

Udara bersih hak setiap orang



Sumber: Dokumen pribadi

PENDAHULUAN



Berkurang 1,4 tahun



Berdasarkan laporan AQLI, polusi udara yang berisi partikel halus PM2.5 berpotensi mengurangi usia harapan hidup rata-rata warga Indonesia hingga 1,4 tahun dibandingkan jika kualitas udara di Indonesia memenuhi standar WHO, yaitu $5 \mu\text{g}/\text{m}^3$ [4].

CLEAN
AIR
FUND

Udara bersih menjadi katalisator untuk mendorong kemajuan agenda 2030 sesuai *Sustainable Development Goals* (SDGs), secara signifikan berpengaruh pada poin 3 dan 7 [6].



Krisis iklim
Bahan bakar fosil
Industri/manufaktur

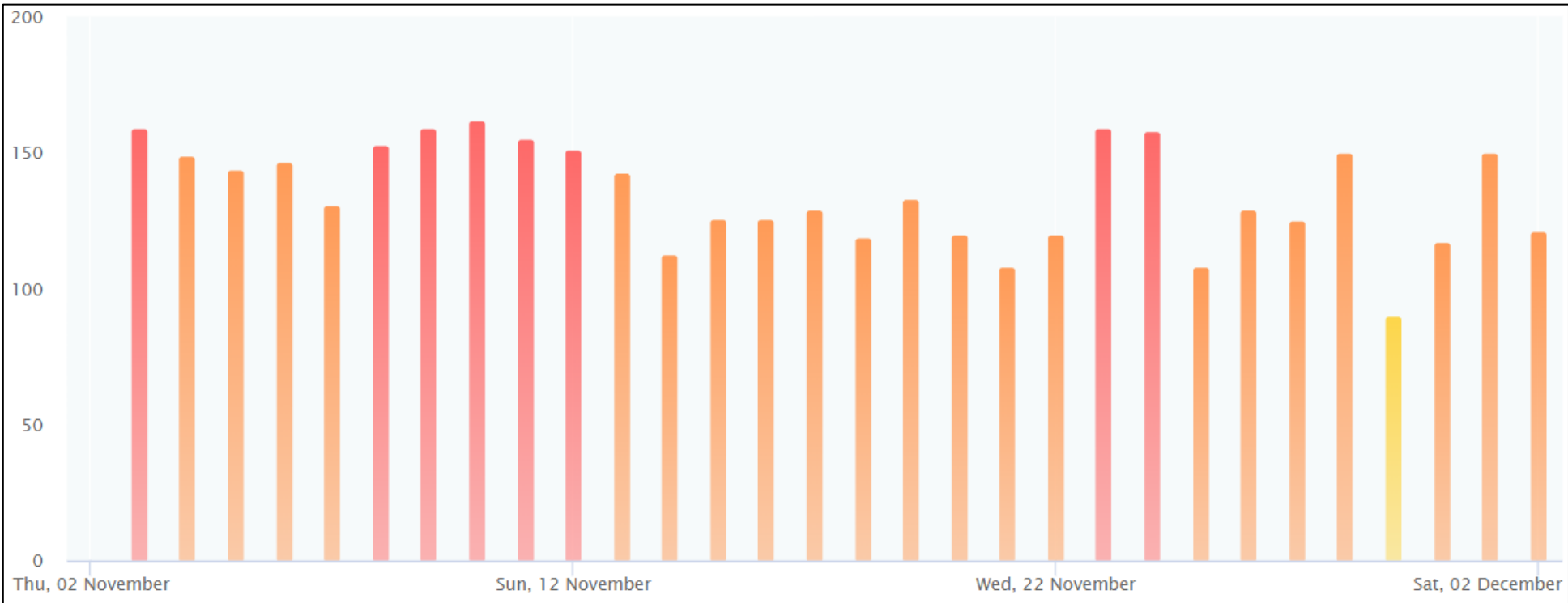
[4] BBC News Indonesia. 2023. *Polusi udara: Indonesia masuk enam negara paling berkontribusi terhadap polusi udara global, warga akan gugat pemerintah dan industri* <https://www.bbc.com/indonesia/articles/c72enp76622o>.

[6] Clear Air Fund. 2023. *Sustainable Development Goals and Air Pollution*. Info@cleanairfund.org.







PENDAHULUAN

#	CITY	AQI US
1	South Tangerang, Provinsi Banten	162
2	Jakarta, Jakarta	157
3	Kabupaten Serang, Provinsi Banten	131
4	Kota Surabaya, Jawa Timur	101
5	Kota Bandung, Jawa Barat	100

5 kota di Indonesia dengan kualitas udara terburuk pada 4 Desember 2023 [9]



Kualitas udara rerata harian Jakarta pada 2 November - 2 Desember 2023 [9]

US AQI Level		PM2.5 (µg/m³)	Health Recommendation (for 24 hour exposure)
WHO PM2.5 (µg/m³) Recommended Guidelines as of September 22, 2021: 0-5.0			
	Good 0-50	0-12.0	Air quality is satisfactory and poses little or no risk.
	Moderate 51-100	12.1-35.4	Sensitive individuals should avoid outdoor activity as they may experience respiratory symptoms.
	Unhealthy for Sensitive Groups 101-150	35.5-55.4	General public and sensitive individuals in particular are at risk to experience irritation and respiratory problems.
	Unhealthy 151-200	55.5-150.4	Increased likelihood of adverse effects and aggravation to the heart and lungs among general public.
	Very Unhealthy 201-300	150.5-250.4	General public will be noticeably affected. Sensitive groups should restrict outdoor activities.
	Hazardous 301+	250.5+	General public at high risk of experiencing strong irritations and adverse health effects. Should avoid outdoor activities.

US AQI level and WHO air quality guidelines 2021 [12]

[9] iqair.com. 2023. *Indeks Kualitas Udara (AQI) Jakarta dan Polusi Udara di Indonesia*. <https://www.iqair.com/id/indonesia/jakarta>. 4 Oktober 2023 (13.40WIB) dan 2 Desember 2023 (10.30WIB).

[12] iqair.com. 2016. *What is the difference between US AQI and WHO air quality guidelines?* <https://www.iqair.com/us/newsroom/what-is-the-difference-between-the-us-aqi-and-who-air-quality-guidelines>.

Di Indonesia, terdapat keterbatasan jumlah dan akurasi sensor kualitas udara yang terpasang di tiap kabupaten/kota yang mengakibatkan deteksi konten dan kualitas udara di suatu tempat selama ini terbatas (tidak ada di semua titik atau tempat) menjadi kurang akurat secara lokal.

7 unit

sensor (stasiun) pemantau kualitas udara per September 2023 di Jakarta [11].

661,5 km^2

luas Jakarta

5 – 11 km^2

jangkauan ideal 1 unit sensor



Pengukuran dan pemantauan kualitas udara di Jakarta sejauh ini masih belum akurat dan ideal

Sistem pemantau kualitas udara yang **efisien, terjangkau, dan akurat** untuk kesehatan manusia dan pengendalian polusi udara sangat penting [1].



Masyarakat dapat mengetahui secara akurat kualitas udara saat itu dan tempat-tempat di mana masyarakat biasa beraktivitas. *Treatment* dan *suggestion* yang lebih tepat di daerah secara lokal.

Alternatif Solusi :

Data, *Internet of Things*,
***Machine Learning* (ML)**
***Artificial Intelligence* (AI)**



Mengestimasi konten dan kualitas udara secara efisien dan lebih terjangkau berbasis gambar langit.

PENDAHULUAN

IMAGE-BASED AIR QUALITY ANALYSIS USING DEEP CONVOLUTIONAL NEURAL NETWORK

Avijoy Chakma¹, Ben Vizona¹, Tingting Cao², Jerry Lin², Jing Zhang¹

¹Computer Science Department, Lamar University, TX, 77710, USA

²Center of Advances in Water and Air Quality, Lamar University, TX, 77710, USA

ABSTRACT

Air pollution may cause many severe diseases. An efficient air quality monitoring system is of great benefit for human health and air pollution control. In this paper, we study image-based air quality analysis, in particular, the concentration estimation of particulate matter with diameters less than 2.5 micrometers (PM_{2.5}). The proposed method uses a deep Convolutional Neural Network (CNN) to classify natural images into different categories based on their PM_{2.5} concentrations. In order to evaluate the proposed method, we created a dataset that contains total 591 images taken in Beijing with corresponding PM_{2.5} concentrations. The experimental results demonstrate that our method are valid for image-based PM_{2.5} concentration estimation.

Index Terms— Particulate Matter, Air Quality Analysis, Deep Convolutional Neural Networks, Image Classification

1. INTRODUCTION

efficient and affordable way to monitor air quality. For example, smartphone users can take photos and estimate the real-time local air quality by themselves.

In recent years, using computer vision and machine learning techniques to analyze the air quality (e.g., haze level and PM_{2.5} concentration) has begun to receive increasing attention and a few papers can be found in the literature.

The existing image-based haze level analysis methods are mainly inspired by the dehazing algorithms [1, 2]. In computer vision and computer graphics areas, the haze image can be modeled as:

$$I(x) = J(x)t(x) + A(1-t(x)), \quad (1)$$

where I is the observed image, J is the haze-free scene image, A is the sky luminance that indicates the lighting condition, and t is transmission matrix that describes the portion of the light that reach the camera. The transmission matrix $t(x)$ can be expressed as

$$t(x) = e^{-\beta d(x)} \quad (2)$$

where $d(x)$ is the scene depth map and β is the atmosphere

Figure 3 shows some images that were correctly classified by the CNN. Each column illustrates 10 images from each class, Good, Moderate, and Severe.

