

# Mini Project Data Science DSLS 2023

Siska Puspita Juita

Kelompok 1A

# Business Understanding

Pada era modern seperti saat ini, transportasi merupakan salah satu kebutuhan penting bagi masyarakat dalam mempercepat dan melancarkan kegiatan sehari-hari. Kemajuan ilmu pengetahuan dan teknologi mendukung perkembangan alat transportasi secara pesat, sehingga menimbulkan peningkatan laju pertumbuhan kendaraan. Perkembangan kendaraan sebagai alat transportasi membawa dampak dalam memenuhi dan meningkatkan kesejahteraan manusia, seperti sebagai mobilisasi dalam memperlancar semua aktivitas bisnis. Namun, hal tersebut juga menimbulkan dampak negatif seperti meningkatnya kemacetan lalu lintas.

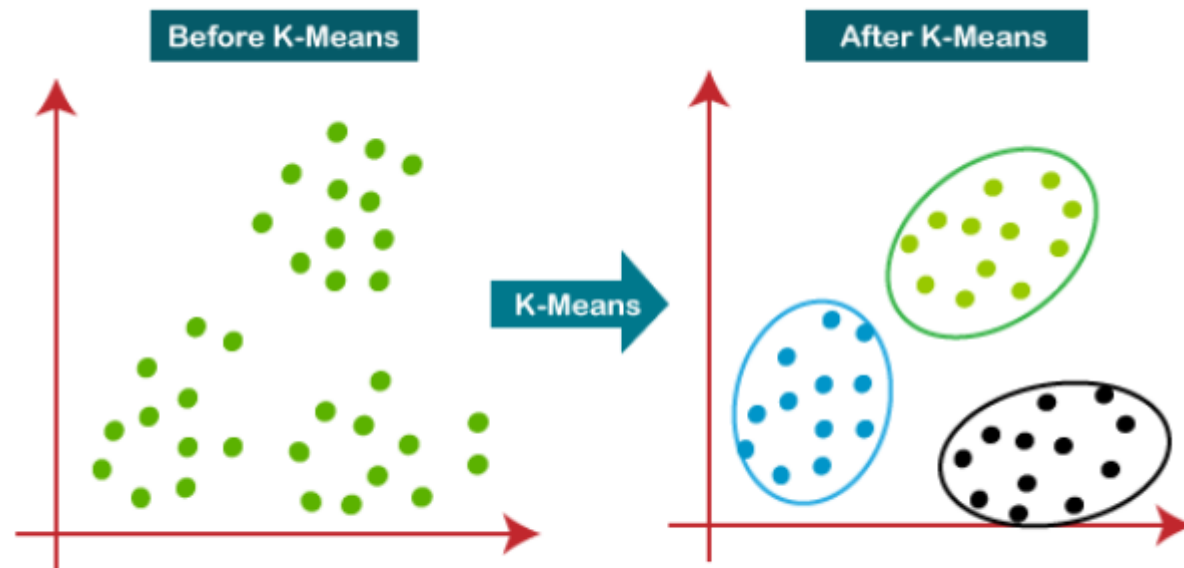
Inrix mengumpulkan data dari 1.360 kota di 38 negara yang mencakup lebih dari 250.000 kilometer persegi jalan dan berfokus pada kemacetan di sepanjang hari dan minggu. Masyarakat Indonesia rata-rata menghabiskan waktu sia-sia di jalan karena macet sampai 47 jam dalam satu tahun. Khususnya Jakarta, kota ini termacet ke-12 dengan waktu yang dibuang pengemudi ketika macet mencapai 63 jam dalam setiap tahun. Berdasarkan kondisi tersebut, maka perlu dilakukan cluster analysis untuk menetapkan tingkat potensi mengalami kemacetan. Kemajuan machine learning memungkinkan akurasi cukup tinggi untuk mengklaster kemacetan.

Waze adalah sebuah aplikasi navigasi yang dapat membantu pengguna dalam merencanakan perjalanan. Dalam proyek kali ini, akan digunakan dataset yang disediakan oleh Waze untuk mengetahui tingkat kemacetan di beberapa daerah, seperti Bandung, Bekasi, Depok, Bogor, dan Tasikmalaya.

