

## Project-based assignment - Pembelajaran mesin Semester ganjil 2022/2023

### Ketentuan Umum:

1. Project-based assignment diberikan untuk memberi kesempatan kepada mahasiswa untuk mengeksplor pembelajaran mesin menggunakan metode-metode ensemble (bagging dan boosting) baik untuk keperluan klasifikasi maupun regresi pada datasets dari dunia nyata.
2. Project-based assignment dikerjakan secara berkelompok:
  - a. 3-4 orang per kelompok untuk kelas regular/internasional;
  - b. 3 orang per kelompok untuk kelas PJJ.
3. Setiap kelompok harus mematuhi etika plagiasi, baik ide, kode, maupun laporan. Segala bentuk pelanggaran akan diberikan sanksi berat, yaitu nilai E untuk mata kuliah Pembelajaran Mesin.
4. Batas akhir (*deadline*) pengumpulan tugas adalah 8 Januari 2023 pukul 23:59 WIB.
5. Presentasi tugas dilakukan menggunakan mode *asynchronous* (melalui video presentasi) saja atau ditambah juga dengan presentasi *synchronous* (waktu nyata di hadapan dosen pengampu) sesuai dengan kebijakan masing-masing dosen pengampu.
6. Tugas dikumpulkan melalui LMS kelas masing-masing.

### Deskripsi Tugas:

1. Setiap kelompok mengerjakan *salah satu* tipe tugas di antara empat tipe tugas yang ada yang merupakan empat kombinasi nilai yang berbeda dari dua variabel tugas yaitu *variable keperluan pembelajaran* (regresi atau klasifikasi) dan variabel *metode yang digunakan* (bagging atau boosting) sebagaimana yang ditunjukkan pada tabel berikut.

	Regresi (dataset: autos MPG)	Klasifikasi (dataset: German credit)
Bagging	Tugas tipe 0	Tugas tipe 1
Boosting	Tugas tipe 2	Tugas tipe 3

Tipe tugas yang dikerjakan oleh suatu kelompok diberikan oleh **NIM terkecil di kelompok tsb modulo 4**. Contoh: NIM terkecil di Kelompok X adalah 101373635. Karena 101373635 modulo 4 adalah 3, kelompok X akan mengerjakan tugas tipe 3 yaitu tugas klasifikasi pada dataset German credit menggunakan metode boosting.

Tugas regresi adalah tugas memprediksi tingkat kehematan bahan bakar kendaraan/MPG (*miles per gallon*: rata-rata jarak tempuh mobil dalam mil untuk setiap galon bahan bakar yang dikonsumsi) berdasarkan profil mobil yang diberikan yang diwakili oleh atribut-atribut seperti silinder, daya (tenaga kuda), tahun keluaran, dll. Dataset autos MPG beserta deskripsi atribut-atributnya bisa diakses di tautan berikut (login SSO):

<https://drive.google.com/drive/folders/1HmKavcNCij76k02nCYQdP012cepM2nmz?usp=sharing>

Tugas klasifikasi adalah tugas menebak kelas/kategori dari calon kreditur (apakah akan menjadi kreditur yang baik atau buruk) berdasarkan profil calon kreditur yang diberikan yang diwakili oleh atribut-atribut-atribut seperti status pekerjaan, status perkawinan, tujuan kredit, usia, jenis kelamin, dll. Dataset German credit beserta deskripsi atribut-atributnya bisa diakses di tautan berikut (login SSO):

[https://drive.google.com/drive/folders/1P7DBFKFolr1CCZbJp6uFg9hCsjMb7Ssk?usp=share\\_link](https://drive.google.com/drive/folders/1P7DBFKFolr1CCZbJp6uFg9hCsjMb7Ssk?usp=share_link)

2. Untuk setiap tipe tugas, mahasiswa diminta melakukan enam hal berikut:
  - a. Formulasi masalah: jelaskan apa permasalahan yang akan diselesaikan.
  - b. Eksplorasi dan pra-pemrosesan data (termasuk *data splitting*): lakukan semua teknik eksplorasi dan pra-pemrosesan data yang menurut Anda perlu dilakukan. Jelaskan mengapa perlu melakukan teknik tersebut, dan lakukan analisis terhadap hasilnya.
  - c. Pemodelan: bangunlah model menggunakan data hasil pra-pemrosesan di poin (b), dan lakukan proses pembelajaran mesin untuk mendapatkan hasil terbaik. Jelaskan secara detail semua proses yang Anda lakukan dilengkapi dengan justifikasi dan analisis hasilnya.
  - d. Evaluasi: pilih metode evaluasi yang sesuai beserta justifikasinya. Lakukan evaluasi terhadap model yang telah dihasilkan. Berikan analisis terhadap hasil evaluasi.
  - e. Eksperimen: lakukan berbagai eksperimen yang melibatkan tahapan eksplorasi dan pra-pemrosesan data, pemodelan, dan evaluasi untuk mendapatkan hasil terbaik. Laporkan semua eksperimen yang Anda lakukan beserta analisis hasil dan perbandingannya.
  - f. Kesimpulan: berikan kesimpulan dari semua proses yang dijalankan beserta hasil akhir dari berbagai eksperimen yang telah dilakukan.
3. Dalam pengerjaan tugas ini, mahasiswa **diperbolehkan** menggunakan library.
4. Bahasa pemrograman yang digunakan adalah **Python**. Mahasiswa **tidak diperbolehkan** menggunakan bahasa pemrograman lain.
5. Pengumpulan tugas untuk setiap kelompok **diwakili oleh salah seorang anggota saja**, yaitu mahasiswa dengan NIM terkecil di kelompok tsb.
6. Berikut artifak-artifak tugas yang wajib dikumpulkan:
  - a. Laporan dalam sebuah file pdf yang diberi nama "Laporan tugas tipe [tipe\_tugas] kelompok [no\_kelompok].pdf";
  - b. *Notebook* (.ipynb) yang siap dieksekusi di Google Colab. Sertakan tautan *notebook* di Google Colab di dalam laporan pada poin (a);
  - c. Video presentasi dengan durasi minimal 10 menit yang diunggah pada akun Youtube. Sertakan tautan video presentasi di dalam laporan pada poin (a);

**Penilaian:**

1. Laporan (30%):
  - a. Kelengkapan bagian laporan (5%);
  - b. Kesesuaian laporan dengan hasil hasil luaran *source code* (5%);
  - c. Analisis pemilihan teknik yang digunakan beserta interpretasi hasil (20%).
2. *Source code* (40%):
  - a. Ketepatan solusi beserta penjelasannya (35%). *Source code* wajib diberikan komentar sebaik-baiknya supaya mudah dipahami oleh dosen dan asisten dosen;
  - b. Performansi model yang dihasilkan (5%).
3. Presentasi (30%):
  - a. Organisasi konten presentasi (5%);
  - b. Pemahaman tentang proyek (25%).