



Analisis Tingkat Polusi Udara DKI Jakarta

Mata Kuliah Pembelajaran Mesin

Anggota Kelompok



23/NPA/528365/19882

Gata Anindhita Zalianingrum



23/NPA/528380/19897

Laila Rahmatin Nisa



SUSTAINABLE CITIES & COMMUNITIES

Indeks Kualitas Udara

Target mengurangi dampak lingkungan perkotaan per kapita yang merugikan, termasuk dengan memberi perhatian khusus pada kualitas udara, karena aktivitas industri, transportasi, dan kegiatan perkotaan lainnya dapat mempengaruhi kualitas udara.

Mata Kuliah Pembelajaran Mesin



GOAL 11 SUSTAINABLE CITIES AND COMMUNITIES

Make cities and human settlements inclusive, safe, resilient and sustainable.



SDGs FACTSHEET



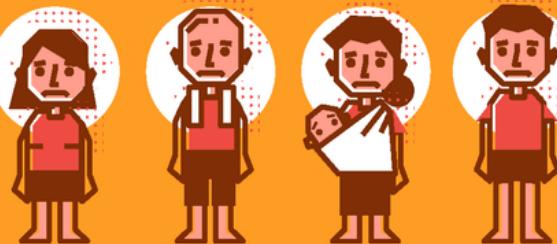
INDONESIA

Half of humanity - 3.5 billion people - lives in cities today. By 2030, almost 60% of the world's population will live in urban areas.



In Indonesia, 134 million people lived in cities in 2014. This represents 53% of Indonesia's population. Indonesia is one of the most urbanized countries in Asia.

However,
828 million people still live in slums today and the number keeps rising.



In Indonesia, the number of people living in slums declined by 1.2% between 2009 and 2014.



The world's cities occupy just 2% of the Earth's land, but account for 60 – 80% of energy consumption and 75% of carbon emissions.

In 2014, 1,559 disasters took place in Indonesia. 490 people died and 2 million people were affected.

Out of 415 Indonesian provinces, 322 have a high disaster risk.

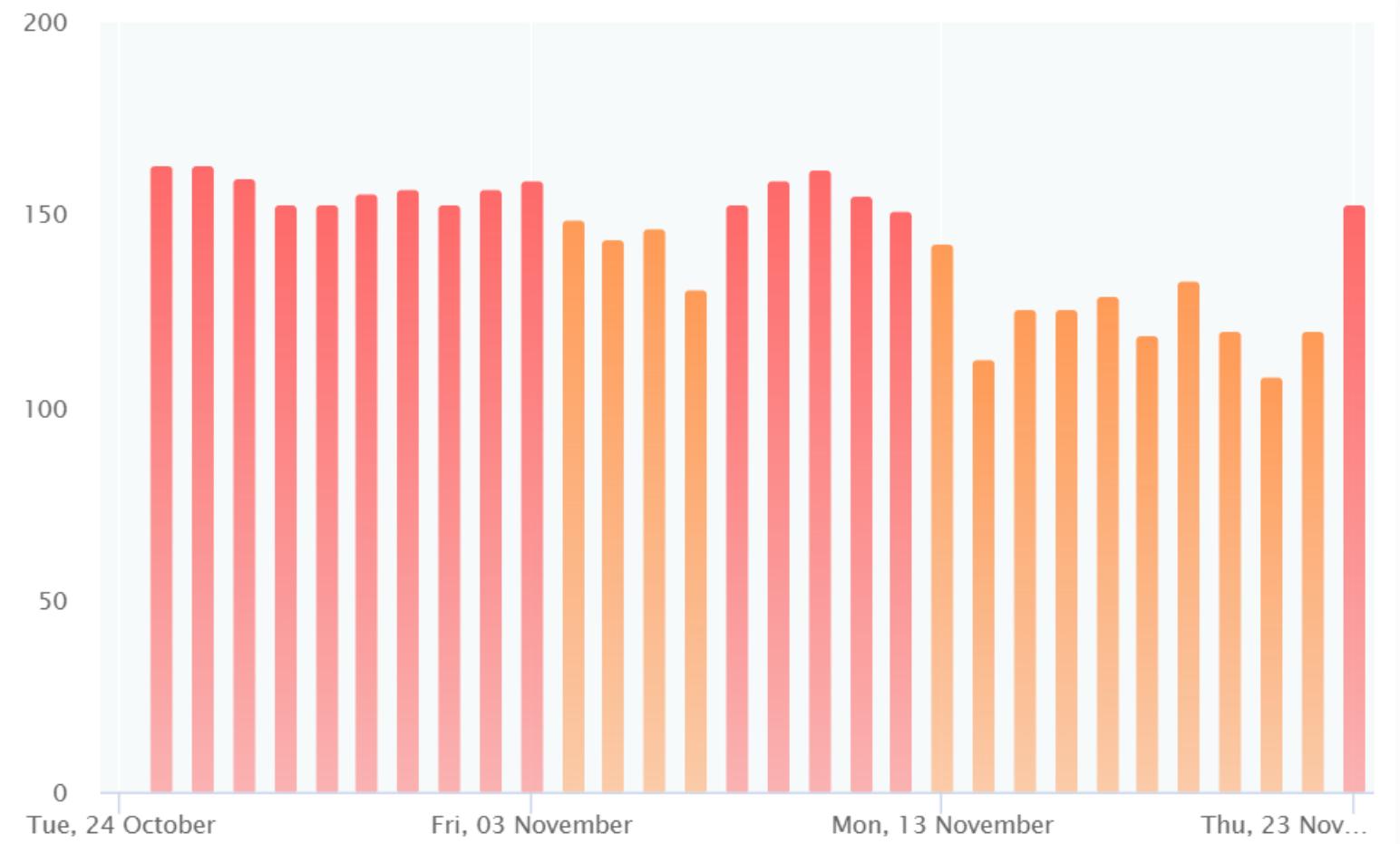


Perkotaan hanya memiliki wilayah 2% pada daratan bumi, namun menyumbang 60-80% konsumsi energi dan 75% emisi karbon

Grafik riwayat kualitas udara untuk Jakarta

PER JAM

PER HARI



Hari	Tingkat polusi	Cuaca	Suhu	Angin
Selasa, Nov 21	Tidak sehat bagi kelompok sensitif 108 AQI US	☁️	33° 26°	▲ 14.4
Rabu, Nov 22	Tidak sehat bagi kelompok sensitif 120 AQI US	☁️	33° 27°	▲ 14.4
Kamis, Nov 23	Tidak sehat 154 AQI US	☁️	33° 27°	▲ 14.4
Hari ini	Tidak sehat 166 AQI US	☁️ 100%	33° 27°	▲ 14.4

