

2014•2015
FACULTEIT BEDRIJFSECONOMISCHE WETENSCHAPPEN
*master in de toegepaste economische wetenschappen:
handelsingenieur in de beleidsinformatica*

Masterproef

Het ontwikkelen van realistische simulatiemodellen door het onderscheiden van entiteitentypes met behulp van process mining

Promotor :
Prof. dr. Benoit DEPAIRE

Copromotor :
De heer Niels MARTIN

Brighid Verstrepen

Scriptie ingediend tot het behalen van de graad van master in de toegepaste economische wetenschappen: handelsingenieur in de beleidsinformatica

2014•2015

FACULTEIT BEDRIJFSECONOMISCHE
WETENSCHAPPEN

*master in de toegepaste economische wetenschappen:
handelsingenieur in de beleidsinformatica*

Masterproef

Het ontwikkelen van realistische simulatiemodellen door
het onderscheiden van entiteitentypes met behulp van
process mining

Promotor :
Prof. dr. Benoit DEPAIRE

Copromotor :
De heer Niels MARTIN

Brighid Verstrepen

*Scriptie ingediend tot het behalen van de graad van master in de toegepaste economische
wetenschappen: handelsingenieur in de beleidsinformatica*

WOORD VOORAF

Deze masterproef kwam tot stand als afsluiting van mijn masteropleiding handelsingenieur in de beleidsinformatica aan de Universiteit Hasselt. Het onderwerp werd reeds gekozen in het eerste masterjaar. In het kader van dit eindwerk heb ik veel van de kennis die ik de voorbije jaren heb opgedaan, kunnen gebruiken. Dit onderwerp brengt dan ook verschillende domeinen samen en laat zien hoe deze samen tot oplossingen van reële problemen kunnen leiden. Deze thesis bevindt zich namelijk in een relatief nieuw onderzoeksdomein dat potentieel oplossingen zou kunnen bieden voor kritieken waarmee simulatietechnieken geconfronteerd worden.

Bij het schrijven van deze thesis werd ik bijgestaan door mijn promotor, Prof. dr. Benoît Depaire en mijn co-promotor Niels Martin. Ik zou hen hierbij graag willen bedanken voor alle feedback en bijsturingen die ik gekregen heb. Hun deskundige adviezen hebben ongetwijfeld geholpen om dit werk tot een goed einde te kunnen brengen.

Daarnaast gaat mijn dank ook uit naar mijn ouders, broer en vriend die mij de voorbije jaren steeds onvoorwaardelijk gesteund hebben. Hen wil ik graag bedanken voor het vele nalezen en het luisteren naar mijn gedachten, problemen en twijfels.

Brighid Verstrepen

SAMENVATTING

In de huidige competitieve bedrijfswereld is een procesgeoriënteerde managementstijl essentieel. Processen worden bijgevolg steeds vaker ondersteund door informatiesystemen zoals onder andere ERP en CRM. Dergelijke systemen genereren event logs die informatie verschaffen over de werking van de bedrijfsprocessen. Deze logs kunnen met behulp van process mining technieken geanalyseerd worden om op die manier vervolgens structurele verbeteringen te kunnen aanbrengen.

Daarnaast zijn er simulatietechnieken beschikbaar die de mogelijkheid bieden om te experimenteren met systemen, wat in de realiteit vaak duur, moeilijk of zelfs onmogelijk is. Zo kunnen de resultaten van verschillende alternatieven vergeleken worden alvorens deze te implementeren. Enkele voordelen van het gebruik van simulatietechnieken zijn de objectieve ondersteuning van bedrijfsbeslissingen, een toegenomen inzicht in het bedrijfsproces en verhoogde creativiteit bij het zoeken naar mogelijke verbeteringen. Toch werden de laatste jaren enkele belangrijke beperkingen van deze technieken aan het licht gebracht. Zo zouden simulatietechnieken onvoldoende ondersteuning bieden voor operationele beslissingen, te weinig gebruik maken van beschikbare data en resources naïef modelleren.

Onderzoek om deze beperkingen aan te pakken heeft onder andere geleid tot het combineren van process mining en simulatietechnieken. Tussen deze twee onderzoeksdomeinen zijn er namelijk raakvlakken aanwezig. Door beide technieken te integreren wordt getracht sneller tot een eerste simulatiemodel te komen. Een andere doelstelling hiervan is het creëren van een simulatiemodel dat dichter aanleunt bij de realiteit omdat dit gebaseerd is op data omtrent de uitvoering van het desbetreffende bedrijfsproces. Dit zou resulteren in een snellere, meer betrouwbare ondersteuning van bedrijfsbeslissingen.