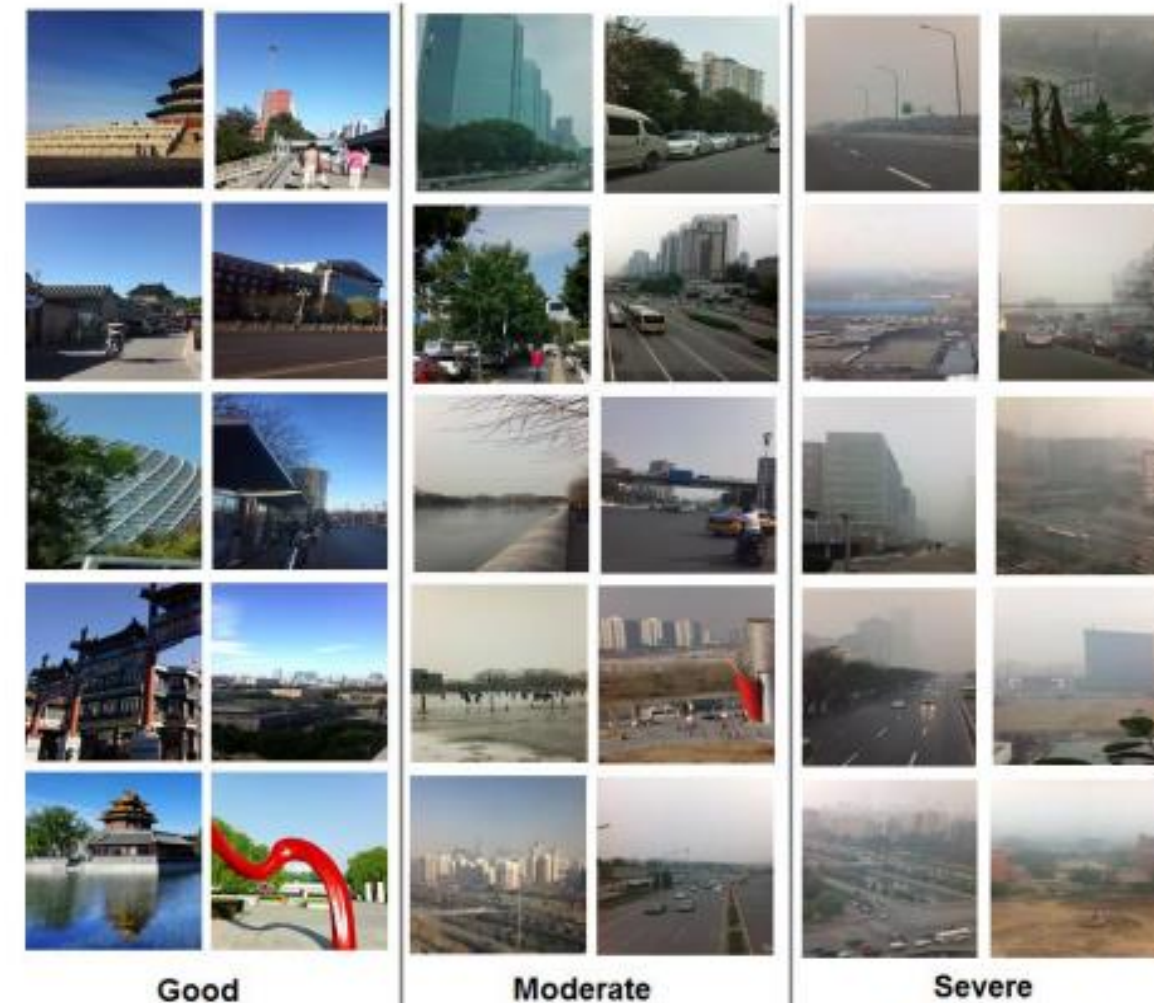


Fig. 3. Example images correctly classified by the CNN

Table 1. The performance of CNN and Random Forest

Methods	Classification Accuracy (Average)
Random Forest	63.62%
CNN	68.74%

Penelitian oleh Chakma et al berhasil mengklasifikasikan gambar-gambar di tempat lokal berdasarkan konsentrasi polutan PM_{2.5} menjadi 3 kelas (*Good*, *Moderate*, dan *Severe*) menggunakan *Deep Convolutional Neural Network* dengan akurasi mencapai 68% [1].

Penelitian untuk prediksi/estimasi kualitas dan konten udara berbasis gambar masih sangat langka.

Preview jurnal IEEE Chakma et al : *Image-based air quality analysis using deep convolutional neural network* [1]



FINAL PAPER COMPETITION

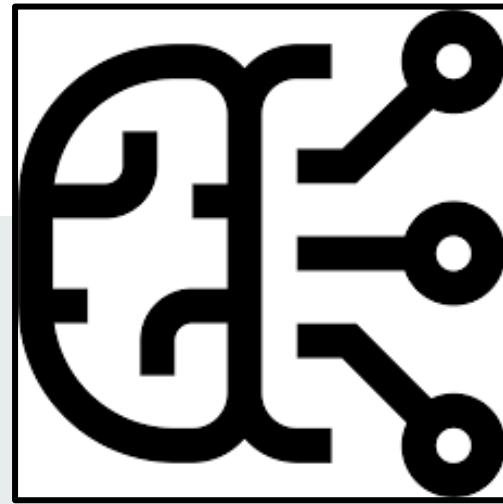
LOGIKA UI 2024

ESTIMASI KONTEN UDARA BERBASIS GAMBAR LANGIT TEMPAT LOKAL MENGGUNAKAN *DEEP CONVOLUTIONAL NEURAL NETWORK*

24-05-015-0

Hendri Erka Setya
Shawn Michael Dayanti Intong

TUJUAN



Membangun (mengembangkan) model prediksi konten udara berbasis gambar langit tempat lokal dengan *Deep Convolutional Neural Network* (CNN).



Mengestimasi konten udara dari gambar-gambar langit tempat lokal.

Studi kasus : FMIPA UI, Depok

METODOLOGI

DATASET

Sumber Dataset :

Data sekunder, diperoleh dari Kaggle
(*Air pollution image dataset from India and Nepal*) [3].

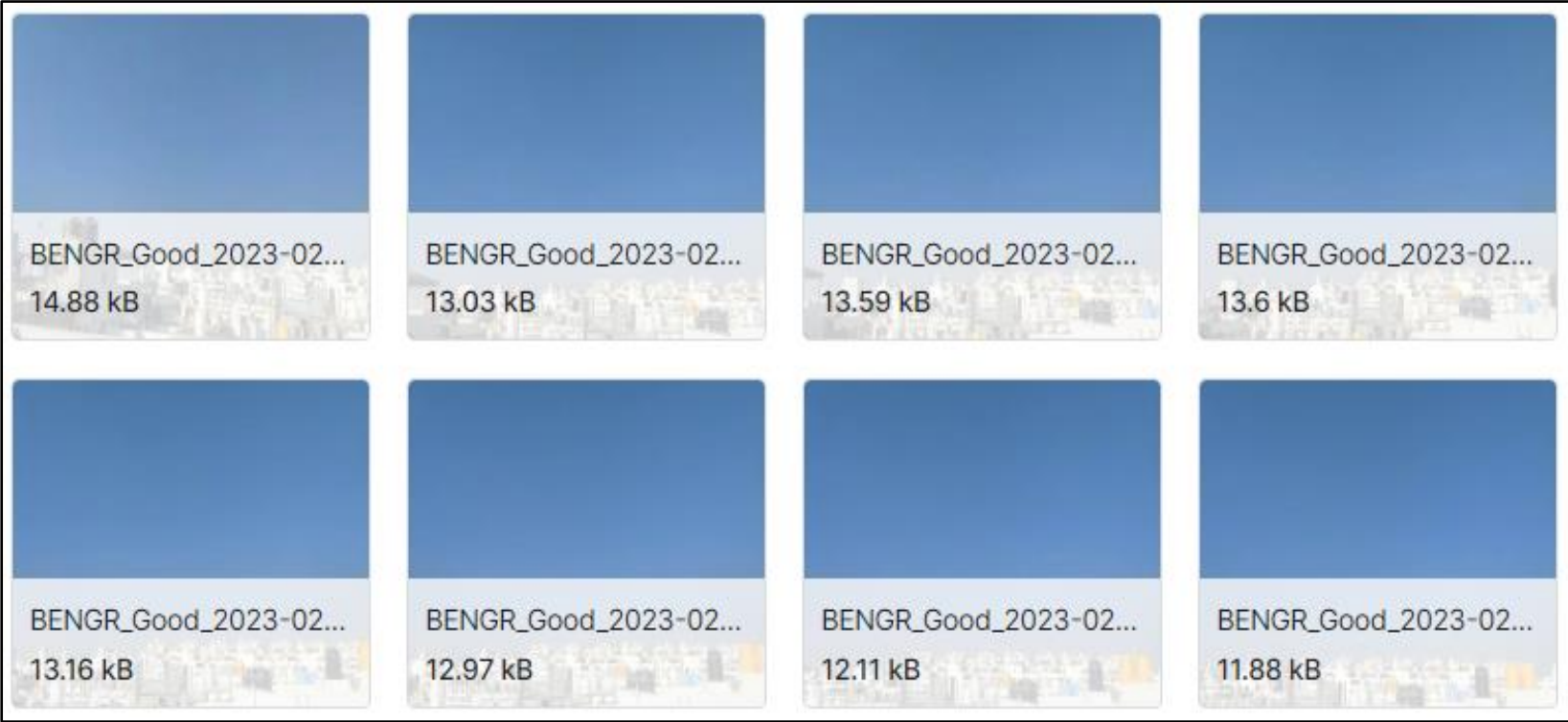
Konten Dataset :

- 1 file csv berisi keterangan masing-masing gambar dan 12240 gambar pemandangan dengan *background* langit dari beberapa kota di India & Nepal.

Sampling Frequency :

Pengambilan gambar per jam selama kurang lebih 3 bulan tidak sekuensial serta pencatatan konten-konten udara berdasarkan website https://app.cpcbcr.com/AQI_India/ (untuk India) dan <https://www.tomorrow.io/weather/NP/4/Biratnagar/079711/hourly/> (untuk Nepal).

Preview Dataset :



