

Mini Project DS

Mini Project ini ditujukan untuk mengasah kemampuan peserta DSLS 2023 terkait ***data science***. Mohon untuk memperhatikan panduan ini dengan saksama.

**Latar Belakang Projek**

**Tools dan Dataset**

**Tools**

**Dataset**

**Petunjuk Pengerjaan**

Bagian 1

Bagian 2

Bagian 3

**Keluaran Projek**

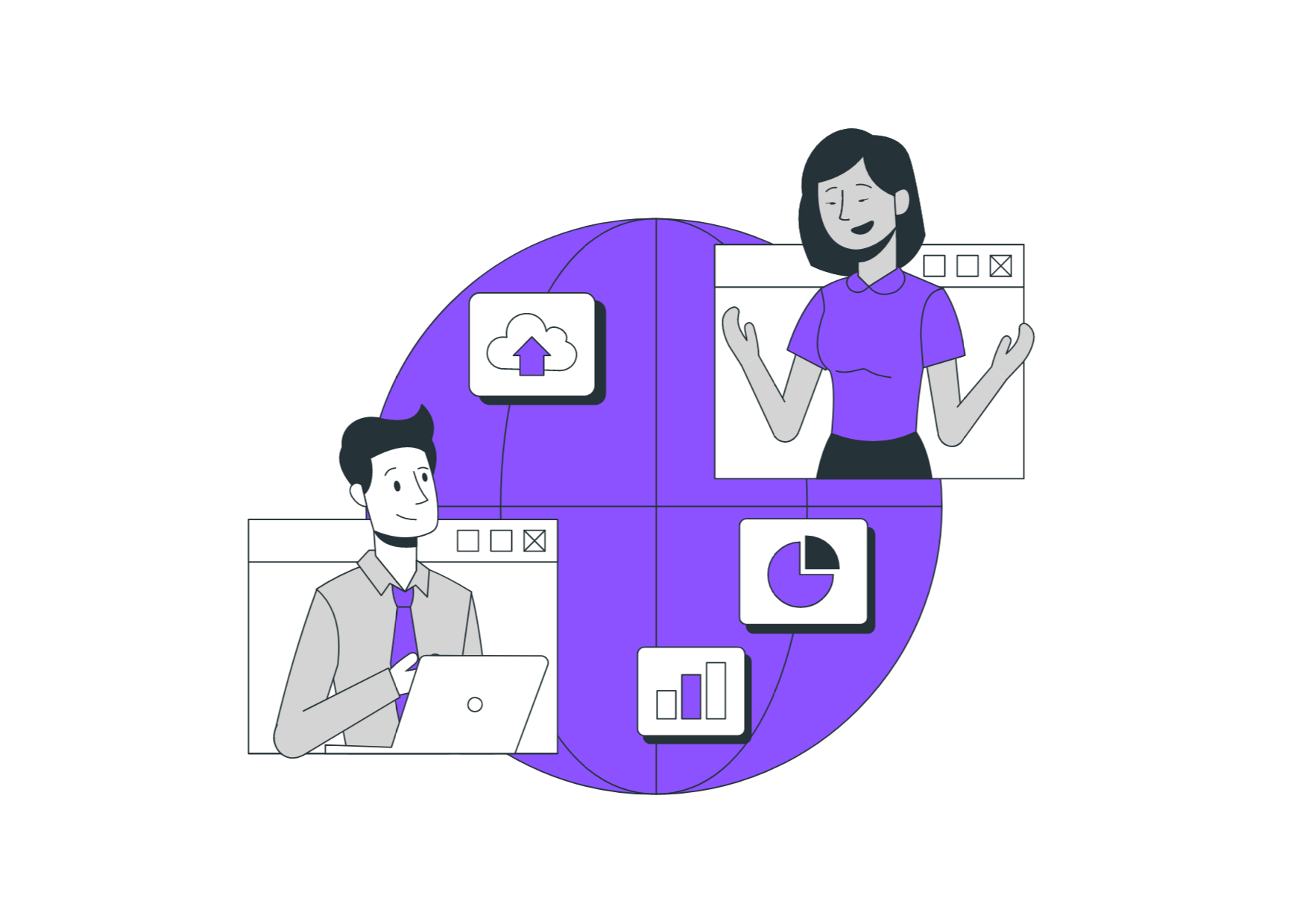
Dokumentasi

Report

**Penilaian Projek**

Mini project ini bersifat **INDIVIDU**.

# Latar Belakang Projek



Tanggung jawab utama seorang *data consultant* adalah membantu perusahaan maupun pemangku kebijakan untuk membuat keputusan berdasarkan data dan mengoptimalkan penggunaan data. Untuk mewujudkan tanggung jawab tersebut, salah satu kemampuan yang harus dikuasai oleh *data consultant* yaitu mampu melakukan pemodelan menggunakan *machine learning*.

Umumnya, *machine learning* terbagi menjadi dua: *supervised learning* dan *unsupervised learning*. *Supervised learning* merupakan cabang *machine learning* dengan tujuan untuk melakukan prediksi pada data yang sudah diberi label. Sedangkan *unsupervised learning* digunakan untuk menemukan pola pada suatu dataset. Pada projek ini, Anda akan berlatih menggunakan *machine learning* untuk memperoleh *insight* yang dapat digunakan sebagai acuan dalam pengambilan keputusan.