Naast resources werd tot nog toe eveneens onvoldoende aandacht besteed aan de entiteiten binnen een systeem. Indien bij entiteiten relevante types onderscheiden kunnen worden, kan het belangrijk zijn om informatie hieromtrent te verwerken bij het opstellen van een simulatiemodel. Relevante entiteitentypes zijn types van entiteiten waarvoor de uitvoering van een proces verschillend is. Zo kunnen het aankomstritme, de duurtijd van activiteiten en gemaakte keuzes in een proces verschillend zijn voor deze types.

In deze masterproef werd gezocht naar een manier om, naast het proces, relevante entiteitentypes te ontdekken vanuit een beschikbare event log. Vervolgens zal de event log gebruikt worden om meer kennis op te doen over de aanwezige types en de procesuitvoering voor elk type. Deze kennis zal vervolgens gebruikt worden bij het opstellen van een simulatiemodel dat idealiter dichter aanleunt bij de realiteit. Er werd gewerkt met een artificiële event log die gecreëerd werd met behulp van een vooropgesteld procesmodel waarin enkele relevante entiteitentypes aanwezig zijn.

INHOUDSOPGAVE

Woord vooraf.....	I
Samenvatting	II
Inhoudsopgave	III
Lijst van figuren	VI
Lijst van tabellen.....	VII
Hoofdstuk 1: Probleemstelling.....	1
1 Inleiding.....	1
2 Onderzoeksvraag.....	2
3 Deelvragen	3
4 Onderzoeksaanpak	4
Hoofdstuk 2: Literatuurstudie	7
1 Bedrijfsprocessen	7
1.1 Definitie	7
1.2 Modelleringstaal	7
2 Simulatie.....	9
2.1 Definitie	9
2.2 Doelstellingen en praktisch nut.....	10
2.3 Onderdelen simulatiemodel.....	11
2.4 Inputs en outputs	13
2.5 Problemen	14
3 Process mining	16
3.1 Doelstellingen	16
3.2 Event logs	17
3.3 Types.....	17
3.4 Process mining tools	19
3.5 Process mining in combinatie met simulatie	19
4 Clustering	22
4.1 Definitie	22
4.2 Types.....	22
4.3 Toepassingen	25
Hoofdstuk 3: Methodologie	29

1 Creatie artificiële event log	29
2 Beschrijvende analyse event log	30
3 Clustering entiteiten	30
3.1 Voorbereidend werk	30
3.2 Input TwoStep clustering	30
3.3 Output TwoStep clustering	31
3.4 Beschrijving gevormde clusters	31
3.5 Controle	32
4 Opstellen simulatiemodellen.....	32
4.1 Input simulatiemodel	32
4.2 Scenario's.....	32
5 Vergelijking simulatiemodellen volgens prestatie maatstaven	32
Hoofdstuk 4: Toepassing methodologie	33
1 Creatie artificiële event log	33
1.1 Het procesmodel	33
1.2 Assumpties	34
2 Beschrijvende analyse event log	40
2.1 Manipulaties log en eerste analyse.....	40
2.2 Process mining	42
2.3 Beschrijvende statistieken	45
3 Clustering entiteiten	47
3.1 Onderscheiden entiteitentypes	47
3.2 Controle	52
4 Opstellen simulatiemodellen.....	53
4.1 Input simulatiemodel	53
4.2 Scenario's.....	57
5 Vergelijking simulatiemodellen volgens prestatie maatstaven	58
5.1 Totale doorlooptijd.....	59
5.2 Totaal gemiddeld aantal entiteiten in het systeem	60
5.3 Gemiddelde wachttijden per type kassa.....	60
5.4 Bezettingsgraad kassa's	61
Hoofdstuk 5: Conclusies	63
Lijst van geraadpleegde werken	65