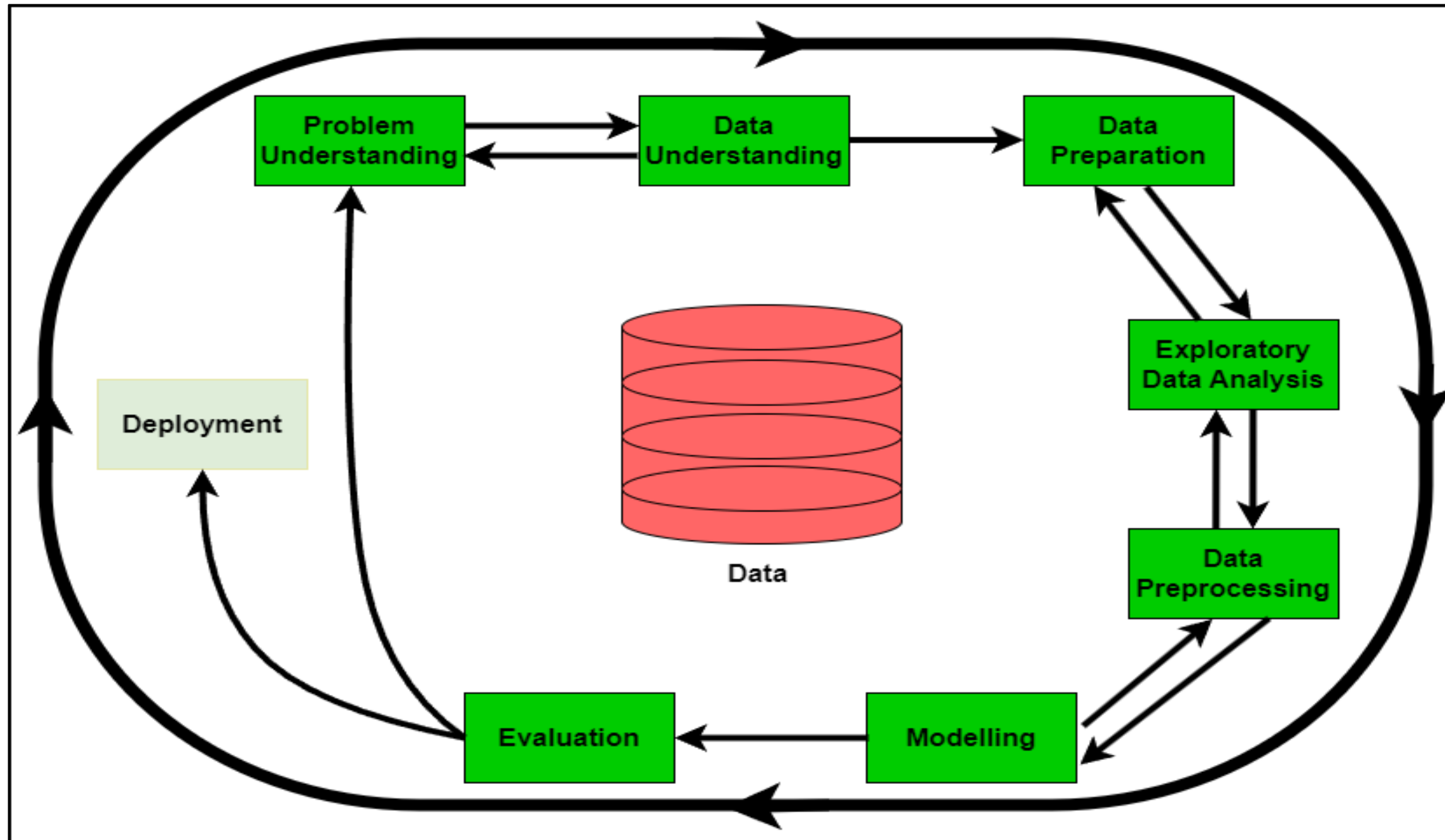
	Location	Filename	Year	Month	Day	Hour	AQI	PM2.5	PM10	O3	CO	SO2	NO2	AQI_Class
0	Biratnagar, Nepal	BRI_Un_2023-02-02- 12.00-9.jpg	2023	2	2	12:00	158	70.08	100.82	58.89	0.49	4.4	1.04	d_Unhealthy
1	Biratnagar, Nepal	BRI_Un_2023-02-02- 12.00-8.jpg	2023	2	2	12:00	158	70.08	100.82	58.89	0.49	4.4	1.04	d_Unhealthy
2	Biratnagar, Nepal	BRI_Un_2023-02-02- 12.00-7.jpg	2023	2	2	12:00	158	70.08	100.82	58.89	0.49	4.4	1.04	d_Unhealthy
3	Biratnagar, Nepal	BRI_Un_2023-02-02- 12.00-6.jpg	2023	2	2	12:00	158	70.08	100.82	58.89	0.49	4.4	1.04	d_Unhealthy
4	Biratnagar, Nepal	BRI_Un_2023-02-02- 12.00-5.jpg	2023	2	2	12:00	158	70.08	100.82	58.89	0.49	4.4	1.04	d_Unhealthy

[3] Adarsh Rouniyar, Sapdo utomo, John A, & Pao-Ann Hsiung. 2023. *Air Pollution Image Dataset from India and Nepal [Dataset]*. Kaggle. <https://doi.org/10.34740/KAGGLE/DS/3152196>.
<https://www.kaggle.com/datasets/adarshrrouniyar/air-pollution-image-dataset-from-india-and-nepal>

10

ALUR PENELITIAN & SOFTWARE

Alur Metode Penelitian :



Software :



Library :



Alur metode penelitian framework *CRISP-DM* (Cross Industry Standard Process for Data Mining) [10]

DATA UNDERSTANDING

Preview Dataset CSV

	Location	Filename	Year	Month	Day	Hour	AQI	PM2.5	PM10	O3	CO	SO2	NO2	AQI_Class
0	Biratnagar, Nepal	BRI_Un_2023-02-02- 12.00-9.jpg	2023	2	2	12:00	158	70.08	100.82	58.89	0.49	4.4	1.04	d_Unhealthy
1	Biratnagar, Nepal	BRI_Un_2023-02-02- 12.00-8.jpg	2023	2	2	12:00	158	70.08	100.82	58.89	0.49	4.4	1.04	d_Unhealthy
2	Biratnagar, Nepal	BRI_Un_2023-02-02- 12.00-7.jpg	2023	2	2	12:00	158	70.08	100.82	58.89	0.49	4.4	1.04	d_Unhealthy
3	Biratnagar, Nepal	BRI_Un_2023-02-02- 12.00-6.jpg	2023	2	2	12:00	158	70.08	100.82	58.89	0.49	4.4	1.04	d_Unhealthy
4	Biratnagar, Nepal	BRI_Un_2023-02-02- 12.00-5.jpg	2023	2	2	12:00	158	70.08	100.82	58.89	0.49	4.4	1.04	d_Unhealthy

