ANALISIS KIMIA UDARA PT PAPERTECH INDONESIA UNIT II MAGELANG," *Jurnal Sains dan Teknologi Lingkungan*, vol. 14, no. 1, hlm. 22–30, Jan 2022, doi: <https://doi.org/10.20885/jstl.vol14.iss1.art5>.
- [6] M. Yasir, "PENCEMARAN UDARA DI PERKOTAAN BERDAMPAK BAHAYABAGI MANUSIA, HEWAN, TUMBUHAN DAN BANGUNAN," Nov 2021, doi: [10.31219/osf.io/nc5rg](https://doi.org/10.31219/osf.io/nc5rg).
- [7] B. K. Hidayatullah, M. Kallista, C. Setianingsih, P. S1, dan T. Komputer, "PREDIKSI INDEKS STANDAR PENCEMAR UDARA MENGGUNAKAN METODE LONG SHORT-TERM MEMORY BERBASIS WEB (STUDI KASUS PADA KOTA JAKARTA)," *eProceedings of Engineering*, vol. 9, no. 3, hlm. 1247–1255, Jun 2022, [Daring]. Available: <https://data.jakarta.go.id/>
- [8] M. J. Sodik, "PERBANDINGAN METODE NAIVE BAYES DAN K-NEAREST NEIGHBOR PADA KLASIFIKASI KUALITAS UDARA DI DKI JAKARTA," Universitas Teknologi Yogyakarta, 2019.

- [9] F. Insani dan S. I. Darlianti, "Pembentukan Model Regresi Linier Menggunakan Algoritma Genetika untuk Prediksi Parameter Indeks Standar Pencemar Udara (ISPU)," *Jurnal CoreIT*, vol. 5, no. 2, hlm. 110–117, Des 2019.
- [10] A. A. H. Kirono, I. Asror, dan Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Menggunakan Algoritma Naive Bayes," *e-Proceeding of Engineering*, vol. 9, no. 3, hlm. 1962–1969, Jun 2022.
- [11] A. Hermawan, "SPKU: SISTEM PREDIKSI KUALITAS UDARA (STUDI KASUS: DKI JAKARTA)," Universitas Teknologi Yogyakarta, 2019.
- [12] R. A. Wulandari, "PENERAPAN METODEORDINARY KRIGINGPADA PENDUGAANKADAR KARBON MONOKSIDA (CO) DI UDARA TAHUN 2020 (STUDIKASUS: PENCEMARAN UDARA KOTA SURABAYA)," Universitas Islam Negeri Sunan Ampel, Surabaya, 2021.
- [13] P. I. Agista, N. Gusdini, dan M. D. D. Maharani, "ANALISIS KUALITAS UDARA DENGAN INDEKS STANDAR PENCEMAR UDARA (ISPU) DAN SEBARAN KADAR POLUTANNYA DI PROVINSI DKI JAKARTA," *Jurnal SEOI –Fakultas Teknik Universitas Sahid Jakarta*, vol. 2, no. 2, hlm. 39–57, 2020, doi: <https://doi.org/10.36441/seoi.v2i2.491>.
- [14] A. Nurkhozin, M. I. Irawan, dan I. Mukhlas, "Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA, Fakultas MIPA," hlm. 33–40, 2011.
- [15] A. N. Hidayah, "PENERAPAN JARINGAN SYARAF TIRUANBACKPROPAGATIONKLASIFI KASI PENYAKIT ISPA," Universitas Sultan Syarif Kasim Riau, 2019.
- [16] A. Toha, P. Purwono, dan W. Gata, "Model Prediksi Kualitas Udara dengan Support Vector Machines dengan Optimasi Hyperparameter GridSearch CV," *Buletin Ilmiah Sarjana Teknik Elektro*, vol. 4, no. 1, hlm. 12–21, 2022, doi: [10.12928/biste.v4i1.6079](https://doi.org/10.12928/biste.v4i1.6079).
- [17] T. Hardoyo dan E. H. Parmadi, "Klasifikasi Usaha Mikro Kecil Menengah Menggunakan Jaringan Syaraf Tiruan Backpropagation," *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 2, no. 1, hlm. 111–123, 2022, doi: <https://doi.org/10.24002/konstelasi.v2i1.5625>.
- [18] A. K. Sharma dkk., "Classification of Indian Classical Music with Time-Series Matching Deep Learning Approach," *IEEE Access*, vol. 9, hlm. 102041–102052, 2021, doi: [10.1109/ACCESS.2021.3093911](https://doi.org/10.1109/ACCESS.2021.3093911).
- [19] M. F. Arifin dan D. Fitriana, "Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada," *InComTech: Jurnal Telekomunikasi dan Komputer*, vol. 8, no. 2, hlm. 87–102, 2018, doi: [10.22441/incomtech.v8i1.2198](https://doi.org/10.22441/incomtech.v8i1.2198).
- [20] A. Damuri, U. Riyanto, H. Rusdianto, dan M. Aminudin, "Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, hlm. 219–225, Des 2021, doi: [10.30865/jurikom.v8i6.3655](https://doi.org/10.30865/jurikom.v8i6.3655).