Pertemuan 8 — UTS: Tugas Besar

Satu masalah nyata, tiga model klasik, satu protokol evaluasi yang jujur

Hendri Karisma

Semester Ganjil 2026/2027

Program Studi Teknik Informatika, STMIK Tazkia

Alur pertemuan hari ini

1. Tujuan dan kaitan dengan CPMK
2. Bentuk tugas dan aturan dasar
3. Pilihan masalah dan *dataset*
4. Kebutuhan wajib dan nilai lebih
5. Protokol evaluasi dan kebocoran data
6. Repositori, laporan, dan presentasi
7. Rubrik penilaian
8. Integritas akademik dan aturan main
9. Jadwal dan *checklist*
10. Pertanyaan yang sering muncul

Weight dan cakupan

UTS bernilai **35%** nilai akhir, mengukur CPMK-1, CPMK-2, dan CPMK-3, dan mencakup materi pertemuan 1–7: *concept learning* dan *inductive bias*, *decision tree*, *ensemble*, Bayes, Naive Bayes, GMM dan EM, kNN dan *K-Means*, serta SVM dan HMM. Seluruh model yang dinilai adalah model klasik; *neural network* adalah wilayah UAS.

Tujuan

Mengapa UTS berbentuk tugas besar

Di tempat kerja, tidak ada yang meminta Anda menuliskan rumus *entropy* dari ingatan. Pertanyaan yang muncul justru seperti ini:

- Masalah ini layak diselesaikan dengan ML atau cukup dengan aturan?
- Model mana yang dipilih, dan apa buktinya?
- Seberapa yakin kita bahwa angka 87% itu akan bertahan pada data besok?
- Kalau model salah, siapa yang menanggung akibatnya?

Yang Anda hasilkan

Satu studi ML utuh atas satu masalah nyata: perumusan masalah, penyiapan data, minimal tiga model klasik yang dibandingkan secara adil, protokol evaluasi yang dapat diulang, dan pembahasan yang jujur tentang keterbatasannya.

Ringkasan

Accuracy tinggi dengan protokol yang bocor bernilai lebih rendah daripada *accuracy* sedang dengan protokol yang benar.

Kaitan dengan CPMK

CPMK	Kemampuan yang diukur	Bukti di tugas besar
CPMK-1	Peta bidang AI, solusi analitik dan numerik, kompleksitas algoritma	Alasan memakai ML; *training cost* dan *inference cost* tiap model diukur dan dikaitkan dengan notasi kompleksitas
CPMK-2	Pembelajaran induktif, *inductive bias*, *generalization*, *overfitting*	Rumusan T, P, E; *inductive bias* tiap model ditulis; kurva validasi dan selisih *training error* terhadap *validation error*
CPMK-3	Merancang, melatih, dan mengevaluasi model klasik, memilih berdasarkan bukti	Tiga model atau lebih, *tuning hyperparameter* di dalam *cross-validation*, *confidence interval*, uji berpasangan, analisis *error*

Setiap komponen rubrik menunjuk ke salah satu baris tabel ini. Bagian yang tidak punya bukti tertulis di *notebook* atau laporan tidak dapat diberi nilai, sebagus apa pun penjelasan lisannya.

Bentuk tugas dan aturan dasar

Menyambung pengumuman Pertemuan 7

Yang sudah diumumkan

- Kelompok **3–4 orang**
- Tiga *output*: repositori kode (*notebook*), laporan **8–12 halaman**, presentasi **12 menit** ditambah tanya jawab
- Minimal **tiga model pembanding** dari Pertemuan 2–7, salah satunya model sederhana sebagai patokan
- Protokol evaluasi dirinci di Pertemuan 8 (buku ajar Bab 10)
- Judul dan data disetujui **satu pekan sebelum pengumpulan**

Yang dirinci hari ini

- Syarat *dataset* dan contoh masalah
- Sepuluh kebutuhan wajib (W1–W10)
- Bentuk tabel hasil dan templat laporan
- Rubrik lengkap beserta tingkat capaiannya
- Jadwal per minggu dan tenggat
- Aturan alat bantu AI dan integritas

Weight rubrik pada slide Pertemuan 7 tetap berlaku. Slide ini menambahkan rinciannya, tanpa mengubah angka.