Features dan Datatype

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 12240 entries, 0 to 12239
Data columns (total 14 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Location    12240 non-null  object
1   Filename    12240 non-null  object
2   Year        12240 non-null  int64
3   Month       12240 non-null  int64
4   Day         12240 non-null  int64
5   Hour        12240 non-null  object
6   AQI         12240 non-null  int64
7   PM2.5       12240 non-null  float64
8   PM10        12240 non-null  float64
9   O3          11938 non-null  float64
10  CO          11660 non-null  float64
11  SO2         10757 non-null  float64
12  NO2         11800 non-null  float64
13  AQI_Class   12240 non-null  object
dtypes: float64(6), int64(4), object(4)
memory usage: 1.3+ MB
```

Jumlah Baris dan Kolom

(12240, 14)

Jumlah Baris Duplikat

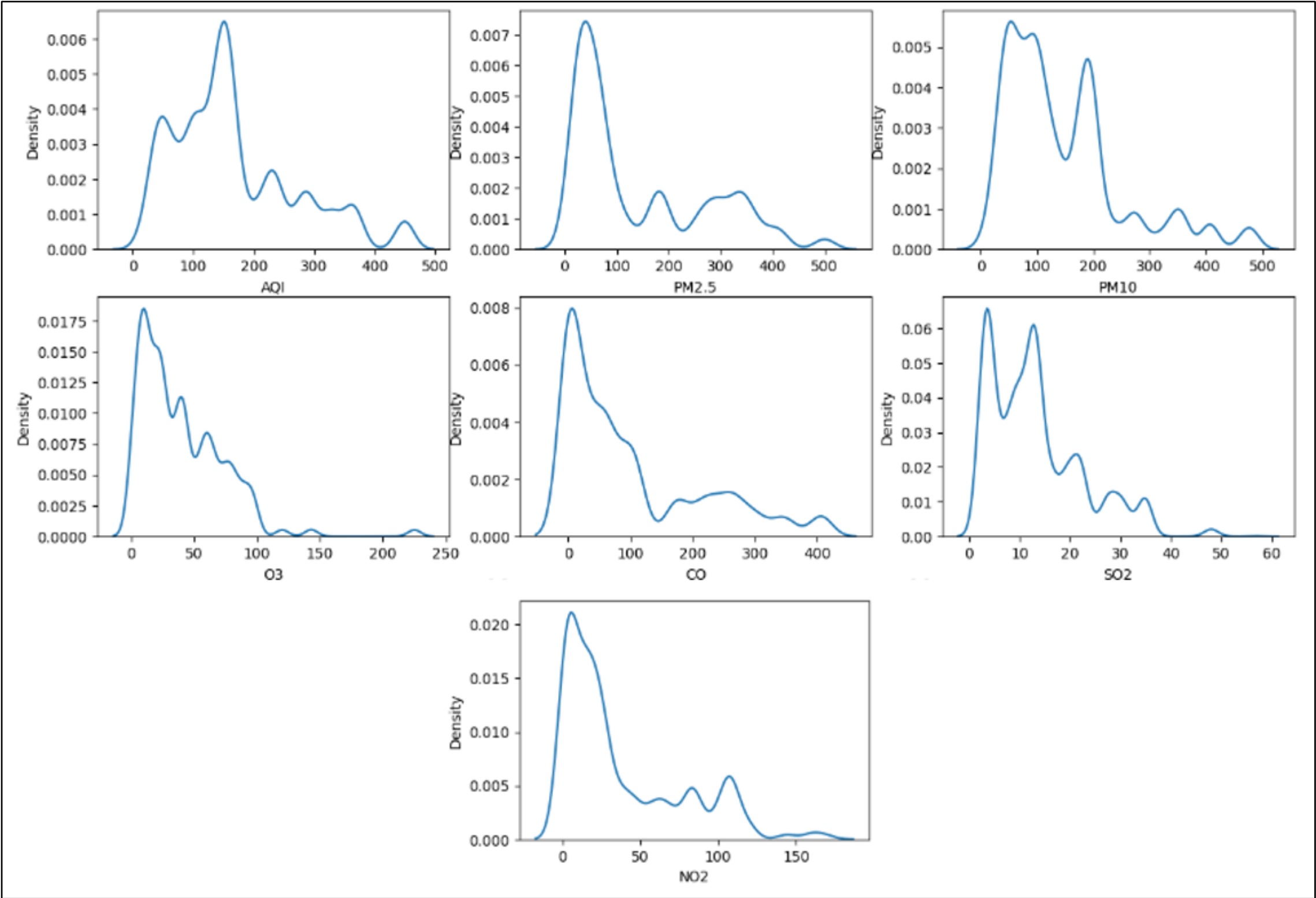
0

DATA UNDERSTANDING

Jumlah Missing Values Tiap Feature

Location	0
Filename	0
Year	0
Month	0
Day	0
Hour	0
AQI	0
PM2.5	0
PM10	0
O3	302
CO	580
SO2	1483
NO2	440
AQI_Class	0
dtype: int64	

Distribusi Feature Numerik



DATA UNDERSTANDING

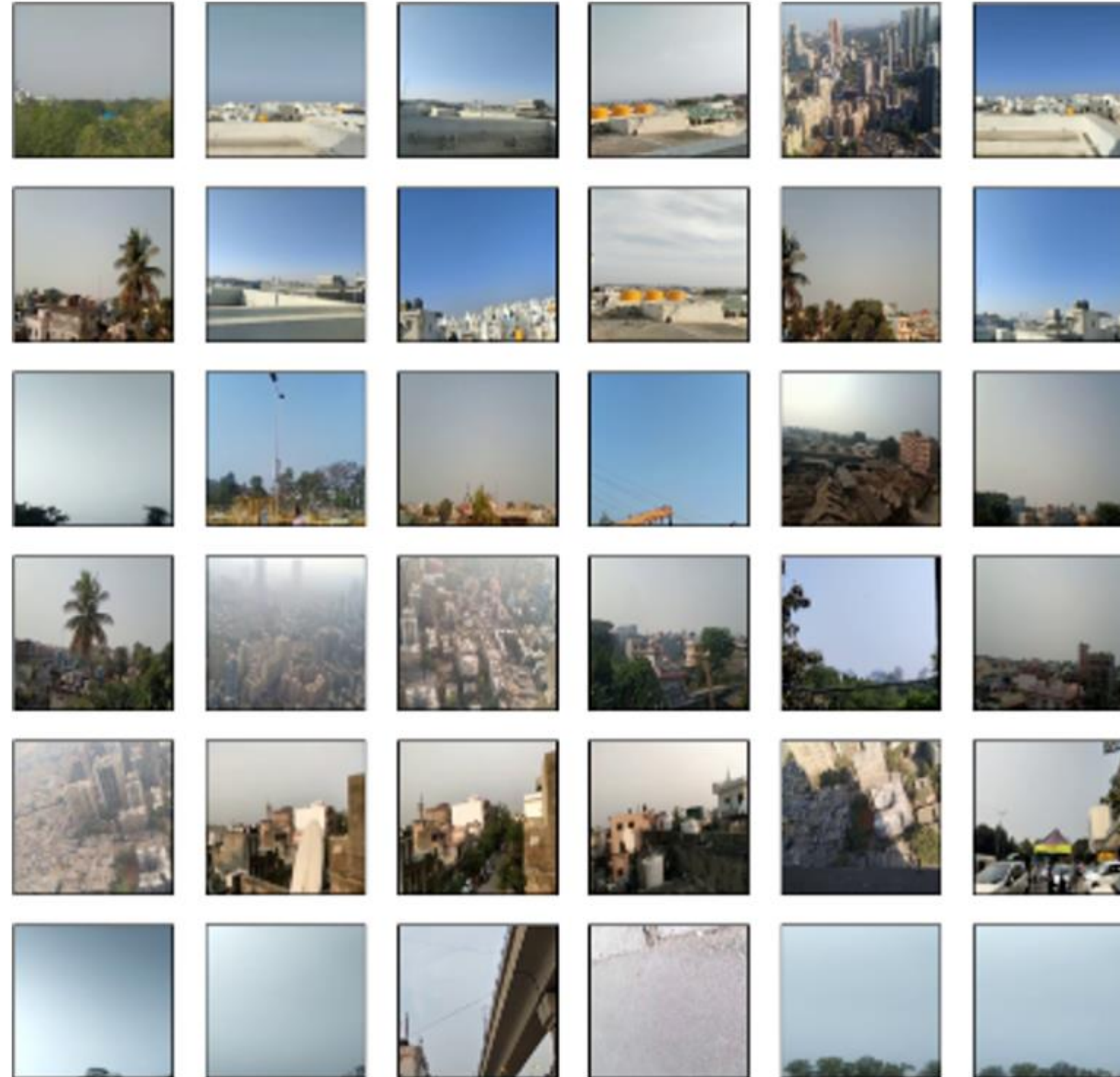
Statistika Deskriptif Feature Numerik

	Year	Month	Day	AQI	PM2.5	PM10	O3	CO	SO2	NO2
count	12240.000000	12240.000000	12240.000000	12240.000000	12240.000000	12240.000000	11938.000000	11660.000000	10757.000000	11800.000000
mean	2022.947631	2.686111	12.114706	167.626797	142.999267	145.663935	39.491436	101.230361	13.239595	37.843576
std	0.222780	1.774345	8.269053	102.818213	130.745815	104.506951	33.342031	115.910128	9.850507	39.551303
min	2022.000000	2.000000	1.000000	15.000000	4.000000	7.000000	1.000000	0.000000	2.000000	0.670000
25%	2023.000000	2.000000	3.000000	97.000000	35.000000	64.000000	13.000000	4.000000	4.400000	7.000000
50%	2023.000000	2.000000	13.000000	152.000000	70.080000	113.000000	31.000000	52.000000	10.000000	20.000000
75%	2023.000000	3.000000	20.000000	230.000000	257.000000	198.000000	59.660000	174.000000	20.000000	64.000000
max	2023.000000	10.000000	28.000000	450.000000	500.000000	480.000000	225.000000	410.000000	57.000000	169.000000

Statistika Deskriptif Feature Kategorik

	Location	Filename	Hour	AQI_Class
count	12240	12240	12240	12240
unique	8	12240	21	6
top	Mumbai	BRI_Un_2023-02-02- 12.00-9.jpg	8:30	c_Unhealthy_for_Sensitive_Groups
freq	2203	1	3300	2863

DATA UNDERSTANDING



Beberapa gambar langit dengan AQI berbeda

DATA PREPARATION

HANDLING MISSING VALUES

- Uji normalitas Shapiro-Wilk pada features dengan missing values, dengan formulasi :

$$T_3 = \frac{[\sum_{i=1}^n a_i(x_{n-i+1} - x_i)]^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1)$$

T_3 : p-value uji normalitas Shapiro Wilk
 a_i : koefisien uji Shapiro Wilk,
 x_{n-i+1} : data ke n+i-1,
 x_i : data ke-i,
 \bar{x} : rata-rata data.

- Dengan ketentuan :

Jika $p - value \leq \alpha$ ----> **Tidak Normal** ----> Imputasi Modus

Jika $p - value > \alpha$ ----> **Normal** ----> Imputasi Mean

- Dengan $\alpha = 0.05$, diperoleh































Tabel 1. Hasil Uji Normalitas *Feature* dengan *Missing Values*

Feature	P-Value ($\alpha = 0,05$)
O3	0,0
CO	0,0
SO2	0,0
NO2	0,0

- Diperoleh bahwa data **tidak normal**, maka imputasi missing values dengan **modus**.

EXPLORATORY DATA ANALYSIS

Beberapa gambar langit dengan AQI Berbeda

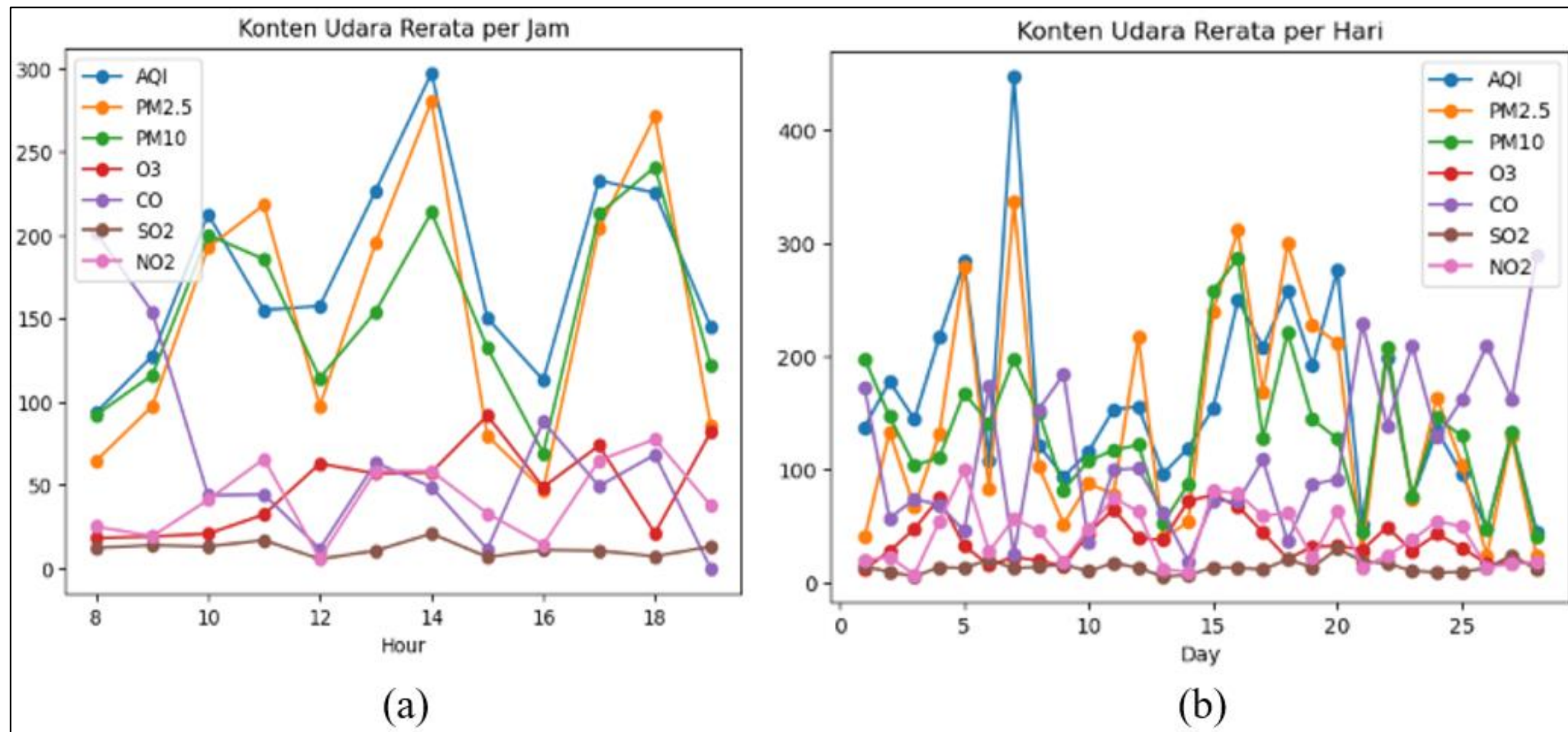
Good						
Moderate						
Unhealthy for Sensitive Group						
Unhealthy						
Very Unhealthy						
Hazardous						

EXPLORATORY DATA ANALYSIS

	Year	Month	Day	Hour	AQI	PM2.5	PM10	O3	CO	SO2	NO2
Year	1.000000	-0.969051	0.287564	0.107495	-0.142615	-0.075521	-0.119981	0.210555	0.102303	0.031558	0.071698
Month	-0.969051	1.000000	-0.320919	-0.169344	0.072730	0.025059	0.090953	-0.226420	-0.029749	-0.028809	-0.085523
Day	0.287564	-0.320919	1.000000	0.010107	-0.063989	0.144846	0.016649	0.042955	0.253177	0.282263	0.189950
Hour	0.107495	-0.169344	0.010107	1.000000	0.432634	0.316370	0.275443	0.603237	-0.488069	-0.024663	0.264029
AQI	-0.142615	0.072730	-0.063989	0.432634	1.000000	0.838827	0.723872	0.158049	-0.360354	0.311084	0.549148
