

LAPORAN ANALISIS DATASET COVID-19 DI INDONESIA



Disusun Oleh:

Nevin Athalla Sukandar (1201221037)

Ferdi Andra Maulana (1201222064)

Aditya Firmansyah A'sing (1201225097)

Syahnauval Dhiaulhaq Ridwan (1201221018)

(SI22001)

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS REKAYASA INDUSTRI

UNIVERSITAS TELKOM KAMPUS JAKARTA

2024

Kata Pengantar

Puji syukur kami panjatkan ke hadirat Tuhan Yang Maha Esa, yang telah melimpahkan rahmat dan karunia-Nya sehingga laporan studi kasus yang berjudul "*Analisis Dataset COVID-19 di Indonesia*" dapat diselesaikan dengan baik. Laporan ini disusun sebagai bagian dari upaya mendalami pemahaman terhadap dampak pandemi COVID-19 di Indonesia melalui analisis data yang komprehensif.

Studi kasus ini bertujuan untuk mengidentifikasi tren, pola, dan faktor-faktor penting yang mempengaruhi penyebaran dan penanganan pandemi. Dengan menggunakan pendekatan berbasis data, laporan ini diharapkan dapat memberikan kontribusi terhadap penyusunan kebijakan yang lebih baik serta meningkatkan kesadaran masyarakat tentang pentingnya pengelolaan data dalam menghadapi krisis kesehatan.

Semoga laporan ini dapat memberikan manfaat bagi pembaca, khususnya dalam memahami dinamika pandemi COVID-19 di Indonesia dan pentingnya analisis data untuk pengambilan keputusan strategis.

Daftar Isi

Kata Pengantar.....	1
Daftar Isi.....	2
Daftar Gambar.....	3
Bab I.....	4
Pendahuluan.....	4
Bab II.....	5
Kerangka Kerja.....	5
2.1 Identifikasi Masalah dan Perumusan Tujuan.....	5
2.2 Pengumpulan dan Pemahaman Data.....	5
2.3 Eksplorasi Data.....	5
2.4 Data Preparation.....	5
1. Scatter Plot Kiri (Tingkat Kematian):.....	11
2. Scatter Plot Kanan (Tingkat Pemulihan):.....	11
2.5 Modelling dan Evaluation.....	11
2.6 Mockup Dashboard.....	12
Bab III.....	13
Hasil dan Pembahasan.....	13
3.1 Unsupervised Analysis: K-Means Clustering.....	13
3.2 Supervised Analysis: Linear Regression.....	13
Kesimpulan.....	14
3.3 Tampilan Dashboard Analisis Data COVID-19 di Indonesia.....	14
BAB IV.....	15
Penutup.....	15
Daftar Pustaka.....	16
Lampiran.....	17

Daftar Gambar

Gambar 2.1 Import Library.....	4
Gambar 2.2 Gathering Data.....	4
Gambar 2.3 Data Info.....	4
Gambar 2.4 Data Shape & Tail.....	5
Gambar 2.5 Data is Null & Sum.....	5
Gambar 2.6 Data Duplicate.....	5
Gambar 2.7 Data Drop.....	6
Gambar 2.8 Exploring Data.....	7
Gambar 2.9 Exploring Data 2.....	7
Gambar 2.10 Data Visualization.....	8
Gambar 2.11 Data Visualization 2.....	9
Gambar 2.12 Clustering.....	9
Gambar 2.13 Regression.....	10
Gambar 2.14 Mockup Dashboard 1.....	10
Gambar 2.15 Mockup Dashboard 2.....	10
Gambar 3.1 Tampilan Dashboard.....	12

Bab I

Pendahuluan

1.1 Latar Belakang Pandemi COVID-19 telah membawa dampak yang signifikan di berbagai aspek kehidupan, termasuk kesehatan, ekonomi, dan sosial. Di Indonesia, keberagaman kondisi geografis, demografis, serta distribusi fasilitas kesehatan di setiap provinsi menjadi tantangan tersendiri dalam pengelolaan pandemi. Untuk menentukan tingkat keketatan kebijakan di setiap wilayah, diperlukan analisis data yang mendalam berdasarkan perkembangan pandemi dan faktor demografis di tingkat provinsi hingga kabupaten/kota. Pentingnya data dalam pengambilan keputusan selama pandemi mendorong tersedianya berbagai sumber data yang diolah menjadi dataset yang terstruktur. Data ini meliputi deret waktu kasus COVID-19, data demografi, hingga penghitungan indikator tertentu yang relevan untuk menggambarkan tingkat risiko di berbagai wilayah. Melalui analisis yang berbasis data, pemerintah dapat menentukan kebijakan yang lebih efektif dan terukur dalam menghadapi pandemi.

1.2 Rumusan Masalah

Beberapa masalah utama yang melatarbelakangi penelitian ini meliputi:

1. Bagaimana perbedaan angka kasus aktif antara pulau?
2. Bagaimana densitas penduduk mempengaruhi tingkat kematian atau pemulihan?

1.3 Tujuan Penelitian

Tujuan utama dari penelitian ini adalah: Mengintegrasikan data COVID-19 dan data demografis untuk menghasilkan insight yang mendukung pengambilan kebijakan. Mengidentifikasi faktor-faktor utama yang mempengaruhi tingkat keketatan kebijakan di setiap provinsi. Mengembangkan analisis berbasis data yang menggambarkan tren pandemi secara komprehensif untuk setiap wilayah di Indonesia.

1.4 State of the Art

Sejauh ini, penelitian terkait COVID-19 banyak berfokus pada analisis data epidemiologis, model prediktif, dan dampak kebijakan pembatasan. Namun, integrasi antara data pandemi dan data demografi untuk menentukan tingkat keketatan wilayah secara spesifik di Indonesia masih memerlukan pengembangan lebih lanjut. Studi ini memberikan kontribusi unik dengan menyusun dataset yang mencakup: Deret waktu kasus COVID-19 di tingkat nasional, provinsi, hingga kabupaten/kota. Data demografis dan indikator yang dihasilkan dari kombinasi kedua jenis data tersebut. Dengan pendekatan ini, penelitian ini tidak hanya memberikan gambaran tren pandemi tetapi juga analisis risiko berdasarkan konteks wilayah masing-masing.

