

Deskripsi Tugas Besar MK Pembelajaran Mesin Lanjut

Ketentuan Umum:

1. Tugas dikerjakan secara berkelompok dua mahasiswa, di mana pembagian kelompok berdasarkan nomor urut presensi. Jika jumlah mahasiswa di suatu kelas adalah ganjil, maka mahasiswa bisa bergabung ke kelompok terakhir atau mengerjakan secara individual (sendirian).
2. Harus mematuhi etika plagiasi baik ide, kode, maupun laporan. Segala bentuk pelanggaran akan diberikan sanksi berat, minimal **nilai E** untuk mata kuliah ini.
3. Deadline pengumpulan tugas adalah **Sabtu, 07 Januari 2023**, pukul **20.00** WIB.

Deskripsi Tugas:

1. *Dataset* yang digunakan dalam tugas besar adalah *dataset* **Breast Cancer Wisconsin (Diagnostic)**.
2. Setiap kelompok harus menyelesaikan sebuah task terhadap *dataset* tersebut dengan menerapkan *Automated Machine Learning*.
3. *Machine Learning Automation Tools* yang boleh digunakan adalah TPOT, Google, atau Microsoft.
4. Bahasa pemrograman yang digunakan adalah Python.
5. Untuk task tersebut, mahasiswa diharapkan setidaknya melakukan hal-hal berikut:
 - a. Formulasi Masalah: mengacu pada 4 tahap *From Problem to ML Solution*.
 - b. Eksplorasi dan Persiapan Data (termasuk split data): Setidaknya menggunakan dua buah teknik eksplorasi dan/atau persiapan data. Jelaskan mengapa menggunakan teknik tersebut.
 - c. Pemodelan: Pilih salah satu dari tiga *Machine Learning Automation Tools* yang akan digunakan. Jelaskan alasan pemilihan tersebut. Lalu, jelaskan dengan detil setiap langkahnya.
 - d. Eksperimen: Wajib melakukan setidaknya sebuah eksperimen dengan menggunakan salah satu tools yang tersedia.
 - e. Evaluasi: Pilih metode evaluasi yang ingin digunakan serta jelaskan alasannya. Lakukan evaluasi terhadap model yang digunakan. Berikan analisis terhadap hasil evaluasi.
 - f. Kesimpulan: Sampaikan kesimpulan dari semua proses yang dijalankan, apa saja lesson learned, dan apa saran untuk improvement ke depan.
6. Delivery tugas berupa:
 - a. Laporan diketik dalam A4 menggunakan times new roman size 12, margin 3cm, space 1.15 (.pdf)
 - b. Kode-kode yang siap dieksekusi (.py/.ipynb)
 - c. *Dataset* hasil eksplorasi

Penilaian:

1. Laporan (30%):
 - a. Sesuai Format (5%)
 - b. Penggunaan Bahasa yang baik dan benar (5%)
 - c. Kelengkapan (5%)
 - d. Akurasi laporan terhadap yang dikerjakan/kode (15%)
2. Kode (40%):
 - a. Kelengkapan semua tahapan (10%)
 - b. Eksekusi kode (20%)
 - c. Kerapian kode, penggunaan fungsi secara tepat, pemberian komentar (10%)
3. Presentasi (30%):
 - a. Organisasi konten presentasi (5%)
 - b. Kemampuan menyampaikan presentasi (5%)
 - c. Pemahaman tentang project (20%)
4. Bonus untuk extraordinary results (20%):
 - a. Melakukan lebih banyak eksperimen yang menarik. Pembuatan kode dengan dokumentasi yang sangat lengkap.
 - b. Penyajian laporan yang menarik, penyertaan ilustrasi yang membantu pemahaman pembaca, analisis yang tajam, usulan yang bermanfaat.
 - c. Presentasi yang menarik disertai penggunaan audio/visual/animasi dan sebagainya yang orisinal.

Dataset:

Dataset dapat diunduh menggunakan akun Telkom University (SSO) di:

<https://drive.google.com/drive/folders/1Rtq5Xcp4TZwDczi3AtWJZn6fBlEjxHAZ?usp=sharing>