Pilihan masalah dan *dataset*

Dua jalur *dataset*

Jalur A — data Indonesia (dianjurkan)

Data terbuka dari sumber Indonesia: Satu Data Indonesia, BPS, portal data provinsi (misalnya Satu Data Jakarta), BMKG, atau korpus teks berbahasa Indonesia yang dirilis terbuka (misalnya IndoNLU).

- Biasanya perlu dibersihkan dan digabung sendiri
- Hasilnya lebih jarang tersedia di internet, sehingga analisis Anda lebih orisinal
- Lisensi dan tanggal unduh wajib dicatat

Jalur B — *dataset* klasik

Dataset acuan dari UCI *Machine Learning Repository*, OpenML, atau Kaggle yang sudah dikenal luas.

- Data lebih bersih, sehingga waktu bisa dipusatkan pada protokol
- Wajib membandingkan hasil Anda dengan minimal satu angka acuan dari makalah atau dokumentasi resmi, lalu menjelaskan selisihnya
- Iris dan Titanic **tidak diterima**: terlalu kecil dan jawabannya terlalu mudah ditemukan

Larangan

Tidak boleh memakai data pribadi sungguhan dari tempat kerja, magang, atau hasil *scraping* yang memuat identitas orang. Data mitra hanya boleh dipakai dengan izin tertulis dan setelah identitas dihapus. Pelanggaran membuat UTS bernilai nol.

Contoh masalah jalur A

Masalah	Sumber data	Jenis	Catatan model
Sentimen ulasan berbahasa Indonesia	Korpus SmSA (IndoNLU)	*classification* teks	TF-IDF lalu Naive Bayes (P5) dan SVM linier (P7); pohon sebagai pembanding
Kategori kualitas udara harian	Data ISPU Satu Data Jakarta	*classification* multikelas	Data berurut waktu: pembagian menurut waktu, jangan acak (Bab 10)
Indeks Pembangunan Manusia kabupaten/kota	Tabel BPS beberapa tahun, digabung	*regression*	*Linear regression* sebagai patokan, kNN dan *random forest*; waspadai multikolinearitas (Bab 10)
Deteksi berita hoaks	Korpus berita hoaks berbahasa Indonesia yang terbuka	*classification* teks	Periksa apakah sumber berita membocorkan *label*
Harga rumah di Jabodetabek	*Dataset* iklan properti terbuka	*regression*	*Outlier* harga dan satuan luas; pertimbangkan log(harga)

Masalah lain boleh diajukan selama memenuhi syarat *dataset* pada slide berikut. Satu *dataset* boleh dipakai paling banyak dua kelompok, dengan pertanyaan yang berbeda; persetujuan diberikan menurut urutan pengajuan.

Contoh jalur B dan syarat *dataset*

Contoh *dataset* klasik yang diterima

- *Adult Census Income*: pendapatan di atas atau di bawah ambang (*classification*, tidak seimbang)
- *Bank Marketing*: nasabah menerima tawaran deposito atau tidak
- *Default of Credit Card Clients*: gagal bayar kartu kredit
- *Wine Quality*: mutu anggur sebagai *regression* atau *classification*
- *SMS Spam Collection*: penyaring *spam* berbasis teks
- *California Housing*: harga rumah per blok sensus (*regression*)

Syarat minimum, berlaku untuk kedua jalur

- Minimal **500 baris** dan **5 *feature*** setelah pembersihan
- *Label* atau target jelas dan tersedia untuk seluruh baris yang dipakai
- Lisensi mengizinkan pemakaian untuk pendidikan
- Tidak memuat identitas orang
- Muat dan dapat dilatih di Google Colab gratis dalam waktu wajar

Data yang terlalu kecil membuat *confidence interval* terlalu lebar untuk membedakan model apa pun. Data di bawah 500 baris boleh diajukan bila ada alasan kuat, misalnya tabel kabupaten/kota yang memang hanya berisi 514 baris.

