



UHASSELT

KNOWLEDGE IN ACTION

Faculteit Bedrijfseconomische Wetenschappen

master handelsingenieur in de beleidsinformatica

Masterthesis

Het verschil tussen de generatie van event logs gebaseerd op imperatieve en declaratieve modellen

Stef Peeters

Scriptie ingediend tot het behalen van de graad van master handelsingenieur in de beleidsinformatica

PROMOTOR :

Prof. dr. Benoit DEPAIRE

BEGELEIDER :

Mevrouw Jessica VAN SUETENDAEL



UHASSELT

KNOWLEDGE IN ACTION

www.uhasselt.be

Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

2024
2025



Faculteit Bedrijfseconomische Wetenschappen

master handelsingenieur in de beleidsinformatica

Masterthesis

Het verschil tussen de generatie van event logs gebaseerd op imperatieve en declaratieve modellen

Stef Peeters

Scriptie ingediend tot het behalen van de graad van master handelsingenieur in de beleidsinformatica

PROMOTOR :

Prof. dr. Benoit DEPAIRE

BEGELEIDER :

Mevrouw Jessica VAN SUETENDAEL

Woord Vooraf

Deze masterproef is de laatste stap in mijn opleiding Handelsingenieur in de Beleidsinformatica aan de UHasselt. Doorheen de opleiding heb ik een interesse voor data-gerelateerde onderwerpen ontwikkeld. Ik ben daarom blij dat ik die interesse heb kunnen gebruiken tijdens mijn masterproef door te werken rond process mining en data generatie.

De motivatie achter dit onderzoek is dus afkomstig van mijn nieuwsgierigheid naar het genereren van artificiële data. Ik wilde verschillende tools uitproberen en onderzoeken welke factoren een invloed kunnen hebben op die generatie. Daarom heb ik concreet onderzoek gedaan naar de invloed van process mining stromingen – imperatief of declaratief – op de generatie van artificiële event logs. Dat onderzoeksonderwerp heeft me namelijk de kans gegeven om nieuwe tools uit te proberen, te leren werken met grote hoeveelheden data en scripts te schrijven om de generatie en analyse te automatiseren. Daarom hoop ik met dit werk een waardevolle bijdrage te leveren aan toekomstig onderzoek rond process mining en artificiële event logs.

Daarnaast wil ik graag mijn oprechte dankbaarheid uiten voor mijn promotor, Prof. dr. Benoît Depaire, en mijn begeleider, Jessica Van Suetendael. Dankzij hun kennis en feedback heb ik alle moeilijkheden kunnen oplossen en een masterproef kunnen ontwikkelen waar ik trots op ben. Verder wil ik graag Manal Laghmouch bedanken die me heeft geholpen bij de ontwikkeling van enkele startpunten van mijn experiment. Daar heeft ze mij en de masterproef ontzettend ver mee vooruit geholpen. Ook de andere leerkrachten en medewerkers van de UHasselt wil ik graag bedanken om mij tijdens mijn opleiding te vormen tot de persoon die ik geworden ben. Ten slotte zou ik graag mijn ouders, mijn vriendin en mijn vrienden bedanken voor hun onvoorwaardelijke steun en hun geloof in mij tijdens het maken van deze masterproef.

Het schrijven van deze thesis heeft me veel bijgeleerd, zowel op persoonlijk als professioneel vlak. Ik heb zo geleerd gedisciplineerd te werken met meer oog voor detail. Ik hoop daarbij dat dit werk een betekenisvolle bijdrage kan vormen voor toekomstig academisch onderzoek naar het genereren van artificiële data voor de evaluatie van process mining algoritmen.

Stef Peeters

Diepenbeek, 3 Juni 2025

Het verschil tussen de generatie van event logs gebaseerd op imperatieve en declaratieve modellen

Peeters, Stef
`stef.peeters@student.uhasselt.be`

UHasselt

Samenvatting Process mining is een onderzoeksgebied dat event logs analyseert om inzichten in bedrijfsprocessen te verkrijgen. Voor de evaluatie van process mining algoritmen hebben onderzoekers grote hoeveelheden event data nodig. Er is echter gebleken dat dat lastig is wanneer er met reële data gewerkt wordt, waardoor artificieel gegenereerde event logs gebruikt worden. Die event logs worden altijd gegenereerd op basis van een onderliggend model. Dit onderzoek heeft daarom als doel om te bepalen of de stroming van dat onderliggende model - imperatief of declaratief - een invloed heeft op de gegenereerde data. Om dat te onderzoeken zijn er zes sets van event logs ontwikkeld, waarvan drie op basis van imperatieve methodes en drie op basis van declaratieve methodes. Die werden gebruikt om opnieuw artificieële event logs te genereren via zowel imperatieve als declaratieve methodes. Die gegenereerde event logs werden vervolgens vergeleken op basis van karakteristieken zoals entropie, de eigenschappen van de paden en lokale patronen in die paden met behulp van Welch's Two Sample t-tests. De resultaten toonden dat er significante verschillen bestonden tussen de artificieële events logs op bijna elk level. Dat betekent dat de process mining stroming een significante invloed heeft op de generatie van event logs. Dit onderzoek toont zo aan dat onderzoekers aandacht moeten besteden aan het kiezen van een process mining stroming voor het genereren van event logs. Daarnaast is dit onderzoek een argument voor de ontwikkeling van een nieuwe generator die modelonafhankelijk kan werken.

Trefwoorden— Process Mining, Declare, BPMN, Data Generatie, Event Log Generatie

1 Introductie

Process mining is een onderzoeksgebied dat data science en process management samenbrengt. Het gebruikt event data uit informatiesystemen om bedrijfsprocessen te ontdekken, te verbeteren en te monitoren [17]. Binnen process mining zijn er twee grote stromingen om processen te modelleren: imperatief en declaratief. Imperatieve process mining wordt daarbij vooral gebruikt bij gestructureerde,

voorspelbare processen zodat de diagrammen leesbaar blijven [2]. Declaratieve process mining is daarentegen ontworpen om tegemoet te komen aan ongestructureerde processen die veel verschillend mogelijk gedrag toelaten. Declaratieve modellen worden daarom ook niet voorgesteld als een process flow diagram, maar als een set regels die niet overschreden mogen worden [15].

