

PENERAPAN MODEL *LONG SHORT-TERM MEMORY* (LSTM) UNTUK PREDIKSI KUALITAS UDARA DAN ANALISIS KORELASI TERHADAP UNSUR CUACA DI JONGKAT KABUPATEN MEMPAWAH

Application of The Long Short-Term Memory (LSTM) Model for Air Quality Prediction and Correlation Analysis with Weather Elements in Jongkat Mempawah Regency

Purnama Arta Uli Sitompul^{1,2*)}, Herry Sujaini²⁾, Redi Ratiandi Yacoub²⁾, Naufal Ananda³⁾

¹⁾ Stasiun Klimatologi Kalimantan Barat. Jl. Raya Sei Nipah km. 20,5, Kalimantan Barat 78351.

²⁾ Jurusan Teknik Elektro, Fakultas Teknik, Universitas Tanjungpura, Pontianak.

³⁾ Balai Besar Meteorologi Klimatologi dan Geofisika Wilayah II, Tangerang Selatan

*E-mail : d2081221003@student.untan.ac.id

Intisari

Kebakaran lahan yang terjadi beberapa waktu lalu, seperti tahun 2023 menyebabkan adanya peningkatan polusi udara di Kabupaten Mempawah dan sekitarnya. Salah satu polutan yang terdeteksi di udara adalah *Particulate Matter* (PM_{2,5}). PM_{2,5} mudah terhirup oleh manusia dan mengganggu proses pertukaran gas di paru-paru. Untuk mitigasi pencemaran udara dan memberikan informasi cepat terkait kualitas udara, diperlukan prediksi yang akurat. Metode *Long Short-Term Memory* (LSTM) dipilih karena kemampuannya menangani data deret waktu yang kompleks berdasarkan perubahan waktu dan polutan udara. LSTM mempelajari pola data historis untuk menghasilkan prediksi yang lebih akurat, membantu pengambilan keputusan dalam mitigasi pencemaran udara. Analisis korelasi PM_{2,5} terhadap unsur cuaca dibuat dalam penelitian ini sebagai informasi dan analisis hubungan hasil prediksi dengan kondisi cuaca lingkungan sekitar untuk mitigasi oleh user. Pemodelan LSTM ini menggunakan max epoch 100, batch size 50, pembagian data training dan testing 90%:10%, optimizer adalah adam, dan mean squared error sehingga diperoleh RMSE terbaik sebesar 3,40. Penelitian ini menghasilkan akurasi prediksi 92,18% untuk klasifikasi data kategori baik dan akurasi 87,75% untuk data kategori sedang hingga tidak sehat. Berdasarkan hasil prediksi tersebut, semakin tinggi konsentrasi dan kategori PM_{2,5} maka persentase akurasi berkurang sehingga butuh data lebih panjang. Analisis Korelasi dihasilkan bahwa adanya hubungan yang kuat PM_{2,5} terhadap parameter cuaca, yaitu kecepatan angin dengan korelasi 0,64 dan curah hujan sebesar -0,49. Hubungan korelasi dikatakan kuat jika berada di rentang $-1 \leq r \leq 1$.

Kata Kunci: PM_{2,5}, Kecerdasan Buatan, LSTM, Prediksi, RMSE

Abstract

Recent land fires, such as those in 2023, have caused an increase in air pollution in Mempawah Regency and its surroundings. One of the pollutants detected in the air is *Particulate Matter* (PM_{2.5}). Humans easily inhale PM_{2.5}, which disrupts the gas exchange process in the lungs. Accurate predictions are needed to mitigate air pollution and provide quick information related to air quality. The *Long Short-Term Memory* (LSTM) method was chosen for its ability to handle complex time series data based on changes in time and air pollutants. LSTM learns historical data patterns to produce more accurate predictions, aiding decision-making in air pollution mitigation. The correlation analysis of PM_{2.5} with weather elements was conducted in this study as information and analysis of the relationship between prediction results and surrounding environmental weather conditions for user mitigation. This LSTM modeling uses a max epoch of 100, a batch size of 50, a training and testing data split of 90%:10%, the optimizer is Adam, and the mean squared error results in the best RMSE of 3.40. This study produced a prediction accuracy of 92.18% for the Good category data classification and 87.75% accuracy for the Moderate to Unhealthy category data. Based on these prediction results, the higher the concentration and category of PM_{2.5}, the lower the accuracy percentage, indicating the need for longer data. The correlation analysis showed a strong relationship between PM_{2.5} and weather parameters, namely wind speed, with a correlation of 0.64 and rainfall of -0.49. The correlation relationship is considered strong if it is in the $-1 \leq r \leq 1$ range.

Keywords: PM2.5, Artificial Intelligence, LSTM, Prediction, RMSE

1. PENDAHULUAN

Pencemaran udara telah menjadi isu global yang mendapatkan perhatian karena semakin meningkat dari waktu ke waktu. Area perkotaan menjadi titik fokus utama dalam isu pencemaran udara ini, dengan berbagai faktor penyebab seperti aktivitas industri, transportasi, pembakaran lahan, serta berbagai aktivitas manusia lainnya (Maharani & Aryanta, 2023). Pencemaran udara juga menjadi masalah yang semakin mengkhawatirkan di Indonesia dan berdampak serius terhadap kesehatan manusia, seperti gangguan pernapasan, penyakit jantung, dan berbagai masalah kesehatan lainnya. Menurut data yang dirilis oleh *World Health Organization* (WHO) pada tahun 2012, pencemaran udara menyebabkan sekitar 7 juta kematian, atau sekitar 1/8 dari semua kematian di seluruh dunia. (Khairani *et al.*, 2023).