Bab II

Kerangka Kerja

2.1 Identifikasi Masalah dan Perumusan Tujuan

Tahapan ini bertujuan untuk memahami masalah utama yang ingin diselesaikan dan merumuskan tujuan bisnis dari analisis data yang dilakukan. Masalah utama dalam konteks ini adalah bagaimana memanfaatkan data COVID-19 dan data demografi untuk mendukung pengambilan keputusan berbasis data yang efektif, seperti menentukan tingkat keketatan kebijakan di berbagai wilayah Indonesia. Tujuan bisnis dari penelitian ini adalah membantu pemerintah atau lembaga kesehatan untuk mengelompokkan wilayah berdasarkan karakteristik pandemi, memprediksi jumlah kasus baru untuk mendukung perencanaan strategis, serta mengidentifikasi wilayah dengan tingkat risiko tinggi yang membutuhkan prioritas penanganan.

2.2 Pengumpulan dan Pemahaman Data

Pada tahap ini, masalah utama yang diidentifikasi adalah bagaimana menganalisis data COVID-19 secara efektif untuk mendukung kebijakan berbasis data. Tujuan bisnisnya adalah membantu pemerintah atau lembaga kesehatan membuat keputusan strategis dengan hasil analisis yang dihasilkan, termasuk clustering wilayah berdasarkan karakteristik pandemi dan prediksi jumlah kasus baru. Dataset yang dianalisis mencakup informasi COVID-19 di Indonesia yang diperoleh dari sumber resmi, seperti covid19.go.id, kemendagri.go.id, bps.go.id, dan bnpb-inacovid19.hub.arcgis.com. Data ini mencakup informasi deret waktu (time-series) tentang kasus harian, total kasus, jumlah kematian, jumlah kesembuhan, populasi, kepadatan penduduk, dan luas wilayah. Eksplorasi awal data menunjukkan adanya pola pertumbuhan kasus yang signifikan di berbagai wilayah, yang memberikan peluang untuk analisis lebih lanjut. Selain itu, data demografi seperti populasi dan kepadatan penduduk dapat menjadi indikator risiko pandemi yang relevan untuk pengambilan keputusan.

2.3 Eksplorasi Data

Eksplorasi awal data mengungkap informasi penting, seperti jumlah kasus baru harian, total kasus, jumlah kematian, kesembuhan, populasi, dan kepadatan penduduk. Data juga menunjukkan pola pertumbuhan kasus di berbagai wilayah, yang dapat dianalisis lebih lanjut untuk menemukan tren dan hubungan antar variabel. Analisis awal mengindikasikan adanya keterkaitan antara kepadatan penduduk dengan jumlah kasus di wilayah tertentu. Visualisasi, seperti grafik tren kasus harian, scatter plot antara kepadatan penduduk dan jumlah kasus, serta heatmap untuk korelasi antar variabel, digunakan untuk memberikan gambaran lebih mendalam mengenai distribusi data dan pola yang signifikan.

2.4 Data Preparation

Data diproses untuk memastikan kualitasnya, langkah-langkah yang dilakukan meliputi:

- Membersihkan data dengan mengisi nilai kosong (*missing values*).
- Mengubah format tanggal menjadi tipe data *datetime* untuk analisis waktu
- Melakukan normalisasi data menggunakan *StandardScaler* untuk clustering agar setiap fitur memiliki skala yang seimbang.
- Membuat fitur tambahan, seperti "Days Since Start", yang mencatat jumlah hari sejak awal pandemi untuk analisis tren.

Import Library

```
import pandas as pd      Import "pandas" could not be resolved from source
import matplotlib.pyplot as plt    Import "matplotlib.pyplot" could not be resolved from source
import seaborn as sns     Import "seaborn" could not be resolved from source
```

✓ 52s Python

Gambar 2.1 Import Library

Dataset yang ditampilkan memuat data dengan beberapa kolom, antara lain: seperti di bawah, Date: Tanggal pencatatan data kasus COVID-19. Location ISO Code: Kode ISO untuk lokasi (misalnya, ID-JK untuk DKI Jakarta). Location: Nama lokasi (provinsi atau negara). New Deaths: Jumlah kematian baru akibat COVID-19 pada tanggal tersebut. Dll seperti di bawah

Gathering Data

```
data = pd.read_csv('covid_19_indonesia_time_series_all.csv')
data.head()
```

✓ 02s Python

	Date	Location ISO Code	Location	New Cases	New Deaths	New Recovered	New Active Cases	Total Cases	Total Deaths	Total Recovered	...	Latitude	New Cases per Million	Total Cases per Million	New Deaths per Million	Total Deaths per Million	Total Deaths per 100rb	Case Fatality Rate	Cases Recovered Ratio
0	3/1/2020	ID-JK	DKI Jakarta	2	0	0	2	39	20	75	...	-6.204699	0.18	3.60	0.0	1.84	0.18	51.28%	192.319
1	3/2/2020	ID-JK	DKI Jakarta	2	0	0	2	41	20	75	...	-6.204699	0.18	3.78	0.0	1.84	0.18	48.78%	182.939
2	3/2/2020	IDN	Indonesia	2	0	0	2	2	0	0	...	-0.789275	0.01	0.01	0.0	0.00	0.00	0.00%	0.009
3	3/2/2020	ID-RI	Riau	1	0	0	1	1	0	1	...	0.511648	0.16	0.16	0.0	0.00	0.00	0.00%	100.009
4	3/3/2020	ID-JK	DKI Jakarta	2	0	0	2	43	20	75	...	-6.204699	0.18	3.96	0.0	1.84	0.18	46.51%	174.429