Om die verschillende soorten modellen op te stellen vanuit event logs, zijn er verschillende algoritmen ontwikkeld. Aangezien er op dit moment heel wat verschillende process discovery algoritmen zijn, is er bovendien meer en meer aandacht voor de evaluatie en vergelijking van die algoritmen [18, 19]. Di Ciccio et al. (2015) en Jouck & Depaire (2019) beargumenteren dat er daarvoor artificiële data nodig is [6, 8]. Enerzijds is het gemakkelijker om aan een grote hoeveelheid artificiële data te komen dan aan reële data. Anderzijds staat artificiële data de onderzoeker toe om de eigenschappen van de data te controleren [8]. Er zijn daarom al verschillende data generators ontworpen. Zo bestaan er generators die kunnen starten vanuit imperatieve modellen, zoals PLG2 van Burattin [5] en generators die starten vanuit een declaratief model, zoals ontworpen door Skydaniienko et al. [16].

Ondanks de beschikbaarheid van verschillende generators, is er op dit moment geen duidelijkheid in de literatuur of de keuze van process mining stroming invloed heeft op de data generatie. De vraag is daarom of artificiële event logs die gegenereerd zijn op basis van een declaratief model verschillen van artificiële event logs die gegenereerd zijn via een imperatief model. Indien er significante verschillen bestaan, moet er immers meer aandacht besteed worden aan de bepaling van de gepaste process mining stroming voordat artificiële event logs gegenereerd kunnen worden. Dat achterhalen, is daarom het hoofddoel van dit onderzoek.

Dit onderzoek zal daarom gebruik maken van een experiment waardoor event logs uit declaratieve generators vergeleken kunnen worden met event logs uit imperatieve generators. Er zullen daarvoor event logs ontwikkeld worden die als startpunten voor het experiment zullen dienen. Die startpunten zullen vervolgens gebruikt worden om zowel imperatieve als declaratieve modellen te maken. Die modellen worden op hun beurt gebruikt om nieuwe event logs te genereren via respectievelijk een imperatieve en declaratieve generator. De imperatief gegenereerde event logs en de declaratief gegenereerde event logs zullen daarna vergeleken worden met elkaar en met de startpunten op basis van specifieke event log karakteristieken.

Deze paper zal dat onderzoek beschrijven door eerste enkele gerelateerde werken te bespreken. Daarna zal de methodologie in detail besproken worden. Dan zullen de bijbehorende resultaten objectief besproken worden, gevolgd door de discussie waarin de resultaten teruggekoppeld zullen worden aan de onderzoeksvraag. Ten slotte zal het werk afgesloten worden met de conclusie.

2 Gerelateerde Werken

Aangezien dit onderzoek gaat over het genereren van event logs, is het belangrijk om enkele mogelijke generatoren te bespreken. Er zijn namelijk al verschillende generatoren ontwikkeld, die zowel werken op basis van imperatieve modellen als declaratieve modellen. PLG2 en de PT and Log Generator zijn voorbeelden van imperatieve generatoren [5, 8]. PLG2 genereert op basis van enkele variabelen BPMN-modellen of neemt BPMN-modellen als input en simuleert die modellen om event logs te bekomen. De gebruiker kan dan onder andere het aantal paden en de gewenste noise-levels instellen waarna PLG2 paden zal opstellen op basis van het BPMN-model. Die paden worden vervolgens toegevoegd aan de event log, die geëxporteerd kan worden in XES-formaat [5].

De PT and Log Generator is nog een voorbeeld van een imperatieve generator [8]. Die generator start echter niet vanuit BPMN-modellen zoals PLG2, maar vanuit specifieke parameters. Voorbeelden van mogelijke parameters zijn het aantal activiteiten, noise en de frequentie van bepaalde patronen zoals loops en parallelismes. Op basis van die parameters worden een gegeven aantal process trees gegenereerd, die vervolgens gebruikt worden als basis om event logs te genereren. Die event logs kunnen dan geëxporteerd worden in XES-formaat [8].

RuM is vervolgens niet alleen een voorbeeld van een declaratieve generator, maar van een veel bredere tool voor declaratieve process mining technieken [1]. De tool staat gebruikers toe om model discovery op event logs uit te voeren, declaratieve modellen aan te passen, de conformance van declaratieve modellen na te kijken en event logs te genereren op basis van declaratieve modellen. Voor dit onderzoek is vooral de module voor event log generatie van belang. Het neemt namelijk declaratieve modellen als input en op basis van de Alloy Generator of de MP MINERful Log Generator werden de modellen gesimuleerd, waarmee event logs opgesteld werden. Daarbij konden parameters zoals het aantal activiteiten per trace en noise aangegeven worden. Opnieuw konden de event logs geëxporteerd worden in XES-formaat [1].

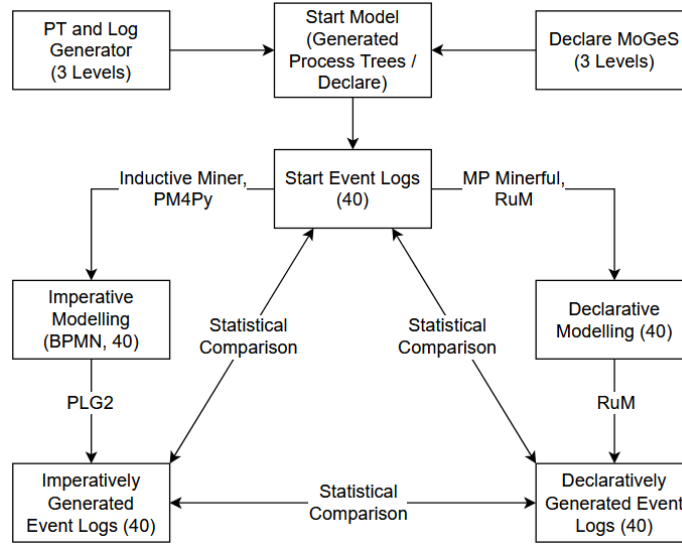
Er zijn dus generatoren ontwikkeld voor zowel imperatieve als declaratieve modellen. Het onderzoeken en testen van die verschillende generatoren, heeft vervolgens geleid tot de onderzoeksvraag van dit werk. Dit onderzoek zal die generatoren daarom gebruiken om te bepalen of de keuze van de stroming een invloed heeft op de generatie van artificiële event logs.

