

# Bogor Traffic Jam Clustering

**Ichfan Kurniawan**

Sebagai seorang yang sering ke Bogor, kemacetan di jalanan kota menjadi hal yang biasa dilalui.

### Bogor Jadi Kota Termacet Kelima di Indonesia, Begini Respons Bima Arya

M. Sholihin - detikNews

Rabu, 12 Jan 2022 15:20 WIB



Potret kemacetan di Kota Bogor. (Foto: M. Sholihin/detikcom)

### Pemkot Sebut 4 Alasan Bogor Jadi Kota Langanan Macet

CNN Indonesia

Selasa, 29 Des 2020 17:19 WIB

Bagikan :  



Kondisi parkir yang tidak tertata menjadi satu dari empat alasan Kota Bogor selalu dipadati wisatawan. Ilustrasi. (CNN Indonesia/Bisma Septalisma).

# Bank Dunia: Indonesia Rugi Rp 56 Triliun Per Tahun Akibat Kemacetan



Oleh **Agatha Olivia Victoria**

3 Oktober 2019, 17:52



ANTARA FOTO/DHEMAS REVIYANTO

Ilustrasi kemacetan. Bank Dunia memproyeksi kemacetan di Indonesia menimbulkan kerugian mencapai Rp 56 triliun per tahun.

Tidak hanya rugi secara tenaga dan moral, namun kerugian juga dapat dialami secara material akibat adanya sebuah kemacetan.

Bayangkan apabila kita dapat mengetahui setiap karakteristik dari kemacetan di tiap-tiap jalan kota di Bogor, tentu akan lebih mudah bukan untuk merencanakan setiap perjalanan kita?

Setiap perjalanan yang akan dilakukan akan lebih ter-planning dengan lebih baik.



Dalam mini project Data Scientist kali ini, akan dilakukan clustering dari setiap traffic jam yang terjadi di jalanan kota Bogor.

Sedemikian sehingga kita dapat mengetahui karakteristik dari tiap-tiap jalan yang sering terjadi kemacetan.



# Overall Process

1. Data Understanding
2. Data Preprocessing
3. Modeling
4. Evaluation
5. Deployment

```
1 jans.info()
2
3 # Irregular df
4 # parse time
5 # fillna street
6 # check if date same with time

[50]
...
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 102322 entries, 0 to 102321
Data columns (total 14 columns):
 #   Column              Non-Null Count  Dtype
---  ---
0   Unnamed: 0          102322 non-null  int64
1   time                102322 non-null  object
2   kemendagri_kabupaten_kode  102322 non-null  float64
3   kemendagri_kabupaten_nama  102322 non-null  object
4   street              100929 non-null  object
5   level               102322 non-null  int64
6   median_length       102322 non-null  float64
7   median_delay         102322 non-null  float64
8   median_speed_kmh     102322 non-null  float64
9   total_records        102322 non-null  int64
10  id                   102322 non-null  int64
11  date                 102322 non-null  object
12  median_level         102322 non-null  float64
13  geometry              102322 non-null  object
dtypes: float64(5), int64(4), object(5)
memory usage: 10.5+ MB
```

	A	B	C	D
1	Field Name	Type	Description	Value Example
2	time	timestamp	Time (every hour)	2020-01-13 9:00:00
3	kemendagri_kabupaten_kode	varchar	Kemendagri code of cities	32,71
4	kemendagri_kabupaten_nama	varchar	Kemendagri name of cities	KOTA BOGOR
5	street	text	Street name	N11 Raden KH Abdullah bin Nuh
6	level	int	Traffic congestion level (median)	5
7	median_length	float	Jam length in meters (median)	17,6
8	median_delay	float	Delay of jam (in seconds) compared to free flow speed (in case of block, -1) (median)	11,2
9	median_speed_kmh	float	Current median speed on jammed seg	78,8
10	total_records	int	Total data recorded in a given time	10
11	id	serial	Row ID	500 342
12	date	date	Date	2020-01-13
13	geometry	geometry	Geometry data type (spatial data)	MULTILINESTRING( ... )
14				
15				
16				
17				
18				

