



Universitas Pembangunan Nasional Veteran Jakarta
Fakultas Ekonomi dan Bisnis

LEMBAR SOAL UJIAN

ETS/UTS SEMESTER GANJIL TA 2026/2027

Mata Kuliah	Analisis Data untuk Bisnis	Kode/SKS	AKT120406/ 2 SKS
Hari/Tanggal/Jam	Sabtu, 10 Oktober 2026: 07.10 sd. 8.55 wib	Kelas	A
Dosen Pengampu	Dr. Donny Maha Putra	Ruang	Daring
Waktu Ujian	100 menit		
Sifat Ujian	Project		
SUB CAPAIAN PEMBELAJARAN MATAKULIAH			
Mahasiswa mampu melakukan analisis dan menjelaskan proses analisis data, teknik visualisasi data, serta pemanfaatan data dalam proses pengambilan keputusan dalam pengembangan riset akuntansi dan bisnis.			Bobot 100%

PETUNJUK DAN KETENTUAN PENYELESAIAN PROYEK

Setiap kelompok WAJIB mengerjakan KETIGA jenis analisis dalam SATU laporan menggunakan SATU DATASET yang sama: (1) Classification, (2) Regression, dan (3) Clustering. Tujuan desain ini adalah membekali seluruh mahasiswa dengan pengalaman praktis lintas teknik analitik serta melatih kemampuan melihat satu dataset dari tiga lensa yang berbeda.

1. Tujuan Umum

Proyek ini bertujuan melatih mahasiswa menerapkan tiga kategori teknik data mining, klasifikasi, regresi, dan clustering, secara terintegrasi pada satu dataset menggunakan Orange Data Mining. Setiap kelompok diharapkan mampu merumuskan tiga pertanyaan analitik yang berbeda dari satu dataset, menerapkan algoritma yang sesuai untuk masing-masing pertanyaan, mengevaluasi kinerja model, serta menyintesis temuan lintas analisis untuk menghasilkan rekomendasi bisnis yang terintegrasi dan bermakna.

2. Instruksi Umum

Setiap mahasiswa wajib memahami dan mengikuti semua instruksi. Kesalahan dalam pengiriman hasil proyek akan berdampak pada penilaian individu. Proyek dikerjakan secara kelompok, namun setiap mahasiswa WAJIB mengunggah laporan dan materi presentasi melalui LEADS secara individual sesuai dengan batas waktu yang ditetapkan.

Setiap kelompok BEBAS menentukan dataset yang akan digunakan, dengan syarat sumber dataset berasal dari **Kaggle** (<https://www.kaggle.com/datasets>). Dataset yang dipilih harus memenuhi SELURUH kriteria teknis pada bagian 5 agar dapat digunakan untuk ketiga jenis analisis.

Setiap kelompok beranggotakan 4 orang (KLP01 sampai KLP13) atau 3 orang (KLP14 dan KLP15). Pada kelompok 4 orang, setiap anggota WAJIB menjadi penanggung jawab (lead) satu bagian: Anggota 1 memimpin EDA & Data Preparation serta Sintesis (Bab 3 dan Bab 7), Anggota 2 memimpin Classification (Bab 4), Anggota 3 memimpin Regression (Bab 5), dan Anggota 4 memimpin Clustering (Bab 6). Pada kelompok 3 orang, setiap anggota memimpin satu analisis, sedangkan EDA dan sintesis dikerjakan bersama-sama.

Peran lead setiap anggota WAJIB dicantumkan pada sampul laporan dan matriks kontribusi. Menjadi lead tidak berarti hanya memahami satu bagian: setiap anggota harus mampu menjelaskan seluruh isi laporan. Nilai individu dapat disesuaikan berdasarkan matriks kontribusi dan kemampuan anggota dalam menjelaskan hasil kerja kelompok.

3. Struktur Laporan Kelompok

Laporan disusun dengan urutan berikut perhatikan bahwa Bab 4, 5, dan 6 merupakan tiga analisis terpisah yang saling melengkapi, diikuti oleh Bab 7 sebagai bab sintesis:

- Sampul (Cover): judul proyek, nama, NIM, peran lead setiap anggota kelompok, nama kelas, mata kuliah, dosen pengampu, dan tahun akademik.
- Bab 1 Pendahuluan: konteks kasus, deskripsi dataset, tiga rumusan masalah (classification, regression, clustering), dan tujuan analisis terintegrasi.
- Bab 2 Tinjauan Teori & Metode: konsep ketiga kategori algoritma serta metodologi analisis (minimal 7 referensi APA 7).
- Bab 3 Exploratory Data Analysis (EDA) & Data Preparation: eksplorasi awal dataset, penanganan missing value, encoding, scaling, serta justifikasi keputusan preprocessing.
- Bab 4 Analisis Klasifikasi: penerapan algoritma klasifikasi, evaluasi kinerja, interpretasi feature importance, dan temuan.
- Bab 5 Analisis Regresi: penerapan algoritma regresi, evaluasi kinerja, interpretasi koefisien/feature importance, dan temuan.
- Bab 6 Analisis Clustering: penerapan algoritma clustering, evaluasi kualitas cluster, interpretasi persona/segmen, dan temuan. (Alternatif: Association Rule jika dataset berformat transaksional.)
- Bab 7 Sintesis & Implikasi Terintegrasi: penyatuan temuan dari ketiga analisis, storyline bisnis yang utuh, serta implikasi manajerial dan akuntansi. **BAGIAN INI MERUPAKAN PEMBEDA UTAMA ANTARA KELOMPOK BIASA DAN KELOMPOK UNGGUL.**
- Bab 8 Kesimpulan & Saran.
- Daftar Referensi (format APA 7th edition).
- Lampiran: screenshot workflow Orange untuk ketiga analisis, screenshot hasil utama, dan matriks kontribusi anggota kelompok (memuat peran lead dan persentase kontribusi tiap anggota).

4. Format Penulisan dan Penamaan File

Format Laporan

- Font: Arial, ukuran 11 pt.

- Spasi: 1,5.
- Margin: 1 inci (2,54 cm) di setiap sisi.
- Nomor halaman: di bagian bawah tengah (kecuali halaman sampul).
- Format ekspor final: PDF (Word → Save As PDF).

Format Presentasi (PowerPoint)

- Font: Aptos (body) dan Aptos Display (judul).
- Warna: latar putih dengan aksen navy (#1E3A5F) dan teal (#14B8A6).
- Ukuran slide: 16:9 (widescreen).
- Jumlah slide: 10 sampai 15 slide.
- Struktur minimal: Cover → Latar Belakang & 3 Rumusan Masalah → Deskripsi Dataset → Metodologi → Hasil Classification → Hasil Regression → Hasil Clustering → Sintesis Terintegrasi → Kesimpulan & Implikasi → Penutup.
- Format ekspor final: PDF (PowerPoint → Save As PDF).

Format Nama File (WAJIB)

- Laporan: KLP[NN]_[NamaMahasiswa]_Laporan.pdf
- Presentasi: KLP[NN]_[NamaMahasiswa]_Presentasi.pdf
- Contoh: KLP01_AhmadFauzi_Laporan.pdf dan KLP01_AhmadFauzi_Presentasi.pdf
- Aturan: NN menggunakan 2 digit (01 sampai 15); nama ditulis tanpa spasi atau menggunakan underscore.

5. Syarat Teknis Dataset

PENTING: Karena setiap kelompok mengerjakan tiga analisis dengan satu dataset, dataset yang dipilih WAJIB memenuhi SELURUH kriteria di bawah ini secara bersamaan. Dataset yang tidak memenuhi semua kriteria tidak diperkenankan untuk digunakan.

Syarat Umum Dataset

- Sumber WAJIB: Kaggle (kaggle.com/datasets). Link dataset harus publik dan masih aktif.
- Lisensi: Public/Open (CC0, CC-BY, atau Kaggle Public Dataset License).
- Format file: CSV, Excel, atau TSV (kompatibel langsung dengan widget File di Orange).
- Dataset tidak boleh sama antarkelompok. Setiap kelompok WAJIB mendaftarkan judul dan tautan dataset melalui formulir yang dibagikan oleh dosen pengampu. Apabila dua kelompok memilih dataset yang sama, kelompok yang lebih dulu mendaftar berhak menggunakannya.
- Dataset yang telah digunakan kelompok UTS Big Data Analytics Semester Genap TA 2025/2026 (Kelas A dan Kelas B) TIDAK boleh digunakan kembali.
- Dataset 'ikonik' DILARANG: Titanic, Iris, Boston Housing, MNIST, Wine Quality, Diamonds, serta dataset default di Orange.
- Dataset harus memiliki deskripsi atribut (data dictionary) yang jelas di halaman Kaggle.
- Label kolom dan kategori dalam bahasa Inggris atau bahasa Indonesia.
- Jumlah observasi: minimum 200 baris, maksimum 100.000 baris.

Syarat Kelayakan untuk Tiga Analisis

Dataset yang dipilih harus memiliki karakteristik berikut sehingga dapat digunakan untuk ketiga jenis analisis secara bersamaan:

- Memiliki MINIMAL 1 variabel kategorik yang bermakna untuk dijadikan target Classification (binary atau multiclass, maksimum 5 kelas, kelas minoritas $\geq 5\%$).
- Memiliki MINIMAL 1 variabel numerik kontinu yang bermakna untuk dijadikan target regresi (bukan ordinal, bukan integer count).
- Memiliki MINIMAL 4 fitur numerik dengan variasi yang memadai untuk digunakan sebagai input clustering.
- Total fitur prediktor: minimal 5 (tidak termasuk ID dan variabel target).
- Disarankan agar dataset memiliki campuran fitur numerik dan kategorik untuk melatih kemampuan preprocessing.