Indeks kualitas udara (AQI) dan polusi udara PM2.5 di Jakarta

AQI US

159

INDEKS AQI LANGSUNG

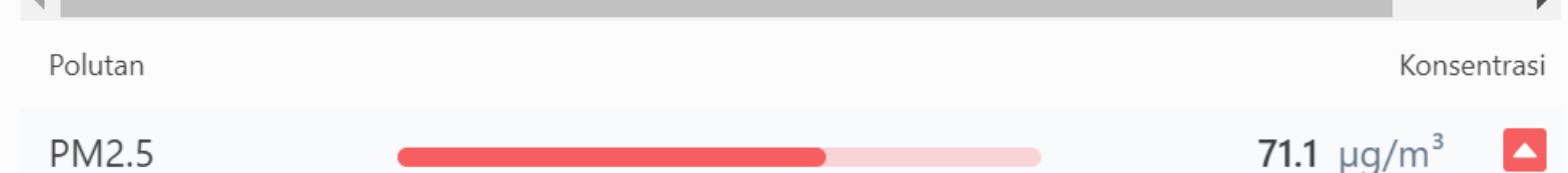
Tidak Sehat



IKHTISAR

Apa kualitas udara saat ini di Jakarta?

Tingkat polusi udara	Indeks kualitas udara	Polutan utama
Tidak sehat	159 AQI US	PM2.5



Konsentrasi PM2.5 di Jakarta saat ini 14.2 kali [nilai panduan kualitas udara tahunan WHO](#)

sumber : <https://www.iqair.com/id/indonesia/jakarta>

Introduction

Udara bersih dan rendah polusi dapat mengurangi risiko penyakit pernapasan dan kesehatan yang terkait dengan polusi udara. Kesehatan masyarakat yang lebih baik dapat mendukung produktivitas tenaga kerja dan menciptakan kondisi yang lebih baik untuk tinggal dan bekerja.

Yang dimana mewujudkan terget dari Sustainable Development Goals (SDGs) dengan meningkatkan kualitas hidup penduduk perkotaan dengan menciptakan lingkungan yang lebih nyaman dan aman.



Problems



Pencemaran udara jadi isu tak berkesudahan, terutama di kota besar seperti Jakarta dengan peningkatan aktivitas industri. Kementerian Lingkungan Hidup dan Kehutanan (KLHK), menyebutkan, emisi kendaraan bermotor menyumbang 70% polutan di wilayah perkotaan, antara lain, nitrogen oksida (NO_x), karbon monoksida (CO), sulfur dioksida (SO_2), dan partikulat (PM) yang dapat mengakibatkan kualitas udara yang buruk, berdampak negatif pada kesehatan masyarakat wilayah perkotaan.

A 3D cartoon character of a director with a brown beret, red-rimmed glasses, and a black beard, wearing a red jacket over a white shirt. He is holding a clapperboard with the words 'DIRECTOR', 'CAMERA', 'SCENE', 'DATE', 'ON', and 'TAKE' on it.

Motivation

Mengambil inspirasi dari permasalahan tersebut dengan menggunakan kumpulan data Indeks Standar Pencemar Udara (ISPU) yang kami gunakan untuk memahami apa yang dimaksud dengan udara yang aman untuk dihirup, serta menerapkan pembelajaran mesin untuk membedakan antara udara yang baik bagi pernafasan dan yang tidak baik bagi pernafasan

Mata Kuliah Pembelajaran Mesin

Method

Random Forest



Random Forest dapat memberikan hasil prediksi yang kuat dan stabil karena memanfaatkan banyak pohon keputusan. Hal ini dapat meningkatkan ketepatan klasifikasi dan meredakan potensi overfitting.

KNN



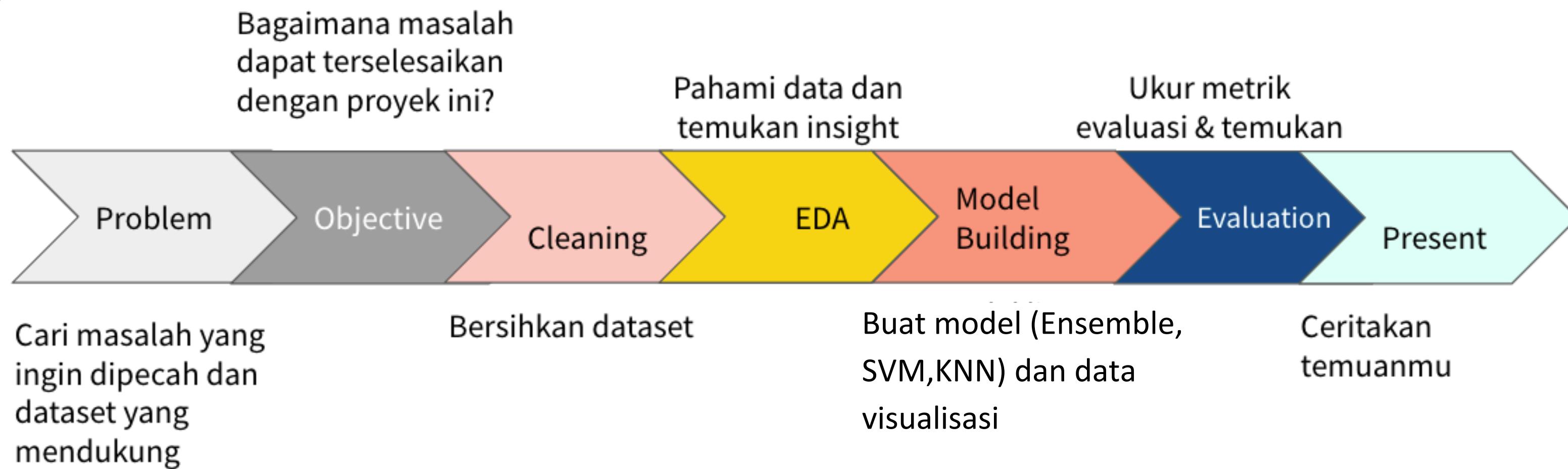
KNN dapat membantu mengidentifikasi pola lokal dalam data polusi udara di Jakarta. Faktor-faktor tertentu, seperti lokasi geografis atau pola lalu lintas, dapat berkontribusi pada polusi udara, dan KNN dapat sensitif terhadap pola ini.