3 Methodologie

Dit onderzoek heeft als doel te achterhalen of de keuze van het type model - imperatief of declaratief - een invloed heeft op het genereren van artificiële event logs. Het onderzoek gebruikt daarvoor de onderzoeksopzet uit figuur 1, die uit 4 concrete stappen bestaat:

1. **Generatie Startpunten:** Er worden via de PT and Log Generator en via Declare MoGeS startmodellen gegenereerd, die gebruikt worden om per iteratie één set van start event logs te genereren [8, 11].
2. **Model Discovery:** De start event logs worden gebruikt om twee modellen op te stellen per event log. Via het Inductive Miner algoritme wordt er een BPMN-model opgesteld en via het MP MINERful algoritme wordt er een declare model opgesteld voor elke start event log in de set [1, 13].
3. **Generatie Artificiële Event Logs:** Elk model wordt vervolgens gebruikt om nieuwe artificiële event logs te genereren. De BPMN-modellen worden via PLG2 gesimuleerd om event logs te bekomen, terwijl de declare modellen gesimuleerd worden met behulp van RuM [1, 5]. Dat leidt per iteratie tot een set imperatief gegenereerde event logs, afkomstig van de BPMN-modellen, en een set declaratief gegenereerde event logs, afkomstig van declare modellen, waarbij beide sets gebaseerd zijn op dezelfde start event logs.
4. **Vergelijking Event Logs:** de start event logs, imperatief gegenereerde event logs en declaratief gegenereerde event logs worden per iteratie statistisch met elkaar vergeleken aan de hand van hun karakteristieken.

In de rest van deze sectie, zullen de design-keuzes en de details van de onderzoeksopzet verder toegelicht worden.



Figuur 1. De Gebruikte Onderzoeksopzet

3.1 Generatie Startpunten

Voor dit onderzoek, zal er gebruik gemaakt worden van 6 startpunten. De startpunten zijn sets van 40 event logs waarvan 3 sets gegenereerd zijn op basis van imperatieve modellen en 3 op basis van declaratieve modellen. Er is gekozen voor 3 startpunten van elke stroming om twee redenen. Ten eerste kunnen zowel event logs uit imperatieve als declaratieve modellen als startpunt gebruikt worden. Imperatieve generatiemethodes zouden bijvoorbeeld een oneerlijk voordeel kunnen hebben als er alleen start event logs gebruikt werden die volgens imperatieve methodes gegenereerd waren. Ten tweede kan zo de complexiteit van de event logs betrokken worden bij het onderzoek. De sets zijn namelijk verdeeld in levels, waarbij level 1 telkens het minst complex is en level 3 het meest complex. De complexiteit wijst daarbij op de hoeveelheid mogelijk gedrag in de event logs. De meest complexe event logs bevatten het meeste mogelijke gedrag. Daarnaast worden er telkens 40 event logs gegenereerd om ruim te voldoen aan de centrale limietstelling, die minstens 30 observaties nodig heeft [10]. Het gebruik van de centrale limietstelling zal verder toegelicht worden in de sectie ‘Vergelijking van Event Logs’.

De imperatieve sets van startpunten zijn ontworpen op basis van Process Trees die gegenereerd zijn aan de hand van de PT and Log Generator [8]. De gebruikte parameters zijn te vinden in tabel 1. De parameters zijn zo gekozen dat er drie sets van startpunten van verschillende complexiteit zijn. Event logs uit Level 1 kunnen bijvoorbeeld slechts 8 activiteiten bevatten, terwijl event logs uit level 3 tot 15 verschillende activiteiten kunnen bevatten. Daarnaast zijn de kansen voor bijvoorbeeld parallele activiteiten en loops hoger bij level 3, wat voor meer mogelijke paden en meer complexiteit zorgt.

Level	Mode	Min	Max	Sequence	Choice	parallel	loop	or	silent	duplicate	lt_dependency	infrequent	no_models	unfold	max	repeat
1	6	4	8	0.9	0.2	0	0	0.2	0	0	0	0	40	0	0	0
2	8	4	12	0.75	0.2	0.2	0	0.2	0	0	0	0	40	0	0	0
3	10	5	15	0.5	0.3	0.3	0.3	0.3	0	0	0	0	40	0	0	0

Tabel 1. PT Log: gebruikte parameters

De declaratieve sets van startpunten zijn ontwikkeld met behulp van DeclareMoGeS [11]. Die tool staat het genereren van declare modellen op basis van enkele parameters toe. Tabel 2 toont de gebruikte parameters. In DeclareMoGeS kan bovendien aangegeven worden welke relaties gebruikt mogen worden. Daarbij zijn de gebruikte relaties de volgende: AlternatePrecedence, AlternateResponse, ChainPrecedence, ChainResponse, Precedence, RespondedExistence, Response. Die relaties zijn geselecteerd omdat ze ook gebruikt worden in de MP MINERful methode om modellen te ontdekken [1]. Opnieuw is er met drie levels gewerkt om startpunten van verschillende complexiteit te bekomen. In dit geval zijn er bijvoorbeeld meer regels en mogelijke activiteiten voor de level 3-modellen dan voor de level 1- en level 2-modellen wat leidt tot complexere modellen.

Level	Alphabet Size	Set Size
1	8	6
2	12	8
3	15	10

Tabel 2. Declare MoGeS: gebruikte parameters

Daarbij is het belangrijk om te vermelden dat er in dit werk alleen gebruik gemaakt is van artificiële data. De argumentatie achter die keuze is vooral gebaseerd op controle over de data. Door artificiële data te gebruiken, konden namelijk makkelijk niveaus van complexiteit gebruikt worden om te onderzoeken of de eventuele invloed van de process mining stroming ook afhankelijk is van de complexiteit van de onderliggende modellen.

3.2 Generatie Modellen

Vervolgens zullen de start event logs gebruikt worden om nieuwe modellen op te stellen. Voor de ontwikkeling van de imperatieve modellen, wordt er via PM4Py gewerkt met het Inductive Miner algoritme [4]. Het Inductive Miner algoritme is namelijk zowel goed in het ontdekken van lineaire processes als processen met meerdere takken [13]. Het is dus algemeen toepasbaar wat geschikt is voor dit onderzoek, aangezien er gewerkt wordt met verschillende levels van complexiteit. In dit werk werd er bovendien voor dat algoritme een noise threshold gehanteerd van 0.2.

Om vervolgens declaratieve modellen te ontwerpen vanuit de start event logs, zal er gebruik gemaakt worden van het MP MINERful algoritme via de RuM-tool [1]. MINERful is geselecteerd omdat het een efficient en schaalbaar algoritme is om declaratieve process modellen te ontdekken [7]. Aangezien het algoritme toegepast zal worden op heel wat verschillende event logs, is de schaalbaarheid zeer handig.