Bijlagen	69
Bijlage 1: Subprocedures	69
1.1 Subprocedure optellen duurtijden	69
1.2 Subprocedure optellen wachttijden	70
1.3 Subprocedure individuele duurtijd activiteiten	71
Bijlage 2: Simulatiemodel realiteit	72
Bijlage 3: Simulatiemodel zonder entiteitentypes	75
Bijlage 4: Simulatiemodel met entiteitentypes	77
Bijlage 5: Onderdeel outputrapporten Arena	80
5.1 Totale doorlooptijd.....	80
5.2 Totaal gemiddeld aantal entiteiten in het systeem	80
5.3 Gemiddelde wachttijden per type kassa	81
5.4 Bezettingsgraad kassa's	81
5.5 Aantal entiteiten per type	82
Bijlage 6: Vergelijkende tabel	83

LIJST VAN FIGUREN

Figuur 1: Schematisch overzicht empirisch onderzoek	5
Figuur 2: Overzicht definitie bedrijfsproces	7
Figuur 3: Voorbeeld procesmodel BPMN.....	8
Figuur 4: Schematische voorstelling simulatiestudie.....	9
Figuur 5: Voorbeeldschema simulatiemodel	11
Figuur 6: Noodzakelijke inputs en outputs simulatiemodel	13
Figuur 7: Business Process modelling life-cycle.....	16
Figuur 8: Basistypes van process mining met respectievelijke input/output.....	17
Figuur 9: Voorbeeld clustering.....	22
Figuur 10: Hiërarchische clustering – Berekende afstanden.....	23
Figuur 11: Hiërarchische clustering – Dendrogram met maximale afstand 10	23
Figuur 12: K-means clustering toegepast op voorbeeld.....	24
Figuur 13: Onderdeel 'spaghettiproces'.....	26
Figuur 14: Werking trace clustering techniek.....	27
Figuur 15: Overzicht methodologie	29
Figuur 16: Procesmodel winkel	33
Figuur 17: Poisson kansverdeling met gemiddelde 60.....	35
Figuur 18: Document 1 – De event log	40
Figuur 19: Document 2 – De attribuutwaarden per entiteit	40
Figuur 20: Tabblad 1 – Event log inclusief duurtijden en wachttijden.....	41
Figuur 21: Tabblad 2 – De eigenschappen per entiteit inclusief duurtijden en wachttijden.....	41
Figuur 22: Koppeling activiteitenummers	41
Figuur 23: Onderdeel tabblad 2 – Opsplitsing duurtijden activiteiten	42
Figuur 24: Procesmodel Disco (Frequentie).....	43
Figuur 25: Procesmodel Disco (Performantie – Gemiddelde duurtijd).....	43
Figuur 26: Procesmodel Disco (Activiteit\\Resource)	44
Figuur 27: Tabblad 2 – De eigenschappen per entiteit inclusief keuze.....	44
Figuur 28: Beschrijvende statistieken continue variabelen	45
Figuur 29: Beschrijvende statistieken categorische variabelen - Gezinsgrootte	45
Figuur 30: Beschrijvende statistieken categorische variabelen – Winkelfrequentie.....	46
Figuur 31: Beschrijvende statistieken categorische variabelen – Ervaring self-scanning	46
Figuur 32: Beschrijvende statistieken categorische variabelen – Nieuwe klant	46
Figuur 33: Beschrijvende statistieken categorische variabelen – Keuze.....	46
Figuur 34: Correlatiematrix met Pearson correlatiecoëfficiënt	47

Figuur 35: Scree plot voor bepaling aantal clusters	48
Figuur 36: Kwaliteit clusteroplossing	48
Figuur 37: Invloed gekozen attributen voor clusteroplossing	49
Figuur 38: Output two-step clustering SPSS - overzichtstabel	50
Figuur 39: Omvang gevormde clusters	51
Figuur 40: Capaciteitbepaling resourceklasse 3	56
Figuur 41: Capaciteitbepaling resourceklasse 5	57
Figuur 42: Berekening betrouwbaarheidsinterval: Totale doorlooptijd - Scenario 1	58
Figuur 43: Berekening betrouwbaarheidsinterval: Aantal entiteiten in het systeem - Scenario 3	59
Figuur 44: Vergelijking resultaten – Gemiddelde totale doorlooptijd	59
Figuur 45: Vergelijking resultaten – Totaal gemiddeld aantal entiteiten in het systeem	60
Figuur 46: Vergelijking resultaten – Gemiddelde wachttijd (Reguliere kassa)	60
Figuur 47: Vergelijking resultaten – Gemiddelde wachttijd (Selfkassa)	61
Figuur 48: Vergelijking resultaten – Gemiddelde wachttijd (Snelkassa)	61
Figuur 49: Vergelijking resultaten – Bezettingsgraad (Kassa 1)	61
Figuur 50: Vergelijking resultaten – Bezettingsgraad (Kassa 2)	62
Figuur 51: Vergelijking resultaten – Bezettingsgraad (Kassa 3)	62
Figuur 52: Vergelijking resultaten – Bezettingsgraad (Kassa 4)	62