Pencemaran udara memiliki dampak ekonomi yang signifikan dalam hal biaya perawatan kesehatan yang tinggi dan potensi kerugian dalam sektor pertanian, industri, dan lainnya. Ada banyak jenis polutan primer dan sekunder di udara ambien yang terbagi atas polutan gas dan partikulat, seperti CO, NO, CO₂, NO₂, SO₂, O₃, PM_{2,5}, PM₁₀, Timah (Pb). Salah satu polutan primer yang sangat berbahaya karena dengan mudah terhirup oleh manusia adalah *Particulate Matter* (PM_{2,5}). PM_{2,5} adalah jenis partikulat yang memiliki diameter kurang dari atau sama dengan 2,5 mikrometer, dapat masuk ke paru-paru dan bahkan masuk ke dalam aliran darah dan mengganggu proses pertukaran gas di paru-paru. (Rafidah *et al.*, 2023). Alat pengukur PM_{2,5} bersifat canggih milik BMKG, salah satunya berada di Stasiun Klimatologi Kalimantan Barat dinamakan *Beta Attenuation Monitor* (BAM) 1020. BAM 1020 dapat mengukur nilai PM_{2,5} dalam radius jarak dekat sekitar beberapa kilometer per meter persegi menurut *Expert from Finland* atau 5 Km berdasarkan hasil Analisa BMKG. BAM1020 mengukur konsentrasi partikulat PM_{2,5} di udara dengan satuan mikrogram per meter kubik ($\mu\text{g}/\text{m}^3$) dengan prinsip atenuasi sinar beta. BAM 1020 ini mengukur PM_{2,5} dengan metode pengambilan *sampling* dengan data terekam dan terkirim setiap jam.

Beberapa faktor, seperti cuaca dan keadaan lingkungan saat itu juga memengaruhi penetapan radius tersebut. Percepatan mitigasi pencemaran udara membutuhkan informasi cepat tentang kualitas udara, yaitu dengan membuat prediksi kualitas udara. Prediksi adalah teknik untuk melakukan analisis perhitungan agar dapat memperkirakan kejadian yang akan terjadi di masa depan dengan menggunakan data historis masa lalu. Pembuatan prediksi ini dapat dilakukan dengan berbagai metode, mulai dari

metode statistik hingga penggunaan pemodelan jaringan syaraf tiruan. Prediksi kualitas udara perlu mempertimbangkan beberapa faktor dalam pembuatannya, seperti perubahan waktu, polutan udara yang ada, dan faktor cuaca (Millah, 2022). Di era digital saat ini, *Artificial Intelligence* (AI) sangat dibutuhkan sebagai sistem kecerdasan buatan yang dituangkan dalam sistem komputer sehingga bekerja seperti pikiran manusia. Selain itu, adanya AI tentu meningkatkan efisiensi untuk melakukan analisis data yang cepat, salah satunya dengan penggunaan *Deep Learning* yang telah dimanfaatkan dalam membuat penelitian yang memudahkan pekerjaan manusia.

Metode *Reccurant Neural Network* adalah bagian dari *Deep Learning*. Model *Long Short-Term Memory* (LSTM) adalah bagian dari RNN yang digunakan untuk membuat prediksi berdasarkan rangkaian data waktu. LSTM memiliki kemampuan untuk menangkap pola-pola kompleks dalam data *time series* yang sering kali sulit diidentifikasi dengan metode regresi linear (Karyadi, 2022) sehingga cocok untuk melakukan prakiraan PM_{2,5} karena datanya bersifat periodisitas dimana terdapat pola berulang dalam interval waktu tertentu dengan nilai cenderung naik di pagi hingga menjelang sore dan menurun di sore di malam hari. Hal ini menunjukkan pola yang jelas dan cocok untuk pemodelan LSTM untuk membuat prakiraan di masa depan berdasarkan *time series*. Selain itu, data PM_{2,5} bersifat *nonlinear* dengan adanya faktor penyebab, seperti polusi lokal dan interaksi antara variabel seperti faktor cuaca, lalu lintas, dan aktivitas manusia. LSTM ini juga membutuhkan data jangka panjang yang sesuai dengan data PM_{2,5} dimana nilai konsentrasi yang terukur memiliki data yang banyak. Penelitian sebelumnya pernah dilakukan untuk model LSTM dan disimpulkan memberikan hasil yang baik dan akurat dalam pembuatan prediksi.

Berdasarkan penjelasan di atas, maka dibuat penelitian dengan judul Penerapan Model *Long Short-Term Memory* (LSTM) untuk prediksi kualitas udara dan analisis korelasi terhadap unsur cuaca di Jongkat Kabupaten Mempawah. Penelitian ini berfokus pada pembuatan prediksi kualitas udara PM_{2,5} dengan menggunakan pemodelan memori jangka panjang (LSTM) dengan mengumpulkan data PM_{2,5} yang ditambahkan kategori indeks pemantauan kualitas udara sesuai standar BMKG. Selain itu, sebagai gambaran dari analisis hubungan antara kondisi cuaca lingkungan sekitar, dilakukan analisis korelasi antara polutan PM_{2,5} dan data cuaca (suhu, kelembaban, curah hujan, dan kecepatan angin). Analisis tersebut dilakukan untuk memperjelas informasi nilai kualitas udara PM_{2,5} yang meningkat pada waktu dan kondisi

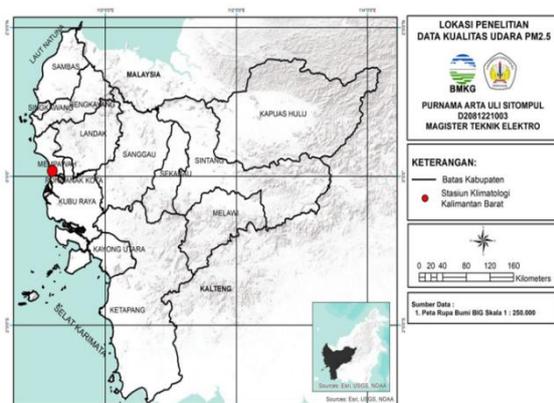
cuaca saat itu. Hasil prediksi berupa informasi kualitas udara PM_{2,5} hingga 7 hari ke depan, Penelitian ini diharapkan dapat memberi informasi yang positif dan cepat dalam upaya menjaga lingkungan dan mengurangi polusi udara bagi kesehatan manusia.

2. METODE

2.1 Lokasi Penelitian

Penelitian ini dilakukan di Stasiun Klimatologi Kalimantan Barat yang terletak di Kecamatan Jongkat, Kabupaten Mempawah, Kalimantan Barat dengan koordinat 0.0753° Lintang Utara dan 109.1908° Bujur Timur serta elevasi setinggi 2 meter di atas permukaan laut dengan gambaran wilayah dapat dilihat pada Gambar 1. Lokasi penelitian tersebut dipilih karena memiliki alat ukur partikulat *Beta Attenuation Monitoring* (BAM) PM_{2,5}. BAM 1020 merupakan alat ukur bersifat canggih dan otomatis yang mengukur dan mencatat tingkat konsentrasi partikulat PM_{2,5} di udara dalam satuan mikrogram per meter kubik (µg/m³) dengan prinsip atenuasi sinar beta yang datanya terekam per jam. Kabupaten Mempawah merupakan salah satu kabupaten di Kalimantan Barat berada dekat dengan pesisir.