PM2.5	-0.075521	0.025059	0.144846	0.316370	0.838827	1.000000	0.841503	0.126404	-0.213667	0.335839	0.748461
PM10	-0.119981	0.090953	0.016649	0.275443	0.723872	0.841503	1.000000	0.217285	-0.210574	0.246111	0.629362
O3	0.210555	-0.226420	0.042955	0.603237	0.158049	0.126404	0.217285	1.000000	-0.357431	-0.197009	0.173887
CO	0.102303	-0.029749	0.253177	-0.488069	-0.360354	-0.213667	-0.210574	-0.357431	1.000000	0.237343	-0.092918
SO2	0.031558	-0.028809	0.282263	-0.024663	0.311084	0.335839	0.246111	-0.197009	0.237343	1.000000	0.313228
NO2	0.071698	-0.085523	0.189950	0.264029	0.549148	0.748461	0.629362	0.173887	-0.092918	0.313228	1.000000

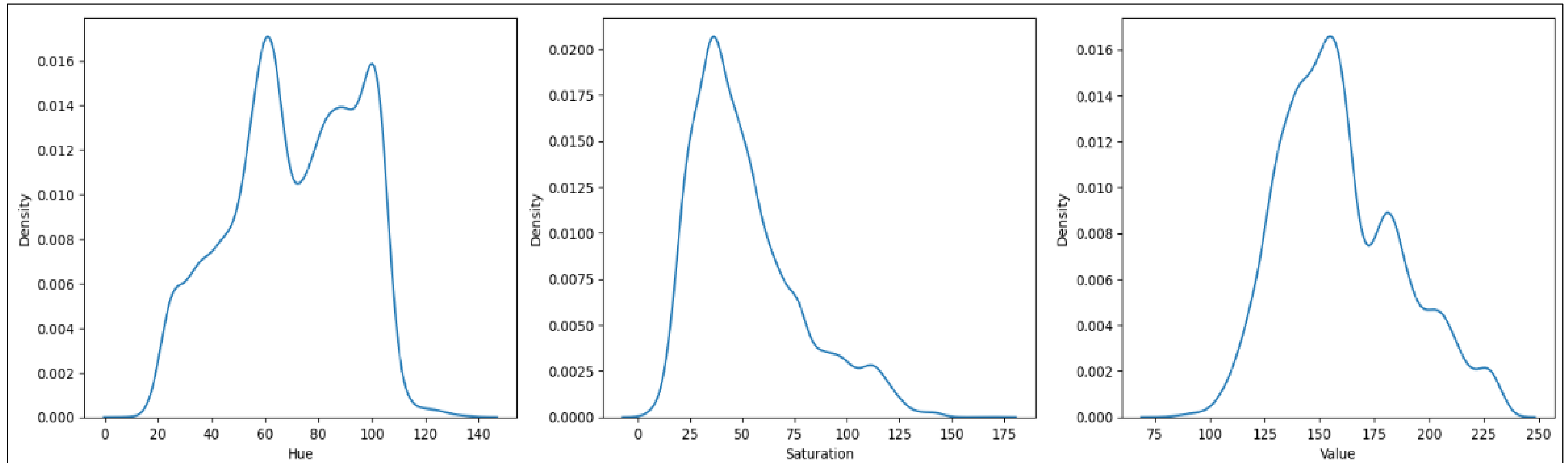
Heatmap korelasi Pearson antar-feature

EXPLORATORY DATA ANALYSIS



Lineplot konten udara rerata per jam (a) dan per hari (b)

EXPLORATORY DATA ANALYSIS



Distribusi *HSV color model* pada dataset gambar

DATA PREPROCESSING

DROP IRRELEVANT FEATURES

AQI_Class Location Minutes

dapat dihilangkan karena problem kami adalah *multi regression problem* dan kurangnya variansi dalam feature-feature tersebut.

IMAGE PREPROCESSING

- Penyesuaian input model, gambar dengan ukuran (224, 224, 3) dikonversi dari color channel RGB ke BGR.
- Kemudian setiap color channel dibuat zero-centered terhadap dataset Imagenet tanpa *scaling*.

MODELLING

- Input yang diterima sendiri berupa gambar dengan *color channel* BGR berukuran 224 dikali 224.
- Gambar-gambar serta konten udara yang bersesuaian dengan gambar tersebut dipisahkan melakukan *train test split* sebesar 80% untuk *training* dan 20% untuk *validation*.
- Adapun *hyperparameter* lain yang didapatkan dari hasil *hyperparameter tuning* seperti tertera pada tabel.

Hyperparameter	Nilai
Optimizer	Adam
Loss Function	Mean Squared Error
Batch Size	16
Epochs	150

SKEMA PENGUJIAN MODEL

Skema uji performa model :

Model	Features
VGG16	Images
VGG16 + Dense Layer	Images, Timeseries
VGG16 + Dense Layer	Images, Timeseries, HSV

Metrik evaluasi :

- RMSE (*Root Mean Squarred Error*) untuk menguji performa model pada kasus regresi (prediksi).
- RMSE juga sering digunakan sebagai metrik standar menguji performa model prediksi penelitian meteorologi, kualitas udara, dan perubahan iklim [13].
- *Multioutput regression* dengan 7 target (feature): “AQI, PM2.5, PM10, O3, CO, SO2, NO2”.

$$RMSE_i = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2} \tag{13}$$

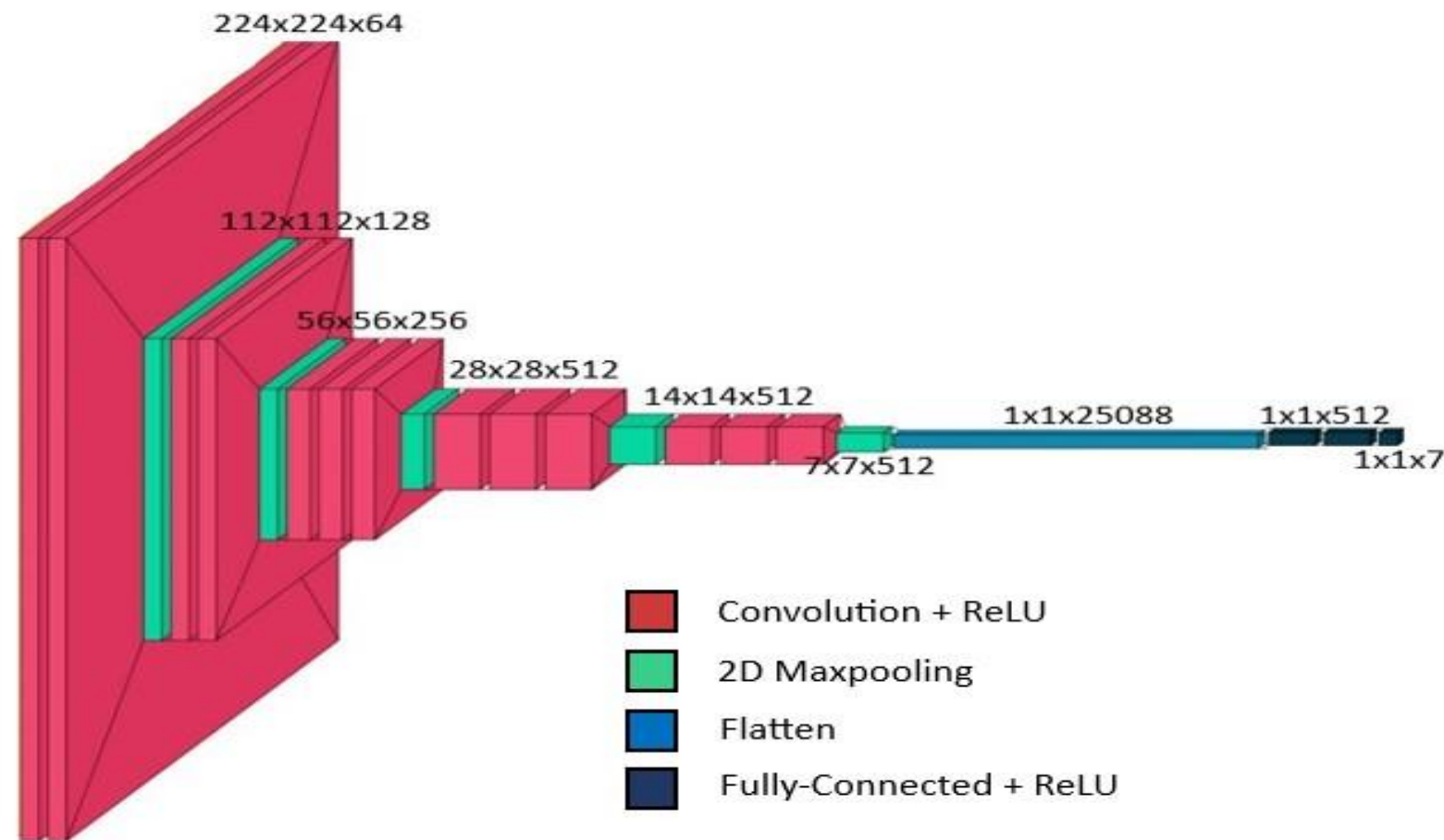
$$RMSE = \frac{\sum_{i=1}^n RMSE_i}{n} \tag{14}$$

Keterangan :

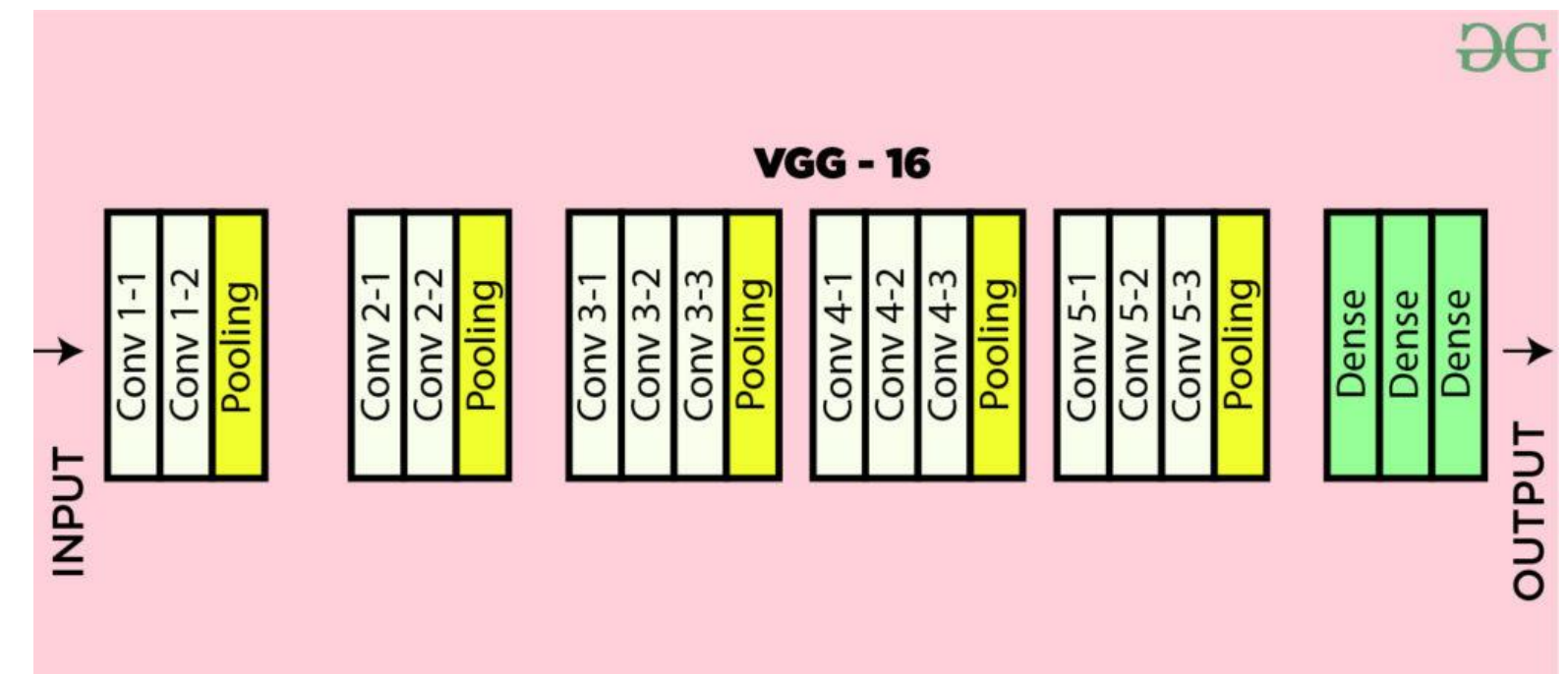
- RMSE* : nilai *root mean squared error*,
- N* : jumlah data,
- $y_i - \bar{y}$: perbedaan data yang diprediksi dengan data sebenarnya
- n* : jumlah feature (7)

[13] Tianfeng, C. & Draxler, R. R., 2014. *Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments Against Avoiding RMSE in The Literature*. Geoscientific Model Development, pp. 1247-1250.

VISUAL GEOMETRY GROUP 16 (VGG16)



Arsitektur Visual Geometry Group-16 (3D)



Arsitektur Visual Geometry Group-16 (2D)

- VGG16 terdiri atas 13 *convolution layer* dan 3 *fully-connected layer* [7]. Input gambar yang diterima berupa gambar BGR hasil preprocessing dengan ukuran "224x224x3".
- Setelah flattening feature gabungan satu dimensi tersebut masuk ke 3 *fully connected layer* dengan yang terakhir memiliki jumlah neuron/units "7", yaitu sesuai dengan jumlah output yang ingin diprediksi dengan activation function "linear" karena merupakan masalah regresi.
- Pada model ini block *convolution* dibiarkan dan dibuat *untrainable* dan menggunakan *pretrained weights* langsung dari *Imagenet* sehingga *layer trainable* hanyalah *fully-connected layer*.

VGG16 + TIME SERIES

- VGG16 digunakan sebagai *backbone* atau *feature extractor* di mana feature map dari blok *convolution* terakhir dilakukan *flattening* yang kemudian digabungkan dengan *feature time series* yang bersesuaian dengan waktu pengambilan tiap gambar yaitu "Year, Month, Day, dan Hour".
- Hal ini **memungkinkan model untuk mempelajari tren serta seasonalitas kondisi langit lokal setempat.**