3.3 Generatie Artificiële Event Logs

Na het ontwerpen van de BPMN-modellen, zullen die modellen gebruikt worden om nieuwe artificiële event logs te genereren. Daarbij zal er gebruik gemaakt worden van PLG2 aangezien die tool BPMN-modellen kan simuleren om event logs te bekomen [5]. De andere variabelen die aangepast kunnen worden binnen PLG2 worden op hun standaardwaarde gehouden, wat overal 0 of 1 is.

Vervolgens zullen de declaratieve modellen gebruikt worden om nieuwe event logs te genereren. Opnieuw zullen RuM en de MP MINERful methode gehanteerd worden om de event logs te genereren. Om logs te genereren moet er aan RuM daarnaast nog aangegeven worden wat het minimum en maximum aantal events per trace moet zijn. Daarvoor worden dezelfde minima en maxima per level gebruikt die bij de generatie van de imperatieve startpunten gebruikt werden en terug te vinden zijn in tabel 1.

Voor zowel de imperatief gegenereerde logs als de declaratief gegenereerde logs, werden er in deze stap, per level, 40 event logs met elk 1000 paden gegenereerd, die vervolgens met elkaar en met de start event logs vergeleken konden worden.

3.4 Vergelijking Event Logs

Voor de vergelijking van de event logs, zullen enkele variabelen gebruikt worden. De gebruikte variabelen zijn de volgende:

- Entropie: de maatstaf die aangeeft hoe complex en variabel een event log is door middel van de frequenties van alle mogelijke paden en subpaden. Een hoge entropie wijst daarbij op een complexe event log. Deze maatstaf komt overeen met de flattened block entropy uit het werk van Back et al. (2019) [3].
- Gem. Lengte Paden: De gemiddelde lengte van de paden die in een event log voorkomen.
- Aantal Paden: Het aantal unieke paden die in een event log voorkomen.
- DF-relaties: Het aantal directly-follows relaties. Een directly-follows relatie is een set van 2 opeenvolgende gebeurtenissen in een pad[9].
- 3-Grams: Het aantal verschillende subpaden van lengte 3. [9].
- 4-Grams: Het aantal verschillende subpaden van lengte 4. [9].

Voor de vergelijkingen van de event logs, zal er gebruik gemaakt worden van de Welch's Two Sample t-tests [20]. Om die statistische test te mogen gebruiken, moeten enkele voorwaarden voldaan zijn. Enerzijds moeten de waarden van de observaties een normaalverdeling volgen. Aangezien er voor elke set 40 event logs gegenereerd zijn, kan er dankzij de centrale limietstelling vanuitgegaan worden dat die voorwaarde voldaan is [10]. Anderzijds moeten de observaties onafhankelijk van elkaar zijn. Hoewel de imperatief en declaratief gegenereerde event logs zich baseren op dezelfde startpunten, worden ze volledig onafhankelijk van elkaar ontwikkeld met andere methodes. De onafhankelijkheidsvoorwaarde is dus ook voldaan [20]. De test is daarnaast geschikt wanneer de varianties van de populaties niet gelijk zijn. Aangezien dit werk onderzoekt of er verschillen tussen de event logs zijn, kan hier niet aangenomen worden dat de varianties gelijk zijn, waardoor de Welch's Two Sample t-test hier aangewezen is. Voor die t-tests zal een betrouwbaarheid van 95% gehanteerd worden. Dankzij de Welch's Two Sample t-tests kunnen zo eventuele significante verschillen tussen de event logs opgespoord worden, op basis van de geselecteerde eigenschappen.

Ten slotte is het belangrijk te vermelden dat er gedurende het onderzoek verschillende scripts gebruikt werden. Die scripts zijn specifiek voor dit onderzoek ontwikkeld en zullen via onderstaande github¹ gedeeld worden. Vooral de java scripts kunnen gebruikt worden in toekomstig onderzoek om modelgeneratie in RuM en de generatie van event logs in PLG2 en RuM te automatiseren.

¹ De GitHub Folder: <https://github.com/Teg-Atreides/Thesis-Bots>

4 Resultaten

In deze sectie worden de resultaten besproken. Voor elk level is een tabel opgesteld waarmee de resultaten getoond worden. In elke tabel zijn de volgende drie vergelijkingen gebruikt:

- Start - Imperatief: de start event logs tegenover de imperatief gegenereerde logs.
- Start - Declaratief: de start event logs tegenover de declaratief gegenereerde logs.
- Imperatief - Declaratief: de imperatief gegenereerde logs tegenover de declaratief gegenereerde logs.

De gehanteerde nulhypothese was daarbij voor elke variabele, bij elke vergelijking, de volgende: Het verschil tussen de variabele van beide sets event logs is gelijk aan 0. Instanties waarbij die hypothese verworpen werd en er dus een significant verschil gevonden werd, zijn aangeduid in het rood. Voor elke test en bijbehorende interpretatie is daarbij een betrouwbaarheidsniveau van 95% gebruikt.

4.1 Imperatieve Startpunten

Level 1 Voor het eenvoudigste level van imperatieve startpunten, tonen de t-testen in tabel 3 dat er geen statistisch bewijs is dat aantoonde dat er verschillen zijn tussen de start event logs en de imperatief gegenereerde event logs. Dat geldt voor alle geselecteerde variabelen. De declaratief gegenereerde event logs zijn daarentegen significant verschillend vergeleken met de basis en imperatief gegenereerde event logs. De declaratief gegenereerde event logs hebben namelijk een hogere entropie, langere paden en meer DF-relaties. Uit de drie vergelijkingen voor dit level blijkt bovendien dat de start event logs zeer gelijkend zijn met de imperatieve event logs. Dat is logisch want het gaat over simpele BPMN-modellen waarbij er niet veel verschillend gedrag mogelijk is. Daarom is de vergelijking Start-Declaratief hier identiek aan Imperatief-Declaratief.

Level 2 Voor het tweede level van de imperatieve startpunten, wordt een verschillend beeld waargenomen, zoals tabel 4 weergeeft. De start event logs vertonen nu namelijk significante verschillen ten opzichte van de imperatieve event logs op vlak van elke geselecteerde variabele behalve het aantal 4-grams. De imperatieve event logs hebben een hogere entropie, met gemiddeld langere paden, meer mogelijke paden, maar minder DF-relaties en minder 3-grams dan de start event logs. Opnieuw verschillen de declaratieve event logs significant voor elke variabele van de start event logs. De declaratieve logs bevatten een hogere complexiteit, langere paden, meer mogelijke paden, meer DF-relaties en meer 3-grams en 4-grams. De imperatieve en declaratieve event logs verschillen daarnaast ook significant op basis van elke variabele buiten de gemiddelde lengte van de paden. De declaratieve logs zijn daarbij het meest complex, met de meeste mogelijke paden, DF-relaties, 3-grams en 4-grams.