5 rows x 38 columns

Gambar 2.2 Gathering Data

`data.info()` Fungsi ini digunakan untuk memberikan informasi tentang dataset yang sedang dianalisis.

```
data.info()
```

✓ 0.0s Python

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 31822 entries, 0 to 31821
Data columns (total 38 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Date                  31822 non-null object
1   Location ISO Code     31822 non-null object
2   Location              31822 non-null object
3   New Cases             31822 non-null int64
4   New Deaths           31822 non-null int64
5   New Recovered         31822 non-null int64
6   New Active Cases     31822 non-null int64
7   Total Cases           31822 non-null int64
8   Total Deaths         31822 non-null int64
9   Total Recovered       31822 non-null int64
10  Total Active Cases    31822 non-null int64
11  Location Level        31822 non-null object
12  City or Regency       0 non-null      float64
13  Province              30893 non-null object
14  Country               31822 non-null object
15  Continent             31822 non-null object
```

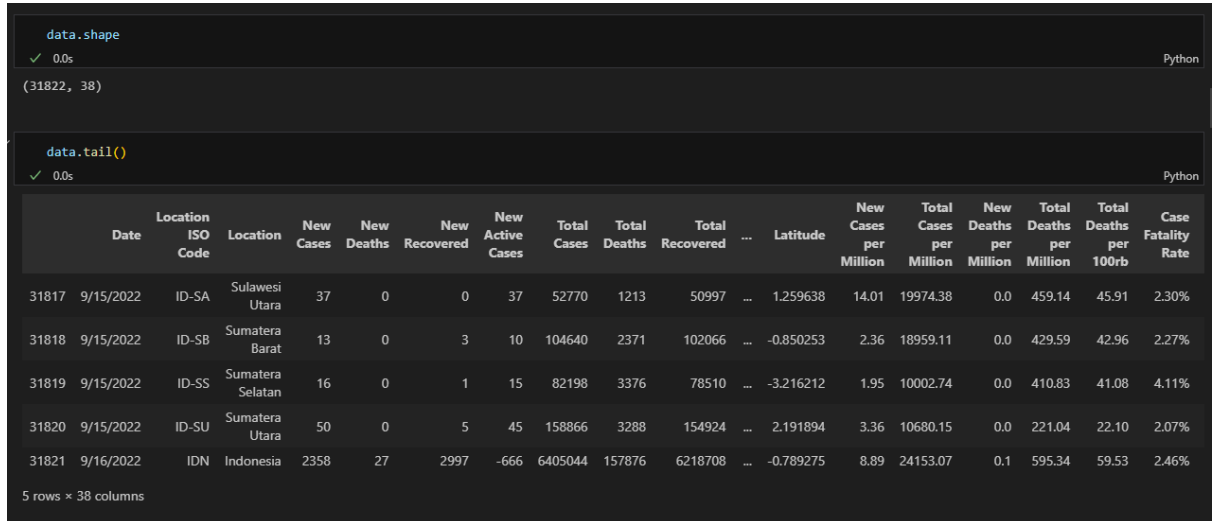
Gambar 2.3 Data Info

Kode yang ditampilkan memanfaatkan fungsi `data.shape` dan `data.tail()` untuk memahami ukuran dataset dan melihat data terakhir

Output berupa tuple dalam format(jumlah baris dan jumlah kolom).

Sedangkan `data.tail` menampilkan lima baris terakhir dari dataset secara default.

Berguna untuk memeriksa data pada bagian akhir dataset, seperti tren atau catatan terbaru.



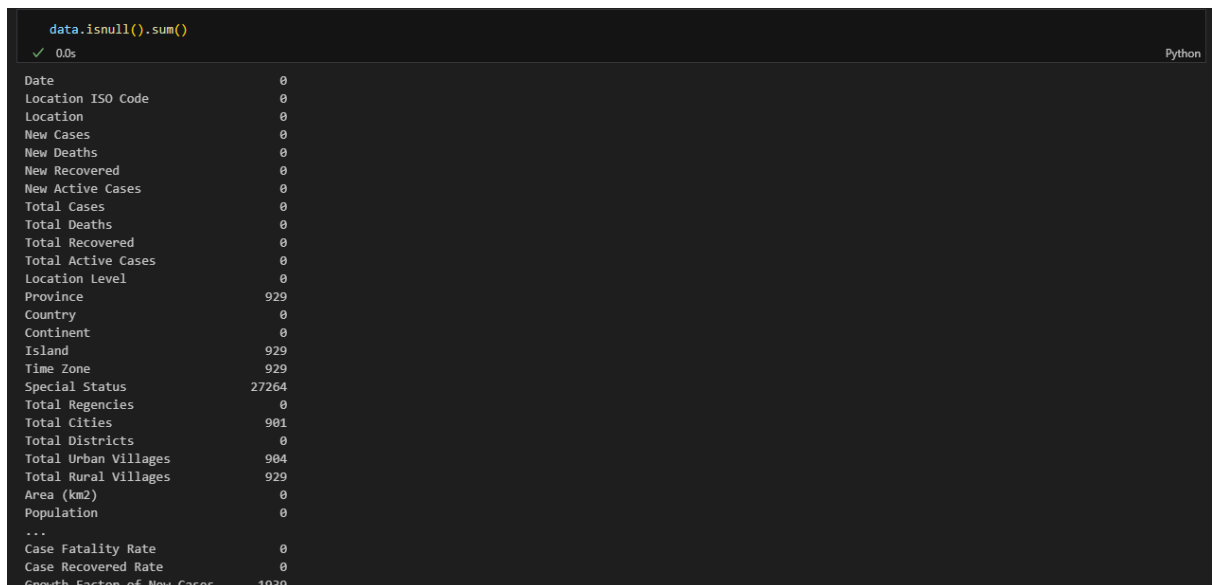
The screenshot shows two code cells in a Jupyter Notebook. The first cell runs `data.shape`, returning the tuple `(31822, 38)`. The second cell runs `data.tail()`, displaying the last five rows of the dataset. The data includes columns for Date, Location (ISO Code, Location), New Cases, New Deaths, New Recovered, New Active Cases, Total Cases, Total Deaths, Total Recovered, Latitude, New Cases per Million, Total Cases per Million, New Deaths per Million, Total Deaths per Million, and Case Fatality Rate.

	Date	Location ISO Code	Location	New Cases	New Deaths	New Recovered	New Active Cases	Total Cases	Total Deaths	Total Recovered	...	Latitude	New Cases per Million	Total Cases per Million	New Deaths per Million	Total Deaths per Million	Total Deaths per 100rb	Case Fatality Rate
31817	9/15/2022	ID-SA	Sulawesi Utara	37	0	0	37	52770	1213	50997	...	1.259638	14.01	19974.38	0.0	459.14	45.91	2.30%
31818	9/15/2022	ID-SB	Sumatera Barat	13	0	3	10	104640	2371	102066	...	-0.850253	2.36	18959.11	0.0	429.59	42.96	2.27%
31819	9/15/2022	ID-SS	Sumatera Selatan	16	0	1	15	82198	3376	78510	...	-3.216212	1.95	10002.74	0.0	410.83	41.08	4.11%
31820	9/15/2022	ID-SU	Sumatera Utara	50	0	5	45	158866	3288	154924	...	2.191894	3.36	10680.15	0.0	221.04	22.10	2.07%
31821	9/16/2022	IDN	Indonesia	2358	27	2997	-666	6405044	157876	6218708	...	-0.789275	8.89	24153.07	0.1	595.34	59.53	2.46%

5 rows x 38 columns

Gambar 2.4 Data Shape & Tail

`data.isnull().sum()` memeriksa setiap elemen dalam dataset untuk melihat apakah nilai tersebut kosong sedangkan `sum` menghitung jumlah nilai kosong



The screenshot shows a code cell running `data.isnull().sum()`, which returns a Series of counts for each column. Most columns have a count of 0, indicating no missing values. Some columns like 'Province', 'Island', 'Time Zone', and 'Special Status' have counts of 929, 929, 929, and 27264 respectively.

Column	Count
Date	0
Location ISO Code	0
Location	0
New Cases	0
New Deaths	0
New Recovered	0
New Active Cases	0
Total Cases	0
Total Deaths	0
Total Recovered	0
Total Active Cases	0
Location Level	0
Province	929
Country	0
Continent	0
Island	929
Time Zone	929
Special Status	27264
Total Regencies	0
Total Cities	901
Total Districts	0
Total Urban Villages	904
Total Rural Villages	929
Area (km2)	0
Population	0
...	
Case Fatality Rate	0
Case Recovered Rate	0
Growth Factor of New Cases	1939

Gambar 2.5 Data is Null & Sum

Memeriksa setiap baris dalam dataset untuk melihat apakah ada baris yang duplikat (artinya semua nilai dalam baris tersebut identik dengan baris lainnya).