- Penambahan feature menimbulkan masalah ketidakseragaman satuan data sehingga perlu dilakukan standarisasi. Standarisasi dapat menyeragamkan skala data dengan melakukan transformasi distribusi dari masing-masing feature menjadi gaussian dengan mean "0" dan standar deviasi "1".

$$Z = \frac{x - \mu}{\sigma}$$

Keterangan :

x : nilai feature,

μ : nilai rata-rata feature,

σ : standar deviasi feature,

z : nilai hasil transformasi atau standarisasi *feature*

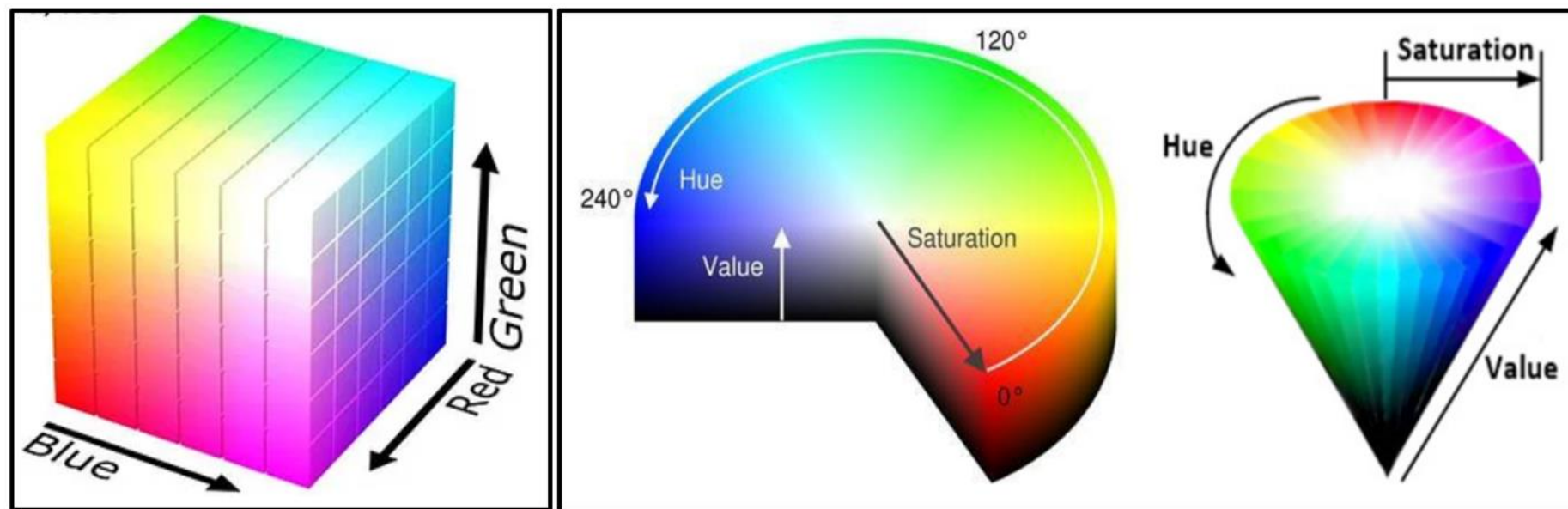
- *Feature* gabungan yang sudah distandarisasi tersebut kemudian di-*feed-forward* ke *dense layer* dengan jumlah neuron yang sama seperti model sebelumnya.

VGG16 + TIME SERIES

	0	1	2	3	4	5	6	7	8	9	...	509	510	511	Year	Month	Day	Hour
0	1.715331	0.000000	1.868870	0.023432	1.181104	0.000000	8.548042	1.019641	0.0	0.000000	...	0.000000	1.232765	0.283906	2023	2	2	12
1	0.000000	1.208342	1.225309	0.644187	0.223882	0.000000	2.286081	2.307400	0.0	0.000000	...	0.000000	0.267492	1.816322	2023	2	2	12
2	1.386775	0.751033	3.194173	0.228653	1.660178	0.007658	1.166687	2.590428	0.0	0.287872	...	0.000000	0.091215	0.027295	2023	2	2	12
3	0.180456	0.000000	0.178854	0.000000	1.615589	0.000000	0.000000	0.066992	0.0	0.000000	...	0.000000	0.149463	0.000000	2023	2	2	12
4	1.607404	0.000000	3.902705	0.000000	1.274675	0.303377	0.000000	1.444000	0.0	0.000000	...	0.000000	2.938054	0.057986	2023	2	2	12
...
12235	0.747594	0.000000	0.264183	0.000000	0.190067	0.000000	0.131433	2.461908	0.0	0.000000	...	0.000000	1.404359	0.330642	2023	2	10	8
12236	0.032757	0.000000	3.284815	1.041903	0.000000	0.716097	0.078727	4.877546	0.0	0.000000	...	1.912802	0.341222	0.000000	2023	2	10	8
12237	0.000000	0.051699	4.570334	0.668476	0.006155	0.820827	0.470621	6.209879	0.0	0.034356	...	1.485656	0.086885	0.000000	2023	2	10	8
12238	0.352296	0.000000	0.406346	0.000000	0.174094	0.000000	0.000000	0.445122	0.0	0.000000	...	0.265741	0.095056	0.732410	2023	2	10	8
12239	0.314347	0.128721	5.518834	0.379400	0.842071	0.000000	0.000000	0.595668	0.0	0.000000	...	0.030936	1.958310	0.000000	2023	2	10	8

VGG16 + TIME SERIES + HSV COLOR CHANNEL

Model ini mirip dengan model *VGG16 with time series*, namun terdapat tambahan fitur rerata nilai-nilai *hue*, *saturation*, dan *value* yang berasal dari konversi *color channel* RGB ke HSV. Hal ini dilakukan agar **model dapat mempelajari pola, tren, atau karakteristik warna dari gambar dengan konten-konten udara tertentu dan kondisi langit tertentu dan termasuk apakah berkaitan secara temporal**. Menurut Acharya & Ray (2005), nilai H, S, dan V dapat dikonversi dari R, G, dan B dengan rumus seperti persamaan (5) – (12) [2].



Ruang warna RGB (kiri) dan HSV (kanan)

- HSV (*Hue Saturation Value*) adalah ruang warna yang cocok untuk mengidentifikasi warna-warna dasar dan salah satu keunggulannya adalah **toleran terhadap perubahan intensitas cahaya** [8].
- Ukuran kemiripan warna dengan salah satu warna yang dapat dilihat disebut hue,
- Tingkat kemurnian suatu warna disebut saturation.
- Ukuran intensitas dari cahaya disebut value.

Konversi HSV dari RGB:

$$H = \tan \left[\frac{3(G-B)}{(R-G) + (R-B)} \right] \quad (5)$$

$$S = 1 - \left[\frac{\min(R,G,B)}{V} \right] \quad (6)$$

$$V = \frac{R+G+B}{3} \quad (7)$$

Normalisasi R, G, B:

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B} \quad (8)$$

Konversi HSV dari RGB normalisasi (rgb)

$$V = \max(r, g, b) \quad (9)$$

$$S = \begin{cases} 0, & \text{jika } V = 0 \\ 1 - \frac{\min(r,g,b)}{V}, & \text{jika } V > 0 \end{cases} \quad (10)$$

$$H = \begin{cases} 0, & \text{jika } S = 0 \\ \frac{60 \cdot (g-b)}{S \cdot V}, & \text{jika } V = r \\ 60 \cdot \left[2 + \frac{(b-r)}{S \cdot V} \right], & \text{jika } V = g \\ 60 \cdot \left[4 + \frac{(r-g)}{S \cdot V} \right], & \text{jika } V = b \end{cases} \quad (11)$$

$$H = H + 360 \quad \text{jika } H < 0 \quad (12)$$

VGG16 + TIME SERIES + HSV COLOR CHANNEL

	0	1	2	3	4	5	6	7	8	9	...	509	510	511	Year	Month	Day	Hour	0	1	2
0	1.715331	0.000000	1.868870	0.023432	1.181104	0.000000	8.548042	1.019641	0.0	0.000000	...	0.000000	1.232765	0.283906	2023	2	2	12	52.463548	47.482422	126.877870