Kartu data: halaman pertama laporan Anda

Isian	Contoh
Nama dan sumber	ISPU DKI Jakarta, Satu Data Jakarta
Lisensi	sebutkan persis
Tanggal unduh	12 Oktober …
Satuan observasi	satu stasiun pada satu hari
Ukuran	*n* baris, *d feature*
Target	kategori ISPU (5 kelas)
Distribusi target	persentase per kelas
Nilai hilang	kolom mana, berapa persen
Rentang waktu	tanggal awal sampai akhir
Potensi bias	stasiun mana yang kurang terwakili

Mengapa kartu data diwajibkan

Kartu ini memaksa Anda menjawab pertanyaan yang paling sering terlewat: apa satu baris itu sebenarnya, dan apakah ada beberapa baris yang berasal dari satu orang, satu lokasi, atau satu hari. Jawabannya menentukan cara membagi data (buku ajar Bab 10).

Satuan observasi yang salah dipahami adalah sumber kebocoran yang paling halus: baris dari stasiun yang sama pada hari berurutan mirip satu sama lain, sehingga pembagian acak membuat *test data* terlihat mudah.

Kebutuhan wajib

Kebutuhan wajib (W1–W10)

W1 **Perumusan masalah** dalam bahasa Mitchell: *T*, *P*, *E*, jenis masalah, dan satu alasan mengapa aturan tulisan tangan tidak cukup.

W2 **Kartu data** lengkap seperti slide sebelumnya.

W3 **Eksplorasi data** yang menjawab pertanyaan, misalnya ketidakseimbangan kelas, nilai hilang, korelasi antar-*feature*, dan *outlier*.

W4 **Praproses di dalam *pipeline***: imputasi, pengkodean kategorikal, *scaling*, dipasang hanya pada *training data*.

W5 **Minimal tiga model** dari Pertemuan 2–7, dari minimal dua keluarga berbeda, satu di antaranya patokan sederhana; ditambah *dummy baseline*.

W6 ***Tuning hyperparameter*** di dalam *cross-validation* pada *training data*, dengan ruang pencarian yang ditulis.

W7 **Protokol evaluasi**: *test data* dikunci sejak awal, *metric* beralasan *cost of error*, *confidence interval*, dan uji berpasangan antara dua model teratas.

W8 ***Cost***: *training time* dan *inference time* tiap model diukur.

W9 **Analisis *error* dan pembahasan**: di mana model salah, *feature* apa yang berpengaruh, risiko bila dipakai sungguhan.

W10 **Reproduktibilitas**: *seed* tetap, daftar paket, dan *notebook* berjalan dari atas ke bawah tanpa *error*.

Bila salah satu kebutuhan wajib tidak dipenuhi, nilai maksimum komponen rubrik yang terkait dibatasi setengahnya.

Model yang dihitung, dan yang tidak

Keluarga	Model (pertemuan)	Yang wajib ditulis di laporan
Berbasis pohon	*Decision tree* (P3), *random forest*, XGBoost (P4)	Kedalaman dan *pruning*; *feature importance* beserta keterbatasannya
Probabilistik	Naive Bayes (P5)	Asumsi independensi bersyarat dan apakah masuk akal pada data Anda; *smoothing*
Margin	SVM linier dan *kernel* (P7)	Pilihan *kernel*, nilai C dan γ, alasan *scaling*
Berbasis contoh	kNN, *locally weighted regression* (P6)	*Distance metric*, k terpilih, *inference cost*
Linier	*Linear*/*logistic regression*, *ridge*, *lasso* (P9)	Tidak dihitung (di luar P2–P7); boleh dipakai sebagai *baseline* tambahan

Tidak dihitung sebagai salah satu dari tiga model: MLP dan jaringan saraf lain, AutoML, serta model bahasa besar. Boleh dilampirkan sebagai catatan, tanpa nilai.

Dummy baseline (kelas mayoritas atau rata-rata target) wajib ada sebagai lantai. Model yang tidak mengalahkannya harus dijelaskan.