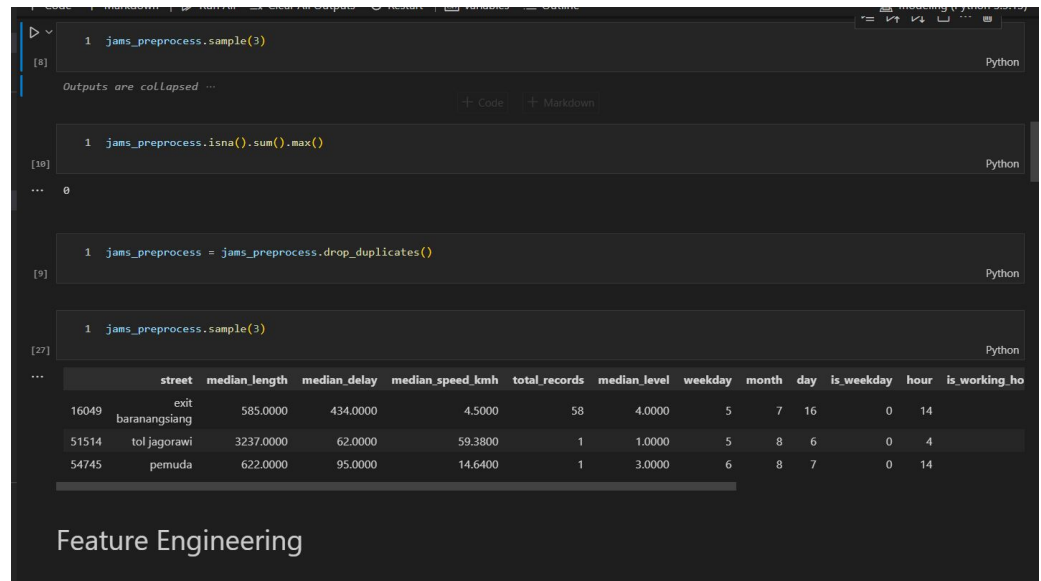
Dalam tahapan data understanding, dilakukanlah pemahaman mengenai garis besar struktur dari data (shape, datatype, jumlah missing values, duplicates) dan memahami makna di tiap-tiap kolom dataset dari data dictionary.

Tahapan ini juga meliputi pemahaman secara non-visual dan visualisasi.

Output dari tahapan ini akan menjadi masukan dalam tahapan data preprocessing.

# Overall Process

1. Data Understanding
- 2. Data Preprocessing**
3. Modeling
4. Evaluation
5. Deployment



The screenshot shows a Jupyter Notebook interface with the following code cells:

- Cell [8]: `jams_preprocess.sample(3)`
- Cell [10]: `jams_preprocess.isna().sum().max()`
- Cell [9]: `jams_preprocess = jams_preprocess.drop_duplicates()`
- Cell [27]: `jams_preprocess.sample(3)`

Below the code cells, a table displays the first three rows of the sampled data:

	street	median_length	median_delay	median_speed_kmh	total_records	median_level	weekday	month	day	is_weekday	hour	is_working_ho
16049	exit baranangsiang	585.0000	434.0000	4.5000	58	4.0000	5	7	16	0	14	
51514	tol jagorawi	3237.0000	62.0000	59.3800	1	1.0000	5	8	6	0	4	
54745	pemuda	622.0000	95.0000	14.6400	1	3.0000	6	8	7	0	14	

Below the table, the text "Feature Engineering" is displayed.