SVM



Jakarta, sebagai kota besar, mungkin memiliki data polusi udara dengan banyak fitur. SVM efektif dalam menangani masalah klasifikasi dalam ruang berdimensi tinggi dan dengan kernelization dapat membantu menangkap kompleksitas dalam data.

Machine Learning



Mata Kuliah Pembelajaran Mesin

Machine Learning

01 Loading dan Preprocessing Data

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

```
# Loading Data  
data = pd.read_csv('/content/drive/MyDrive/dataISPUDKI.csv')
```

	Unnamed: 0	tanggal	stasiun	pm10	so2	co	o3	no2	max	critical	pm25	kategori
0	0	2021-01-01	DKI1 (Bunderan HI)	38.0	29.0	6.0	31.0	13.0	53.0	PM25	53.0	SEDANG
1	1	2021-01-02	DKI1 (Bunderan HI)	27.0	27.0	7.0	47.0	7.0	47.0	O3	46.0	BAIK
2	2	2021-01-03	DKI1 (Bunderan HI)	44.0	25.0	7.0	40.0	13.0	58.0	PM25	58.0	SEDANG
3	3	2021-01-04	DKI1 (Bunderan HI)	30.0	24.0	4.0	32.0	7.0	48.0	PM25	48.0	BAIK
4	4	2021-01-05	DKI1 (Bunderan HI)	38.0	24.0	6.0	31.0	9.0	53.0	PM25	53.0	SEDANG
...
1720	1720	2021-12-27	DKI5 (Kebon Jeruk)	54.0	36.0	14.0	21.0	47.0	76.0	PM25	76.0	SEDANG
1721	1721	2021-12-28	DKI5 (Kebon Jeruk)	44.0	20.0	11.0	21.0	33.0	68.0	PM25	68.0	SEDANG
1722	1722	2021-12-29	DKI5 (Kebon Jeruk)	34.0	28.0	8.0	25.0	29.0	54.0	PM25	54.0	SEDANG
1723	1723	2021-12-30	DKI5 (Kebon Jeruk)	53.0	25.0	15.0	23.0	44.0	75.0	PM25	75.0	SEDANG
1724	1724	2021-12-31	DKI5 (Kebon Jeruk)	60.0	28.0	19.0	30.0	53.0	87.0	PM25	87.0	SEDANG

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 1725 entries, 0 to 1724  
Data columns (total 11 columns):  
 #   Column      Non-Null Count  Dtype     
---    
 0   tanggal    1725 non-null   datetime64[ns]  
 1   stasiun    1725 non-null   object  
 2   pm10       1725 non-null   float64  
 3   so2        1725 non-null   float64  
 4   co          1725 non-null   float64  
 5   o3          1725 non-null   float64  
 6   no2        1725 non-null   float64  
 7   max         1725 non-null   float64  
 8   critical    1725 non-null   object  
 9   pm25       1725 non-null   float64  
 10  categori   1725 non-null   object  
dtypes: datetime64[ns](1), float64(7), object(3)  
memory usage: 148.4+ KB
```

(1725, 12) artinya pada dataset ini kita memiliki 1725 baris dan 12 kolom/feature

Machine Learning

02 Cleaning Data

Menghapus kolom 'Unnamed: 0' yang dianggap tidak penting

```
data = data.drop(columns=['Unnamed: 0'])
```

03 Exploratory Data Analysis

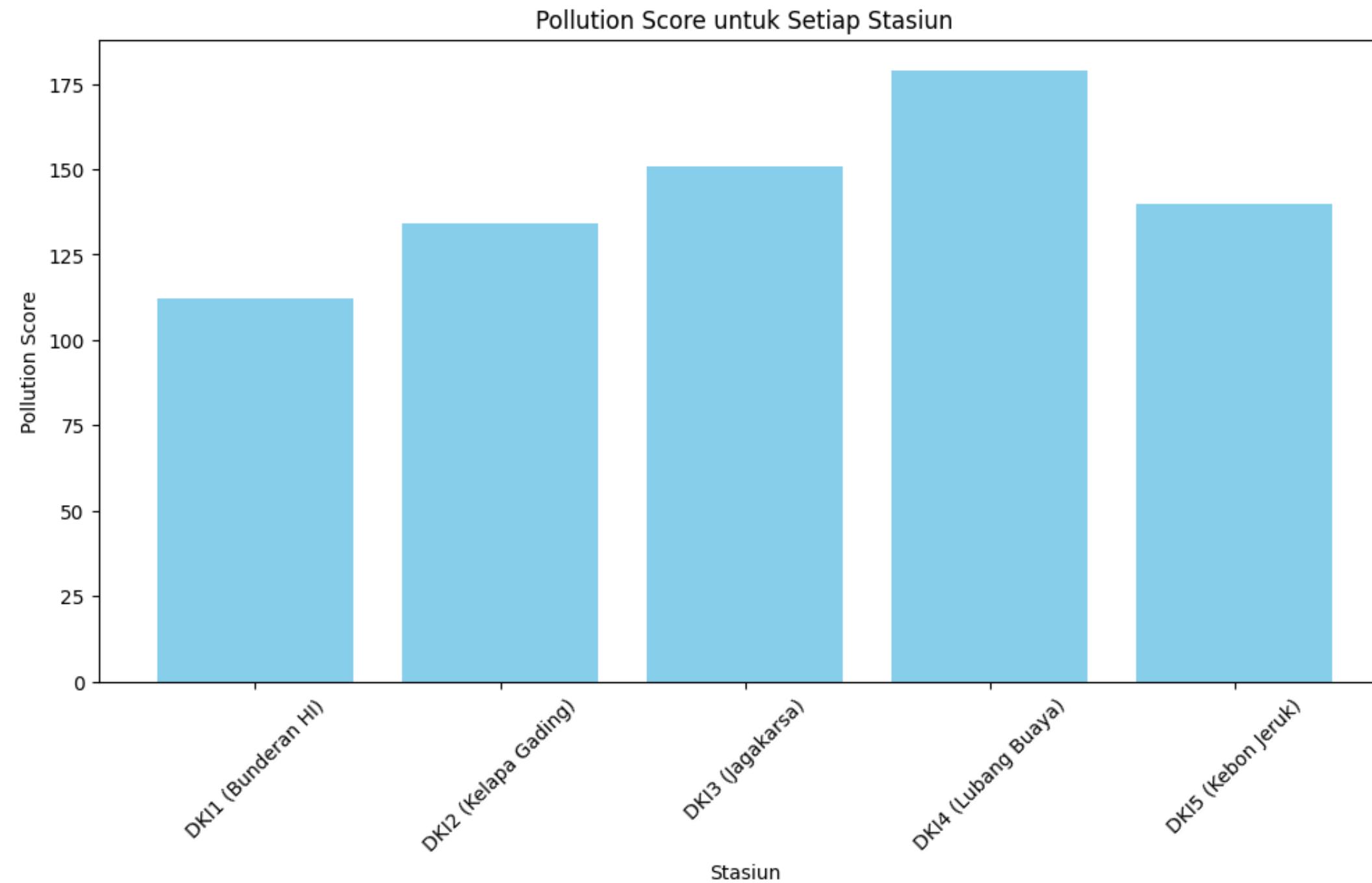
Mengelompokkan data berdasarkan stasiun dan menghitung skor polusi untuk setiap stasiun.