Contoh dataset yang umumnya memenuhi tiga syarat sekaligus (hanya contoh, mahasiswa tetap bebas memilih): Telco Customer Churn (target Classification: churn; target Regression: monthly_charges atau total_charges; fitur numerik demografi untuk Clustering), HR Employee Attrition (target Classification: attrition; target Regression: monthly_income; fitur untuk segmentasi), Insurance Premium Data, Online Retail (untuk fitur level pelanggan), dan sejenisnya. Mahasiswa wajib memverifikasi sendiri kelayakan dataset yang dipilih.

6. Instruksi Teknis Analisis

6.1 Analisis Klasifikasi

- Pilih satu variabel kategorik dari dataset sebagai variabel target.
- Jika distribusi kelas tidak seimbang secara ekstrem, wajib menerapkan penyeimbangan data dan mendokumentasikannya: undersampling melalui widget Data Sampler, atau SMOTE melalui widget Python Script (SMOTE tidak tersedia sebagai widget bawaan di Orange).
- Algoritma WAJIB: K-Nearest Neighbour (KNN), Decision Tree, dan Random Forest.
- Algoritma OPSIONAL (untuk nilai plus): Naïve Bayes, AdaBoost, atau Gradient Boosting.
- Metrik evaluasi wajib: Confusion Matrix, Accuracy, Precision, Recall, F1-Score, dan ROC-AUC.
- Wajib menyajikan tabel perbandingan metrik antar-algoritma serta justifikasi pemilihan model terbaik.

6.2 Analisis Regresi

- Pilih satu variabel numerik kontinu dari dataset sebagai variabel target.
- Jika distribusi target sangat skewed, wajib melakukan transformasi (log/Box-Cox) dengan justifikasi.
- Algoritma WAJIB: Linear Regression dan Random Forest Regression.
- Algoritma OPSIONAL (untuk nilai plus): Regression Tree, Gradient Boosting, atau SVR.
- Metrik evaluasi wajib: R^2 , MAE, RMSE, dan MAPE.

- Wajib menyajikan plot Predicted vs Actual, Residual Plot, serta interpretasi 3 variabel yang paling berpengaruh.

6.3 Analisis Clustering

- Gunakan fitur numerik dataset (kategorik boleh di-encode). Target variabel TIDAK digunakan sebagai input.
- Algoritma WAJIB: K-Means dan Hierarchical Clustering.
- Algoritma OPSIONAL (untuk nilai plus): DBSCAN.
- Evaluasi wajib: penentuan K optimal menggunakan Silhouette Score untuk beberapa nilai K (tersedia pada widget k-Means; Elbow Method bersifat pelengkap), visualisasi scatter plot cluster, dan interpretasi persona tiap cluster.
- Alternatif: Jika dataset berformat transaksional, Clustering boleh diganti dengan Association Rule (Apriori), minimum 10 rules bermakna disertai Support, Confidence, dan Lift.

6.4 Sintesis Terintegrasi (Bab 7)

Pada bagian ini, kelompok wajib menjawab pertanyaan-pertanyaan integratif berikut:

- Apa kesamaan temuan penting yang muncul dari ketiga analisis tersebut? Apakah fitur yang penting dalam Classification juga penting dalam Regression?
- Apakah cluster yang terbentuk pada Clustering memiliki kecenderungan ke kelas tertentu pada Classification atau ke nilai target tertentu pada Regression?
- Bagaimana ketiga hasil analisis tersebut secara bersama-sama membentuk satu cerita bisnis atau insight manajerial yang utuh?
- Apa rekomendasi praktis yang dapat ditawarkan kepada pemangku kepentingan (manajer, auditor, atau analis) berdasarkan sintesis ini?

7. Bobot Penilaian

Penilaian difokuskan pada kedalaman insight, ketajaman interpretasi, dan kemampuan sintesis lintas analisis. Komponen Sintesis Terintegrasi (Bab 7) merupakan differentiator utama dengan bobot 15%.