LIJST VAN TABELLEN

Tabel 1: Capaciteiten resources	34
Tabel 2: Relevante variabelen in verband met entiteiten en hun mogelijke waarden	35
Tabel 3: Koppeling attributen aan entiteitentypes	37
Tabel 4: Aankomst ritme en tussenaankomsttijd per entiteitentype	37
Tabel 5: Logica beslissingspunten	38
Tabel 6: Bepaling duurtijd activiteiten	39
Tabel 7: Geobserveerde procesuitvoeringen in event log	42
Tabel 8: Overeenkomsten clusters voor controledoelinden	52
Tabel 9: Confusion matrix	52
Tabel 10: Gepaste kansverdelingen tussenaankomsttijd	53
Tabel 11: Gemaakte keuzes beslissingspunten	54
Tabel 12: Gepaste kansverdelingen duurtijd activiteiten	55
Tabel 13: Overzicht capaciteit resourceklassen	57

HOOFDSTUK 1: PROBLEEMSTELLING

1 INLEIDING

De toenemende concurrentie tussen bedrijven heeft de voorbije decennia een evolutie teweeg gebracht van een productiegeoriënteerde naar een klantgerichte bedrijfsstructuur (Aguilar, Rautert, & Pater, 1999). Hierdoor is procesgeoriënteerd management steeds belangrijker geworden. Bij deze managementvorm ligt de focus op het creëren van waarde voor de klant door middel van het optimaliseren van bedrijfsprocessen. Deze procesoriëntatie vereist een goede samenwerking tussen de verschillende delen van de organisatie en zorgt voor meer flexibiliteit en een groter aanpassingsvermogen aan een veranderende omgeving. De nood aan hoge flexibiliteit is niet meer weg te denken in de huidige bedrijfswereld.

Vroeger was het begrijpen, voorspellen en optimaliseren van het gedrag van bedrijfsprocessen gebaseerd op theoretische analyses. Een alternatieve mogelijkheid is simulatie. Hierbij worden bedrijfsprocessen indirect bestudeerd door het opbouwen van een model dat gelijkenissen vertoont met het werkelijke proces. Analyses die vervolgens op het model worden uitgevoerd, leveren idealiter informatie op over de werking en de prestaties van het werkelijke proces (Paul, Hlupic, & Giaglis, 1998). Daarnaast kunnen simulatietechnieken ook gebruikt worden om onbestaande systemen optimaal te ontwerpen alvorens over te gaan tot een implementatie.

Simulatietechnieken zijn toepasbaar in meerdere domeinen waaronder productie en logistiek en bieden de mogelijkheid om te experimenteren met bedrijfsprocessen (Hlupic & Robinson, 1998). Zo kunnen verschillende scenario's bijvoorbeeld vergeleken worden door het aanpassen van de inputparameters van het model. De resultaten kunnen vervolgens met elkaar vergeleken worden aan de hand van key performance indicators (KPI's) die bepaald kunnen worden op basis van vooropgestelde doelstellingen (van der Aalst, 2015). Enkele voorbeelden van KPI's zijn de doorlooptijd, kost en bezettingsgraad van resources. Bijkomende voordelen van het gebruik van simulatietechnieken zijn het toegenomen inzicht in de bedrijfsprocessen en een betere communicatie tussen belanghebbenden (Paul, Hlupic, & Giaglis, 1998; Aguilar, Rautert, & Pater, 1999).

Ondanks deze voordelen, werden simulatietechnieken de voorbije jaren bekritiseerd. Zo zouden simulatietechnieken geen ondersteuning bieden voor operationele beslissingen, onvoldoende gebruik maken van de beschikbare data en human resources te naïef modelleren (van der Aalst, Nakatumba, Rozinat, & Russell, 2008; van der Aalst, 2010). Door deze kritieken heerst de overtuiging dat het gebruik van simulatietechnieken minder betrouwbare resultaten oplevert en zo tot suboptimale beslissingen zou kunnen leiden. Momenteel worden deze technieken dan ook beperkt op een gestructureerde en effectieve manier gebruikt voor het ondersteunen van bedrijfsbeslissingen (van der Aalst, 2010). De potentiële voordelen die deze technieken te bieden hebben, komen dus momenteel niet volledig tot hun recht.