Wilayah sekitar Kabupaten Mempawah didominasi lahan bergambut dan beriklim tropis digunakan oleh masyarakat setempat sebagai lahan pertanian. Potensi hujan rendah di waktu tertentu seharusnya tidak menjadikan kekeringan secara cepat namun ketika hari tanpa hujan cukup panjang menyebabkan adanya titik *hotspot* dan ditambah kebakaran lahan yang di sengaja oleh oknum. Selain itu, aktivitas industri menyebabkan polusi udara meningkat di lingkungan sekitar. Stasiun Klimatologi Kalimantan Barat dulunya bernama Stasiun Klimatologi Mempawah berubah namanya mulai tanggal 18 April 2022 berdasarkan PerKa BMKG No. 1 Tahun 2022.



Gambar 1. Lokasi penelitian.

2.2 Data Penelitian

Data dalam penelitian ini berupa konsentrasi PM_{2,5} dengan satuan µg/m³ dengan format *comma separated values* (csv) dan bentuk data historis *time series* periode September 2021 hingga Desember 2023. Data tersebut diunduh langsung dari *database* alat BAM1020 dengan pengukuran alat *real time*. Data PM_{2,5} dan Indeks Kategori Pemantauan akan menjadi variabel dalam penelitian ini dan akan dimodelkan menggunakan LSTM. Rentang Indeks Kategori Pemantauan Kualitas Udara dapat dilihat pada Gambar 2 yaitu baik (0–15,5 µg/m³), sedang(15,6–55,4 µg/m³), tidak sehat (55,5–150,4 µg/m³), sangat tidak sehat (150,5–250,4 µg/m³), dan berbahaya (>250,4 µg/m³).

Selain itu, akan dilakukan analisis korelasi PM_{2,5} terhadap data meteorologi, seperti suhu udara, kelembaban udara, kecepatan angin, dan curah hujan. Data meteorologi ini diperoleh dari *Automatic Weather Station* (AWS) yang merupakan sebuah alat pengukuran otomatis yang menghasilkan data *real time*. Hasil korelasi ini akan digunakan sebagai analisis dan informasi tambahan untuk hasil prediksi kualitas udara berdasarkan kondisi cuaca saat itu.

2.3 Metode Prediksi dengan LSTM

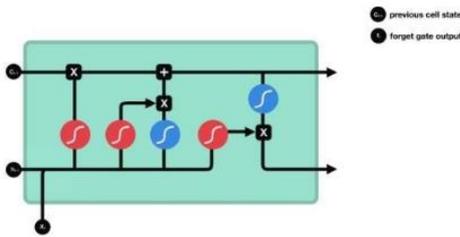
Penelitian bersifat kuantitatif menggunakan pendekatan pengambilan data dari alat secara langsung di lapangan, pengambilan data dari aplikasi, dan melakukan pengolahan data mentah. Selain itu, penelitian ini menggunakan aplikasi *open source Anaconda Navigator* yang menyediakan *Jupyter Notebook* untuk pemodelan LSTM. Dalam metode LSTM ini, tidak ada syarat khusus pemilihan waktu yang digunakan karena berlaku untuk semua waktu asal datanya *time series*. Cara kerja *Long Short-Term Memory* (LSTM) menambahkan sel memori yang dapat menyimpan informasi untuk jangka waktu yang panjang. Hal ini membantu mengatasi *vanishing gradient* pada RNN saat memproses data sekuensial yang panjang.

Informasi yang dibuang dari sel akan ditentukan oleh LSTM dalam *layer forget gate* untuk melakukan keputusan ini yang dapat dilihat pada Gambar 2. Ini digunakan untuk memutuskan informasi mana yang harus dibuang dan mana yang harus disimpan.

$$ft = (Wf[ht - 1 * xt] + bf) \tag{1}$$

Dimana :

- ft = fungsi *sigmoid*
- Wf = nilai *weight* untuk *forget gate*
- $ht-1$ = nilai *output* sebelum orde ke t
- xt = nilai *input* pada orde ke t
- bf = nilai bias pada *forget gate*



Gambar 2. Forget gate.

Nilai yang akan diperbaharui ditentukan oleh *input gate* dan vektor nilai baru yang dibuat oleh lapisan *tanh* akan dimasukkan ke dalam keadaan sel. Hasil akhir akan dilakukan pembaharuan ke *cell state*, persamaan rumus sebagai berikut:

$$it = \sigma(Wi * [ht - 1 * xt] + bi) \quad (2)$$

Dimana:

- it* = *input gate*
- σ = fungsi *sigmoid*
- Wi* = nilai *weight* untuk *input gate*
- ht-1* = nilai *output* sebelum orde ke *t*
- xt* = nilai *input* pada orde ke *t*
- bi* = nilai bias pada *input gate*

Output gate menjalankan lapisan *sigmoid* untuk menentukan bagian *cell state* mana yang akan menjadi *output*, yang kemudian akan ditempatkan melalui *tanh*.

$$t = \sigma(Wo * [ht - 1 * xt] + bo) \quad (3)$$

Dimana:

- Ot* = *output gate*
- σ = fungsi *sigmoid*
- Wo* = nilai *weight* untuk *output gate*
- ht-1* = nilai *output* sebelum orde ke *t*
- xt* = nilai *input* pada orde ke *t*
- bo* = nilai bias pada *output gate*

Dalam melakukan pemodelan LSTM ada lapisan tambahan yang digunakan untuk melatih model sehingga diperoleh nilai *loss* dan *RMSE* bernilai kecil. Dua lapisan tambahan adalah *epoch* dan *batch size*. *Epoch* menunjukkan bahwa seluruh dataset telah menjalani satu putaran proses pelatihan pada *Neural Network*. *Batch size* menunjukkan jumlah data yang didistribusikan ke *Neural Network*. Alur penelitian dapat dilihat pada Gambar 3 .