Setelah mendapatkan pemahaman yang cukup dari tahapan 1, dilakukanlah preprocessing data yang meliputi:

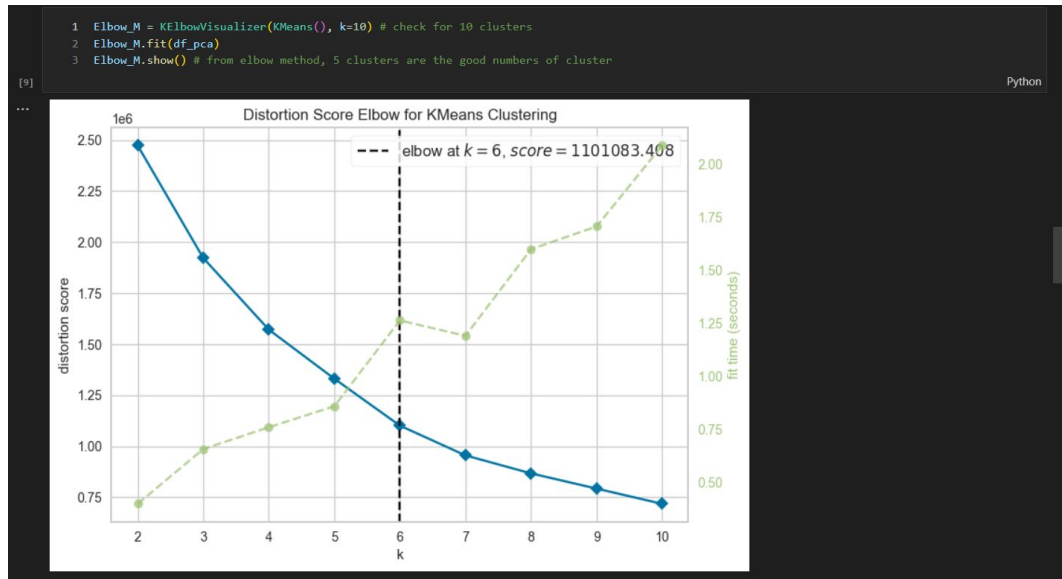
- Mengecek data lewat statistical descriptive
- Menangani missing value
- Menangani duplicate value
- Menangani outlier

Terakhir dilakukanlah feature engineering, untuk mendapatkan feature turunan dari kolom utama dataset.



# Overall Process

1. Data Understanding
2. Data Preprocessing
- 3. Modeling**
4. Evaluation
5. Deployment



Setelah mendapatkan preprocessed data yang memadai dilakukanlah pemodelan menggunakan metode k-means clustering.

Pada trial 1, diinginkan granularity pada level per jalan, namun didapatkan data yang sparse. Sedemikian sehingga hasil clustering juga sparse

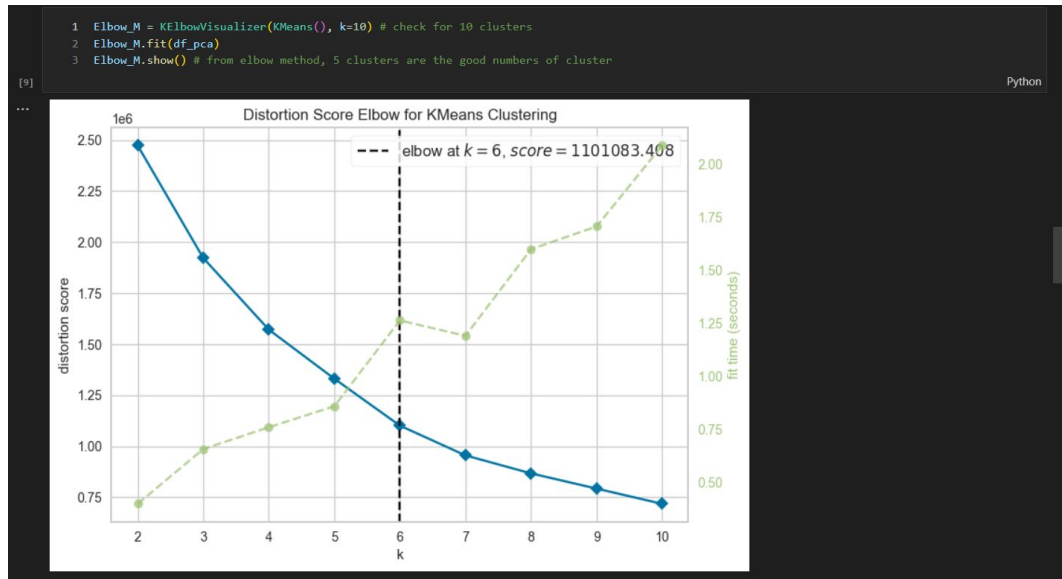
Pada trial 2, digunakanlah granularity per tiap event dari traffic jam dan mendapatkan hasil yang lebih baik dari trial 1.