# Tools dan Dataset



Tugas modeling menggunakan *machine learning* (baik supervised maupun unsupervised learning atau yang tidak termasuk ke dalam dua termilogi terebut) dilakukan menggunakan bahasa pemrograman python atau R pada editor **Visual Studio Code**. Pastikan python dan R sudah terinstall pada perangkat Anda.



Dataset yang digunakan pada projeck ini adalah *crowdsourced/user generated* data program [Waze for Cities](https://www.waze.com/wazeforcities/) dari Waze. Terdapat tiga tabel yang dapat Anda gunakan dalam mengerjakan *case study*, yaitu aggregate\_median\_jams, aggregate\_median\_irregularities, dan aggregate\_alerts. Tabel-tabel tersebut merupakan hasil agregat dari *raw data* yang dirilis oleh Waze (jams, alerts, dan irregularities). Ketiga tabel agregat tersebut memiliki rentang waktu 6 Juli 2022 hingga 6 September 2022 dan dapat diakses di [sini](https://drive.google.com/drive/folders/1612sKseH8t3Pc1Mb9fTYfbV2VnbSAvGn?usp=share_link).

Anda juga dapat menggunakan *supporting data* untuk mengerjakan *case study*, misalnya [data geospasial](https://opendata.jabarprov.go.id/id/dataset), data [Humanitarian Data Exchange (HDX) terbitan Facebook](https://dataforgood.facebook.com/dfg/docs/high-resolution-population-density-maps-demographic-estimates-documentation), [data sinyal internet](https://github.com/teamookla/ookla-open-data), *weather data* (salah satunya [OpenWeather](https://openweathermap.org/api)) untuk mendukung analisis maupun modeling Anda.

# Petunjuk Pengerjaan

**Case Study**

Pilih satu *use case* yang akan menjadi *portfolio* Anda menggunakan data Waze. Beberapa pilihan *use case* sebagai berikut, akan tetapi Anda juga dipersilakan untuk mengeksplor *use case* lain sesuai dengan minat Anda:

* *Forecasting* (ex: memprediksi lalu lintas pada *weekend* di Kota Bogor)
* *Prediction* (ex: traffic flow prediction, congestion prediction)
* *Clustering* (ex: clustering jalan berdasarkan delay time)

Lakukan *modeling* dengan menggunakan beberapa algoritma, kemudian lakukan *hyperparameter tuning* dan periksa performa model berdasarkan metrik yang Anda telah pelajari!

# Keluaran Projek (Deliverables)

Dokumentasi

* Satukan code dan *report* dalam satu folder. Pastikan code Anda dilengkapi comment (*clean code*) sehingga dapat dibaca dengan jelas.
* Masukkan satu folder tersebut ke Github sesuai [panduan](https://bit.ly/GitHubGuide).
* Silakan kumpulkan link Github pada form [ini](https://bit.ly/SubmissionMiniProject).

Report

* **Case Study**, yang merangkum:
  + Business Understanding
  + Data Understanding
  + Data Cleansing and Preprocessing (termasuk uji statistik, jika menggunakan)
  + Modeling (merangkum tentang penjelasan model, performa model, dan hyperparameter tuning)
  + Evaluation (melakukan evaluasi pada model)
  + Deployment, *optional*
  + Referensi
  + Buat dengan format PPT dan masukkan dalam satu link Github dengan code. Bayangkan PPT ini digunakan sebagai paparan kepada para penentu kebijakan publik (Dishub, Gubernur, Walikota, Satpol PP), perusahaan penyedia jasa layanan transportasi umum, *urban planner* atau perusahaan logistik.

# Penilaian Projek

1. Business Understanding 5%
2. Data Understanding 10%
3. Data Cleansing and Preprocessing 20%
4. Modelling 35%
5. Evaluation 15%
6. Dokumen Report 15%

# Referensi

[Incorporating User-Generated Waze Data and Machine Learning into Traffic Analysis: A Case-Study in Louisville, Kentucky](https://www.tesc.psu.edu/assets/docs/user-generated-waze-data.pdf)

[Quality of location-based crowdsourced speed data on surface streets: A case study of Waze and Bluetooth speed data in Sevierville, TN](https://www.researchgate.net/publication/342967447_Quality_of_location-based_crowdsourced_speed_data_on_surface_streets_A_case_study_of_Waze_and_Bluetooth_speed_data_in_Sevierville_TN)

[Big Data Analytics and Prediction of Traffic in Los Angeles](https://koreascience.kr/article/JAKO202011161036290.view?orgId=anpor&hide=breadcrumb,journalinfo)

[Menilik Prediksi Arus Mudik Idulfitri 2022 di Jawa Barat](https://opendata.jabarprov.go.id/id/artikel/menilik-prediksi-arus-mudik-idulfitri-2022-di-jawa-barat)

[Menilik Mudik 2022: Jumlah Pemudik Melesat, Lalu Lintas Tersendat](https://opendata.jabarprov.go.id/id/artikel/menilik-mudik-2022-jumlah-pemudik-melesat-lalu-lintas-tersendat)

[Menilik Tren Mudik Sebelum dan Ketika Pandemi](https://opendata.jabarprov.go.id/id/artikel/menilik-tren-mudik-sebelum-dan-ketika-pandemi)

[Waze for Cities Case Study](https://www.waze.com/wazeforcities/casestudies/)

Selamat mengerjakan ya teman-teman 🙂.

Jika ada pertanyaan mengenai mini project, silakan tanyakan di discord thread **#mini-project**