1. Dataset Penelitian: menetapkan data yang digunakan untuk penelitian prediksi kualitas udara. Data PM2,5 yang terukur dan tersimpan pada alat BAM 1020 berupa data per jam direalisasi menjadi data harian. Historis data dimulai tanggal 20 September 2021 hingga 31 Desember 2023 sehingga didapatkan 797 data harian. Setelah reanalisis harian dilakukan, ditambahkan Indeks Kategori PM2,5 berdasarkan nilai PM2,5 hasil reanalisis. Pengolahan data

reanalisis harian PM2,5 ini secara manual dilakukan dengan *pivot table* di Ms. Excel untuk mencari nilai PM2,5 rata-rata harian.

2. *Preprocessing* data: melakukan *import* data prediktor PM2,5 yang telah diunduh sebelumnya ke aplikasi *jupyter notebook*. Kemudian melakukan proses *coding* dan penyesuaian *datetime* untuk menampilkan dataset pada lembar kerja *jupyter notebook* visualisasikan data yang berhasil di *input* dalam bentuk tabel atau grafik.
3. Normalisasi data: melakukan normalisasi data pada dataset dengan metode *max-min* yang berguna untuk mengubah seluruh nilai variabel yang dipakai sebagai input dengan rentang nilai 0 sampai 1 sehingga dengan cepat dan stabil model tersebut untuk mencapai nilai minimum dalam fungsi objektif. *Coding* untuk normalisasi dapat dilihat pada Gambar 4. Persamaan normalisasi data dengan *MinMaxScaler* sebagai berikut :

$$Scaler = MinMaxScaler \dots\dots\dots (4)$$

$$X_{norm} = (X - X_{min}) / (X_{max} - X_{min}) \dots\dots\dots (5)$$

Keterangan :

- X_{norm} = nilai normalisasi
- X_{min} = nilai minimum data *training*
- X_{max} = nilai maksimum dari data *training*

```
In [4]: tf.random.set_seed(7)
ds = np.array(df_data).reshape(-1,1)
scaler = MinMaxScaler(feature_range=(0, 1))
dataset = scaler.fit_transform(ds)
dataset
```

Gambar 4. Coding Normalisasi

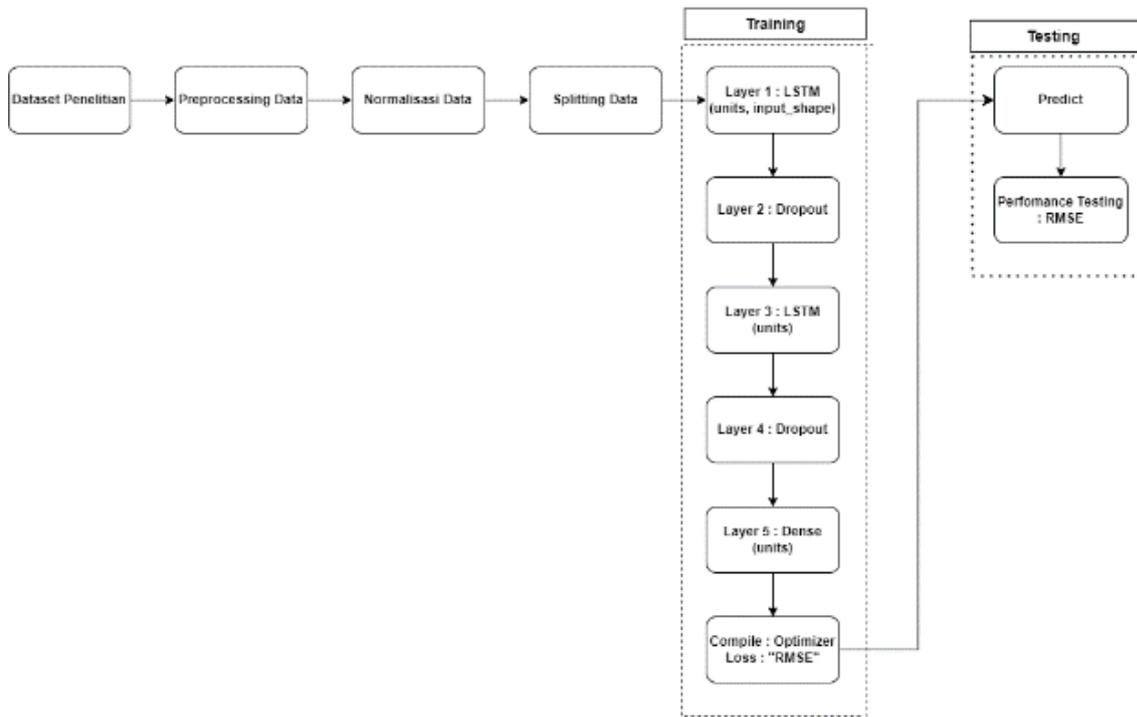
Setelah data dinormalisasi, maka akan digunakan sebagai *input* untuk model LSTM. Untuk mendapatkan nilai prediksi yang sebenarnya, keluaran LSTM akan di denormalisasi. Persamaan 4 dan 5 ditambahkan pada *coding* dan jika berhasil maka akan lanjut pada proses *coding* berikutnya.

4. *Splitting Data* : proses membagi data menjadi 2 (dua), yaitu *training* data dan *testing* data. Proses pembagian tersebut dilakukan dengan beberapa skenario, *training* : *testing* sebesar 50%:10%, 60%:40%, 70%:30%, 80%:20% dan 90%:10%. Beberapa skenario yang dilakukan pada proses LSTM sampai didapatkan RMSE untuk penelitian ini.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n}} \quad (6)$$

Keterangan :

- RMSE : *Root Mean Square Error*
- n* : banyak data
- y_i : nilai sebenarnya
- \hat{y} : nilai hasil prediksi



Gambar 3. Diagram alur penelitian.