```
# Kelompokkan data berdasarkan stasiun
grouped_data = data.groupby('stasiun').agg({
    'pm10': 'max',
    'so2': 'max',
    'co': 'max',
    'o3': 'max',
    'no2': 'max',
    'pm25': 'max'
})

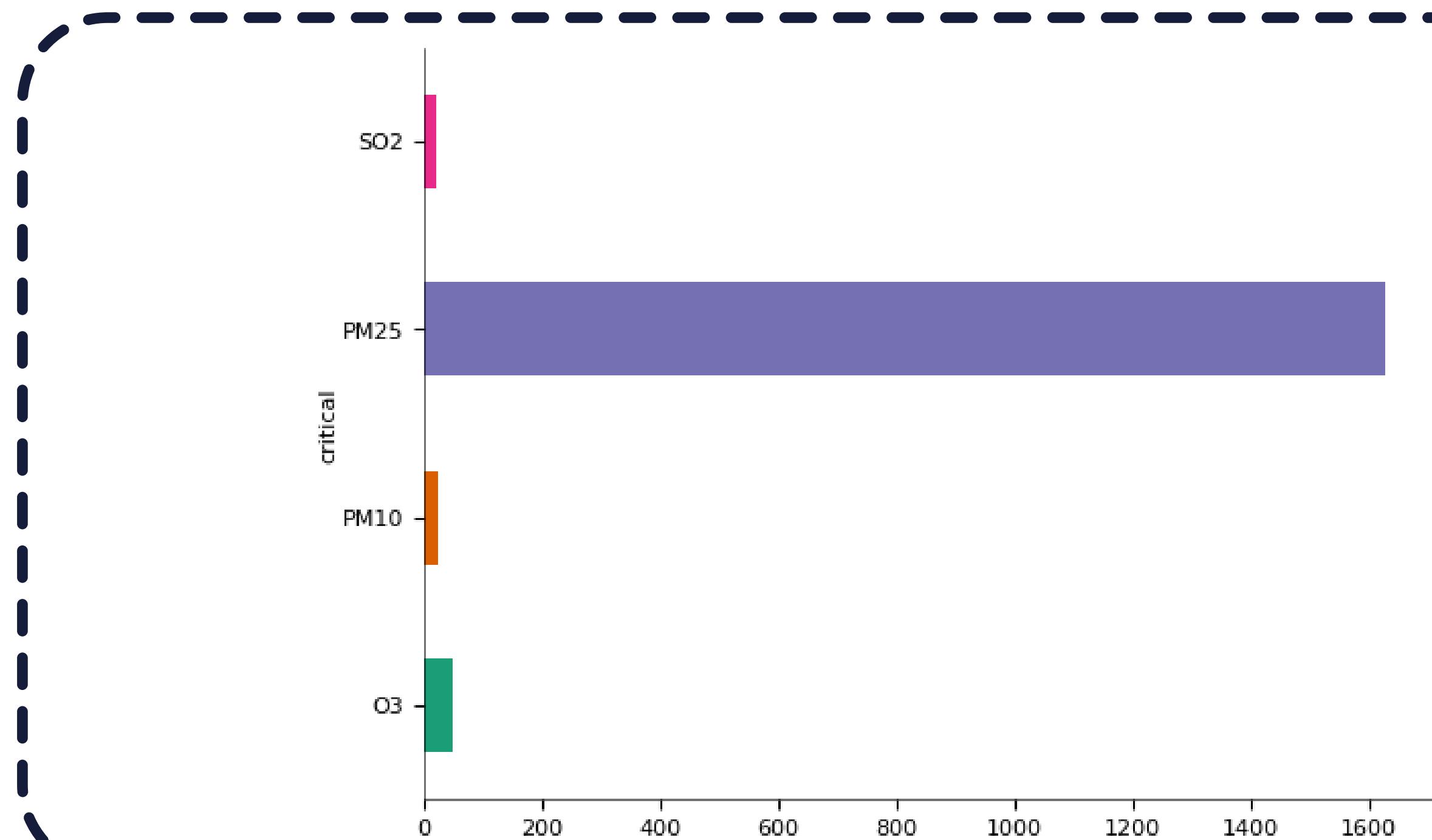
# Hitung nilai rata-rata atau maksimum dari keenam kolom
grouped_data['pollution_score'] = grouped_data[['pm10', 'so2', 'co', 'o3', 'no2', 'pm25']].max(axis=1)