Intussen werden reeds meermaals pogingen ondernomen om deze probleempunten aan te pakken. Zo wordt voortdurend gezocht naar manieren om simulatietechnieken te verbeteren opdat deze

dichter zouden aanleunen bij de realiteit. Deze inspanningen dragen bij aan het verhogen van de betrouwbaarheid van simulatiemodellen met als doelstelling een frequenter gebruik, onder andere in het bedrijfsleven.

Een onderzoeksdomein dat meer en meer betrokken wordt bij het verbeteren van simulatietechnieken is process mining of het afleiden van informatie uit procesdata. Heden ten dage worden bedrijfsprocessen meer en meer ondersteund door Process Aware Information Systems (PAIS). Deze informatiesystemen houden een event log bij waarin alle uitgevoerde activiteiten en het tijdstip van uitvoering opgeslagen worden (Nakatumba & van der Aalst, 2010). Bovendien kunnen dergelijke logs ook informatie bevatten over de uitvoerders, de cases waarop de activiteiten worden uitgevoerd en eventueel aanvullende informatie. Process discovery technieken zijn in staat om de informatie uit dergelijke logs om te zetten in een procesmodel (van der Aalst, 2011). Dit model zou vervolgens gebruikt kunnen worden als vertrekpunt bij het opstellen van een simulatiemodel. Op die manier zou het simulatiemodel meer gebaseerd worden op reële data en niet zozeer op subjectieve inschattingen. Als gevolg hiervan zal het simulatiemodel idealiter nauwer aansluiten bij de werkelijke bedrijfsprocessen, wat tot meer betrouwbare resultaten en betere ondersteuning voor bedrijfsbeslissingen zal leiden.

Onderzoekers zien dan ook de grote potentiële voordelen van het combineren van kennis uit beide domeinen (Rozinat, Mans, Song, & van der Aalst, 2009; Nakatumba, Westergaard, & van der Aalst, 2012). Ondanks het geloof en vertrouwen in deze nieuwe aanpak, is het aantal uitgevoerde studies waarin deze domeinen gecombineerd worden momenteel echter beperkt. Deze studies hebben voornamelijk betrekking op de activiteiten en hun duurtijd, het definiëren van resourceklassen en de beslissingslogica in een proces (Martin, Depaire, & Caris, 2014). Uit de beschikbare logs kan echter veel meer informatie geëxtraheerd worden, wat nog veel mogelijkheden biedt om simulatiemodellen realistischer te maken. Er is hier met andere woorden ruimte voor verder onderzoek.

2 ONDERZOEKSVRAAG

In dit onderzoek werd gekozen om dieper in te gaan op een onderdeel van een simulatiemodel, meer bepaald de entiteiten. Met entiteiten worden onder andere objecten of personen bedoeld die het systeem binnenkomen, een bepaalde weg afleggen doorheen het proces en verwerkt of behandeld worden door resources om tenslotte het systeem te verlaten (Kelton, Sadowski, & Swets, 2010). Voorbeelden hiervan zijn klanten in een winkelproces, postpakketjes in een sorteerproces, patiënten in een ziekenhuis of aanvragen tot een lening bij een bank. Entiteiten beschikken over bepaalde unieke eigenschappen die attributen worden genoemd. Zo hebben klanten een bepaalde leeftijd, postpakketjes een bepaald gewicht, patiënten bepaalde symptomen of ziektes en aanvragen tot lening een bepaald bedrag en een bepaalde rentevoet.

Momenteel wordt in simulatiemodellen vaak onvoldoende onderscheid gemaakt tussen de verschillende entiteiten die aanwezig kunnen zijn binnen het systeem. Alle entiteiten worden dan gelijkaardig behandeld en bij het opstellen van een simulatiemodel wordt weinig rekening gehouden met eventuele onderlinge verschillen op het vlak van procesuitvoeringen. Het kan hierbij

gaan om een verschillend aankomst ritme, verschillende doorlooptijden of verschillende beslissingen. Het negeren van deze verschillen lijkt dan ook een onterechte vereenvoudiging of abstractie van de realiteit te zijn. Daarom zouden we in deze masterproef willen onderzoeken of het onderscheiden van relevante entiteitentypes een simulatiemodel realistischer kan maken. Entiteiten behoren tot relevante entiteitentypes wanneer deze types betekenisvolle verschillen vertonen in de manier waarop het proces doorlopen wordt. Dit brengt ons tot de volgende onderzoeksvraag:

"Kunnen simulatiemodellen de realiteit beter benaderen door verschillende entiteitentypes te integreren die geëxtraheerd worden uit event logs en op welke wijze?"