5. Proses *compile* menggunakan *Adam Optimizer* untuk pengoptimalan parameter model dengan mengubah kecepatan pembelajaran secara adaptif berdasarkan gradien yang dihitung selama *training*. Dengan menggunakan RMSE dan fungsi *loss*, model LSTM akan berusaha untuk meminimalkan perbedaan antara prediksi dan nilai sebenarnya.
6. Proses *testing*: setelah hasil model *training* telah didapatkan, dilanjutkan *testing data* untuk menguji efektifitas dari model tersebut. Dalam proses pelatihan, data pengujian ini tidak digunakan. Namun dijadikan data untuk mengevaluasi seberapa baik model dapat membuat prediksi pada data yang belum pernah dilihat sebelumnya sehingga diperoleh RMSE terbaik. Selanjutnya, persamaan yang didapat dari hasil *training* dan *testing* yang telah dilakukan akan dipakai untuk memprediksi kualitas udara dengan data terbaru yang ada. Uji performa dilakukan dengan membandingkan prediksi model dengan nilai sebenarnya dari alat ukur.
7. Analisis korelasi: proses menghitung nilai korelasi data polutan PM2,5 dengan data meteorologi (suhu, kelembaban, curah hujan, dan kecepatan angin) untuk menentukan seberapa kuat dan lemah pengaruh cuaca terhadap penyebaran PM2,5 di lingkungan sekitar. Analisis korelasi ini juga akan membantu dalam melakukan analisis prediksi kualitas udara untuk beberapa waktu kedepan.

$$r_{xy} = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{\{n \sum x^2 - (\sum x)^2\} \{n \sum y^2 - (\sum y)^2\}}} \quad (7)$$

Keterangan:

x, y = variabel data

r = koefisien korelasi

n = jumlah data

Nilai yang menunjukkan seberapa kuat hubungan antara huruf x dan y dalam fungsi linier disebut koefisien korelasi. Nilai ini berkisar antara -1 dan 1, sehingga jika $r =$ koefisiensi korelasi, nilai r dinyatakan antara $-1 \leq r \leq 1$. Jika $r = 1$, hubungan x dan y sempurna dan positif (mendekati 1, yang berarti hubungan sangat kuat dan positif), dan jika $r = -1$, hubungan x dan y sempurna dan negatif.

3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data PM2,5 dari pengukuran alat otomatis BAM1020 kemudian dilakukan reanalisis menjadi data periode harian. Selanjutnya, penambahan indeks kategori pemantauan PM2,5 berdasarkan acuan BMKG. Data yang digunakan sebanyak 797 data harian yang terlihat pada Tabel 1.

Tabel 1. Data PM2,5 dan kategorinya tahun 2021-2023.

Kategori PM2,5	Jumlah Klasifikasi Nilai Indeks		
	2021	2022	2023
BAIK	91	296	195
SEDANG	10	54	131
TIDAK SEHAT	0	3	18

Pada tahun 2021, data yang tersedia dimulai dari tanggal 20 September hingga 31 Desember 2021 dengan jumlah data sebanyak 91 data dengan kategori baik, 10 data dengan kategori sedang dan 0 data dengan kategori tidak sehat.

Tahun 2022, jumlah data dengan kategori baik sebanyak 296 data yang lebih banyak dibanding tahun 2023 dengan kategori baik sebanyak 195 data sedangkan jumlah data dengan ketegori sedang sebanyak 54 data dan tidak sehat sebanyak 3 data. Pada tahun 2023 juga terjadi peningkatan kualitas udara dalam ketegori sedang dan tidak sehat dengan jumlah data sebanyak 131 untuk kategori sedang dan 18 data untuk kategori tidak sehat. Peningkatan PM2,5 di tahun 2023 bisa disebabkan oleh beberapa faktor, seperti cuaca, hari tanpa hujan yang panjang, kebakaran lahan, dan lainnya. Data meteorologi, contohnya data angin dan suhu udara mempengaruhi sebaran dan tinggi rendahnya konsentrasi polutan di udara ambien termasuk PM2,5 (Wark & Warner, 1981).

3.1 Analisis Prediksi dengan Model LSTM

Skenario dilakukan dalam penelitian ini adalah menganalisis dampak *splitting data training* dan *testing data*. Pengujian dilakukan dengan beberapa skenario pembagian *training* dan *testing* data 50%:50% hingga 90%:10% dapat dilihat pada Tabel 2. Selain itu, akan dilakukan uji coba dengan mengubah nilai *hidden layer* dan *epoch*. Pengaturan nilai *Batch size* = 50, *epoch* = 100, dan *dropout* 0,5 dilakukan sebagai nilai yang tepat saat uji coba dalam mendapat nilai *loss* dan RMSE yang terkecil dalam penelitian ini sehingga nilai prediksi menjadi akurat. *Batch size* merupakan banyaknya sampel data yang dipakai dalam proses satu kali literasi dan diatur ke 50 saat uji coba karena berhasil membantu model untuk

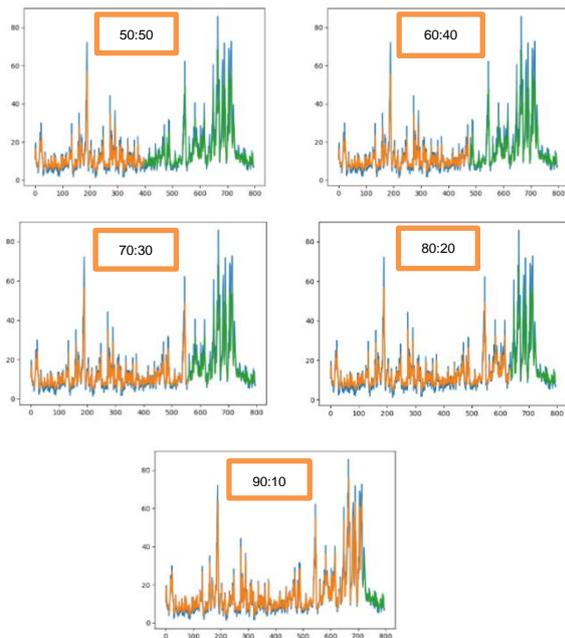
belajar lebih stabil dan mengurangi fluktuasi dalam gradien sehingga penggunaan memori lebih efisien dibandingkan dengan menambah ukuran *batch size*. *Epoch* diartikan ketika satu *epoch* diperoleh setelah melalui proses pelatihan satu kali penuh dengan saat pengujian didapatkan RMSE terbaik saat nilai *epoch* bernilai 100. Jumlah *epoch* yang tinggi akan menyebabkan terjadinya *overfitting* karena akan lebih banyak data yang dipelajari. *Dropout* sebesar 0,5 digunakan umumnya pada pemodelan dan terbukti dapat digunakan untuk RMSE karena adanya keseimbangan yang sesuai antara *underfitting* dan *overfitting*.