# Tentukan stasiun dengan nilai tertinggi sebagai kota paling berpolusi
most_polluted_station = grouped_data['pollution_score'].idxmax()
```

Skor Kota Polusi Tertinggi

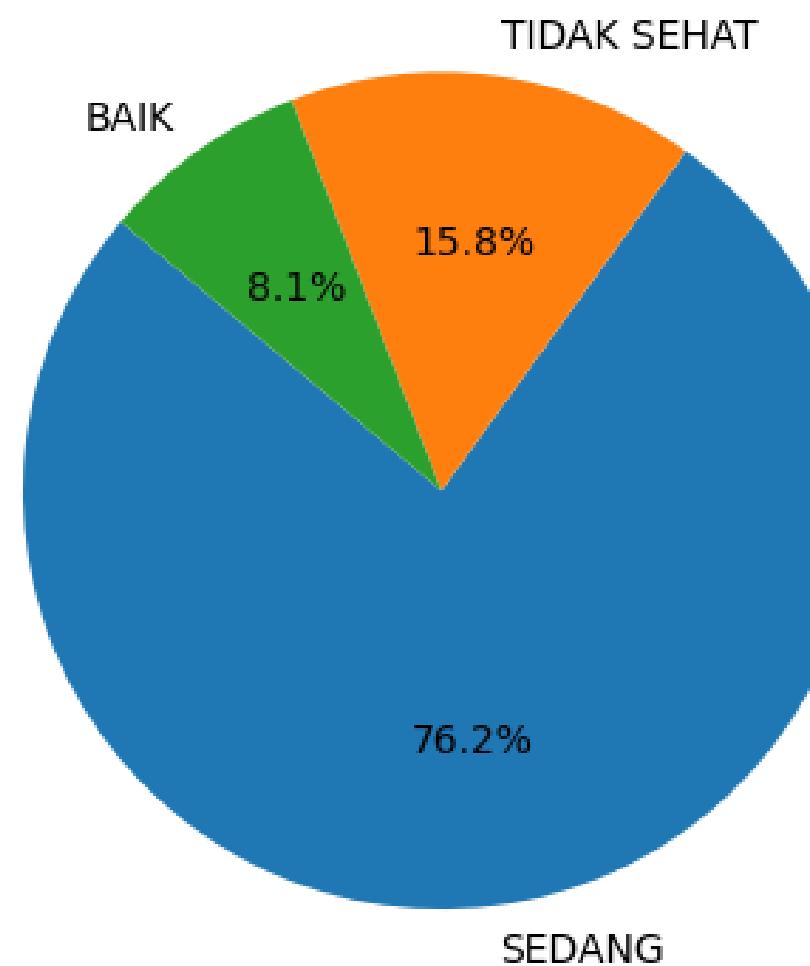


Skor Polutan Tertinggi



Persentase Kategori

Pie Chart Kategori



kategori	Jumlah Data
BAIK	139
SEDANG	1314
TIDAK SEHAT	272

Machine Learning

04 Encoding

Menggunakan LabelEncoder dari Scikit-Learn untuk mengubah variabel kategori menjadi format numerik.

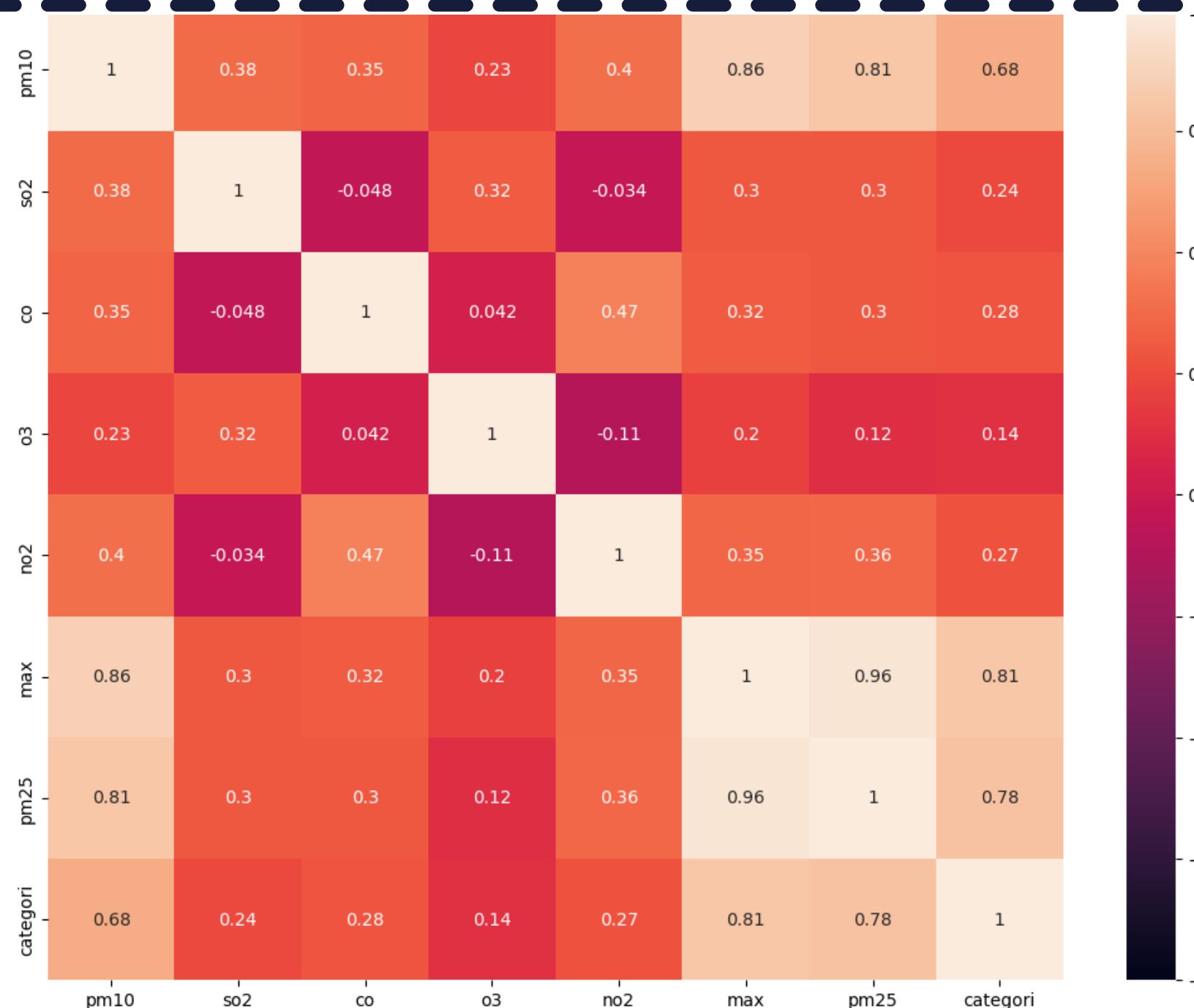
```
from sklearn.preprocessing import LabelEncoder  
label_encoder = LabelEncoder()  
data['kategori'] = label_encoder.fit_transform(data['kategori'])
```

05 Pemilihan Fitur dan Target

Memilih kolom-kolom tertentu sebagai fitur model dan kolom 'kategori' sebagai variabel target.

```
# memilih kolom untuk digunakan sebagai fitur model -- menghilangkan target variabel dan kolom tidak penting lainnya  
x = data.drop(columns=['kategori','max','critical','tanggal','stasiun'])  
# memilih kolom untuk digunakan sebagai variabel target  
y = data['kategori']
```

Correlation Heatmap



Heatmap
korelasi
antar
variabel
numerik
dalam
dataset.

Machine Learning

06 Pembagian Data

Menggunakan `train_test_split` untuk membagi data menjadi set pelatihan dan pengujian.

