

## PROJECT DM 2023

1. Pilih salah satu topik:

- industri otomotif / industri 4.0 (manufaktur) / smart factory
- bisnis otomotif / terkait otomotif
- smart / autonomous vehicle

Pilih salah satu task: Klasifikasi, Regresi, Forecasting, Clustering, ARM

2. Dataset harus fresh. Bukan dataset dari Kaggle, Weka, atau dataset lain yang sudah sering digunakan (Titanic, iris flower, PIMA diabetes, etc). Dataset harus terkini (misal data yang direkam dalam tahun 2023. Jangan data jadul seperti tahun < 2017. Boleh menggunakan data lawas asalkan tetap berjalan sampai 2023, misal: data penjualan mobil 2010-2023). Dataset tidak boleh sama antar kelompok. Data harus riil, bukan dummy. Jumlah data minimal 100.

3. Gunakan metode2 DM. Tidak harus yang diajarkan di kelas.

4. Simpan model dan deploy dalam sebuah aplikasi web (Streamlit, Flask, Django / lainnya). Akurasi model minimal 75% (klasifikasi). Jika regresi, MAPE max 10%.

Task	Kemampuan aplikasi
Klasifikasi, Regresi	Menampilkan tabel data Visualisasi data Memprediksi dan menampilkan hasil prediksi
Clustering	Menampilkan tabel data Visualisasi data Visualisasi hasil cluster "Memprediksi" sebuah data baru termasuk di cluster yang mana (untuk kasus profiling) Keterangan tindak lanjut untuk masing2 cluster
ARM	Menampilkan tabel data Visualisasi data Menampilkan aturan2 yang ditemukan dan metriknya (confidence, support, lift) Memberikan rekomendasi item jika diketahui seseorang (telah) menggunakan item / itemset tertentu

5. Buat laporan singkat dengan panjang 8-12 halaman. Sampaikan hanya hal yang penting. Isi laporan:

- Judul (tujuan, metode, objek). Contoh: Forecasting permintaan oli Federal di Bengkel X menggunakan XGBOOST.
- Latar belakang masalah dan tujuan
- Data: Cara mendapatkan data. Penjelasan mengenai data. Contoh (sampel) data. Data pre-processing (jika ada).
- Metode: Penjelasan setting parameter dari metode DM yang digunakan.
- Hasil. Pembahasan akurasi model. Contoh hasil prediksi. Pembahasan kelebihan dan kekurangan model yang telah dibuat. Screenshot program.
- Pola yang terlihat / insight yang dihasilkan / informasi berguna yang didapatkan.
- Kesimpulan. Kelebihan dan kekurangan / keterbatasan. Saran perbaikan berdasarkan kekurangan / keterbatasan.
- Referensi: daftar referensi (URL) sumber data, code, tutorial, artikel, jurnal, dsb yang membantu / menginspirasi Anda dalam penyelesaian project ini

#### 6. Upload di Drive:

<https://drive.google.com/drive/folders/1B3figlZRkELawslzWqh-OJ-l4iQtYTPj?usp=sharing>

Daftar file yang di-upload:

- Code untuk (data-preprocessing dan) training model (IPYNB)
- Data (CSV / ZIP)
- Model (H5 / PKL / SAVEDMODEL, ETC)
- Code program (PY) dan resources lain yang dipakai, misal: gambar, javascript, CSS.
- requirements.txt (file text berisi daftar library yang digunakan)