Pertama, kombinasi 50%:50% dengan jumlah 398 data *training* dan 399 data *testing* dan diuji coba dengan beberapa *neuron hidden* didapatkan RMSE terbaik sebesar 8,07 dengan jumlah *neuron hidden* 100. Percobaan kedua, dengan kombinasi 60%:40% dengan pembagian dengan jumlah 478 data *training* dan 319 data *testing* didapatkan RMSE terbaik sebesar 8,94 dengan *neuron hidden* 100. Percobaan ketiga, kombinasi 70%:30% dengan 557 data *training* dan 240 data *testing* diperoleh RMSE terbaik sebesar 9,34 dengan *neuron hidden* 100. Percobaan keempat, kombinasi *training* dan *testing* 80%:20% dengan jumlah data *training* 637 dan 160 data *testing* dan didapatkan RMSE terbaik 10,47 dengan *neuron hidden* 50. Percobaan terakhir dengan *splitting data training* dan data *testing* 90%:10% dengan penggunaan jumlah data *training* sebanyak 717 data dan 81 *testing data*. Percobaan *neuron hidden* ini dilakukan dengan cara coba-coba sampai didapatkan hasil yang terbaik dikarenakan tidak ada aturan baku dalam penggunaan *neuron hidden* yang optimal untuk menghasilkan RMSE terbaik. Penelitian ini menggunakan 3 lapisan dalam proses pelatihan karena saat melanjutkan ke lapisan berikutnya tidak ada perubahan sehingga dapat menghemat memori.

Setelah skenario yang telah dilakukan, dapat disimpulkan bahwa skenario pembagian *training* dan *testing* dalam menghasilkan RMSE terbaik untuk data tersebut pada penelitian ini adalah 90%:10% dengan nilai RMSE sebesar 3,40 dan fungsi *loss* dihasilkan 0,0058. Skenario yang dibuat divisualisasikan dalam bentuk grafik untuk melihat perbandingan dari data *training*, *testing*, dan data aktual dapat dilihat pada Gambar 5.

Tabel 2. Hasil pengujian dari kombinasi *training* dan *testing data*

Max Epoch	Neuron Hidden	RMSE				
		50%:50%	60%:40%	70%:30%	80%:20%	90%:10%
100	5	10,28	9,52	9,73	10,76	3,43
100	10	8,89	9,17	9,63	10,61	3,41
100	20	8,27	9,05	9,40	10,55	3,41
100	50	8,13	8,98	9,38	10,47	3,40
100	70	8,12	8,97	9,37	10,48	3,42
100	100	8,07	8,94	9,34	10,51	3,42

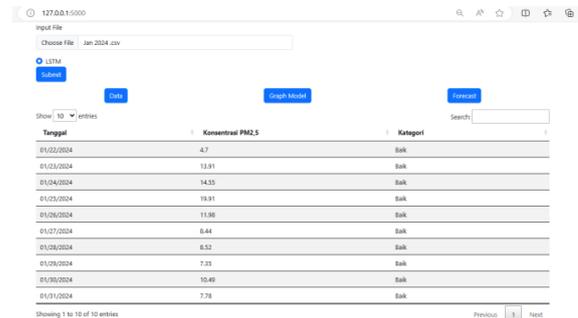


Gambar 5. Visualisasi grafik skenario *training* dan *testing data*.

3.2 Uji Peforma Prediksi Kualitas Udara dengan LSTM

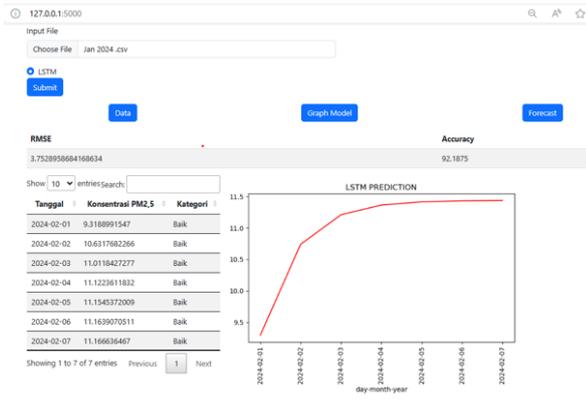
Pemodelan data prediksi kualitas udara dengan LSTM telah dilakukan dan didapatkan persamaan dari proses *training* dan *testing*. Selanjutnya, dilakukan uji performa menggunakan persamaan tersebut dengan data PM2,5 harian yang mewakili data di tahun 2023 dan tahun 2024. Tampilan data tersebut di visualisasikan dalam *Graphical User Interface* (GUI) menggunakan *visual studio code*.

Pengaplikasian *flask* di *vscode* sehingga menghasilkan IP *local* tanpa harus terkoneksi internet, yg menampilkan menu-menu dapat dilihat pada Gambar 6.



Gambar 6. Tampilan GUI prediksi kualitas udara Januari 2024.

Pertama, pengujian peforma untuk data harian di tanggal 22 hingga 31 Januari 2024. Berdasarkan data tersebut, dihasilkan konsentrasi PM2,5 prediksi untuk 7 hari ke depan dimulai tanggal 1 hingga 7 Februari 2024 (Gambar 7). Hasil prediksi menunjukkan bahwa akurasi prediksi sebesar 92,18% dengan RMSE 3,75289. Hasil prediksi selanjutnya dibandingkan dengan data PM2,5 dari alat secara langsung pada Tabel 3.



Gambar 7. Tampilan hasil prediksi tanggal 1-7 Februari 2024.

Tabel 3. Perbandingan data PM2,5 dari alat dan hasil prediksi Februari 2024.

Tanggal	PM2,5 (µg/m³)	Kategori	Prediksi PM2,5 (µg/m³)	Kategori	Koreksi
02/01/2024	6,31	Baik	9,31	Baik	-3
02/02/2024	4,93	Baik	10,63	Baik	-5,7
02/03/2024	10,6	Baik	11,01	Baik	-0,41
02/04/2024	10,02	Baik	11,122	Baik	-1,102
02/05/2024	11,87	Baik	11,154	Baik	0,716
02/06/2024	8,35	Baik	11,163	Baik	-2,813
02/07/2024	12,05	Baik	11,166	Baik	0,884