Variabele	t-waarde	Vrijheids- graden	P-waarde	95% Betrouwbaarheids- interval
Start - Imperatief				
Entropie	0.00	78.00	1.00	[-0.208; 0.208]
Gem. Lengte Paden	0.00	78.00	1.00	[-0.657; 0.657]
Aantal Paden	0.00	78.00	1.00	[-0.355; 0.355]
DF Relaties	0.00	78.00	1.00	[-0.542; 0.542]
3-Grams	0.00	78.00	1.00	[-0.576; 0.576]
4-Grams	0.00	78.00	1.00	[-0.550; 0.550]
Start - Declaratief				
Entropie	-4.00	56.06	0.00	[-1.025; -0.341]
Gem. Lengte Paden	-3.62	62.36	0.00	[-1.514; -0.437]
Aantal Paden	-1.65	39.00	0.11	[-114.505; 11.555]
DF Relaties	-2.35	40.21	0.02	[-6.828; -0.522]
3-Grams	-1.75	39.04	0.09	[-36.177; 2.577]
4-Grams	-1.57	39.00	0.13	[-155.542; 19.792]
Imperatief - Declaratief				
Entropie	-4.00	56.06	0.00	[-1.025; -0.341]
Gem. Lengte Paden	-3.62	62.36	0.00	[-1.514; -0.437]
Aantal Paden	-1.65	39.00	0.11	[-114.505; 11.555]
DF Relaties	-2.35	40.21	0.02	[-6.828; -0.522]
3-Grams	-1.75	39.04	0.09	[-36.177; 2.577]
4-Grams	-1.57	39.00	0.13	[-155.542; 19.792]

Tabel 3. Imperatief - Level1: Vergelijking

Level 3 Bij level 3, het meest complexe level van de imperatieve startpunten, komt er een gelijkend beeld als het vorige level naar boven. Er zijn slechts twee variabelen waarbij niet bewezen kon worden dat ze verschillend zijn. Het gaat dan over de 3-grams en 4-grams bij de vergelijking van de start event logs en de imperatief gegenereerde event logs. Voor alle andere vergelijkingspunten tussen de 3 sets van event logs, zijn er significante verschillen gevonden, zoals tabel 5 aantoont. Opnieuw hebben de imperatieve logs een hogere entropie dan de start event logs, met gemiddelde langere paden en meer mogelijke paden. De start event logs hebben daarentegen wel meer DF-relaties. De declaratieve event logs zijn daarnaast het meest complex met de langste paden, de meeste mogelijke paden, de meeste mogelijke DF-relaties, 3-grams en 4-grams.

4.2 Declaratieve Startpunten

Level 1 Het eerste level van de declaratief gegenereerde startpunten, toont vervolgens evenzeer verschillen tussen alle sets van event logs, zoals opgemerkt kan worden in tabel 6. De imperatief gegenereerde logs verschillen significant op vlak van alle variabelen buiten het aantal DF-relaties. De imperatieve logs zijn daarbij complexer, bevatten langere en meer mogelijke paden met ook meer 3-grams en 4-grams dan de start event logs. De start event logs bevatten daarnaast

Variabele	t-waarde	Vrijheids- graden	P-waarde	95% Betrouwbaarheids- interval
Start - Imperatief				
Entropie	-11.82	60.14	0.00	[-4.476; -3.180]
Gem. Lengte Paden	-10.15	63.26	0.00	[-7.019; -4.710]
Aantal Paden	-12.11	39.74	0.00	[-655.194; -467.706]
DF Relaties	8.69	47.22	0.00	[22.802; 36.548]
3-Grams	5.25	75.94	0.00	[28.340; 63.010]
4-Grams	0.66	66.13	0.51	[-21.691; 43.291]
Start - Declaratief				
Entropie	-21.34	63.13	0.00	[-7.104; -5.887]
Gem. Lengte Paden	-12.09	78.00	0.00	[-5.844; -4.191]
Aantal Paden	-28.99	40.68	0.00	[-962.785; -837.365]
DF Relaties	-2.92	69.75	0.00	[-27.861; -5.239]
3-Grams	-8.05	40.22	0.00	[-538.437; -322.463]
4-Grams	-10.39	39.09	0.00	[-3241.308; -2184.792]
Imperatief - Declaratief				
Entropie	-6.89	77.49	0.00	[-3.438; -1.897]
Gem. Lengte Paden	1.47	63.22	0.15	[-0.307; 2.001]
Aantal Paden	-6.11	67.88	0.00	[-449.260; -227.990]
DF Relaties	-9.70	43.05	0.00	[-55.840; -36.610]
3-Grams	-8.93	39.88	0.00	[-583.905; -368.345]
4-Grams	-10.42	39.22	0.00	[-3252.487; -2195.213]

Tabel 4. Imperatief - Level2: Vergelijking

een hogere complexiteit, meer paden, meer DF-relaties, meer 3-grams en 4-grams dan de declaratieve event logs. Voor de gemiddelde lengte van de paden, werd er geen significant verschil waargenomen tussen de declaratieve event logs en de start event logs. De imperatieve event logs verschillen daarnaast op elk vlak van de declaratieve event logs, waarbij de imperatieve logs complexer zijn met langere paden, meer mogelijke paden, meer DF-relaties, 3-grams en 4-grams.

Level 2 Voor level 2 van de declaratieve startpunten, worden er opnieuw veel verschillen waargenomen, zoals tabel 7 toont. Ondanks het feit dat de startpunten vanuit declaratieve modellen opgesteld zijn, zijn er toch 3 variabelen waarbij er geen bewijs voor een verschil tussen de start event logs en de imperatieve event logs waargenomen werd. Dat zijn namelijk het aantal paden, het aantal DF-relaties en het aantal 3-grams. De declaratieve logs verschillen daarnaast op vlak van elke variabele van de start event logs, behalve voor de gemiddelde lengte van de paden. De start event logs zijn daarbij complexer, met meer mogelijke paden en meer DF-relaties, 3-grams en 4-grams. De imperatieve en declaratieve logs verschillen ook op bijna elk vlak. De imperatieve logs zijn daarbij wel het meest complex, met de langste gemiddelde paden, het grootste aantal DF-relaties en het grootste aantal 3- en 4-grams.

