Tóth István

J0P7MF

2025.04.12

# Folyamatmodellezés és folyamatbányászat

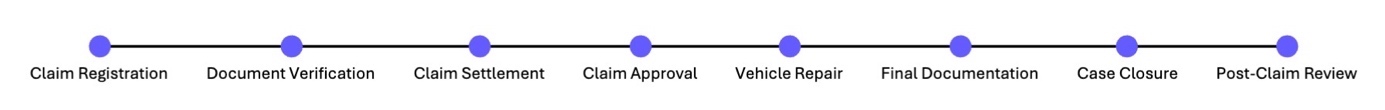
## A feladathoz használt folyamat

Az event logot a <https://processminingdata.com/> oldalról töltöttem le, a Car Insurance Claims adatállományt használtam. Ez 441550 eventet tartalmaz 49873 claim-hez, 2022.01.01 és 2023.12.31 között. Az adatok tanulási céllal készült szintetikus adatok, amelyek valós folyamatokat próbálnak mintázni. Az adathalmazt csv formátumban töltöttem le és autóbiztosításhoz benyújtott kártérítési igények elbírálási folyamatát írják le.

## Logok előkészítése

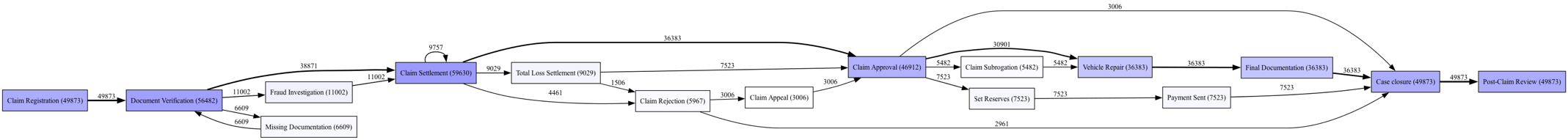
Mivel a logok csv formátumban vannak, ezért a logok előkészítésével kezdtem, amely magában foglalja a xes fájl létrehozásához szükséges mező átalakításokat (pl. dátum) és a mezők átnevezését, hogy könnyebben lehessen azonosítani őket. Ezek alapján elkészítettem a xes fájlt.

## Folyamatábra



Ezt a folyamatábrát az adathalmaz mellé mellékeli a weboldal, de ez egy nagyon leegyszerűsített, csak a leggyakoribb lépéseket leíró ábra. A feladatban az is volt a célom, hogy az itt nem szereplő eseményeket feltérképezzem.

## DFG



A directly folows gráf nagyon pontosan szemlélteti az eredeti folyamatot és azonnal látszanak is rajta a további események, a fő csomópontok teljesen megegyeznek.

A képen szöveg, képernyőkép, diagram látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

A heatmap elkészítésével ellenőriztem DFG átmeneteit és a számosságukat.

## Alpha Miner

Az alpha miner-rel nagyon furcsa Petri-hálót kaptam, ami számomra nem tűnik áttekinthetőnek sem és helyesnek sem. Olyan, mintha két különálló folyamat lenne, amik közt nincs kapcsolat, ami természetesen nem valós.

A képen diagram, sor, vázlat, fehér látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

A következő ábrán már az átmenetek száma is látszik, de nem jutottam vele közelebb a megoldáshoz, hogy miért tűnik helytelennek.

A képen sor, diagram, Diagram, szöveg látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

Összehasonlítottam az AM footprint és a DFG footprintet is, de ez sem vezetett megoldáshoz.

Alpha Miner footprint

A képen nyugta, sor látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

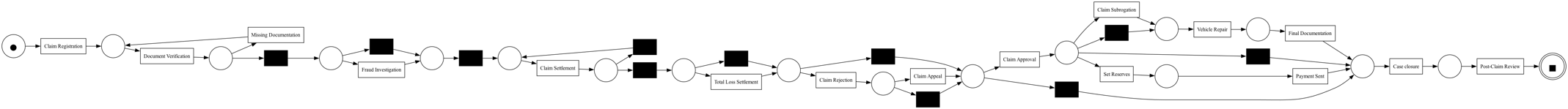
DFG footprint

A képen nyugta, sor látható

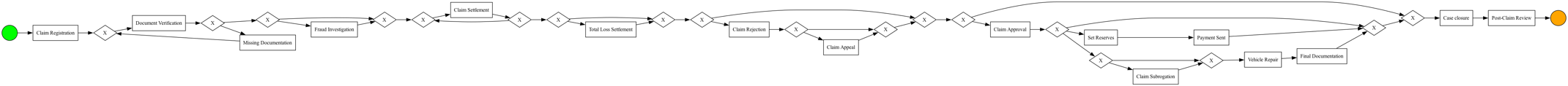
Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

## Inductive Miner

Az inductive miner-re készült Petri-háló eredményei már sokkal meggyőzőbbek voltak, egy folyamatot ábrázolnak, nincs kétrészre szakadva és a sorrend is megfelelő. A Petri-hálóban a fekete téglalapok (telített átmenetek) általában láthatatlan átmeneteket jelölnek, amelyeket az inductive miner algoritmus vezet be a folyamatstruktúra modellezése során. Ezek az átmenetek nem megfigyelhető tevékenységek az eseménynaplóban, hanem a vezérlési folyamat logikáját segítik.