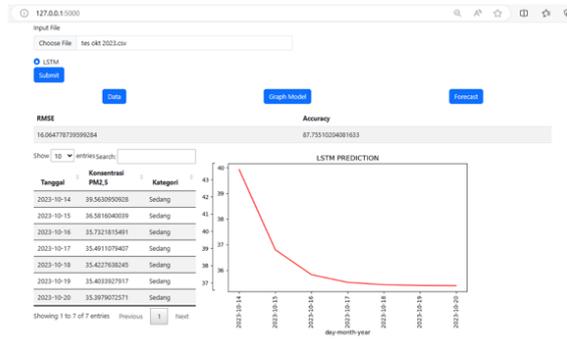
Hasil prediksi konsentrasi PM2,5 terendah adalah 9,31 µg/m³ dan tertinggi sebesar 11,166 µg/m³. Nilai koreksi terendah sebesar -5,7 di tanggal 2 Februari dan koreksi tertinggi sebesar 0,884 pada tanggal 7 Februari 2024. Nilai koreksi sebagian besar bernilai negatif artinya hasil prediksi PM2,5 bernilai lebih tinggi dari nilai sebenarnya dari alat. Secara keseluruhan hasil prediksi PM2,5 dan nilai sebenarnya berada dalam kategori yang sama yaitu baik yang menunjukkan bahwa nilai keduanya masih berada di rentang yang sama dan dapat disimpulkan informasi kualitas udara hasil prediksi cukup akurat saat kondisi normal atau saat kualitas udara dalam kondisi Baik. Selanjutnya dilakukan pengujian data PM2,5 untuk kategori sedang dan tidak sehat untuk melihat keakuratan data saat nilai PM2,5 meningkat.

Data berikutnya untuk pengujian adalah data PM2,5 tanggal 5 hingga 13 Oktober 2023. Data ini sebenarnya masuk dalam *training* untuk pembelajaran model namun karena pada bulan ini kategori PM2,5 berada di rentang sedang dan tidak sehat maka dicoba untuk melakukan prediksi terhadap model yang telah didapatkan.

Tanggal	Konsentrasi PM2,5	Kategori
10/05/2023	49.9	Tidak Sehat
10/06/2023	72.7	Tidak Sehat
10/07/2023	57	Tidak Sehat
10/08/2023	26.93	Sedang
10/09/2023	19.75	Sedang
10/10/2023	21.32	Sedang
10/11/2023	23.71	Sedang
10/12/2023	23.27	Sedang
10/13/2023	36.73	Sedang

Gambar 8. Tampilan GUI prediksi kualitas udara Oktober 2023.

Gambar 8 menunjukkan bahwa pada bulan Oktober 2023 konsentrasi PM2,5 cenderung berada di kategori sedang hingga tidak sehat dengan curah hujan menengah di rentang 151-200 mm dengan sifat hujan bawah normal di rentang 51–84% (Buletin Iklim Edisi November 2023). Data PM2,5 di *running* untuk mendapatkan hasil prediksi selama 7 hari ke depan di tanggal 14 hingga 20 Oktober 2023. Hasil prediksi dapat dilihat pada Gambar 9 menunjukkan bahwa di tanggal tersebut PM2,5 berada dalam kategori sedang dengan nilai sekitar 35, 397 µg/m³ hingga 39,563 µg/m³.



Gambar 9. Tampilan hasil prediksi tanggal 14-20 Oktober 2023.

Selanjutnya, dilakukan uji performa dengan membandingkan data hasil prediksi dan pengukuran alat di tanggal yang terlihat pada Tabel 4.

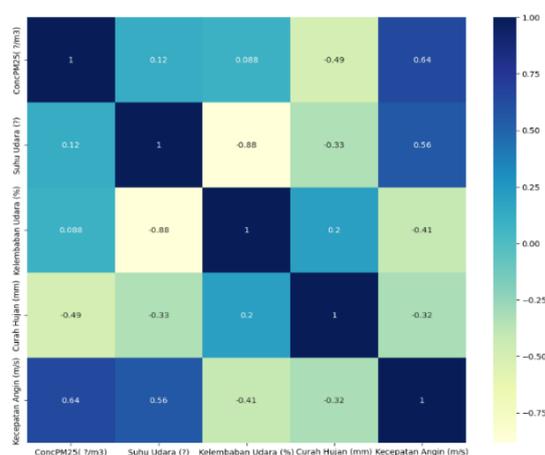
Tabel 4. Perbandingan data partikulat PM2,5 dari alat dan hasil prediksi Oktober 2023.

Tanggal	PM2,5 (µg/m³)	Kategori	Prediksi PM2,5 (µg/m³)	Kategori	Koreksi
10/14/2023	39,52	Sedang	39,563	Sedang	-0,043
10/15/2023	33,69	Sedang	36,581	Sedang	-2,891
10/16/2023	30,06	Sedang	35,732	Sedang	-5,672
10/17/2023	27,22	Sedang	35,491	Sedang	-8,271
10/18/2023	16,46	Sedang	35,422	Sedang	-18,962
10/19/2023	17,68	Sedang	35,403	Sedang	-17,723
10/20/2023	14,13	Baik	35,397	Sedang	-21,267

Nilai koreksi terendah sebesar -21,267 di tanggal 20 Oktober 2023 dan koreksi tertinggi sebesar -0,043 pada tanggal 14 Oktober 2023. Nilai koreksi keseluruhan bernilai negatif yang artinya hasil prediksi PM2,5 bernilai lebih tinggi dari nilai pengukuran alat. Dari hasil prediksi 7 hari tersebut terdapat 6 data dengan kategori sedang dan 1 data yang berbeda yaitu kategori baik di tanggal 20 Oktober 2023 sehingga dapat dikatakan prediksi kualitas udara dengan kondisi nilai PM2,5 meningkat. Dengan kategori sedang–tidak sehat masih dapat memberikan hasil prediksi yang cukup akurat dengan sebagian besar kategori masih sama walaupun ada kategori yang berbeda. Nilai RMSE yang cukup besar dibanding uji data yang pertama, yaitu sebesar 16,06, dan akurasi untuk uji data tersebut sebesar 87,75%.

3.3 Analisis Korelasi

Gambar 10 menunjukkan hasil perhitungan korelasi antar variabel dari polutan PM_{2,5} terhadap data meteorologi seperti suhu udara, kelembaban udara, curah hujan, dan kecepatan angin. Faktor meteorologi yang mempunyai korelasi searah yang bernilai positif dengan konsentrasi PM_{2,5} adalah kecepatan angin, suhu udara, dan kelembaban udara. Korelasi searah hubungan yang paling kuat adalah hubungan korelasi antara polutan PM_{2,5} dengan kecepatan angin sebesar 0,64. Hal ini menunjukkan bahwa ketika kecepatan angin bernilai tinggi maka nilai konsentrasi PM_{2,5} meningkat karena angin mengangkat sehingga adanya pengumpulan partikel polutan debu di sekitarnya. Hasil tersebut didukung dengan penelitian yang dilakukan oleh (Chen *et al.*, 2021) bahwa saat angin kencang maka polusi udara meningkat dari data korelasi 24 jam dan 48 jam yang lalu dari penelitiannya.