```
data.duplicated().sum()
✓ 0.0s
np.int64(0)
```

Gambar 2.6 Data Duplicate

`data.dropna()` Fungsi ini digunakan untuk memeriksa apakah ada nilai kosong (NaN) dalam dataset.

Fungsi ini digunakan untuk menghapus baris yang mengandung nilai kosong (NaN) dari dataset. Secara default, fungsi ini akan menghapus semua baris di mana salah satu kolom memiliki nilai NaN.

Hasilnya adalah dataset baru yang hanya berisi baris dengan data lengkap (tidak ada nilai kosong).

```
data = data.dropna()
✓ 0.0s

data.info()
✓ 0.0s

<class 'pandas.core.frame.DataFrame'>
Index: 2319 entries, 100 to 31812
Data columns (total 37 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Date                  2319 non-null  object
1   Location ISO Code     2319 non-null  object
2   Location               2319 non-null  object
3   New Cases             2319 non-null  int64
4   New Deaths           2319 non-null  int64
5   New Recovered         2319 non-null  int64
6   New Active Cases      2319 non-null  int64
7   Total Cases           2319 non-null  int64
8   Total Deaths          2319 non-null  int64
9   Total Recovered       2319 non-null  int64
10  Total Active Cases     2319 non-null  int64
11  Location Level         2319 non-null  object
12  Province               2319 non-null  object
13  Country                2319 non-null  object
14  Continent              2319 non-null  object
15  Island                 2319 non-null  object
16  Time Zone              2319 non-null  object
17  Special Status         2319 non-null  object
```

Gambar 2.7 Data Drop

Exploring Data seperti gambar di bawah untuk code nya dimana menampilkan Output beberapa kolom, yaitu:

1. Total Active Cases
2. Island Nama pulau tempat data dicatat (misalnya, "Jawa").
3. Cluster Hasil clustering yang menunjukkan kelompok data berdasarkan pola dalam fitur Total Active Cases dan Island.

Contoh Output:

Baris pertama:

Total Active Cases: 100.

Island: "Jawa".

Cluster: 1 (artinya data ini termasuk dalam cluster ke-1 berdasarkan analisis K-Means).

Exploring Data

```

from sklearn.preprocessing import LabelEncoder    Import "sklearn.preprocessing" could not be resolved from source
from sklearn.cluster import KMeans              Import "sklearn.cluster" could not be resolved from source

# Encode the 'Island' column
label_encoder = LabelEncoder()
data['Island_encoded'] = label_encoder.fit_transform(data['Island'].astype(str))

# Select relevant columns for clustering
clustering_data = data[['Total Active Cases', 'Island_encoded']]

# Apply KMeans clustering
kmeans = KMeans(n_clusters=4, random_state=42)
data['Cluster'] = kmeans.fit_predict(clustering_data)

