

TUGAS BESAR (PROJECT AKHIR)

Mata Kuliah: Process Mining

Program Studi: S1 Informatika

Platform: Google Colab (Python)

Tools: pm4py, pandas, matplotlib

Dataset: event_log_rumah_sakit_sintetis.csv

Tujuan Pembelajaran

Melalui tugas besar ini, mahasiswa mampu:

1. Mengolah **event log** menjadi **model proses bisnis**
 2. Menerapkan **process discovery, conformance checking, dan performance analysis**
 3. Menginterpretasikan hasil process mining untuk **perbaikan proses layanan rumah sakit**
 4. Menyusun analisis berbasis **kode Python yang reproducible**
-

Deskripsi Dataset

Dataset merepresentasikan **alur pelayanan rumah sakit**, mulai dari:

Pendaftaran → Triage → Pemeriksaan → (Lab/Radiologi/Tindakan) → Pembayaran → Selesai

Dataset memiliki variasi proses:

- rawat jalan
 - rawat inap
 - tindakan medis
 - pembatalan layanan
-

Ketentuan Pengerjaan

- Dapat dikerjakan **berkelompok (maksimal 3 mahasiswa)** atau boleh juga **sendiri**.
 - Semua analisis **WAJIB menggunakan Python di Google Colab**
 - Semua output harus bisa direproduksi dengan **Run All**
-

STRUKTUR SOAL & PETUNJUK Pengerjaan

BAGIAN A — Setup & Data Understanding (10%)

Soal A

1. Upload dataset CSV ke Google Colab dan tampilkan 10 baris pertama
2. Tampilkan:
 - jumlah case
 - jumlah event
 - jumlah aktivitas unik
 - jumlah resource unik
3. Jelaskan secara singkat **alur proses rumah sakit** berdasarkan dataset

Petunjuk Teknis

- Pastikan kolom timestamp bertipe datetime
 - Urutkan data berdasarkan case_id dan timestamp
-

BAGIAN B — Event Log Preparation (15%)

Soal B

1. Lakukan data cleaning:
 - hapus missing value pada kolom utama
 - hapus event duplikat
2. Konversi dataframe menjadi **event log PM4Py**
3. Validasi bahwa setiap case memiliki minimal 2 aktivitas

Output Wajib

- Event log berhasil dikonversi (pm4py.objects.log.log.EventLog)
 - Ringkasan statistik event log
-

BAGIAN C — Exploratory Process Analysis (EDA) (15%)

Soal C

Lakukan analisis eksplorasi proses dengan Python:

1. Visualisasi:
 - Top 10 aktivitas paling sering
 - Distribusi jumlah event per case
 - Distribusi durasi case (jam)
2. Hitung:
 - Rata-rata dan median durasi case
 - 5 case terlama dan tercepat

Petunjuk

- Gunakan matplotlib
 - Semua grafik diberi judul dan label sumbu
-

BAGIAN D — Process Discovery (25%)

Soal D

1. Bangun model proses menggunakan **dua algoritma**:
 - **Inductive Miner**
 - **Heuristic Miner**
2. Visualisasikan masing-masing model
3. Bandingkan kedua model berdasarkan:
 - kompleksitas model
 - kemudahan dibaca (readability)
 - kesesuaian dengan log

Output Wajib

- 2 model proses (Petri Net / DFG / Process Tree)
 - Penjelasan singkat perbandingan model
-

BAGIAN E — Conformance Checking (20%)

Soal E

1. Pilih salah satu model discovery sebagai **model referensi**
2. Lakukan **token-based replay** atau **alignment**

3. Laporkan:

- nilai fitness keseluruhan
- aktivitas yang sering menyebabkan deviasi
- 3 contoh trace dengan fitness terendah

Petunjuk

- Gunakan `pm4py.algo.conformance.tokenreplay`
-

BAGIAN F — Performance & Bottleneck Analysis (15%)

Soal F

1. Identifikasi:
 - aktivitas dengan waktu tunggu paling lama
 - bottleneck utama proses
 2. Analisis resource:
 - resource paling sibuk
 - perbandingan durasi case antar minimal 2 resource
 3. Visualisasikan performance process (DFG atau grafik lain)
-

BAGIAN G — Rekomendasi Perbaikan Proses (Bonus 0–10%)

Soal G

Berdasarkan hasil analisis, berikan:

- **3 rekomendasi perbaikan proses** berbasis data
 - Setiap rekomendasi harus menyebutkan:
 - masalah
 - bukti (hasil analisis)
 - solusi yang diusulkan
-

OUTPUT YANG DIKUMPULKAN

1. **Notebook Google Colab (.ipynb)**
2. **File CSV (dataset)**

3. **Laporan PDF ringkas (5–7 halaman)**
4. Slide presentasi (maks. 10 slide)