Variabele	t-waarde	Vrijheids- graden	P-waarde	95% Betrouwbaarheids- interval
Start - Imperatief				
Entropie	-7.69	77.90	0.00	[-3.599; -2.118]
Gem. Lengte Paden	-7.28	67.47	0.00	[-5.477; -3.119]
Aantal Paden	-9.89	70.12	0.00	[-620.718; -412.432]
DF Relaties	5.99	71.06	0.00	[22.793; 45.557]
3-Grams	1.92	77.84	0.06	[-1.869; 105.419]
4-Grams	-0.86	73.85	0.39	[-200.567; 79.817]
Start - Declaratief				
Entropie	-19.24	75.33	0.00	[-7.112; -5.777]
Gem. Lengte Paden	-14.16	74.38	0.00	[-8.404; -6.330]
Aantal Paden	-25.19	46.02	0.00	[-855.160; -728.590]
DF Relaties	-5.02	74.15	0.00	[-52.203; -22.547]
3-Grams	-10.54	42.86	0.00	[-1125.458; -763.792]
4-Grams	-15.61	40.08	0.00	[-6541.128; -5041.272]
Imperatief - Declaratief				
Entropie	-10.48	74.33	0.00	[-4.268; -2.904]
Gem. Lengte Paden	-4.80	75.26	0.00	[-4.342; -1.796]
Aantal Paden	-6.31	42.52	0.00	[-363.299; -187.301]
DF Relaties	-10.64	62.18	0.00	[-84.988; -58.112]
3-Grams	-11.14	42.52	0.00	[-1176.899; -815.901]
4-Grams	-15.38	40.76	0.00	[-6483.553; -4978.097]

Tabel 5. Imperatief - Level3: Vergelijking

Level 3 Level 3 van de declaratieve startpunten toont opnieuw veel verschillen tussen de sets van event logs, zoals in tabel 8 weergegeven wordt. De imperatieve event logs verschillen op elke variabele buiten het aantal DF-relaties en het aantal 3-grams van de start event logs. Zoals in het vorige level van de declaratieve startpunten, zijn de imperatieve logs complexer met gemiddeld langere paden en meer 4-grams. De start event logs hebben wel meer mogelijke paden dan de imperatieve event logs. De declaratief gegenereerde logs tonen daarbij minder significante verschillen vergeleken met de start event logs. Voor de entropie, de gemiddelde lengte van de paden en het aantal mogelijke paden werd namelijk onvoldoende bewijs gevonden om te spreken van significante verschillen. Daarentegen bevatten de start event logs wel nog meer DF-relaties en meer 3- en 4-grams dan de declaratieve event logs. Verder verschillen de imperatieve logs opnieuw van de declaratief gegenereerde logs op vlak van bijna elke variabele. De imperatief gegenereerde logs zijn daarbij complexer, met gemiddeld langere paden, meer mogelijke DF-relaties en meer mogelijke 3-grams en 4-grams.

Variabele	t-waarde	Vrijheids- graden	P-waarde	95% Betrouwbaarheids- interval
Start - Imperatief				
Entropie	-11.39	63.48	0.00	[-3.575; -2.508]
Gem. Lengte Paden	-17.41	49.95	0.00	[-5.779; -4.583]
Aantal Paden	-4.21	76.55	0.00	[-158.672; -56.728]
DF Relaties	-1.91	77.46	0.06	[-12.769; 0.269]
3-Grams	-2.07	77.31	0.04	[-116.502; -2.248]
4-Grams	-2.64	70.29	0.01	[-678.079; -94.121]
Start - Declaratief				
Entropie	2.88	60.92	0.01	[0.235; 1.296]
Gem. Lengte Paden	-1.79	75.57	0.08	[-0.538; 0.028]
Aantal Paden	2.94	56.28	0.00	[38.652; 204.587]
DF Relaties	4.40	72.35	0.00	[8.426; 22.379]
3-Grams	3.83	75.22	0.00	[52.059; 165.183]
4-Grams	3.57	74.80	0.00	[199.184; 702.206]
Imperatief - Declaratief				
Entropie	11.78	75.98	0.00	[3.163; 4.451]
Gem. Lengte Paden	16.74	47.96	0.00	[4.335; 5.518]
Aantal Paden	5.69	52.02	0.00	[148.420; 310.219]
DF Relaties	5.97	74.48	0.00	[14.428; 28.877]
3-Grams	5.65	76.00	0.00	[108.780; 227.212]
4-Grams	5.51	72.72	0.00	[534.275; 1139.315]

Tabel 6. Declaratief - Level1: Vergelijking

5 Discussie

Het onderzoek is opgezet om te achterhalen of de keuze voor process mining stroming - imperatief of declaratief - een invloed heeft op de generatie van artificiële event logs. Om dat te onderzoeken zijn drie sets van start event logs ontwikkeld volgens imperatieve modellen en drie sets volgens declaratieve modellen. Die sets zijn gebruikt om nieuwe imperatieve en declaratieve modellen te ontdekken die vervolgens gebruikt werden om artificiële event logs te genereren. De imperatief en declaratief gegenereerde event logs, die uiteindelijk telkens gebaseerd waren op dezelfde start event logs, konden dan vergeleken worden op basis van entropie [3], de gemiddelde lengte van de paden, het aantal verschillende paden, het aantal DF-relaties [9] en het aantal 3- en 4-grams [9]. Die vergelijking werd bovendien opgesteld met behulp van Welch's Two Sample t-tests waarbij een betrouwbaarheidsniveau van 95% gebruikt werd [20].

De resultaten toonden dat de process mining stroming wel degelijk een invloed had op de generatie van de artificiële event logs. Voor de imperatieve start event logs, waren de declaratief gegenereerde event logs telkens het meest complex met langere paden, meer mogelijke paden, meer DF-relaties en meer 3- en 4-grams. Dat beeld komt overeen met de logica rond declaratieve modellen. Ze zijn name-

Variabele	t-waarde	Vrijheids- graden	P-waarde	95% Betrouwbaarheids- interval
Start - Imperatief				
Entropie	-7.37	65.19	0.00	[-3.371; -1.933]
Gem. Lengte Paden	-10.83	51.57	0.00	[-9.368; -6.438]
Aantal Paden	0.98	42.40	0.33	[-25.620; 73.916]
DF Relaties	-0.48	74.74	0.63	[-14.913; 9.143]
3-Grams	-1.69	68.35	0.10	[-315.263; 26.128]
4-Grams	-4.36	49.27	0.00	[-3046.117; -1123.683]
Start - Declaratief				
Entropie	2.13	63.67	0.04	[0.051; 1.530]
Gem. Lengte Paden	-1.38	76.83	0.17	[-1.395; 0.254]
Aantal Paden	3.18	41.00	0.00	[34.427; 154.074]
DF Relaties	6.59	64.47	0.00	[33.562; 62.770]
3-Grams	4.93	70.67	0.00	[241.013; 568.468]
4-Grams	3.62	65.66	0.00	[513.755; 1781.163]
Imperatief - Declaratief				
Entropie	7.89	75.88	0.00	[2.573; 4.312]
Gem. Lengte Paden	9.99	52.40	0.00	[5.860; 8.805]
Aantal Paden	1.86	73.37	0.07	[-4.974; 145.179]
DF Relaties	6.63	70.23	0.00	[35.691; 66.411]
3-Grams	5.71	75.71	0.00	[357.750; 740.865]
4-Grams	6.23	61.98	0.00	[2194.805; 4269.913]