Nilai lebih (tidak wajib, maksimal +10%)

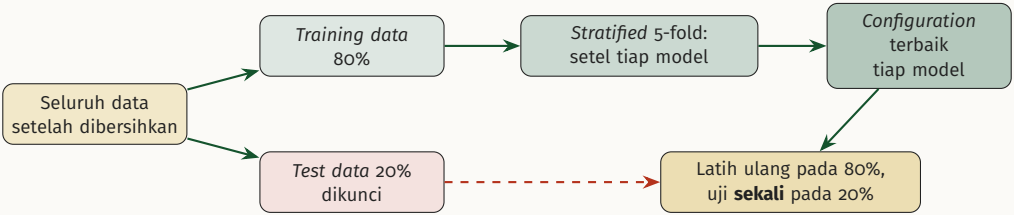
- Satu algoritma ditulis dari nol (ID3, Naive Bayes, atau kNN) dengan NumPy, lalu hasilnya dicocokkan dengan scikit-learn
- Menggabungkan dua sumber data Indonesia dengan kunci yang dijelaskan
- Kurva belajar: kinerja terhadap jumlah *training data*
- *Nested cross-validation* untuk taksiran yang tidak bias oleh *tuning*
- Analisis kalibrasi peluang (*reliability diagram*)
- Analisis keadilan: kinerja dipisah per wilayah, jenis kelamin, atau kelompok lain yang relevan

Ringkasan

Nilai lebih hanya dihitung bila W1–W10 sudah terpenuhi. Tambahan di atas fondasi yang rapuh tidak menambah nilai.

Protokol evaluasi

Alur protokol yang diminta



Pembagian harus sesuai satuan observasi. *Stratified* hanya untuk *classification*; *regression* memakai `KFold`, data berurut waktu `TimeSeriesSplit`, data dengan banyak baris per orang atau lokasi `GroupKFold`. Pembagian acak hanya sah bila baris saling bebas.

Simpan indeks *test data* ke *file* pada Minggu 6, sertakan *hash*-nya di repositori, dan catat tanggal pertama kali *test data* dievaluasi. Mengevaluasi berulang lalu memilih yang terbaik adalah *tuning* pada *test data*.

Kebocoran yang paling sering, dan cara mencegahnya

- ***Scaling* sebelum dibagi.** Rata-rata dan simpangan baku dihitung dari seluruh data, termasuk *test data*.
- **Imputasi dengan statistik global.** Sama dengan di atas, untuk nilai hilang.
- ***Feature selection* pada seluruh data.** *Feature* dipilih karena berkorelasi dengan target di *test data*.
- ***Oversampling* sebelum dibagi.** Salinan baris yang sama muncul di *training* dan *test*.
- ***Feature* dari masa depan.** Kolom yang baru tersedia setelah target terjadi, misalnya “tanggal pelunasan” untuk meramal gagal bayar.

Satu cara mencegah sebagian besar kebocoran: masukkan semua langkah ke dalam satu *pipeline*, lalu berikan *pipeline* itu ke *cross-validation*.

```
pre = ColumnTransformer([
  (”num”, Pipeline([
      (”imp”, SimpleImputer(strategy=”median”)),
      (”sc”,  StandardScaler())]), kol_num),
  (”kat”, OneHotEncoder(handle_unknown=”ignore”),
          kol_kat)])
model = Pipeline([(”pre”, pre),
                  (”clf”, SVC(kernel=”rbf”))])
cari = GridSearchCV(model,
  {”clf__C”: [0.1, 1, 10],
   ”clf__gamma”: [”scale”, 0.01, 0.1]},
  cv=StratifiedKFold(5, shuffle=True,
                     random_state=42),
  scoring=”f1_macro”)
cari.fit(X_latih, y_latih)   # X_uji belum disentuh
```

Metric dan ketidakpastian

Pilih *metric* dari *cost of error*

- Kelas seimbang dan *cost* setara: *accuracy* boleh, disertai F1
- Kelas langka (penipuan, hoaks, gagal bayar): *precision*, *recall*, PR-AUC
- *Regression*: RMSE atau MAE dalam satuan target, ditambah R^2
- Tuliskan satu kalimat: kesalahan jenis mana yang lebih mahal, dan bagi siapa