Komponen Penilaian	Bobot	Kategori
Problem framing & tiga rumusan masalah	5%	Thinking
EDA & Data Preparation	10%	Thinking + Doing
Analisis Classification (workflow + evaluasi + interpretasi)	15%	Doing + Thinking
Analisis Regression (workflow + evaluasi + interpretasi)	15%	Doing + Thinking
Analisis Clustering (workflow + evaluasi + interpretasi)	15%	Doing + Thinking
Sintesis Terintegrasi lintas analisis (Bab 7)	15%	Thinking (KUNCI)
Implikasi praktis/bisnis/akuntansi	10%	Thinking
Originalitas & kedalaman insight	5%	Thinking
Kualitas laporan (struktur + APA 7 + penulisan)	5%	Format
Kualitas PPT (desain + narasi)	5%	Format

8. Etika Akademik dan Integritas

Proyek ini menekankan integritas akademik sebagai pilar utama. Mahasiswa wajib menyajikan pekerjaan yang merupakan hasil pemikiran dan analisisnya sendiri.

Larangan Penggunaan AI Generatif

Dilarang keras menggunakan AI generatif (ChatGPT, Claude, Gemini, Copilot, Perplexity, DeepSeek, atau layanan sejenis) untuk menulis atau menghasilkan isi laporan dan materi presentasi. Seluruh narasi, interpretasi, pembahasan, dan kesimpulan harus merupakan hasil pemikiran mahasiswa sendiri.

Yang Diperbolehkan

- Penggunaan Orange Data Mining sebagai alat analisis (sesuai dengan tujuan mata kuliah).
- Penggunaan mesin pencari (Google, Google Scholar) untuk mencari referensi.
- Penggunaan grammar/spell checker (Microsoft Editor, Grammarly free tier) untuk koreksi teknis bahasa.

Yang Dilarang

- Generate paragraf, bab, kesimpulan, atau interpretasi menggunakan chatbot AI.
- Parafrase otomatis menggunakan alat AI (Quillbot AI mode, dsb.).
- Generate slide PPT secara otomatis menggunakan AI (Gamma, Beautiful.ai AI mode, Tome, dll.).
- Copy-paste output AI kemudian diedit seadanya.

Mekanisme Pengecekan

Seluruh laporan akan dipindai menggunakan AI Detection Tools (GPTZero, Turnitin AI Detection, atau ZeroGPT). Laporan dengan skor AI-generated content di atas 30% akan dikenakan sanksi sebagaimana diatur berikut ini.

Konsekuensi Pelanggaran

1. Pelanggaran ringan (30% sampai 50% AI content): nilai laporan 0, wajib mengulang (remedial) dengan bobot maksimum 70.
2. Pelanggaran berat (>50% konten AI): nilai UTS 0, disertai surat peringatan akademik.

Tips Menulis Laporan yang Aman dari AI Detection

1. Tulis dalam bahasa Indonesia yang natural dan personal; hindari gaya bahasa yang terlalu formal dan kaku.
2. Masukkan contoh spesifik dari dataset Anda (angka aktual, nama kolom, dan pola yang ditemukan).
3. Variasikan panjang kalimat, AI cenderung memproduksi kalimat panjang yang seragam.
4. Tambahkan refleksi pengalaman kelompok selama proses analisis.

5. Hindari frasa klise AI seperti: "Dalam era modern...", "Sebagai kesimpulan, dapat dikatakan bahwa...", "Hal ini menunjukkan pentingnya..."
6. Selipkan konteks lokal (Indonesia, UPNVJ, perusahaan/industri Indonesia yang Anda kenal).
7. Gunakan sitasi APA 7 yang benar, AI sering memfabrikasi referensi fiktif.

Larangan Plagiat

Dilarang keras melakukan copy-paste dari kelompok lain, kelas lain, laporan UTS semester sebelumnya, atau sumber mana pun tanpa sitasi yang benar. Pelanggaran dapat mengakibatkan kelompok yang bersangkutan dinyatakan gagal dalam mata kuliah ini.

9. Pengumpulan Hasil Proyek

- Batas waktu pengumpulan: Sabtu, 10 Oktober 2026, pukul 23.59 WIB.
- Platform pengumpulan: LEADS (setiap mahasiswa WAJIB mengunggah secara individual meskipun dikerjakan secara kelompok).
- File yang dikumpulkan: (1) Laporan dalam PDF dan (2) Materi Presentasi dalam PDF.
- Format penamaan file mengikuti ketentuan Bagian 4.
- Keterlambatan pengumpulan akan mengurangi nilai akhir.

10. Catatan Penting

Kepatuhan terhadap seluruh instruksi, mulai dari pemilihan dataset, penerapan ketiga analisis, format penulisan, hingga ketepatan waktu pengumpulan akan memengaruhi penilaian akhir. Proyek orange data mining ini dirancang untuk melatih kemampuan data mining lintas teknik, kolaborasi tim, penalaran analitis, serta komunikasi ilmiah dalam menyajikan analisis yang terintegrasi dan bermakna dalam konteks akuntansi dan bisnis.