

Folyamatbányászati feldolgozási útvonal

1. Adat-előkészítés

- **Cél:** a nyers logfájl (CSV) átalakítása érvényes eseménynaplóvá.
 - **Feladatok:**
 - Azonosítsd a három fő oszlopot:
 - *Case ID* → folyamatpéldány azonosító
 - *Activity* → esemény neve
 - *Timestamp* → időbélyeg
 - Tisztítsd az adatokat (hiányzó, hibás, duplikált rekordok kezelése).
 - **Eszközök:**
 - *Python (pandas + pm4py)*: adattisztítás, formátum-ellenőrzés
 - *ProM Import Plugins* vagy *Disco*: automatikus oszlopfelismerés
-

2. Folyamatmodell felfedezése

- **Cél:** elsődleges folyamatára előállítás.
 - **Algoritmusok:**
 - *Inductive Miner* → stabil, jól olvasható modellekhez
 - *Heuristics Miner* → zajtűrés, valós adatokhoz gyakori
 - *Alpha Miner* → egyszerű, inkább oktatási célra
 - **Eszközök:**
 - *ProM*: klasszikus algoritmusok, sok beállítás
 - *pm4py*: Pythonban könnyen szkriptelhető
 - *Disco*: gyors, felhasználóbarát vizualizáció
-

3. Konformancia-ellenőrzés

- **Cél:** összevetni a bányászott modellt az elvárt / szabályozott folyamattal.
 - **Módszerek:**
 - *Token-based replay* (ProM)
 - *Alignments* (pm4py, pontosabb, de lassabb)
 - **Kimenet:** hol vannak eltérések, mennyire felel meg a gyakorlat a szabályoknak.
-

4. Teljesítmény-elemzés

- **Cél:** szűk keresztmetszetek és késések azonosítása.
- **Módszerek:**
 - Átfutási idők mérése (case duration, activity duration).
 - Bányászott modell **időadatokkal színezve** → hőtérképek.

- **Eszközök:**
 - *Disco*: beépített teljesítménynézet
 - *ProM Performance Analysis plugin*
 - *pm4py*: `performance_diagnostics` modul
-

5. Erőforrás-elemzés

- **Cél:** megérteni, hogyan dolgoznak a résztvevők.
 - **Módszerek:**
 - *Social Network Miner* (ProM) → ki kivel dolgozik együtt
 - Terhelés elemzése: ki mennyi eseményt dolgozott fel
 - **Eszközök:**
 - *ProM SNA plugin*
 - *pm4py social network* funkció
-

6. Variáns-elemzés

- **Cél:** a leggyakoribb és ritka folyamatútvonalak feltárása.
 - **Módszerek:**
 - Top-k variánsok listázása
 - Hasonlósági elemzés különböző útvonalak között
 - **Eszközök:**
 - *Disco*: automatikusan csoportosítja a variánsokat
 - *pm4py*: `case_statistics.get_variant_statistics`
-

7. Prediktív elemzés és fejlettebb technikák

- **Cél:** jövőbeli események előrejelzése (pl. átfutási idő, hibavalószínűség).
 - **Módszerek:**
 - Gépi tanulás *process prediction* keretrendszerekkel
 - Recurrent Neural Network / XGBoost modellek attribútumokból
 - **Eszközök:**
 - *pm4py ML modulok*
 - *sklearn* / *PyTorch* integráció
-

8. Kommunikáció és dashboard

- **Cél:** az eredmények üzleti felhasználása.
- **Módszerek:**
 - Dashboard építés (Power BI, Tableau) → folyamatmodellek, KPI-k

- Időszakos riportok → folyamatbetartás és teljesítmény trendek
 - **Eszközök:**
 - *pm4py export* → *csv/json* → BI eszközök
 - *ProM reporting*
 - *Celonis* (ha enterprise szintű szoftvert is használtak)
-

Összefoglalva:

Az útvonal így néz ki:

[Adat-előkészítés] → [Modellépítés] → [Konformancia] → [Teljesítmény] → [Erőforrás] → [Variánsok] → [Predikció] → [Dashboard/riportálás].

Minden lépéshez van kipróbált eszköz, amivel fokozatosan mélyítheted az elemzést.