Door het onderscheiden van de verschillende entiteiten die het proces binnenkomen, zou meer informatie geëxtraheerd kunnen worden uit event logs. Deze informatie kan vervolgens gebruikt worden bij het opstellen van een simulatiemodel. Op die manier draagt dit onderzoek bij aan het realistischer maken van simulatiemodellen wat zorgt voor een hogere betrouwbaarheid voor het beter ondersteunen van bedrijfsbeslissingen.

3 DEELVRAGEN

Om de onderzoeksvraag te kunnen beantwoorden, zullen vervolgens deelvragen geformuleerd worden. De antwoorden op deze vragen dragen bij aan het beantwoorden van de onderzoeksvraag.

Wat is een entiteit?

Het is belangrijk om te weten wat een entiteit juist is en welke kenmerken deze heeft. De kenmerken of attributen die een entiteit kan hebben zijn grotendeels domeinafhankelijk en zijn van belang voor het groeperen en/of beschrijven van de entiteiten.

Waar kunnen verschillende entiteitentypes een invloed hebben op het simulatiemodel?

Het is eveneens belangrijk om te bepalen hoe een entiteit zich kan gedragen in een simulatiemodel, zowel op zichzelf als in combinatie met andere entiteiten. Zo moet er onderzocht worden op welke vlakken de procesuitvoering verschilt bij verschillende entiteitentypes.

Hoe kunnen entiteitentypes geleerd worden uit event logs?

Om een onderscheid te kunnen maken tussen entiteitentypes, is er een analyse nodig van de event log om de attributen die verschillen van entiteit tot entiteit te identificeren. Een deel van de attributen zal steeds aanwezig zijn in de event log. Zo zal aan elk event een activiteitsnaam en een tijdstip gerelateerd zijn. Andere attributen kunnen optioneel of domein-specifiek zijn. Vervolgens zullen we op zoek moeten gaan naar een manier om verschillende entiteitentypes te ontdekken vanuit de beschikbare data in de event log.

Hoe kan de kennis over de aanwezige entiteitentypes ingebracht worden in het simulatiemodel?

Hierbij zullen we op zoek gaan naar de parameters en parameterwaarden waarmee de verschillende entiteitentypes ingebracht kunnen worden in het simulatiemodel. Hierbij moet bijvoorbeeld bestudeerd worden welke verdelingen gehanteerd moeten worden voor de aankomst en verwerking van de verschillende entiteitentypes in het systeem.

Hoe kan gemeten worden hoe realistisch een simulatiemodel is?

Hiervoor zullen we in de literatuur op zoek gaan naar maatstaven die gebruikt worden om opgestelde simulatiemodellen met de werkelijkheid te vergelijken. Indien dergelijke maatstaven niet bestaan zullen bedrijfskundige toepassingen van simulatietechnieken gezocht moeten worden om hieruit veelgebruikte vergelijkingsmaatstaven af te leiden.

Is het simulatiemodel realistischer geworden door deze inbreng?

Om deze laatste deelvraag te kunnen beantwoorden zullen de resultaten van een simulatiemodel zonder entiteitentypes vergeleken moeten worden met deze van een simulatiemodel met entiteitentypes als input. Voor beide simulatiemodellen moet aan de hand van vooropgestelde maatstaven bepaald worden hoe sterk en op welke vlakken deze afwijken van de realiteit, zoals opgenomen in de event log.

Idealiter zullen de resultaten van het model met verschillende entiteitentypes de realiteit beter benaderen. Dit zou aantonen dat het maken van een onderscheid tussen relevante types entiteiten in het systeem kan leiden tot kwalitatief waardevollere modellen.