Tabel 7. Declaratief - Level2: Vergelijking

lijk ontworpen om tegemoet te komen aan meer ongestructureerde processen en zouden dus meer gedrag moeten toelaten dan imperatieve modellen [15]. De declaratieve startpunten toonden daarentegen een ander beeld. Het waren namelijk de imperatief gegenereerde event logs die daar meer mogelijk gedrag leken te bevatten. Over de drie declaratieve levels heen, bevatten de imperatief gegenereerde event logs gemiddeld een hogere entropie, langere paden en meer verschillende lokale patronen dan de declaratieve event logs en de start event logs. Bij elk level waren er dus significante verschillen tussen de imperatieve en declaratieve event logs. Die verschillen waren sterker nog afhankelijk van de stroming van de start event logs. Dat wijst erop dat de onderliggende process mining stroming een significante invloed heeft op de generatie van event logs.

Daarnaast tonen de resultaten aan dat de verschillende event logs vaak ook significante verschillen vertonen ten opzichte van de start event logs. Zo vertoont het merendeel van de variabelen voor elke vergelijking met de start event logs - behalve bij de basis-imperatieve vergelijking van de imperatieve startpunten van level 1 uit tabel 3 - significante verschillen ten opzichte van de andere event logs. Dat lijkt aan te tonen dat de karakteristieken van de start event logs niet altijd goed behouden worden tijdens de event log generatie, zeker wanneer de onderliggende modellen complexer worden.

Variabele	t-waarde	Vrijheids- graden	P-waarde	95% Betrouwbaarheids- interval
Start - Imperatief				
Entropie	-4.60	56.35	0.00	[-2.709; -1.066]
Gem. Lengte Paden	-7.64	48.47	0.00	[-10.533; -6.146]
Aantal Paden	2.58	39.15	0.01	[12.351; 101.499]
DF Relaties	-0.38	76.10	0.71	[-21.918; 14.968]
3-Grams	-1.28	63.82	0.20	[-545.502; 118.752]
4-Grams	-4.36	48.94	0.00	[-5380.846; -1985.804]
Start - Declaratief				
Entropie	1.12	66.65	0.27	[-0.290; 1.027]
Gem. Lengte Paden	-1.49	77.39	0.14	[-1.862; 0.270]
Aantal Paden	1.88	39.30	0.07	[-2.138; 61.038]
DF Relaties	5.22	66.83	0.00	[35.587; 79.713]
3-Grams	3.97	69.95	0.00	[293.690; 886.910]
4-Grams	2.23	68.43	0.03	[121.268; 2161.732]
Imperatief - Declaratief				
Entropie	4.89	72.39	0.00	[1.336; 3.176]
Gem. Lengte Paden	6.84	50.24	0.00	[5.328; 9.758]
Aantal Paden	-1.02	70.25	0.31	[-81.279; 26.329]
DF Relaties	5.25	72.76	0.00	[37.907; 84.343]
3-Grams	4.29	76.12	0.00	[430.324; 1177.026]
4-Grams	5.36	59.52	0.00	[3022.860; 6626.790]

Tabel 8. Declaratief - Level3: Vergelijking

Die bevindingen tonen zo toekomstige onderzoekspistes. Aangezien de process mining stroming een invloed heeft op de datageneratie, is er nood aan een model-onafhankelijke generator. Die zou bijvoorbeeld kunnen werken aan de hand van bepaalde karakteristieken van event logs. Dit werk, samen met de werken van Maldonado et al. (2023), Back et al. (2019) en Kabierski et al. (2025) [3, 9, 12], bevatten voorbeelden van dergelijke karakteristieken. Totdat zo een generator ontwikkeld wordt, zullen onderzoekers meer aandacht moeten schenken aan de bepaling van de geschikte process mining stroming. De foute stroming kiezen, kan namelijk onnodige complexiteit toevoegen aan de artificiële event logs.

6 Conclusie

Dit werk had als doel om te bepalen of de keuze voor een process mining stroming - imperatief of declaratief - een invloed heeft op het genereren van artificiële event logs. Daar was namelijk nog geen onderzoek naar gedaan in de huidige literatuur. Het is echter belangrijk voor onderzoekers om op de hoogte te zijn van een eventuele invloed, zodat daar rekening mee gehouden kan worden tijdens onderzoek. Om die invloed te bepalen, is er gebruik gemaakt van een experiment op basis van zes sets event logs die als startpunt gebruikt werden. Drie

sets werden ontwikkeld vanuit imperatieve modellen. De andere drie sets werden ontwikkeld vanuit declaratieve modellen. Die startpunten werden gebruikt om telkens nieuwe sets van event logs te genereren, zowel via imperatieve als via declaratieve methodes. Daardoor werden telkens imperatieve en declaratieve event logs gegenereerd op basis van dezelfde startsituatie waardoor ze met elkaar vergeleken konden worden. Op die manier kon dan achterhaald worden of er verschillen waren tussen de gegenereerde event logs, waardoor bepaald kon worden of de achterliggende process mining stroming invloed had op de event log generatie.

Het experiment leidde tot een duidelijke conclusie. De stroming van het onderliggende model heeft een invloed op de generatie van event logs. Er werd namelijk gezien dat er voor elk level significante verschillen waren tussen de imperatief en declaratief gegenereerde event logs. Die verschillen waren bovendien niet éénzijdig. Wanneer de startpunten ontwikkeld waren volgens imperatieve modellen, waren de bijbehorende declaratief gegenereerde event logs complexer. Bij de declaratieve startpunten daarentegen, vertoonden de imperatief gegenereerde event logs dan weer een complexer beeld. Daarnaast waren er op bijna elk level significante verschillen tussen de gegenereerde event logs en de startpunten. Die bevindingen vormen een sterke argumentatie voor de ontwikkeling van een nieuwe generator die bijvoorbeeld werkt op basis van de karakteristieken van een event log, zonder dat er een onderliggend model nodig is.