Pemilihan nilai k, didapatkan dari metode elbow.



# Overall Process

1. Data Understanding
2. Data Preprocessing
3. Modeling
- 4. Evaluation**
5. Deployment



Dalam hal pemilihan nilai elbow dilakukanlah trial 2 kali,

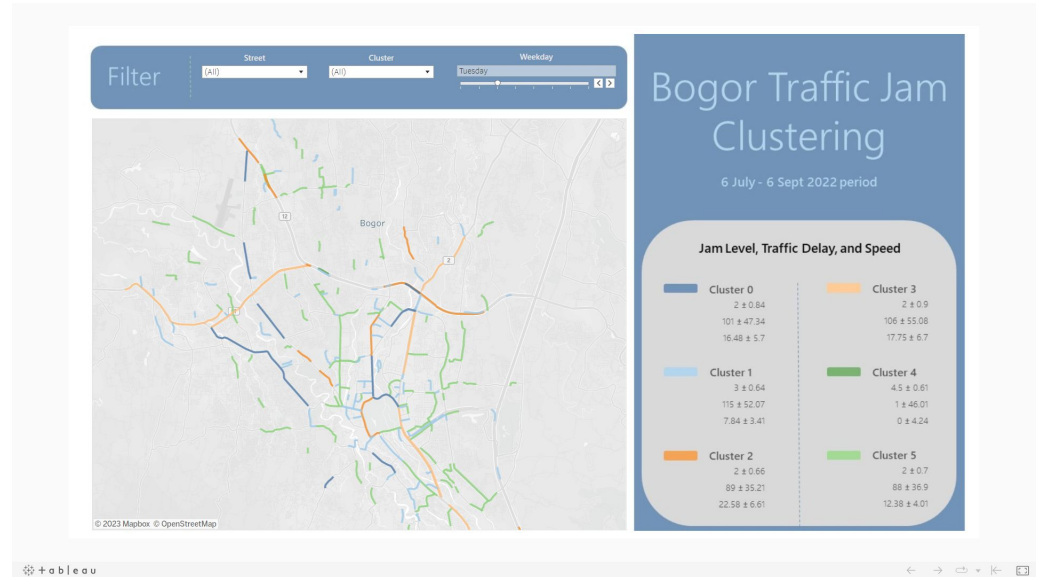
Pada trial 1, dengan menggunakan nilai  $k = 6$

Pada trial 2, dengan menggunakan nilai  $k = 3$

Dari 2 trial tersebut diambil nilai  $k = 6$ , dikarenakan pada nilai  $k = 3$  tidak terjadi perbedaan yang cukup signifikan diantara kluster yang terbentuk.

# Overall Process

1. Data Understanding
2. Data Preprocessing
3. Modeling
4. Evaluation
5. **Deployment**



Hasil dari clustering disajikan dalam dashboard public Tableau pada [link berikut](#).

Dengan harapan publik, utamanya yang sering melewati kota Bogor, dapat mengakses dengan bebas.

# Bogor Traffic Jam Clustering

**Ichfan Kurniawan**



# Bogor Traffic Jam Clustering

**Ichfan Kurniawan**

## Referensi

<https://news.detik.com/berita-jawa-barat/d-5894932/bogor-jadi-kota-termacet-kelima-di-indonesia-begini-respons-bima-arya>

<https://www.cnnindonesia.com/ekonomi/20201229133443-92-587383/pemkot-sebut-4-alasan-bogor-jadi-kota-langganan-macet>

<https://katadata.co.id/agustiyanti/finansial/5e9a4e616e420/bank-dunia-indonesia-rugi-rp-56-triliun-per-tahun-akibat-kemacetan>

<https://www.freepik.com/>