# Display the first few rows of the dataframe with the new 'Cluster' column
data[['Total Active Cases', 'Island', 'Cluster']].head()

```

✓ 2.0s Python

	Total Active Cases	Island	Cluster
100	-1	Jawa	1
111	-1	Jawa	1
123	-1	Jawa	1
152	-1	Jawa	1
167	-1	Jawa	1

Gambar 2.8 Exploring Data

```

from sklearn.linear_model import LinearRegression    Import "sklearn.linear_model" could not be resolved from source
import numpy as np    Import "numpy" could not be resolved

df = data[['Population Density', 'Total Deaths', 'Total Recovered', 'Total Cases']].dropna()

# Hitung tingkat kematian dan pemulihan
df['Case Fatality Rate'] = (df['Total Deaths'] / df['Total Cases']) * 100
df['Case Recovered Rate'] = (df['Total Recovered'] / df['Total Cases']) * 100

# Variabel independen dan dependen
X = df[['Population Density']]
y_fatality = df['Case Fatality Rate']
y_recovery = df['Case Recovered Rate']

# Regresi Linear untuk tingkat kematian
model_fatality = LinearRegression()
model_fatality.fit(X, y_fatality)
y_pred_fatality = model_fatality.predict(X)

# Regresi Linear untuk tingkat pemulihan
model_recovery = LinearRegression()
model_recovery.fit(X, y_recovery)
y_pred_recovery = model_recovery.predict(X)

```

✓ 0.0s Python

Gambar 2.9 Exploring Data 2

Data visual jawaban no.1 yang ada di rumusan masalah di mana

Sumbu X (Island):Mewakili nama pulau, misalnya Jawa, Kalimantan, Papua, dll.

Sumbu Y (Total Active Cases): Menunjukkan jumlah total kasus aktif di setiap pulau

Warna (Cluster):

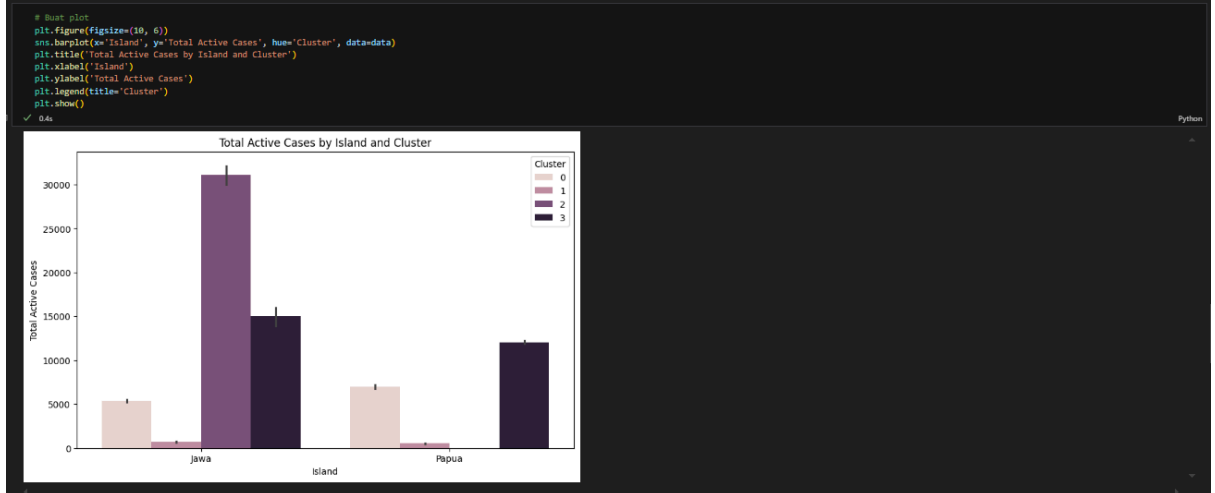
Setiap warna mewakili cluster tertentu (0, 1, 2, 3) yang dikelompokkan berdasarkan analisis clustering sebelumnya.

Perbedaan Kasus Aktif Antar Pulau:

- Pulau Jawa memiliki jumlah kasus aktif yang jauh lebih tinggi dibandingkan pulau lainnya, khususnya di cluster tertentu (misalnya, cluster 0).
- Pulau lain seperti Papua memiliki jumlah kasus aktif yang lebih rendah dan tersebar di beberapa cluster.

Data Visualization

1. Bagaimana perbedaan angka kasus aktif antara pulau?



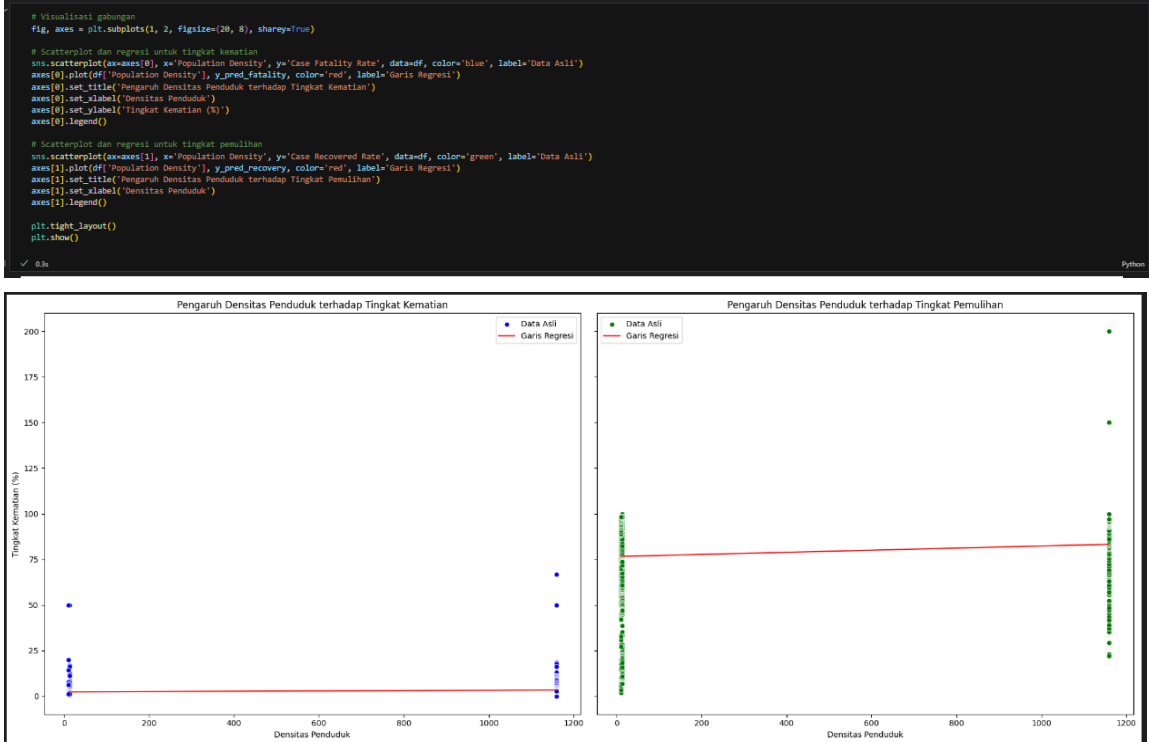
Gambar 2.10 Data Visualization