Elkészítettem a BPMN ábráját is a folyamatnak, ami szintén helyes, pontosnak tűnik.



## Heuristic Miner

A heuristic miner-rel először elkészítettem a heurisztikus hálót több különböző dependency tresholddal (0.25, 0.5, 0.75, 0.99). Ezek nem mutatnak különbséget, az ábrák megegyeznek. Ennek az lehet az oka, hogy az adatok nem hiányosak vagy hibásak, a folyamatokban nincs zaj, minden eseménynek van elődje és utódja, nincsenek ciklusok és tulajdonképpen a folyamat egyszerű.

A képen diagram, rajz, sor, tervezés látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.A képen diagram, rajz, sor, tervezés látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.A képen diagram, rajz, sor, tervezés látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.A képen diagram, rajz, sor, tervezés látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

Ezek után elkészítettem különböző minimal activity count-al (1000, 5000, 10000, 25000). Itt az elvárt eredményt kaptam, a mac növelésével az ábrán megjelenő események száma csökken, de az ábrák végig helyesek maradnak.

A képen diagram, rajz, clipart, tervezés látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.A képen diagram, rajz, tervezés, illusztráció látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.A képen képernyőkép, diagram, sor, szöveg látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.A képen szöveg, képernyőkép, diagram, sor látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

Mac = 25000 értéknél gyakorlatilag pontosan visszakaptam azt az ábrát, amit az adathalmazhoz melléklenek, ezen csak egyetlen plusz átmenet van (Claim Approval -> Case closure).

Ugyanezt a vizsgálatot elvégeztem a min dfg segítségével is (1000, 5000, 10000, 25000), ennek eredménye is vártnak megfelelően alakult, tehát a gráf egyre kevesebb eseményt tartalmaz és a végén szintén előállt a már ismert folyamat.

A képen diagram, rajz, tervezés, illusztráció látható

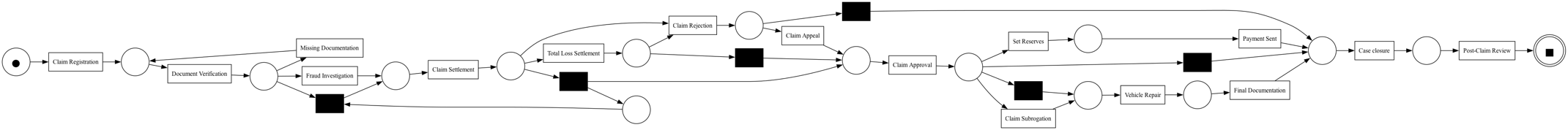
Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.A képen diagram, képernyőkép, tervezés látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.A képen szöveg, képernyőkép, diagram, sor látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.A képen szöveg, képernyőkép, tervezés, nyugta látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

Végül a dth=0.5 (alapérelmezet) érték mellet az alábbi Petri-hálót állítottam elő.



Az inductive minerhez képest kevesebb telített átmenet jelenik meg a képen, így számomra ez áttekinthetőbbnek tűnik, de ettől eltekintve nagyon hasonlít arra.

## Az egyes módszerek kiértékelése

A fentiek kiértékelésére az alábbi mátrixot készítettem.

A képen szöveg, képernyőkép, szám, Téglalap látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

Az ábrán az alpha miner precisionje nekem nagyon alacsony érték (a log és a modell viselkedése nagyon eltér), és gondolom ennek köszönhető az, hogy a generalization értéke magas (a modell által leírt lehetséges, várható viselkedések). A fitness értéke is a legalacsonyabb (kevésbé fedi le a modell a naplóban megfigyelhető eseményeket). Figyelembe véve, hogy a Petri-hálója sem sikerült egyértelműen értelmezhetőre én ezt a modellt ennél az adathalmaznál nem használnám.

Érdekes még az inductive modell 1-es fitness értéke (teljes mértékben lefedi a modell a naplóban megfigyelt eseményeket).

Én mégis ebben az adott esetben a heuristic modell-t alkalmaznám, annak bár alcsonyabb a fitness értéke, de a precision értéke jelentősen magasabb, egyszerűbb a modell és a generalization értéke hasonlóan magas.

## Folyamat bányászat

Azt vizsgáltam, hogy a különböző algoritmusok esetében hogyan befolyásolja a minimum support értéke az eredményeket.

A képen szöveg, képernyőkép, Betűtípus, diagram látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

A képen szöveg, képernyőkép, Betűtípus, diagram látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.

A képen szöveg, képernyőkép, Betűtípus, diagram látható

Előfordulhat, hogy a mesterséges intelligencia által létrehozott tartalom helytelen.