1	0.000000	1.208342	1.225309	0.644187	0.223882	0.000000	2.286081	2.307400	0.0	0.000000	...	0.000000	0.267492	1.816322	2023	2	2	12	60.740653	35.299585	133.714704
2	1.386775	0.751033	3.194173	0.228653	1.660178	0.007658	1.166687	2.590428	0.0	0.287872	...	0.000000	0.091215	0.027295	2023	2	2	12	92.904875	38.562540	114.578703
3	0.180456	0.000000	0.178854	0.000000	1.615589	0.000000	0.000000	0.066992	0.0	0.000000	...	0.000000	0.149463	0.000000	2023	2	2	12	70.483359	48.696747	141.293188
4	1.607404	0.000000	3.902705	0.000000	1.274675	0.303377	0.000000	1.444000	0.0	0.000000	...	0.000000	2.938054	0.057986	2023	2	2	12	82.662847	44.858658	127.722816
...
12235	0.747594	0.000000	0.264183	0.000000	0.190067	0.000000	0.131433	2.461908	0.0	0.000000	...	0.000000	1.404359	0.330642	2023	2	10	8	32.504843	100.424864	138.865035
12236	0.032757	0.000000	3.284815	1.041903	0.000000	0.716097	0.078727	4.877546	0.0	0.000000	...	1.912802	0.341222	0.000000	2023	2	10	8	45.029895	61.471660	155.266960
12237	0.000000	0.051699	4.570334	0.668476	0.006155	0.820827	0.470621	6.209879	0.0	0.034356	...	1.485656	0.086885	0.000000	2023	2	10	8	45.288744	58.704998	159.301140
12238	0.352296	0.000000	0.406346	0.000000	0.174094	0.000000	0.000000	0.445122	0.0	0.000000	...	0.265741	0.095056	0.732410	2023	2	10	8	44.059670	78.627531	124.348414
12239	0.314347	0.128721	5.518834	0.379400	0.842071	0.000000	0.000000	0.595668	0.0	0.000000	...	0.030936	1.958310	0.000000	2023	2	10	8	66.636858	49.236567	145.248107

HASIL DAN PEMBAHASAN

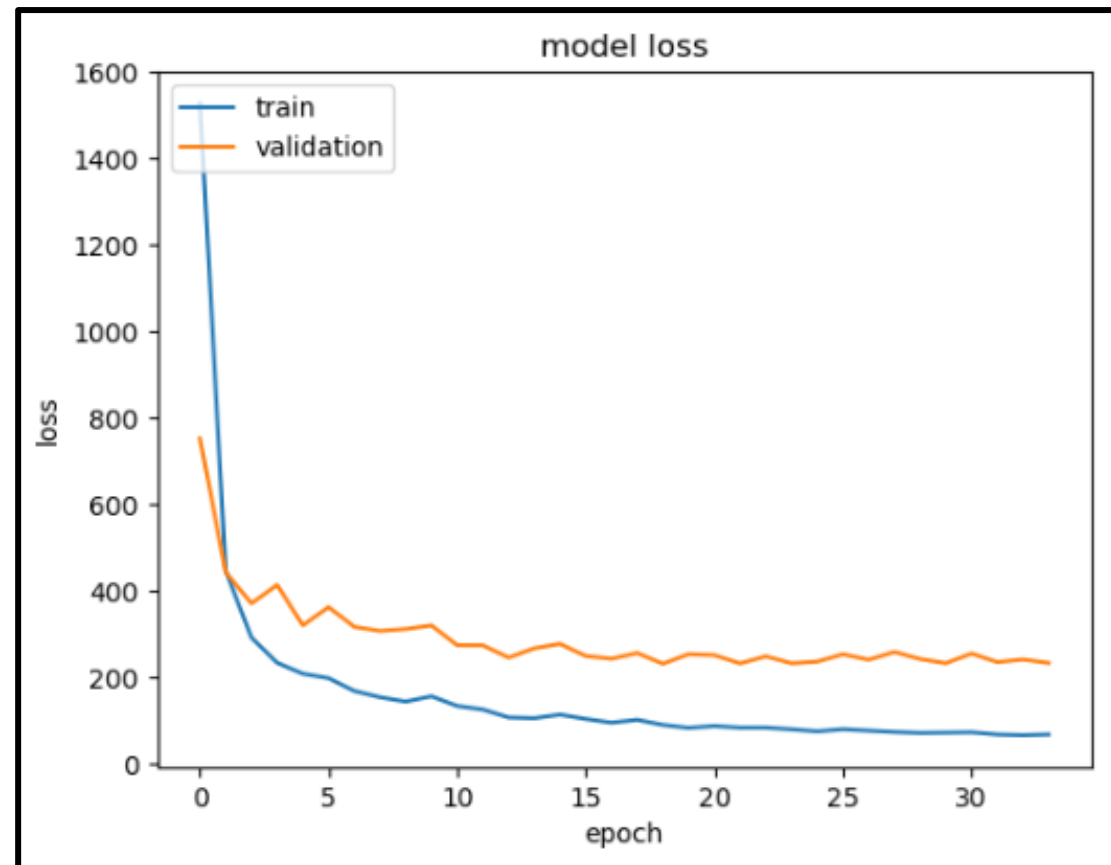
HASIL UJI PERFORMA MODEL

Tabel 4. Perbandingan RMSE model dengan *features* berbeda

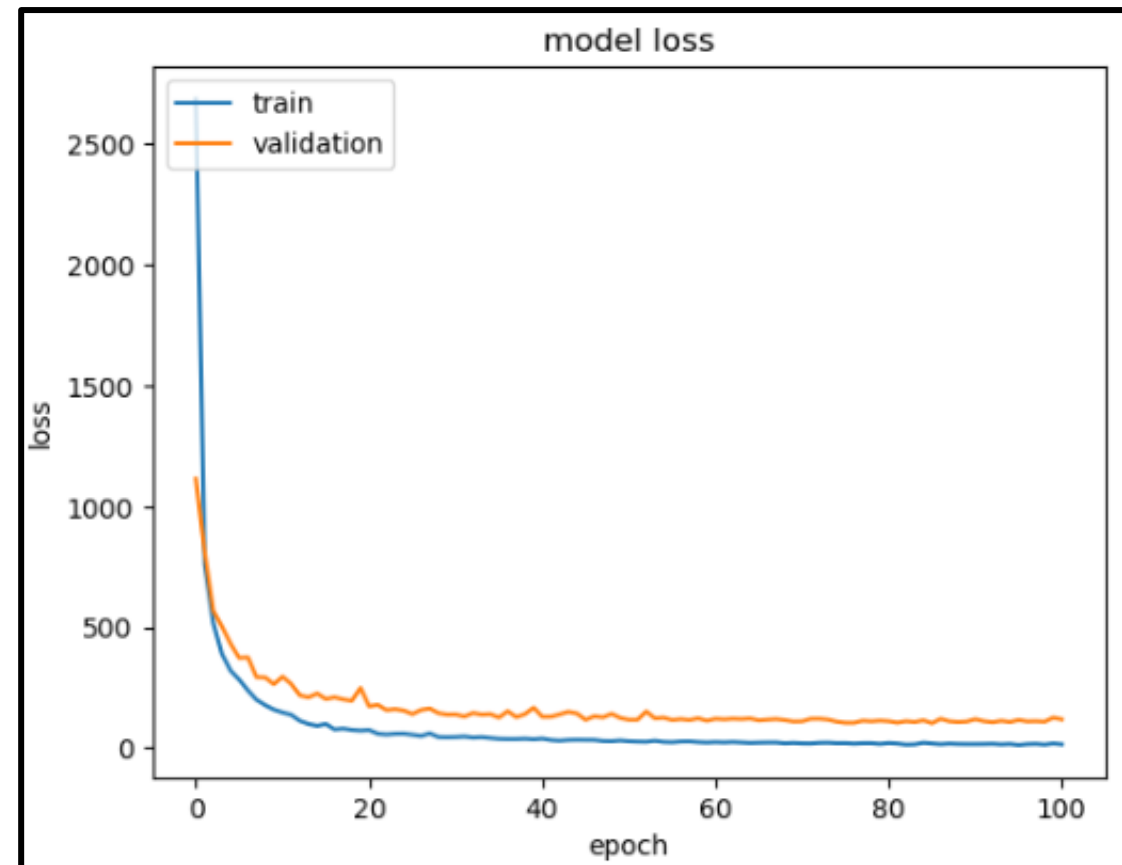
Model	Features	RMSE
VGG16	Images	15,282184595653366
VGG16 + Dense Layer	Images, Timeseries	11,204401132426852
VGG16 + Dense Layer	Images, Timeseries, HSV	10,451656768734122

- ❑ Penambahan *feature timeseries* berdasarkan waktu pengambilan gambar selain *feature* hasil *feature extraction* dari *layer convolution* VGG16 membuat model mampu mempelajari tren dan seasonalitas kondisi langit lokal setempat dibanding tanpa *timeseries*. Pengurangan RMSE sebesar 4.
- ❑ Penambahan *feature timeseries* dan HSV (*Hue, Saturation, Value*) membuat RMSE berkurang sampai dengan ± 5 .
- ❑ Pambahan fitur nilai rerata *HSV color channel* membuat model mampu mempelajari tren dan karakteristik warna gambar input[8] sesuai dengan konten udaranya dibanding tanpa HSV.

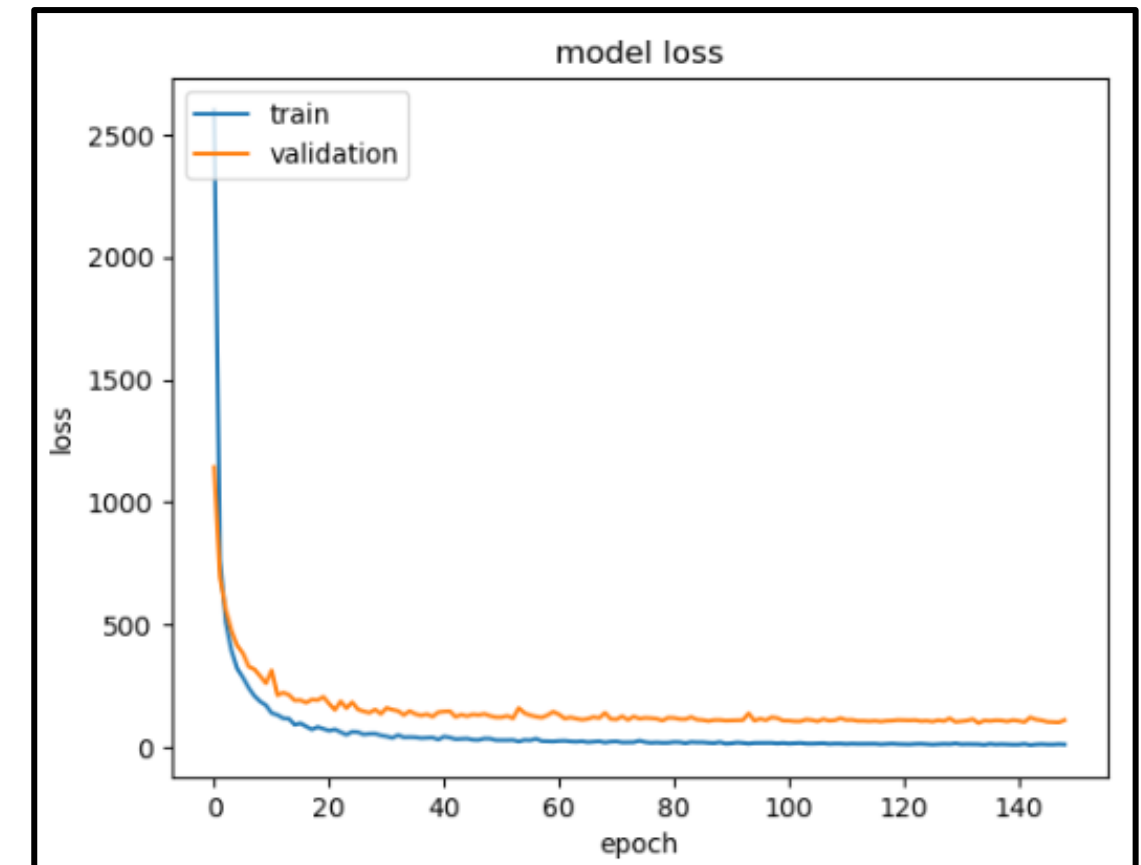
KURVA LOSS PERFORMA MODEL



Kurva Loss untuk VGG16

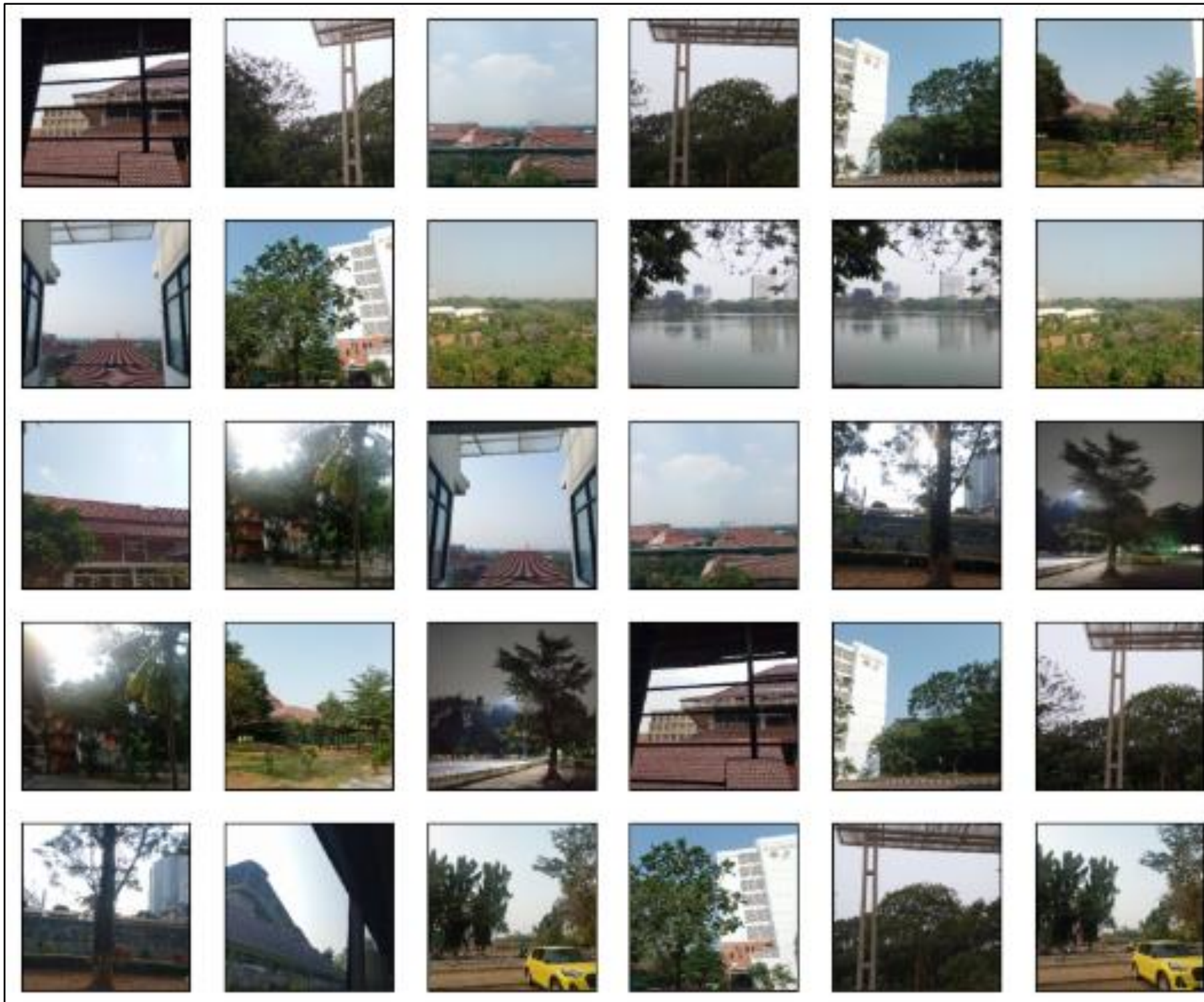


Kurva Loss untuk VGG16 + Timeseries



Kurva Loss untuk VGG16 + Timeseries + HSV

ESTIMASI KONTEN UDARA TEMPAT LOKAL



Model terbaik pada pengujian
feature **VGG16 + Timeseries + HSV** :

Diuji menggunakan gambar dengan *background* langit dari beberapa tempat di sekitar FMIPA UI.

Tanggal pengambilan gambar :

Diambil pada rentang waktu mulai tanggal 1-7 Oktober 2023, total 32 gambar.

Sampling Frequency :

Tidak sekuensial per hari per sekian, dan paling banyak diambil pada pukul 11.00-14.00 WIB dan 17.00-18.00 WIB.

Gambar langit sekitar FMIPA UI Depok yang diambil pada 1-7 Oktober 2023

HASIL ESTIMASI KONTEN UDARA FMIPA UI & SEKITAR

Statistika deskriptif hasil seluruh gambar yang diestimasi

	AQI	PM2.5	PM10	O3	CO	SO2	NO2
count	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000	32.000000
mean	168.635681	130.077866	109.860847	52.553383	60.497662	21.545452	45.366798
std	45.054710	47.835903	28.651346	25.298229	36.988552	9.759649	22.185972
min	87.433411	41.462658	70.663513	17.199795	1.312261	5.777575	5.323960
25%	128.732796	89.555529	83.740292	34.886041	31.464930	13.180787	31.342475
50%	167.534698	127.431328	106.190434	49.125887	64.544622	21.245168	43.681759
75%	201.422726	161.130219	132.258518	63.932923	84.812355	27.689995	59.380404
max	269.771027	209.305801	176.896454	112.488983	129.433243	42.178459	97.948029

- Model (*VGG16 with Timeseries and HSV color channel*) dapat dengan baik memprediksi 7 target konten udara sekaligus bersamaan dengan *AQI Class* yang bersesuaian. AQI tertinggi mencapai 269 dan AQI rata-rata sekitar 168 dan standar deviasi 45, kualitas udara di FMIPA UI masuk kategori “Very Unhealthy”.