Data visual jawaban no.2 dari rumusan masalah

Penjelasan Visualisasi:

1. Scatter Plot Kiri (Tingkat Kematian):
 - Data Asli (Titik Biru): Menunjukkan distribusi tingkat kematian di berbagai wilayah dengan densitas penduduk berbeda.
 - Garis Regresi (Garis Merah): Garis linear yang menunjukkan tren hubungan antara densitas penduduk dan tingkat kematian.
2. Scatter Plot Kanan (Tingkat Pemulihan):
 - Data Asli (Titik Hijau): Menunjukkan distribusi tingkat pemulihan di berbagai wilayah dengan densitas penduduk berbeda.
 - Garis Regresi (Garis Merah): Garis linear yang menunjukkan tren hubungan antara densitas penduduk dan tingkat pemulihan.

2. Bagaimana densitas penduduk memengaruhi tingkat kematian atau pemulihan?

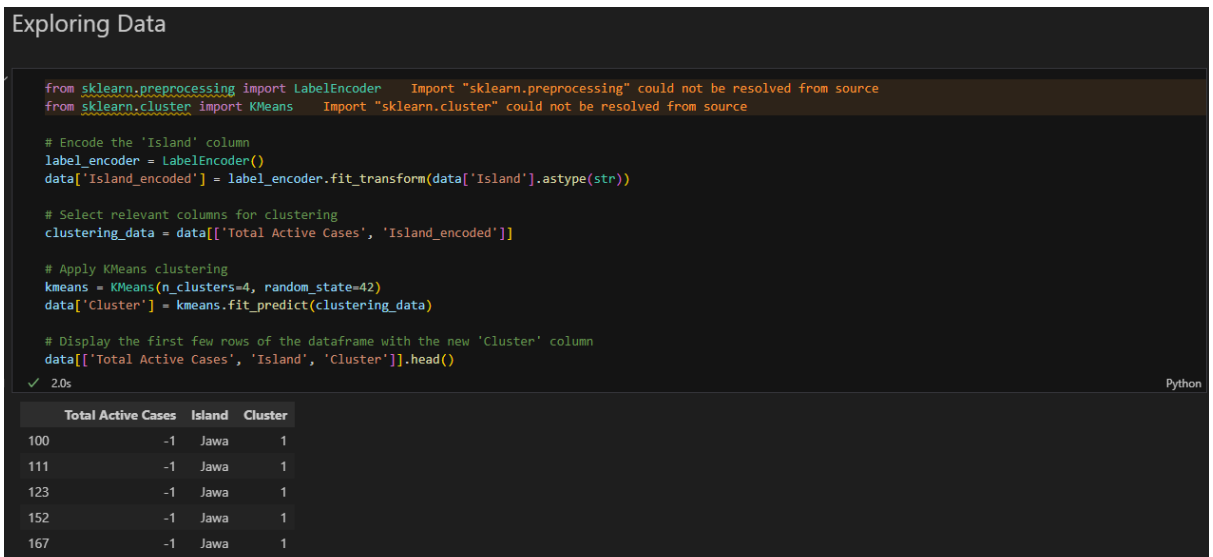


Gambar 2.11 Data Visualization 2

2.5 Modelling dan Evaluation

Beberapa model digunakan dalam analisis ini:

1. **Clustering:** Metode ini termasuk unsupervised learning yang digunakan untuk mengelompokkan data berdasarkan pola atau kesamaan pada dataset sehingga data yang akan digunakan tidak bercampur aduk dengan yang lainnya.



Gambar 2.12 Clustering

2. **Regression:** Metode ini termasuk kedalam metode supervised learning yang digunakan untuk memprediksi nilai kontinu berdasarkan hubungan antar variabel

independen dan dependen-nya, sehingga data yang diolah dapat dengan mudah diketahui.

```
from sklearn.linear_model import LinearRegression    Import "sklearn.linear_model" could not be resolved from source
import numpy as np    Import "numpy" could not be resolved

df = data[['Population Density', 'Total Deaths', 'Total Recovered', 'Total Cases']].dropna()

# Hitung tingkat kematian dan pemulihan
df['Case Fatality Rate'] = (df['Total Deaths'] / df['Total Cases']) * 100
df['Case Recovered Rate'] = (df['Total Recovered'] / df['Total Cases']) * 100

# Variabel independen dan dependen
X = df[['Population Density']]
y_fatality = df['Case Fatality Rate']
y_recovery = df['Case Recovered Rate']

# Regresi Linear untuk tingkat kematian
model_fatality = LinearRegression()
model_fatality.fit(X, y_fatality)
y_pred_fatality = model_fatality.predict(X)