Laporkan ketidakpastian (Bab 10)

- Pada *cross-validation*: rata-rata $\pm$ simpangan baku dari lima lipatan
- Pada *test data*: *confidence interval* 95% untuk *error*,

$$e \pm 1{,}96\sqrt{\frac{e(1-e)}{n}}$$

  atau *bootstrap* untuk *metric* lain
- Dua model teratas: uji t berpasangan atas skor per lipatan, dengan catatan keterbatasannya

Ringkasan

Selisih 0,8 poin dengan *confidence interval* yang saling tumpang tindih belum menjadi bukti. Kalimat “belum ada bukti bahwa model A lebih baik” adalah kesimpulan yang sah dan dinilai baik.

Bentuk tabel hasil yang diminta

Model	F1 CV (rata $\pm$ sb)	F1 uji	CI 95% uji	*Training time*	Inferensi / 1000 baris
Dummy (mayoritas)	…	…	…	…	…
Decision tree (kedalaman $= 6$)	…	…	…	…	…
Naive Bayes	…	…	…	…	…
SVM RBF ($C = 10$, $\gamma = 0{,}01$)	…	…	…	…	…
XGBoost (300 pohon)	…	…	…	…	…

- Tulis *hyperparameter* terpilih di dalam nama model
- Waktu diukur pada mesin yang sama, sebutkan CPU dan RAM-nya
- Untuk *regression*, ganti F1 dengan RMSE dan tambahkan R^2
- Sertakan *confusion matrix* model terbaik pada *test data*
- Sertakan kurva validasi untuk satu *hyperparameter* penting (misalnya k atau kedalaman pohon)
- Angka pada tabel harus dapat dihasilkan ulang dari *notebook*

Repositori, laporan, dan presentasi

Struktur repositori

```
uts-ml-kelompok-XX/
  README.md                       cara jalan, pembagian kerja
  requirements.txt                versi paket disematkan
  data/README.md                  cara unduh, lisensi
  data/split_uji.csv              indeks test data
  notebooks/01_eda.ipynb          W2, W3
  notebooks/02_model.ipynb        W4–W8
  notebooks/03_analisis.ipynb     W9
  src/                            fungsi bersama (opsional)
  hasil/                          tabel dan grafik
  laporan/laporan.pdf             8–12 halaman
  AI-USAGE.md                     catatan alat bantu AI
```

- Data besar tidak di-*commit*. Sediakan skrip atau tautan unduhan beserta *checksum*-nya.
- Setiap *notebook* diuji dengan ***Restart and Run All*** di Colab sebelum dikumpulkan. Waktu jalan total sebaiknya di bawah 30 menit.
- Riwayat *commit* memperlihatkan kontribusi setiap anggota. Satu *commit* raksasa di malam terakhir adalah tanda bahaya.
- Repositori boleh di GitHub atau GitLab, privat dengan akses untuk dosen.

Templat laporan (8–12 halaman)

1. **Ringkasan** — masalah, model terbaik, angka utama dengan *confidence interval* (setengah halaman)
2. **Perumusan masalah** — *T*, *P*, *E*; alasan memakai ML
3. **Data** — kartu data, eksplorasi, keterbatasan
4. **Praproses** — tiap langkah dan cara mencegah kebocoran
5. **Model** — tiga model atau lebih, *inductive bias* masing-masing, ruang *hyperparameter*
6. **Protokol evaluasi** — pembagian, *cross-validation*, *metric*, uji statistik
7. **Hasil** — tabel hasil, kurva validasi, *confusion matrix* atau plot *residual*, *training cost* dan *inference cost*
8. **Pembahasan** — mengapa pemenang menang, analisis *error*, risiko dan etika
9. **Kesimpulan dan keterbatasan**
10. **Lampiran** — pembagian kerja, ringkasan `AI-USAGE.md`, tabel lengkap

Ringkasan

Tulis analisis. Kalimat “kami memakai SVM” tidak menambah nilai; kalimat “SVM RBF unggul karena batas kelasnya melengkung, terlihat dari kurva validasi γ, dengan *cost* inferensi seperlima kNN karena hanya 20% data menjadi *support vector*” menambah nilai.