# K-Means Clustering

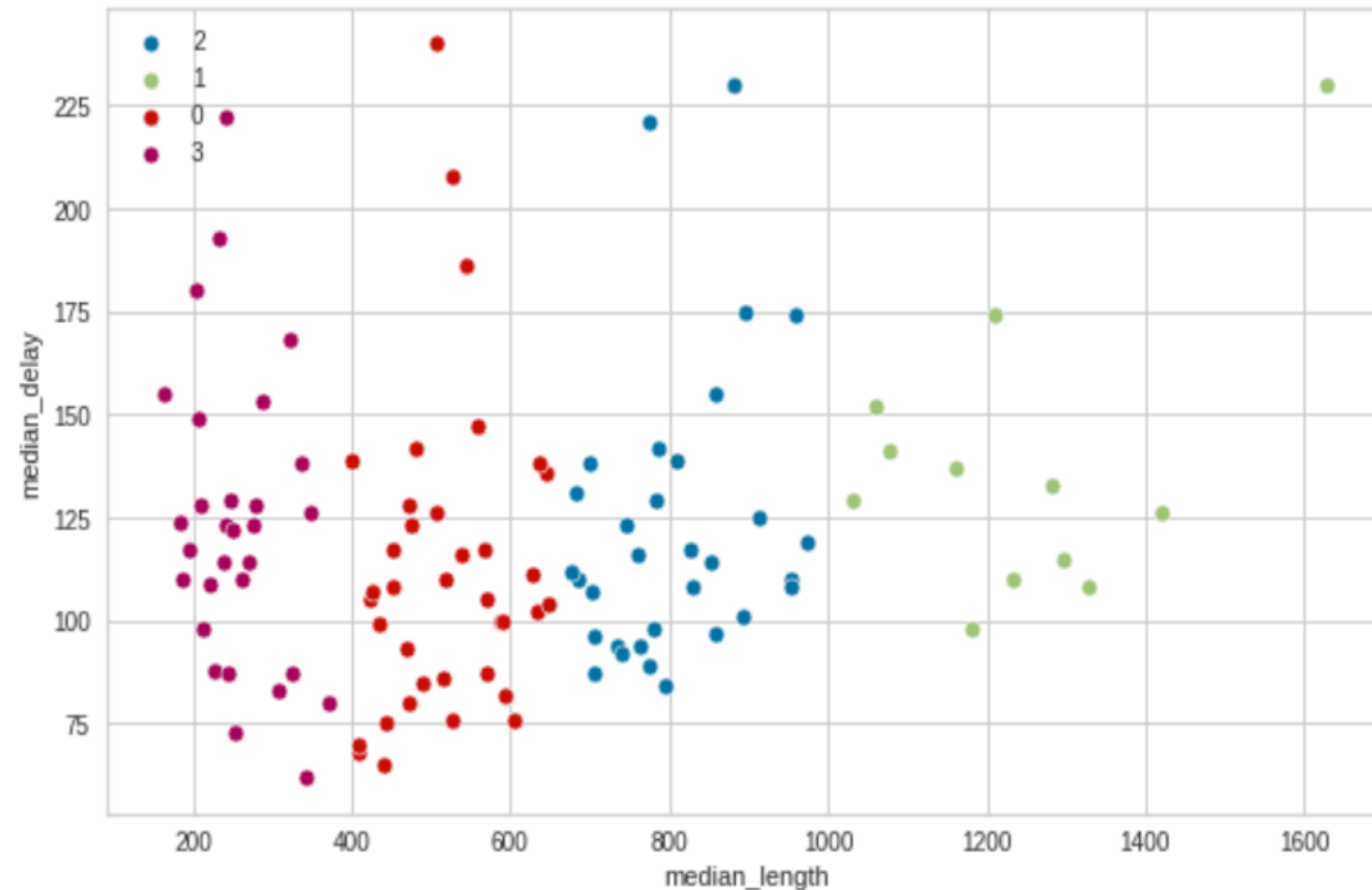
K-Means clustering adalah algoritma unsupervised learning yang dipakai untuk mengelompokkan dataset yang belum dilabel ke dalam kluster yang berbeda. Simbol K pada K-means clustering menandakan jumlah kluster yang digunakan.

Diagram di bawah ini adalah ilustrasi cara kerja Algoritma Clustering K-means



Sumber: [javatpoint.com](http://javatpoint.com)

# Cluster



## Cluster 0

Cluster 0 adalah cluster yang memiliki waktu dan panjang kemacetan yang pendek.

Terdapat 37 jalan di Sukabumi masuk dalam cluster 0.

## Cluster 2

Cluster 2 adalah cluster yang memiliki waktu dan panjang kemacetan yang pendek hingga menengah kebawah.

Terdapat 32 jalan di Sukabumi masuk dalam cluster 2.

## Cluster 1

Cluster 1 adalah cluster yang memiliki waktu dan panjang kemacetan menengah hingga menengah keatas.

Terdapat 12 jalan di Sukabumi masuk dalam cluster 1.

## Cluster 3

Cluster 3 adalah cluster yang memiliki waktu dan panjang kemacetan yang paling lama.

Terdapat 30 jalan di Sukabumi masuk dalam cluster 3.

# Evaluation

Model mendapatkan nilai DB Score

DB Score K-Means : 0.52

DB Score Agglom : 0.59

Dimana angka tersebut mengarah ke nilai 1, sehingga dapat dikatakan bahwa model tersebut memiliki performa yang baik.