

# LAMPIRAN

Lampiran 1. Dataset yang diambil dari portal satudata.jakarta.go.id

Organization



Provinsi DKI Jakarta

Informasi Dataset

Dataset Dipublikasi

21-06-2024

INFORMASI: Data berikut ini masih dalam proses pemenuhan Prinsip SDI.

Data Indeks Standar Pencemar Udara (ISPU) di Provinsi DKI Jakarta 2023

adalah Data Indeks Standar Pencemar Udara (ISPU) yang diukur melalui Stasiun Pemantau Kualitas Udara (SPKU) di Provinsi DKI Jakarta

Data and Resources

Data Indeks Standar Pencemar Udara (ISPU) di Provinsi DKI Jakarta 2023

Download

ISPU

SDI

kualitasudara

udara

Metadata

Lampiran 2. Jumlah data dan variabel yang ada pada dataset

|      | periode_data | tanggal    | stasiun                        | PM10 | PM2.5 | so2 | co  | o3  | no2 | max | critical | category |
|------|--------------|------------|--------------------------------|------|-------|-----|-----|-----|-----|-----|----------|----------|
| 0    | 202302       | 2/25/2023  | DKI5 Kebon Jeruk Jakarta Barat | 35   | -     | 13  | 12  | 31  | 18  | 35  | PM10     | BAIK     |
| 1    | 202302       | 2/26/2023  | DKI5 Kebon Jeruk Jakarta Barat | 23   | -     | 14  | 9   | 32  | 11  | 32  | O3       | BAIK     |
| 2    | 202302       | 2/27/2023  | DKI5 Kebon Jeruk Jakarta Barat | 20   | -     | 13  | 8   | 33  | 13  | 33  | O3       | BAIK     |
| 3    | 202302       | 2/28/2023  | DKI5 Kebon Jeruk Jakarta Barat | 30   | -     | 21  | 11  | 28  | 18  | 30  | PM10     | BAIK     |
| 4    | 202303       | 3/1/2023   | DKI1 Bunderan HI               | 38   | 44    | 50  | 8   | 19  | 27  | 50  | 3        | BAIK     |
| ...  | ...          | ...        | ...                            | ...  | ...   | ... | ... | ... | ... | ... | ...      | ...      |
| 1820 | 202311       | 11/26/2023 | DKI5 Kebon Jeruk               | 48   | 71    | 33  | 21  | 44  | 20  | 71  | PM25     | SEDANG   |
| 1821 | 202311       | 11/27/2023 | DKI5 Kebon Jeruk               | 51   | 76    | 32  | 18  | 53  | 19  | 76  | PM25     | SEDANG   |
| 1822 | 202311       | 11/28/2023 | DKI5 Kebon Jeruk               | 56   | 88    | 33  | 20  | 48  | 21  | 88  | PM25     | SEDANG   |
| 1823 | 202311       | 11/29/2023 | DKI5 Kebon Jeruk               | 56   | 88    | 32  | 18  | 56  | 18  | 88  | PM25     | SEDANG   |
| 1824 | 202311       | 11/30/2023 | DKI5 Kebon Jeruk               | 30   | 57    | 32  | 15  | 61  | 17  | 61  | O3       | SEDANG   |