Presentasi (Pertemuan 8)

Format per kelompok: 18 menit

Masalah dan data	3 menit
Protokol dan model	4 menit
Hasil dan pembahasan	5 menit
Tanya jawab	6 menit

Sesi UTS berlangsung 100 menit. Bila jumlah kelompok melebihi kapasitas sesi, sebagian presentasi dijadwalkan pada sesi tambahan di minggu yang sama.

Yang wajib tampil

1. Rumusan *T*, *P*, *E* dalam satu slide
2. Diagram protokol evaluasi Anda
3. Tabel hasil lengkap dengan *confidence interval*
4. Dua contoh kesalahan model beserta dugaan penyebabnya
5. Satu sel *notebook* dijalankan langsung atas permintaan dosen

Ringkasan

Setiap anggota wajib berbicara dan menjawab minimal satu pertanyaan. Waktu dibatasi ketat; slide yang melewati 12 menit tidak dinilai.

Rubrik penilaian

Rubrik penilaian

Komponen	*Weight*	Nilai penuh bila
Perumusan masalah dan pemahaman data	15%	*T*, *P*, *E* jelas; kartu data lengkap; eksplorasi menjawab pertanyaan yang relevan (W1–W3)
Praproses dan *feature engineering*	15%	Semua langkah di dalam *pipeline*; pilihan imputasi, pengkodean, dan *scaling* beralasan (W4)
Ragam model dan ketepatan penerapannya	20%	Tiga model dari dua keluarga atau lebih ditambah *dummy*; *tuning* di dalam CV; *inductive bias* dijelaskan (W5, W6)
Kebenaran protokol evaluasi	25%	*Test data* dikunci; pembagian sesuai satuan observasi; *metric* beralasan; CI dan uji berpasangan; *cost* terukur (W7, W8, W10)
Pembahasan, keterbatasan, dan etika	15%	Analisis *error* dengan contoh; pengaruh *feature*; risiko dan pihak yang terdampak (W9)
Laporan dan presentasi	10%	Templat diikuti, $\leq$ 12 halaman, tepat waktu, semua anggota menjawab dengan paham
Nilai lebih	+10%	*Algoritma dari nol, gabungan data, kurva belajar, nested CV, kalibrasi, keadilan*

Weight sama dengan yang diumumkan pada Pertemuan 7. Nilai akhir UTS dibatasi 100. UTS bernilai 35% dari nilai mata kuliah.

Tingkat capaian untuk tiga komponen terbesar

Komponen	Sangat baik (85–100)	Cukup (60–84)	Kurang (<60)
Protokol evaluasi (25%)	Pembagian sesuai satuan observasi; *test data* dipakai sekali dan terbukti dari *hash*; CI dan uji berpasangan ditafsirkan dengan benar	Pembagian benar, tetapi CI atau uji statistik tidak ada atau salah tafsir	Ada kebocoran, atau *test data* dipakai untuk memilih model; nilai komponen maksimal setengah
Ragam model (20%)	Tiga model atau lebih dari keluarga berbeda, disetel dengan ruang pencarian yang masuk akal, *inductive bias* dikaitkan dengan hasil	Tiga model ada, tetapi sebagian memakai nilai bawaan tanpa *tuning*	Kurang dari tiga model, atau model dari luar Pertemuan 2–7 dihitung
Pembahasan (15%)	Kesalahan dianalisis dengan contoh dan dugaan penyebab; risiko pemakaian dikaitkan dengan *cost of error*; keterbatasan jujur	Pembahasan umum, tanpa contoh kesalahan yang konkret	Hanya mengulang angka tabel; tidak ada keterbatasan

Komponen lain dinilai dengan pola yang sama: nilai penuh bila seluruh isian pada kolom “nilai penuh bila” terpenuhi dan dapat ditunjukkan buktinya.