Gambar 10. Korelasi data PM_{2,5} dengan data meteorologi.

Selain itu, korelasi searah dengan hubungan yang lemah adalah korelasi PM_{2,5} terhadap suhu udara yaitu dengan nilai korelasi 0.12. Ini menandakan bahwa ketika suhu udara meningkat, konsentrasi partikel PM_{2,5} juga cenderung meningkat, meskipun hubungannya lemah atau tidak saling mempengaruhi. Kelembaban udara memiliki hubungan yang sangat lemah dengan PM_{2,5} dengan nilai korelasi 0,088, yang berarti tidak ada hubungan yang signifikan antara kedua variabel tersebut.

Hubungan korelasi yang tidak searah atau bernilai negatif antara PM_{2,5} dengan curah hujan bernilai -0,49 artinya adanya hubungan yang cukup kuat dan berbanding terbalik jika nilai curah hujan tinggi di suatu wilayah maka nilai konsentrasi polutan PM_{2,5} kecil karena hujan membantu dalam proses pembersihan udara dari partikel polutan di udara ambien. Hasil korelasi ini didukung oleh penelitian (Syafaati *et al.*, 2023) yang mendapatkan nilai korelasi PM_{2,5} dengan curah hujan sebesar -0,7 sehingga adanya hubungan kuat dan bernilai negatif.

4. KESIMPULAN

Penelitian ini berhasil menggunakan model LSTM untuk memprediksi kualitas udara PM_{2,5} di Stasiun Klimatologi Kalimantan Barat, Kecamatan Jongkat, Kabupaten Mempawah. Data yang digunakan berupa data historis dimulai dari tanggal 20 September 2021 sampai dengan 31 Desember 2023 dengan pembagian 90% untuk *training* data dan 10% untuk *testing* data dan menghasilkan RMSE terbaik 3,40 dan fungsi loss 0,0058. Pengujian dilakukan dalam dua skenario: kondisi udara baik dan kondisi sedang hingga tidak sehat, dengan hasil prediksi yang cukup akurat.

Analisis korelasi menunjukkan hubungan searah yang paling kuat yakni antara PM_{2,5} dan kecepatan angin sebesar 0,64 dimana saat kecepatan angin bernilai tinggi sehingga konsentrasi PM_{2,5} meningkat, sedangkan adanya hubungan negatif dengan korelasi tidak searah antara PM_{2,5} dan curah hujan sebesar -0,49 yakni ketika nilai curah hujan tinggi maka konsentrasi PM_{2,5} menurun. Sedangkan korelasi hubungan yang lemah terjadi pada hubungan suhu udara dengan PM_{2,5} dengan sebesar 0,12 dan hubungan PM_{2,5} dengan kelembaban udara yang sangat lemah sebesar 0,088.

Adapun saran untuk penelitian ini adalah menggunakan data dengan historis lebih panjang dan variatif sehingga model RNN-LSTM ini dapat mempelajari, melatih, dan menguji data prediktor dengan berbagai kondisi dan kategori sehingga dihasilkan prediksi yang lebih akurat. Selain itu, menambahkan parameter lain untuk memprediksi kualitas udara seperti data meteorologi yang *running* bersama parameter PM_{2,5} untuk mendapatkan prediksi secara langsung terkait kualitas udara berdasarkan cuaca yang mempengaruhinya.

5. DAFTAR PUSTAKA

Anggraini, M. (2023). Literatur: Pengertian, Tujuan, Ciri, Fungsi, Jenis, & Contoh. Cerdika. Diakses pada 23 November 2024, dari Cerdika.

Arkadia, A., Hananto, B., & Prasvita, D. S. (2022). Optimasi Long Short Term Memory Dengan Adam Menggunakan Data Udara Kota DKI Jakarta. pp. 559–568.

Stasiun Klimatologi Kelas II Kalimantan Barat. (2023). Buletin Iklim Kalbar November 2023. BMKG.

Battelle. (2001). Environmental Technology Verification Report: Met One Instruments BAM 1020 Particle Monitor. U.S. Environmental Protection Agency. Retrieved from EPA Archive.

Chen, H., Guan, M., & Li, H. (2021). Air quality prediction based on integrated dual

- LSTM model. *IEEE Access*, 9, 93285-93297.
- Karyadi, Y. (2022). Prediksi Kualitas Udara Dengan Metoda LSTM, Bidirectional LSTM, dan GRU. *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)* 9(1), pp. 671–684. doi: 10.35957/jatisi.v9i1.1588.
- Khairani, R., Anggiane, M., Woro, D. A., & Widowati, H. (2023). Peningkatan Pengetahuan Tentang Kesehatan Paru Dan Polusi Udara Pada Pengemudi Ojek Online. *J. ABDIMAS Kesehat. TERPADU*, 2(1).
- Maharani, S., & Aryanta, W. R. (2023). Dampak Buruk Polusi Udara Bagi Kesehatan Dan Cara Meminimalkan Risikonya. *J. Ecocentrism*, 3(2), pp. 47–58.
- Millah, H. R. (2022). Hubungan Faktor Meteorologis Dan Kepadatan Lalu Lintas Dengan Kualitas Udara Di Kota Tabanan. Poltekkes Kemenkes Denpasar Jurusan Kesehatan Lingkungan.
- Rafidah, R., Rivai, A., & Kasim, K. P. (2023). Edukasi Mengurangi Paparan Debu PM2,5 Dengan Penggunaan Alat Pelindung Diri Bagi Pedagang Kaki Lima Dan Warga Di Wilayah Jalan Rappocini Raya Kota Makassar. *Media Implementasi Ris. Kesehatan*, (4)1, pp. 34–38.
- Syafaati, A. D., Utami, S. N. N., & Arifin, S. (2023). Analisis Kualitas Udara Parameter PM2.5 di Wilayah Kota Sorong Berbasis ISPU. *Megasains*, 14(2), 6-13. Stasiun Pemantau Atmosfer Global Puncak Vihara Klademak Sorong doi: 10.46824/megasains.v14i2.131
- World Meteorological Organization. (2018). *Guide to Instruments and Methods of Observation (WMO-No. 8)*. Geneva: WMO