Verder heeft dit onderzoek gevolgen voor toekomstige onderzoeken waarbij artificiële event logs gegenereerd worden. Onderzoekers moeten namelijk aandacht besteden aan de keuze van hun process mining stroming. Start event logs moeten daarom onderzocht worden of ze best binnen de imperatieve of declaratieve stroming passen voordat er event logs gegenereerd kunnen worden. Dit werk heeft namelijk aangetoond dat de artificiële data complexer kan worden wanneer de verkeerde process mining stroming gebruikt wordt.

Voor deze studie werden echter wel twee limitaties geïdentificeerd. Ten eerste is er enkel gebruik gemaakt van artificiële data. Het zou daarom interessant zijn om te onderzoeken of dezelfde verschillen gevonden worden bij reële startpunten. Ten tweede zijn er telkens specifieke methodes gehanteerd. Voor de imperatieve modellering werd er bijvoorbeeld gebruik gemaakt van BPMN-modellen en het Inductive Miner algoritme. Voor toekomstig onderzoek is het daarom interessant om andere methodes te gebruiken om te bepalen of die tot andere resultaten leiden. Voor de imperatieve modellen kan zo bijvoorbeeld gebruik gemaakt worden van Petri netten en het Heuristics Miner algoritme.

Referenties

- [1] ALMAN, A., DI CICCIO, C., HAAS, D., MAGGI, F. M., AND NOLTE, A. Rule mining with rum. In *2020 2nd International Conference on Process Mining (ICPM)* (2020), IEEE, pp. 121–128.
- [2] BACK, C. O., DEBOIS, S., AND SLAATS, T. Towards an empirical evaluation of imperative and declarative process mining. In *International conference on conceptual modeling* (2018), Springer, pp. 191–198.
- [3] BACK, C. O., DEBOIS, S., AND SLAATS, T. Entropy as a Measure of Log Variability. *Journal on Data Semantics* 8, 2 (June 2019), 129–156.
- [4] BERTI, A., VAN ZELST, S. J., AND VAN DER AALST, W. Process Mining for Python (PM4Py): Bridging the Gap Between Process- and Data Science, May 2019. arXiv:1905.06169 [cs].
- [5] BURATTIN, A. PLG2: Multiperspective Processes Randomization and Simulation for Online and Offline Settings, July 2016. arXiv:1506.08415 [cs].
- [6] DI CICCIO, C., BERNARDI, M. L., CIMITILE, M., AND MAGGI, F. M. Generating event logs through the simulation of declare models. In *Enterprise and Organizational Modeling and Simulation: 11th International Workshop, EOMAS 2015, Held at CAiSE 2015, Stockholm, Sweden, June 8-9, 2015, Selected Papers 11* (2015), Springer, pp. 20–36.
- [7] DI CICCIO, C., SCHOUTEN, M. H., DE LEONI, M., MENDLING, J., ET AL. Declarative process discovery with minerful in prom. In *CEUR Workshop Proceedings* (2015), vol. 1418, CEUR-WS. org, pp. 60–64.
- [8] JOUCK, T., AND DEPAIRE, B. Generating Artificial Data for Empirical Analysis of Control-flow Discovery Algorithms. *Business & Information Systems Engineering* 61, 6 (Dec. 2019), 695–712.
- [9] KABIERSKI, M., RICHTER, M., AND WEIDLICH, M. Quantifying and relating the completeness and diversity of process representations using species estimation. *Information Systems* 130 (2025), 102512.
- [10] KWAK, S. G., AND KIM, J. H. Central limit theorem: the cornerstone of modern statistics. *Korean journal of anesthesiology* 70, 2 (2017), 144.
- [11] LAGHMOUCH, M., DEPAIRE, B., GIGANTE, N., JANS, M., AND MONTALI, M. Declare moGES: Model generator and specializer. In *ICPM Doctoral Consortium/Demo* (2023).
- [12] MALDONADO, A., TAVARES, G. M., OYAMADA, R. S., CERAVOLO, P., AND SEIDL, T. Feed: Feature extraction from event data. In *ICPM Doctoral Consortium/Demo* (2023).
- [13] NURITHA, I., AND MAHENDRAWATHI, E. Structural similarity measurement of business process model to compare heuristic and inductive miner

- algorithms performance in dealing with noise. *Procedia Computer Science* 124 (2017), 255–263.
- [14] OPENAI. Chatgpt: Ai language model, 2025. Accessed March 2025.
 - [15] PESIC, M., SCHONENBERG, H., AND VAN DER AALST, W. M. DECLARE: Full Support for Loosely-Structured Processes. In *11th IEEE International Enterprise Distributed Object Computing Conference (EDOC 2007)* (Oct. 2007), pp. 287–287. ISSN: 1541-7719.
 - [16] SKYDANIENKO, V., DI FRANCESCO MARINO, C., GHIDINI, C., AND MAGGI, F. M. A tool for generating event logs from multi-perspective declare models. *BPM (Dissertation/Demos/Industry) 2196* (2018), 111–115.
 - [17] VAN DER AALST, W., ADRIANSYAH, A., DE MEDEIROS, A. K. A., ARCIERI, F., BAIER, T., BLICKLE, T., BOSE, J. C., VAN DEN BRAND, P., BRANDTJEN, R., BUIJS, J., ET AL. Process mining manifesto. In *Business Process Management Workshops: BPM 2011 International Workshops, Clermont-Ferrand, France, August 29, 2011, Revised Selected Papers, Part I 9* (2012), Springer, pp. 169–194.
 - [18] VANDEN BROUCKE, S. K. L. M., DELVAUX, C., FREITAS, J., ROGOVA, T., VANTHIENEN, J., AND BAESSENS, B. Uncovering the Relationship Between Event Log Characteristics and Process Discovery Techniques. In *Business Process Management Workshops* (Cham, 2014), N. Lohmann, M. Song, and P. Wohed, Eds., Springer International Publishing, pp. 41–53.
 - [19] WANG, J., WONG, R. K., DING, J., GUO, Q., AND WEN, L. On Recommendation of Process Mining Algorithms. In *2012 IEEE 19th International Conference on Web Services* (June 2012), pp. 311–318.
 - [20] WELCH, B. L. The Generalization of ‘Student’s’ Problem when Several Different Population Variances are Involved. *Biometrika* 34, 1/2 (1947), 28–35. Publisher: [Oxford University Press, Biometrika Trust].