Nilai individu dan penilaian sejawat

Nilai kelompok menjadi titik awal. Nilai individu dapat disesuaikan dari tiga sumber:

1. **Tanya jawab.** Setiap anggota ditanya tentang bagian yang tidak ia kerjakan.
2. **Riwayat *commit*** dan pembagian kerja di `README.md`.
3. **Penilaian sejawat**: formulir rahasia, setiap anggota menilai kontribusi anggota lain pada skala 1–5.

Aturan penyesuaian

- Penyesuaian maksimal ± 15 poin dari nilai kelompok
- Anggota yang tidak dapat menjelaskan kode kelompoknya sama sekali dapat dikurangi lebih jauh setelah wawancara terpisah
- Konflik kelompok dilaporkan ke dosen paling lambat Minggu 7, sebelum tenggat

Integritas akademik dan aturan main

Integritas akademik

Alat bantu AI dan LLM

- Boleh dipakai, misalnya untuk menjelaskan pesan *error* atau merapikan kode.
- Tuliskan di `AI-USAGE.md`: alat apa, untuk bagian mana, dan apa yang Anda periksa atau ubah.
- Setiap anggota harus mampu menjelaskan baris mana pun dari kode kelompoknya. Bagian yang tidak dapat dijelaskan tidak dinilai.
- Teks laporan, terutama pembahasan, ditulis sendiri. Angka dan kutipan yang disarankan alat bantu wajib diperiksa ke sumbernya.

Dinilai nol

- Menyalin *notebook* kelompok lain atau *notebook* Kaggle tanpa menyebut sumber
- Memalsukan atau mengarang angka hasil
- Menyembunyikan bahwa *test data* sudah dipakai berulang kali

Boleh, dengan menyebut sumber

Potongan kode dari dokumentasi scikit-learn, buku, atau forum. Tandai di komentar kode dan cantumkan di daftar pustaka laporan.

Aturan main

- Kelompok 3–4 orang, terbentuk paling lambat Minggu 5. Kelompok yang kurang dari tiga orang digabung oleh dosen.
- Judul, masalah, dan *dataset* disetujui dosen paling lambat dua hari setelah Pertemuan 7. Pergantian *dataset* sesudahnya harus seizin dosen.
- Bahasa laporan: Indonesia. Istilah asing dicetak miring.
- Python 3 dengan NumPy, pandas, scikit-learn, dan XGBoost. *Library* lain boleh dipakai bila dicantumkan di `requirements.txt`.
- Keterlambatan pengumpulan: dikurangi 10 poin per hari, maksimal 3 hari. Setelah itu tidak diterima.
- Repositori dibekukan dengan *tag* `uts-final`. Perubahan sesudah tenggat tidak dinilai.
- *Notebook* yang tidak dapat dijalankan ulang: dikurangi 10 poin, dan angka yang tidak dapat dihasilkan ulang dianggap tidak ada.
- Kelompok yang berhalangan presentasi karena alasan sah menghubungi dosen sebelum hari presentasi untuk jadwal pengganti.

Jadwal



Rencana tugas besar beserta protokol evaluasinya (Minggu 7) dinilai sebagai tugas dan masuk komponen “tugas dan kuis” (20%), terpisah dari nilai UTS. Persetujuan judul pada Minggu 7 + 2 hari memberi jarak lebih dari satu pekan sebelum tenggat. Bila *dataset* berganti saat persetujuan, *test data* dikunci ulang dan tanggalnya dicatat.

Checklist pengumpulan

- ☐ Tautan repositori dapat diakses dosen
- ☐ `README.md`: cara menjalankan dan pembagian kerja
- ☐ Kartu data dan cara mengunduh data
- ☐ Indeks *test data* dan *hash*-nya
- ☐ Semua *notebook* lulus *Restart and Run All*
- ☐ `requirements.txt` dengan versi paket
- ☐ Tabel hasil dengan CI, uji berpasangan, dan *cost*
- ☐ Laporan PDF 8–12 halaman
- ☐ `AI-USAGE.md` terisi
- ☐ Formulir penilaian sejawat dikirim tiap anggota
- ☐ *Slide* presentasi
- ☐ *Tag* `uts-final` dibuat sebelum tenggat