1825 rows x 12 columns

### Lampiran 3. Melakukan pengujian di Google Colab

1. Melakukan import kode python untuk digunakan pada tahap selanjutnya

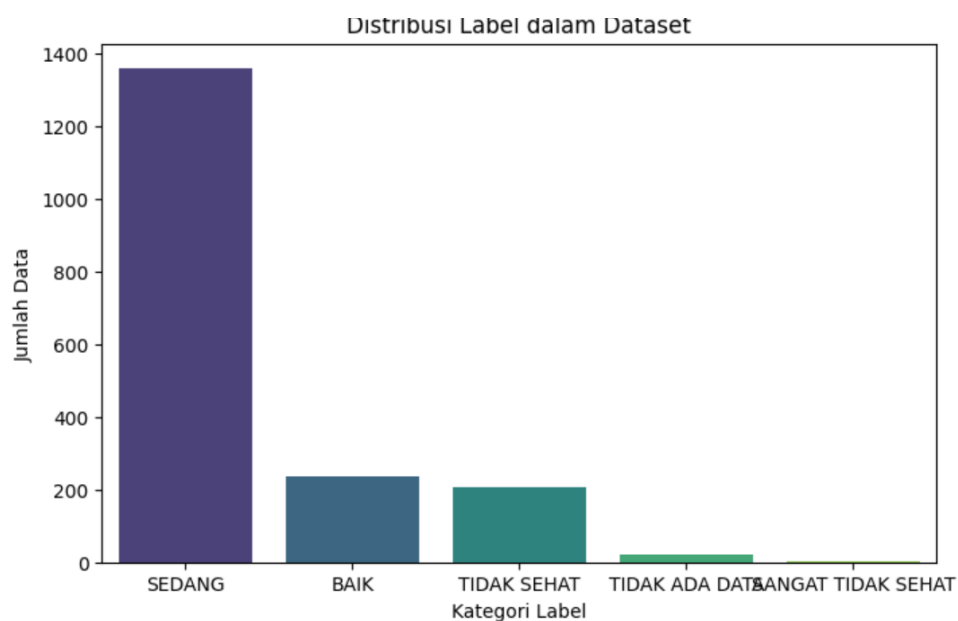
```
[ ] import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
```

2. Melakukan distribusi label

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

|
label_counts = df['category'].value_counts()

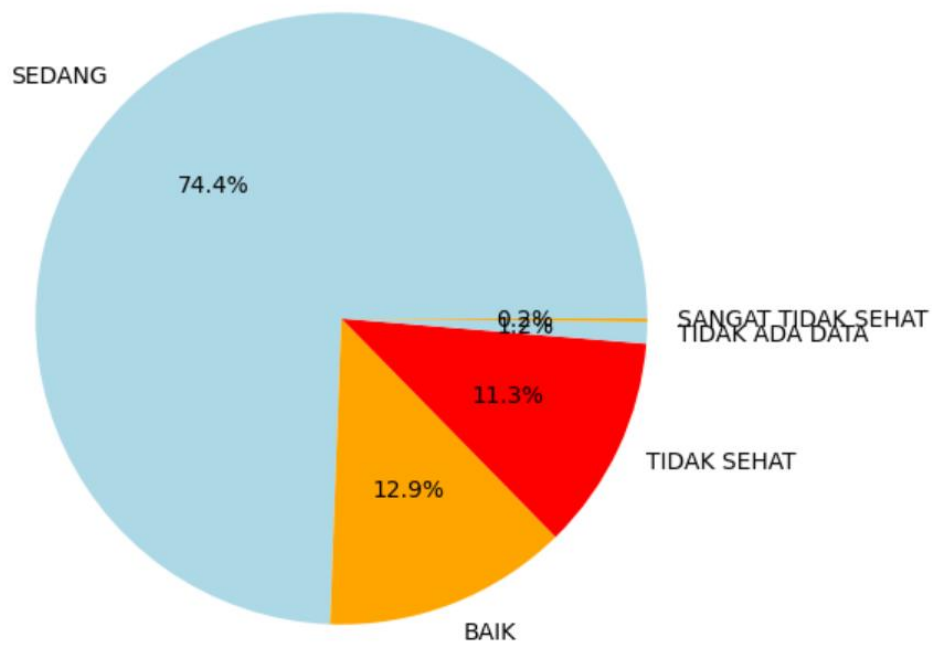
# Plot distribusi label
plt.figure(figsize=(8, 5))
sns.barplot(x=label_counts.index, y=label_counts.values, palette="viridis")
plt.xlabel("Kategori Label")
plt.ylabel("Jumlah Data")
plt.title("Distribusi Label dalam Dataset")
plt.show()
```



### 3. Persentase label dalam dataset

```
df['category'].value_counts().plot.pie(autopct='%1.1f%%', figsize=(6,6), colors=['lightblue', 'orange', 'red'])  
plt.title("Persentase Label dalam Dataset")  
plt.ylabel("")  
plt.show()
```

Persentase Label dalam Dataset



4. Menampilkan perbandingan nilai akurasi pelatihan dan akurasi validasi untuk melihat apakah terjadinya overfitting.