```
# membagi data ke training dan testing data
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.2)

print("X_train : ", x_train.shape)
print("X_test : ", x_test.shape)
print("Y_train : ", y_train.shape)
print("Y_test : ", y_test.shape)
```

```
X_train : (1380, 6)
X_test : (345, 6)
Y_train : (1380,)
Y_test : (345,)
```

07 SMOTE

Di sini kita melihat bahwa label target tidak dibagi dengan setara. Ini akan menyebabkan masalah ketidakseimbangan data. Oleh karena itu menggunakan SMOTE (Synthetic Minority Over-sampling Technique) untuk oversampling agar menyeimbangkan kelas target.

```
from imblearn.over_sampling import SMOTE
oversample = SMOTE()
x_train,y_train = oversample.fit_resample(x_train,y_train)
print('Classes and number of values in trainset after SMOTE:',Counter(y_train))
```

```
Kelas dan jumlah nilai di trainset Counter({1: 1042, 2: 225, 0: 113})
```

```
Classes and number of values in trainset after SMOTE: Counter({1: 1042, 0: 1042, 2: 1042})
```

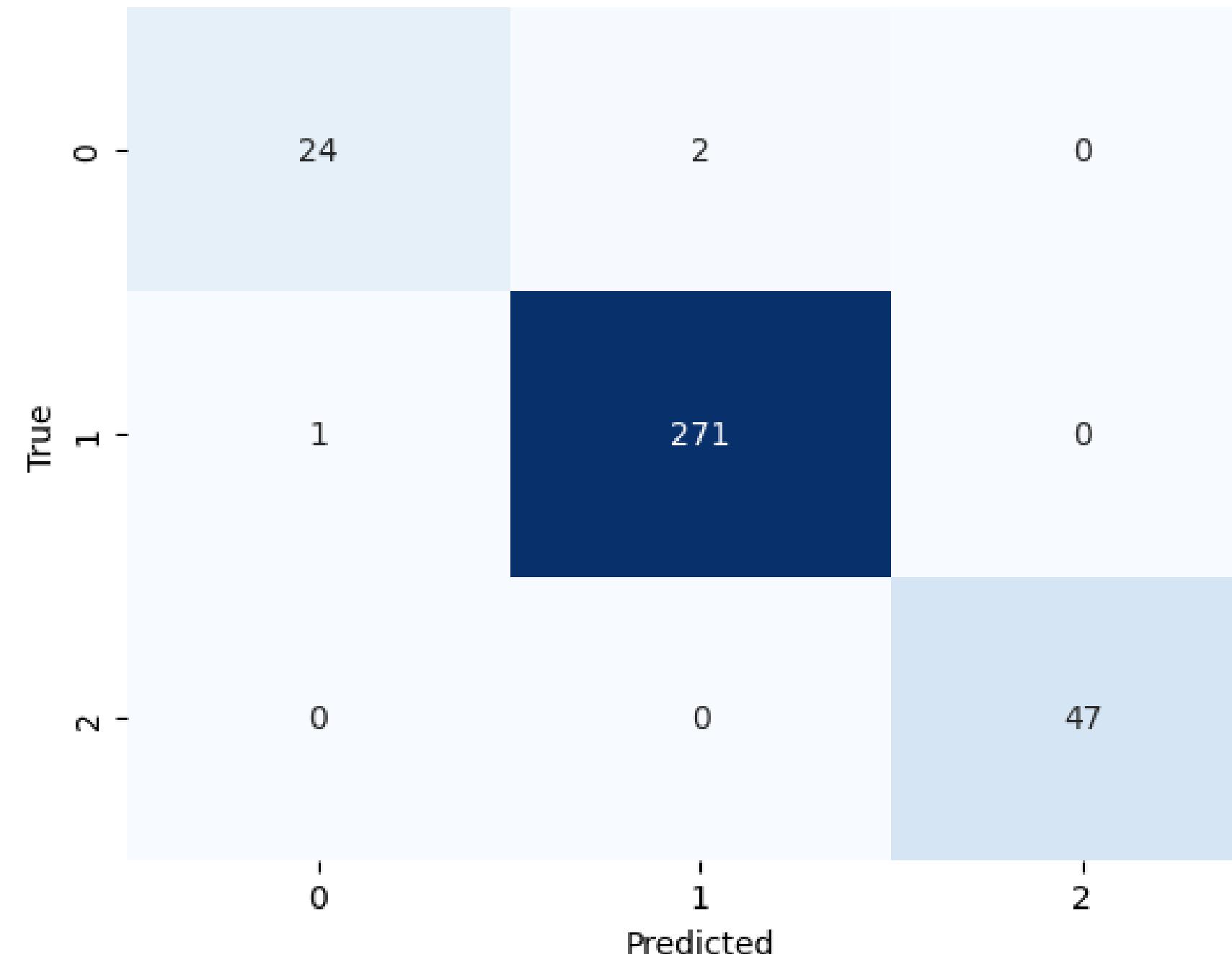
Machine Learning

08 Random Forest Classifier

Menggunakan RandomForestClassifier dari Scikit-Learn untuk membuat model Random Forest.
Melakukan fitting model pada data pelatihan dan membuat prediksi pada data pengujian.

```
rf = RandomForestClassifier(n_estimators=20, random_state=23)
rf.fit(x_train, y_train)
rf_predict=rf.predict(x_test)
rf_predict1=rf.predict(x_train)
rf_conf_matrix = confusion_matrix(y_test, rf_predict)
rf_acc_score = accuracy_score(y_test, rf_predict)
print("confussion matrix")
print(rf_conf_matrix)
print("\n")
print("accuracy",rf_acc_score*100)
```

Result Random Forest



confussion matrix