Pertanyaan yang sering muncul

Pertanyaan yang sering muncul (1)

- ***Accuracy* kami rendah. Apakah nilainya ikut rendah?** Belum tentu. Protokol (25%) dan pembahasan (15%) menilai kejujuran dan ketepatan cara, dan masalah yang sulit memang menghasilkan angka yang sedang. Yang dinilai rendah adalah angka yang tidak dapat dipercaya.
- **Boleh memakai LightGBM atau CatBoost?** Boleh, dihitung dalam keluarga *gradient boosting* bersama XGBoost. Jelaskan perbedaannya dengan XGBoost.
- **Boleh menambahkan MLP?** Boleh dilampirkan, tetapi tidak dihitung sebagai salah satu dari tiga model. *Neural network* dinilai di UAS.
- **Kelas kami tidak seimbang.** Pakai pembagian *stratified*, *metric precision*, *recall*, dan PR-AUC, serta `class_weight`. Bila memakai *oversampling*, lakukan di dalam *pipeline* agar hanya menyentuh lipatan latih.
- **Apakah algoritma harus ditulis dari nol?** Tidak. scikit-learn sudah cukup. Menulis satu algoritma dari nol masuk nilai lebih.

Pertanyaan yang sering muncul (2)

- **Boleh memakai data tempat kerja atau magang?** Hanya dengan izin tertulis, setelah identitas orang dihapus, dan bila lisensinya mengizinkan publikasi laporan. Bila ragu, pilih data terbuka.
- **Notebook atau skrip `.py`?** Keduanya boleh. Yang wajib: satu perintah atau satu *Restart and Run All* menghasilkan ulang seluruh angka di laporan.
- ***Test data* terlanjur kami pakai untuk memilih model.** Laporkan dengan jujur, bentuk pembagian baru dari data yang belum pernah dilihat bila memungkinkan, dan jelaskan di bagian keterbatasan. Kejujuran ini dinilai jauh lebih baik daripada menyembunyikannya.
- **Bagaimana dengan data teks?** Ubah menjadi *feature* dengan *bag of words* atau TF-IDF di dalam *pipeline*. *Embedding* dari model bahasa besar tidak dipakai pada UTS.
- **Satu anggota menghilang.** Laporkan ke dosen sebelum Minggu 7. Pembagian kerja disesuaikan dan penilaian sejawat mencatatnya.

Pertanyaan yang akan muncul saat presentasi

1. Apa satu baris data Anda, dan mengapa pembagian datanya begitu?
2. Langkah praproses mana yang paling berisiko bocor, dan bagaimana Anda mencegahnya?
3. Apa *inductive bias* model pemenang, dan mengapa cocok dengan data ini?
4. *Confidence interval* dua model teratas tumpang tindih. Apa kesimpulan Anda?
5. Mengapa *metric* ini, dan siapa yang menanggung kesalahan jenis yang paling mahal?
6. Tunjukkan dua contoh yang salah ditebak. Apa yang sama dari keduanya?
7. Model mana yang Anda pilih bila waktu jawab dibatasi 10 milidetik?
8. Tunjukkan satu baris kode yang Anda tulis sendiri dan jelaskan.

Ringkasan

Daftar ini sengaja dibagikan sekarang. Kelompok yang menyiapkan jawabannya dengan bukti dari *notebook* hampir pasti mendapat nilai baik.

Setelah UTS

Paruh pertama memakai model yang dapat dilatih dengan rumus atau dengan pencarian serakah: pohon, peluang, campuran Gaussian, rantai Markov, tetangga, dan *margin*.

Paruh kedua mencari parameter dengan menuruni *error surface*. Dari sana kita sampai ke *neural network*, CNN, *transformer*, dan LLM. Model terbaik tugas besar UTS Anda dapat menjadi patokan yang harus dikalahkan pada UAS.

Pertemuan 9

Linear regression, evaluasi hipotesis secara statistik, lalu *loss function*, *gradient descent*, dan SGD.

Selamat mengerjakan

Protokol yang jujur dulu, angka yang tinggi kemudian.