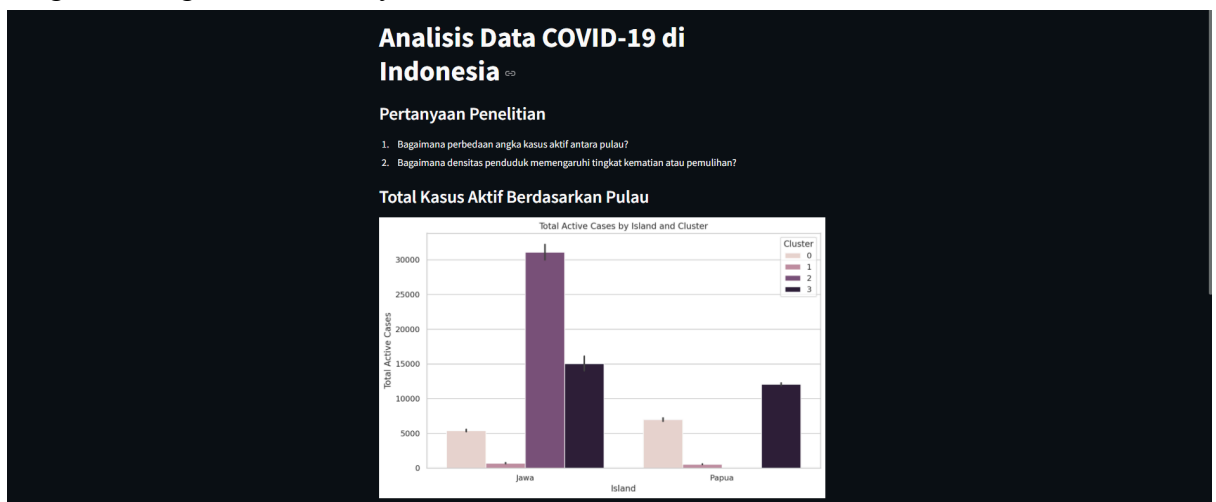
# Regresi Linear untuk tingkat pemulihan
model_recovery = LinearRegression()
model_recovery.fit(X, y_recovery)
y_pred_recovery = model_recovery.predict(X)
```

Gambar 2.13 Regression

2.6 Mockup Dashboard

Sebagai bagian dari hasil analisis, sebuah *mockup dashboard* dirancang untuk menampilkan data secara interaktif. Fitur utama dari dashboard ini meliputi:

- Peta interaktif yang menunjukkan hasil clustering wilayah.
- Grafik yang menampilkan tren kasus harian secara visual.
- Ringkasan tingkat risiko wilayah berdasarkan hasil analisis klasifikasi.



Gambar 2.14 Mockup Dashboard 1



Gambar 2.15 Mockup Dashboard 2

Bab III

Hasil dan Pembahasan

3.1 Unsupervised Analysis: K-Means Clustering

Metode dan Hasil:

- Model yang digunakan: K-Means Clustering.
- Input data: Dataset yang mencakup total kasus aktif, kategori pulau, dan kepadatan penduduk.
- Evaluasi model: Jumlah cluster optimal ditentukan menggunakan argumen manual dengan nilai `n_clusters=4`.
- Hasil clustering:
 1. Cluster 1: Wilayah dengan jumlah kasus aktif yang tinggi dan kepadatan penduduk besar (misalnya Jawa).
 2. Cluster 2: Wilayah dengan kasus aktif sedang dan kepadatan menengah.
 3. Cluster 3: Wilayah dengan jumlah kasus aktif rendah dan kepadatan rendah (seperti Papua).
 4. Cluster 4: Wilayah dengan kasus aktif fluktuatif tetapi kepadatan menengah.

Pembahasan:

- Hasil clustering memberikan informasi penting untuk menentukan prioritas wilayah dalam pengendalian pandemi.
- Wilayah dalam cluster dengan kepadatan penduduk besar (misalnya, Cluster 1) memerlukan perhatian lebih besar terkait distribusi fasilitas kesehatan dan kebijakan pembatasan.

3.2 Supervised Analysis: Linear Regression

Prediksi Tingkat Kematian dan Pemulihan

- Model yang digunakan: Linear Regression.
- Input data: Variabel independen adalah kepadatan penduduk (Population Density), sedangkan variabel dependen adalah:
 1. Tingkat kematian (Case Fatality Rate).
 2. Tingkat pemulihan (Case Recovered Rate).

Hasil Evaluasi:

1. Prediksi Tingkat Kematian:

Hubungan antara kepadatan penduduk dan tingkat kematian sangat lemah, terlihat dari garis regresi yang hampir mendatar.
Model tidak sepenuhnya mampu menjelaskan pola kematian berdasarkan kepadatan penduduk saja.
2. Prediksi Tingkat Pemulihan:

Hubungan antara kepadatan penduduk dan tingkat pemulihan juga sangat kecil.

Model regresi menunjukkan keterbatasan dalam menangkap pengaruh variabel independen.

Pembahasan:

- Faktor lain seperti tingkat vaksinasi, fasilitas kesehatan, dan kebijakan kesehatan publik kemungkinan memiliki pengaruh lebih besar dibandingkan kepadatan penduduk.
- Model dapat ditingkatkan dengan menambahkan variabel-variabel tambahan yang lebih relevan untuk analisis.

Kesimpulan

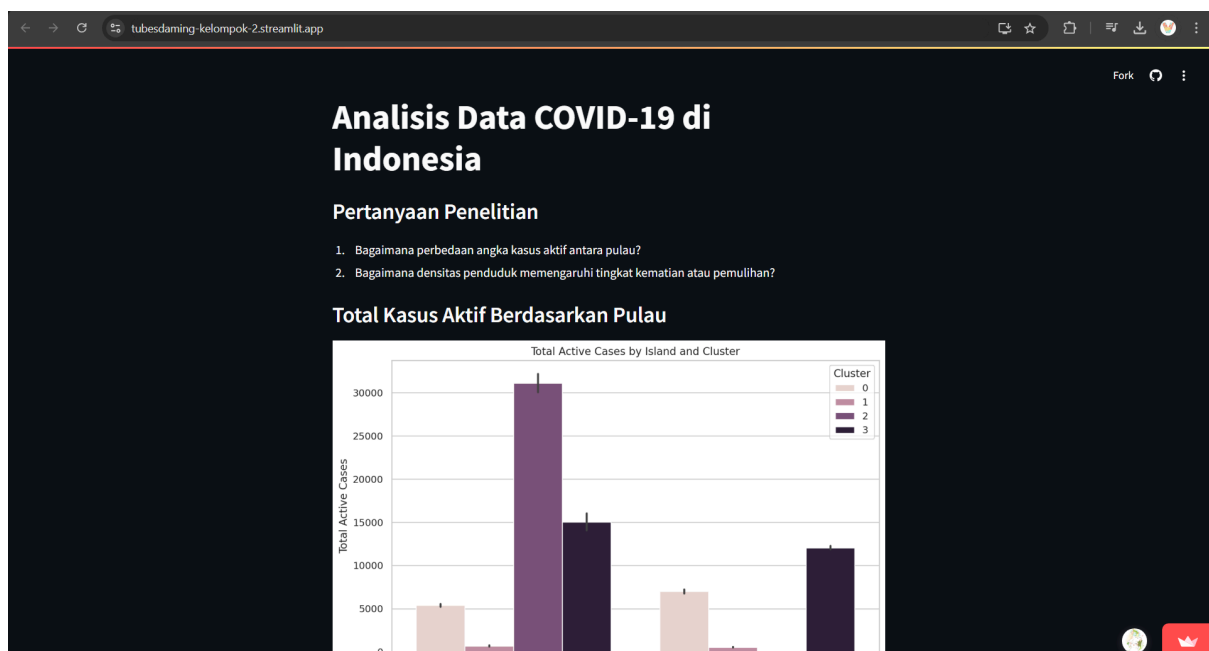
1. K-Means Clustering:

- Hasil clustering memberikan pembagian wilayah berdasarkan karakteristik kasus aktif dan kepadatan penduduk, membantu menentukan strategi pengendalian yang efektif.