```
[[ 24  2  0]
 [  1 271  0]
 [  0  0  47]]
```

accuracy 99.1304347826087

	precision	recall	f1-score	support
0	0.96	0.92	0.94	26
1	0.99	1.00	0.99	272
2	1.00	1.00	1.00	47
accuracy				0.99
macro avg	0.98	0.97	0.98	345
weighted avg	0.99	0.99	0.99	345

Machine Learning

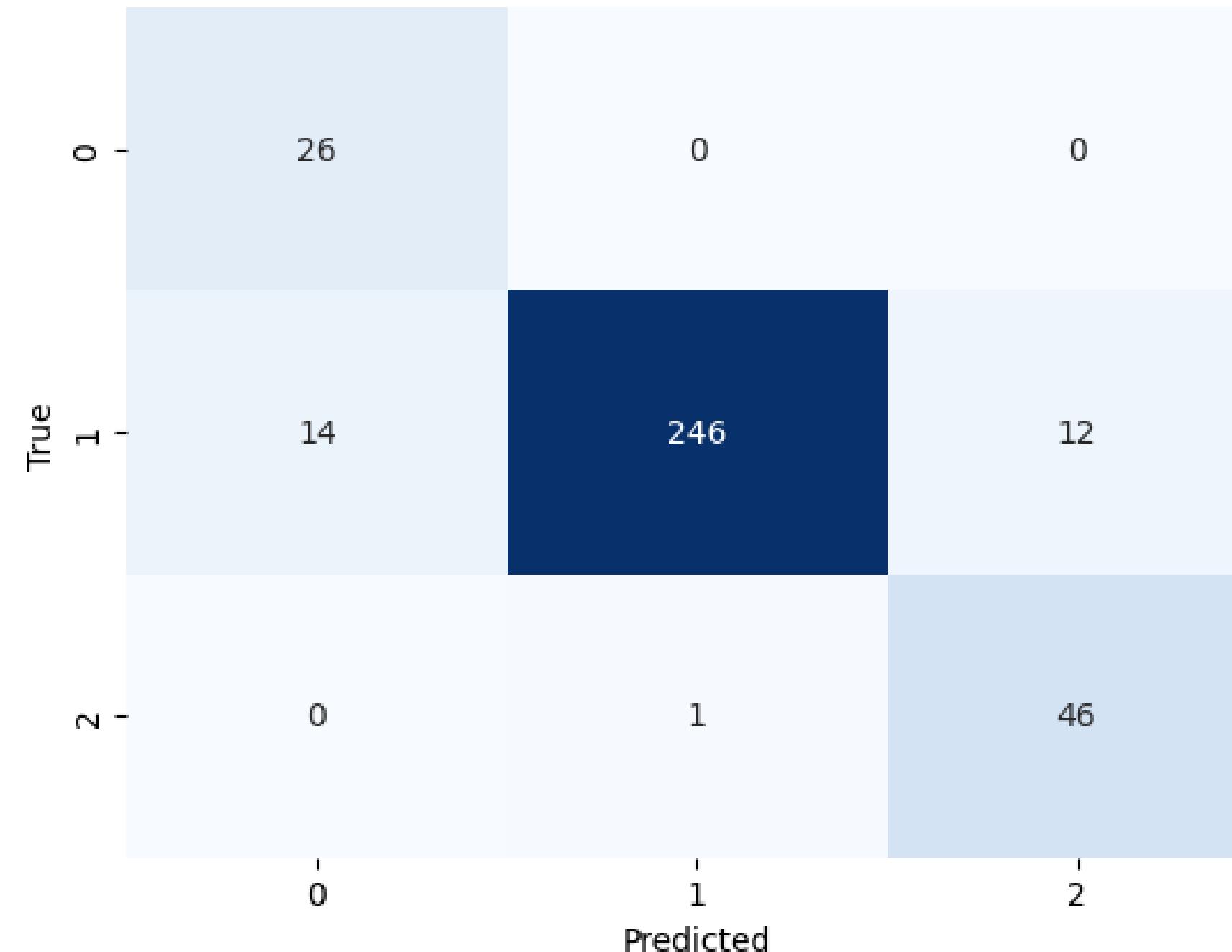
09 K-Nearest Neighbors (KNN) Classifier

Menggunakan KNeighborsClassifier dari Scikit-Learn untuk membuat model KNN.

Melakukan fitting model pada data pelatihan dan membuat prediksi pada data pengujian.

```
from sklearn.neighbors import KNeighborsClassifier  
knn = KNeighborsClassifier(n_neighbors=5)  
knn.fit(x_train, y_train)  
knn.score(x_test, y_test)  
y_pred = knn.predict(x_test)  
cm = confusion_matrix(y_test, y_pred)
```

Result KNN



confussion matrix

```
[[ 26  0  0]
 [ 14 246 12]
 [  0   1 46]]
```

accuracy 92.17391304347827

	precision	recall	f1-score	support
0	0.65	1.00	0.79	26
1	1.00	0.90	0.95	272
2	0.79	0.98	0.88	47
accuracy				0.92
macro avg	0.81	0.96	0.87	345
weighted avg	0.94	0.92	0.93	345

Machine Learning

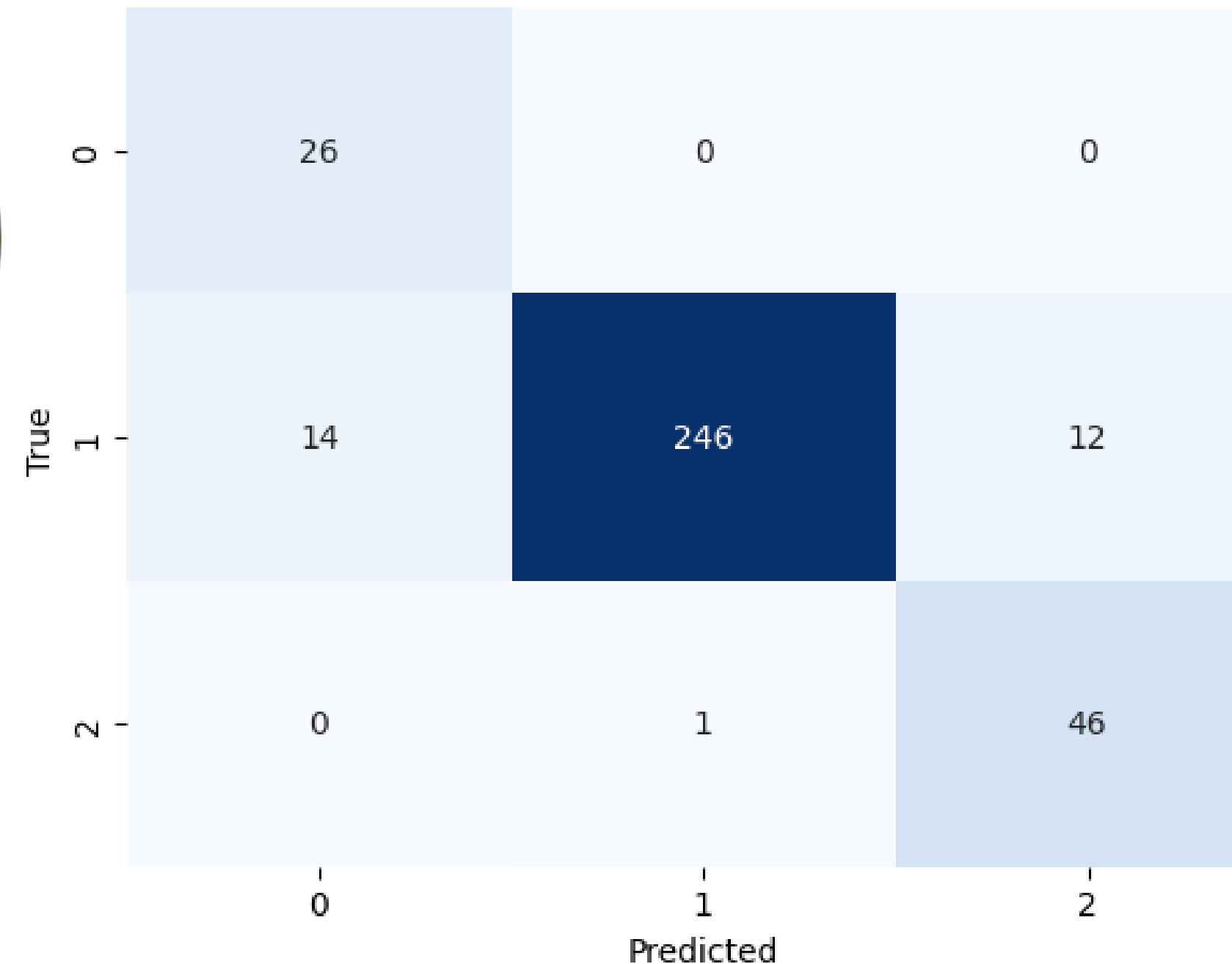
10 Support Vector Machine (SVM)

Menggunakan menggunakan kelas SVC dari Scikit-Learn untuk membuat model SVM.

Melakukan fitting model pada data pelatihan dan membuat prediksi pada data pengujian.