```
import matplotlib.pyplot as plt

# Contoh data akurasi (Gantilah dengan data asli dari model kamu)
epochs = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] # Jumlah epoch
train_acc = [0.75, 0.80, 0.85, 0.87, 0.90, 0.92, 0.93, 0.94, 0.95, 0.96] # Training Accuracy
val_acc = [0.72, 0.78, 0.82, 0.85, 0.88, 0.90, 0.91, 0.92, 0.93, 0.94] # Validation Accuracy

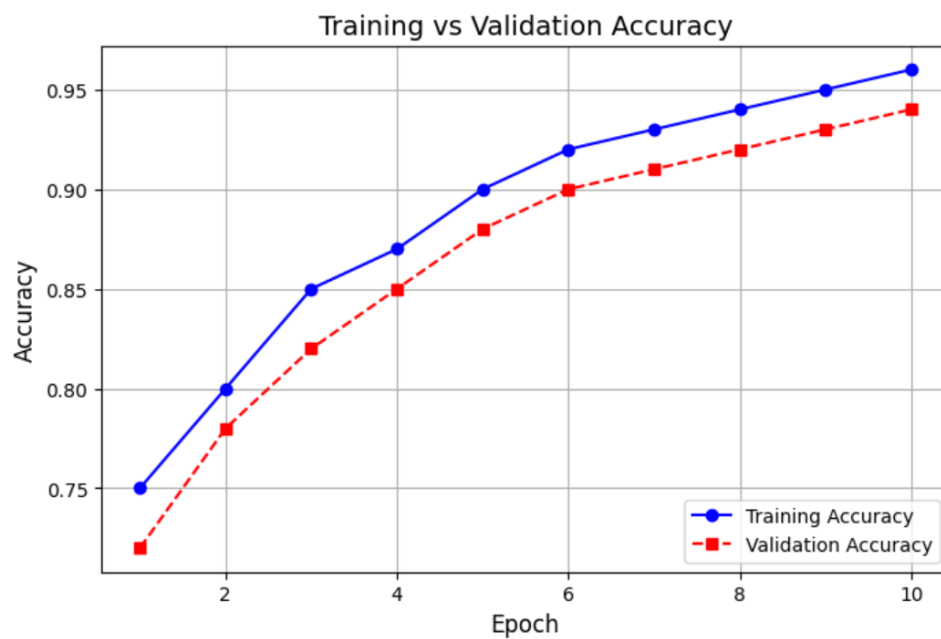
# Membuat plot
plt.figure(figsize=(8,5))
plt.plot(epochs, train_acc, label='Training Accuracy', marker='o', linestyle='-', color='blue')
plt.plot(epochs, val_acc, label='Validation Accuracy', marker='s', linestyle='--', color='red')

# Menambahkan judul dan label sumbu
plt.title("Training vs Validation Accuracy", fontsize=14)
plt.xlabel("Epoch", fontsize=12)
plt.ylabel("Accuracy", fontsize=12)

# Menampilkan legend
plt.legend(loc='lower right')

# Menampilkan grid
plt.grid(True)

# Menampilkan grafik
plt.show()
```



5. Menampilkan grafik penurunan nilai loss (kerugian) selama proses pelatihan model machine learning dengan 50 epoch.

```
import numpy as np
import matplotlib.pyplot as plt

# Simulasi data loss (gantilah dengan data asli dari model)
epochs = np.arange(1, 51) # 50 epoch
train_loss = np.linspace(0.9, 0.2, 50) # Contoh penurunan training loss
val_loss = np.linspace(1.0, 0.3, 50) # Contoh penurunan validation loss

plt.figure(figsize=(10, 6))
plt.plot(epochs, train_loss, label='Training Loss', marker='o', linestyle='--')
plt.plot(epochs, val_loss, label='Validation Loss', marker='s', linestyle='--')

plt.xlabel('Epochs')
plt.ylabel('Loss')
plt.title('Training vs Validation Loss (50 Epochs)')
plt.legend()
plt.grid()
plt.show()
```