- Hasil ini cukup representatif, validitas diperkuat dengan waktu pengambilan gambar pada pukul 11.00-14.00 WIB yang mana kantin Dallas FMIPA UI padat dan aktivitas memasak sangat tinggi sehingga terus-menerus mengemisikan gas-gas polutan, serta pada 17.00-18.00 WIB di mana mobilisasi warga kampus dengan kendaraan bermotor sangat tinggi dan hal ini turut mempengaruhi kualitas udara.

KESIMPULAN

Best model performance :
VGG16 + Timeseries + HSV

Model dengan *feature* temporal *Timeseries* serta karakteristik warna dari *feature* rerata HSV memungkinkan model mempelajari tren seasonal serta karakteristik warna dari gambar lokal langit setempat sesuai konten-konten udaranya, dengan RMSE 10,45.

Kualitas Udara FMIPA UI :
“Very Unhealthy”
dan AQI Rerata 1 68

Pada rentang 1-7 Oktober 2023 kualitas udara di sekitar FMIPA UI ada pada kategori “Very Unhealthy” dengan rerata AQI sekitar 168 dan standar deviasi mencapai 45. Hasil ini valid dan representatif dari aktivitas keseharian orang-orang di lingkungan FMIPA UI.

DAFTAR PUSTAKA

- [1] A. Chakma, B. Vizona, T. Cao, J. Lin and J. Zhang. 2017. *Image-based air quality analysis using deep convolutional neural network*. IEEE International Conference on Image Processing (ICIP), Beijing, China, 2017, pp. 3949-3952, doi: 10.1109/ICIP.2017.8297023.
- [2] Acharaya, T., & Ray, A. K. 2005. *Image Processing Principles and Application*. New Jersey: A John Wiley & Sons.
- [3] Adarsh Rouniyar, Sapdo utomo, John A, & Pao-Ann Hsiung. 2023. *Air Pollution Image Dataset from India and Nepal [Dataset]*. Kaggle. <https://doi.org/10.34740/KAGGLE/DS/3152196>. <https://www.kaggle.com/datasets/adarshrouniyar/air-pollution-image-dataset-from-india-and-nepal>
- [4] BBC News Indonesia. 2023. *Polusi udara: Indonesia masuk enam negara paling berkontribusi terhadap polusi udara global, warga akan gugat pemerintah dan industri* <https://www.bbc.com/indonesia/articles/c72enp76622o>. 2 Desember 2023 (11.25 WIB).
- [5] Brownlee, Jason. 2020. *A Gentle Introduction to Rectified Linear Unit (ReLU)*. 8 Desember 2023 (22.00 WIB).
- [6] Clear Air Fund. 2023. *Sustainable Development Goals and Air Pollution*. Info@cleanairfund.org. 2 Desember 2023 (10.20 WIB).
- [7] E. Tanuwijaya dan A. Roseanne. 2021. *Modifikasi Arsitektur VGG16 untuk Klasifikasi Citra Digital Rempah-Rempah Indonesia*. Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer. Vol. 21, No. 1, November 2021, pp. 189-196, doi: 10.30812/matrik.v21i1.1492.
- [8] I. F. Ashari, E. Rachmawati, dan F. Sthevanie. 2020. *Pengenalan Jenis Buah pada Citra Menggunakan HOG dan Histogram HSV*. Jurnal Tugas Akhir Fakultas Informatika, e-Proceeding of Engineering : Vol.8, No.1 Februari 2021, pp. 793 – 808.
- [9] iqair.com. 2023. *Indeks Kualitas Udara (AQI) Jakarta dan Polusi Udara di Indonesia*. <https://www.iqair.com/id/indonesia/jakarta>. 4 Oktober 2023 (13.40 WIB) dan 2 Desember 2023 (10.30 WIB).
- [10] J. S. Saltz. 2021. *CRISP-DM for Data Science: Strengths, Weaknesses and Potential Next Steps*. IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 2337-2344, doi: 10.1109/BigData52589.2021.9671634.
- [11] M. P. Rosalina, S. P. Wisanggeni, dan A. Krisna. 2023. *Alat Pemantauan Terbatas, Data yang Diperoleh Tidak Ideal*. <https://www.kompas.id/baca/investigasi/2023/09/21/buta-tanpa-data-kualitas-udara>. 3 Desember 2023 (Pukul 15.00 WIB).
- [12] Staf Penulis IQAir. 2016. *What is the difference between US AQI and WHO air quality guidelines?* <https://www.iqair.com/us/newsroom/what-is-the-difference-between-the-us-aqi-and-who-air-quality-guidelines>. 4 Oktober 2023 (15.30 WIB).
- [13] Tianfeng, C. & Draxler, R. R., 2014. *Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments Against Avoiding RMSE in The Literature*. Geoscientific Model Development, pp. 1247-1250.
- [14] Wijaya. (2001). *Analisis Statistika dengan Program SPSS 10.0*. Bandung: Alfabeta.



TERIMA KASIH

24-05-015-0

- Hendri Erka Setya
 - Shawn Michael Dayanti Intong
- 