```
cls=SVC()
cls.fit(x_train,y_train)
svmpred=cls.predict(x_test)
svmpred
cm=confusion_matrix(y_test,svmpred)
print("confussion matrix")
print(cm)
print("\n")
accuracy=accuracy_score(y_test,svmpred)
print("accuracy",accuracy*100)
```

Result SVM



confussion matrix

```
[[ 26  0  0]
 [ 23 241  8]
 [  0  0 47]]
```

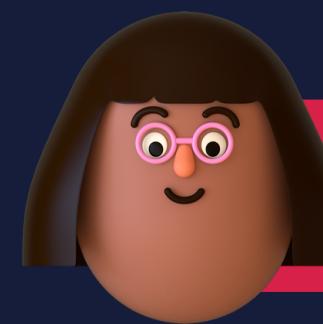
accuracy 91.01449275362319

	precision	recall	f1-score	support
--	-----------	--------	----------	---------

0	0.53	1.00	0.69	26
1	1.00	0.89	0.94	272
2	0.85	1.00	0.92	47

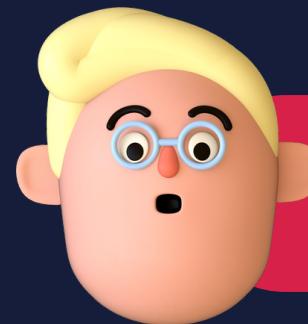
accuracy			0.91	345	
macro avg		0.80	0.96	0.85	345
weighted avg		0.94	0.91	0.92	345

Result & Discussion



Polusi Tertinggi

Dihasilkan pada daerah Provinsi Ibukota DKI Jakarta yaitu terdapat pada daerah DKI4 (Lubang Buaya) dengan kandungan polutan tertinggi yaitu PM2.5



Akurasi Tertinggi

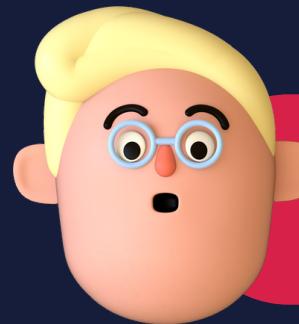
Pada pengklasifikasian dengan 3 metode yang digunakan menghasilkan tingkat akurasi tertinggi dengan hasil akurasi 99.1304347826087 menggunakan metode random forest

Conclusion

Menggunakan tiga metode klasifikasi berbeda, yaitu Random Forest Classifier, K-Nearest Neighbors (KNN), dan Support Vector Machine (SVM) untuk mengidentifikasi kategori polusi udara berdasarkan sejumlah fitur yang diamati. Setiap metode memberikan hasil prediksi berdasarkan pendekatan dan prinsipnya masing-masing. Hasil evaluasi ini meghasilkan random forest sebagai metode terbaik yang dapat memprediksi kategori polusi udara pada dataset yang digunakan.



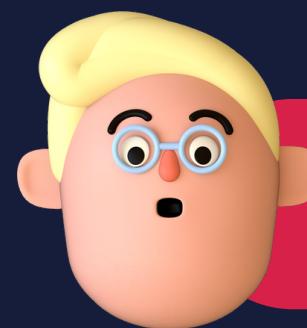
Link Video Present



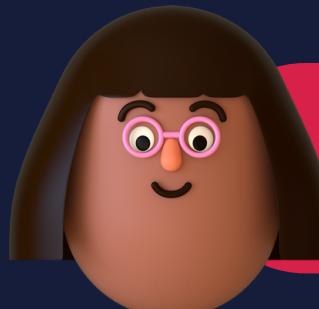
<https://youtu.be/uyESCxxEgMk?si=bkgXrs2m780HnFpe>

Mata Kuliah Pembelajaran Mesin

Link Code Data Cleaning



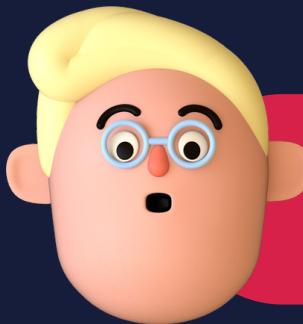
[Google Colab Data Cleaning](#)



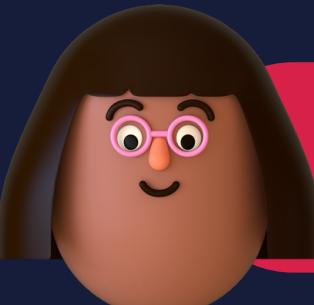
[Dataset ISPU 5 Daerah](#)

Mata Kuliah Pembelajaran Mesin

Link Code Analisis Udara



[Google Colab Machine Learning](#)



[Dataset ISPU Gabungan 5 Daerah](#)

Mata Kuliah Pembelajaran Mesin



thank
you

Mata Kuliah Pembelajaran Mesin