2. Linear Regression:

- Hubungan antara kepadatan penduduk dengan tingkat kematian dan pemulihan cenderung lemah. Analisis lebih lanjut diperlukan untuk mengidentifikasi faktor-faktor yang lebih signifikan.

3.3 Tampilan Dashboard Analisis Data COVID-19 di Indonesia



Gambar 3.1 Tampilan Dashboard

BAB IV

Penutup

Penelitian yang dilakukan dalam laporan ini mengangkat analisis dataset COVID-19 di Indonesia dengan menggunakan metode seperti K-Means Clustering untuk analisis tanpa pengawasan (unsupervised learning) dan Linear Regression untuk analisis terarah (supervised learning). Hasil penelitian ini berhasil memberikan wawasan mengenai pola penyebaran, pengaruh densitas penduduk terhadap tingkat kematian dan pemulihan, serta pembagian wilayah berdasarkan karakteristik pandemi. Ditemukan bahwa wilayah dengan kepadatan penduduk tinggi, seperti Pulau Jawa, memiliki jumlah kasus aktif yang lebih besar, sehingga memerlukan perhatian khusus dalam distribusi sumber daya kesehatan dan penerapan kebijakan yang lebih ketat. Meskipun hubungan antara kepadatan penduduk dengan tingkat kematian dan pemulihan cenderung lemah, analisis ini memberikan pijakan awal untuk studi lebih lanjut dengan mempertimbangkan faktor-faktor lain, seperti tingkat vaksinasi, fasilitas kesehatan, dan kebijakan publik. Kami menyadari bahwa laporan ini memiliki keterbatasan, terutama dalam cakupan data dan variabel yang digunakan. Oleh karena itu, pengembangan lebih lanjut diharapkan dapat mencakup data tambahan serta metode yang lebih canggih untuk meningkatkan akurasi dan relevansi hasil. Semoga hasil penelitian ini dapat menjadi referensi yang bermanfaat bagi pemerintah, peneliti, maupun masyarakat umum dalam memahami dinamika pandemi COVID-19 dan mengembangkan strategi yang lebih efektif dalam menghadapi krisis serupa di masa depan.

Daftar Pustaka

Amna, Wahyuddin S, I Gede Iwan Sudipa, Tri Andi E. Putra, Ahmad Junaidi Wahidin, Wara

Alfa Syukrilla, Anindya Khrisna Wardhani, Nono Heryana, Tutuk Indriyani, Leo

Willyanto Santoso. *DATA MINING*. Koto Tengah, PT GLOBAL EKSEKUTIF

TEKNOLOGI, 2023,

[https://repository.uinjkt.ac.id/dspace/bitstream/123456789/71093/1/DATA%20MININ](https://repository.uinjkt.ac.id/dspace/bitstream/123456789/71093/1/DATA%20MINING.pdf)

G.pdf. Accessed 1 11 2024.

“COVID-19 Indonesia Dataset.” *Kaggle*,

<https://www.kaggle.com/datasets/hendratno/covid19-indonesia>. Accessed 22

December 2024.

Lampiran

Nama	Tugas
Nevin Athalla Sukandar	Dashboard Python, Mockup, Laporan dan PPT
Ferdi Andra Maulana	Data Cleaning, Laporan dan PPT
Syahnaulval Dhiaulhaq Ridwan	Program Jupyter, Laporan, dan PPT
Adit Firmansyah	Laporan dan PPT

Link :

- Kaggle (dataset) : <https://www.kaggle.com/datasets/hendratno/covid19-indonesia>
- Presentasi : https://www.canva.com/design/DAGaBcaLgX4/hEIuREj2_cGlsCVQbmRtzw/edit
- Dashboard : <https://tubesdaming-kelompok-2.streamlit.app/>