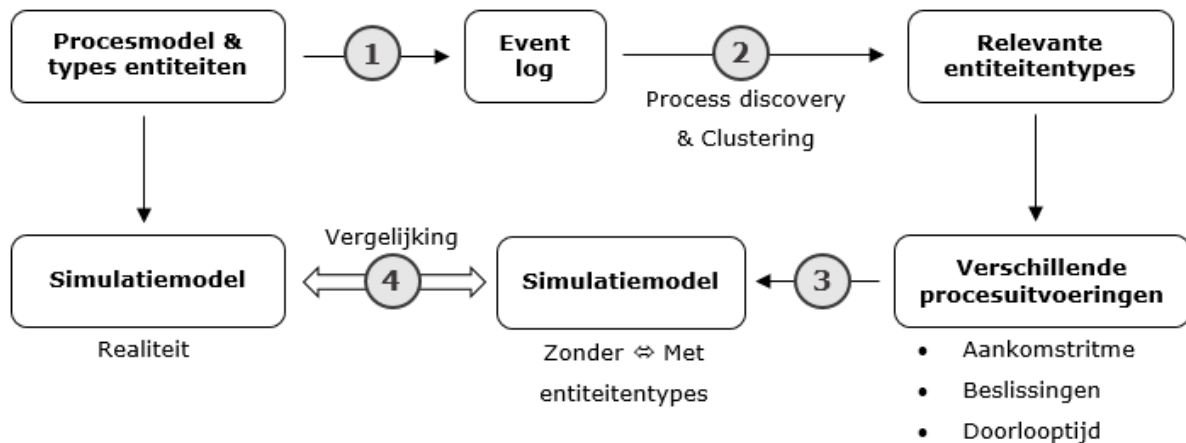
4 ONDERZOEKSAANPAK

In deze masterproef zal vertrokken worden vanuit een grondige literatuurstudie. In dit eerste onderdeel zullen de volgende vier onderwerpen uitvoerig behandeld worden:

- Bedrijfsprocessen
- Simulatie
- Process mining
- Clustering

Het doel van deze literatuurstudie is voornamelijk om een beter beeld te geven van de onderzoeksdomeinen waaraan deze masterproef raakt. Bovendien zullen verschillende begrippen en onderzoeksmethodes aan bod komen die relevant zijn voor het begrijpen van het volgende onderdeel.

Het tweede onderdeel van deze masterproef bestaat uit een empirisch onderzoek waarbij met artificiële data gewerkt zal worden. Een overzicht van de stappen in het empirische onderzoek van deze masterproef wordt weergegeven in figuur 1.



Figuur 1: Schematisch overzicht empirisch onderzoek

Een eerste stap is het bedenken en uittekenen van een proces waarin verschillende procesuitvoeringen mogelijk zijn. Dit wil zeggen dat er beslissingspunten aanwezig moeten zijn waardoor verschillende entiteiten een verschillend pad kunnen afleggen binnen het proces. Vervolgens wordt gezocht naar mogelijke verschilpunten tussen de aanwezige entiteiten. Daarna wordt een keuze gemaakt naar het aantal entiteitentypes toe en zullen de verschillen tussen deze types nader gedefinieerd worden. Hierbij zal erop gelet worden dat ook de uitvoering van het proces voldoende moet verschillen voor de gedefinieerde entiteitentypes. Entiteiten van deze types zullen bijgevolg een ongelijk aankomstrijtme hebben, andere beslissingen nemen en de doorlooptijd kan verschillend zijn. Op basis van deze informatie zal een event log geproduceerd worden die in de volgende stappen beschouwd zal worden als de werking van het reële systeem. Door het creëren van deze artificiële setting kunnen we zeker zijn dat er in dit proces relevante entiteitentypes aanwezig zijn. Het doel is vervolgens om deze types zonder enige voorkennis af te leiden uit de aangemaakte event log en informatie hieromtrent te gebruiken bij het opstellen van een simulatiemodel.

In een tweede stap zullen methodes zoals process discovery technieken en clusteranalyses toegepast worden op de event log. Allereerst zullen process discovery technieken gebruikt worden om het achterliggende proces correct in kaart brengen. Vervolgens worden clusteranalyses uitgevoerd om alle entiteiten in te delen in entiteitentypes die onderling gelijkenissen vertonen. De variatie tussen de entiteiten binnen een welbepaalde cluster dient hierbij zo klein mogelijk te zijn.

Na het onderscheiden van een aantal entiteitentypes die verschillen vertonen in het doorlopen van het proces, kan overgegaan worden tot een derde stap in het onderzoek. Hierin zullen de ontdekte types met bijhorende eigenschappen zoals de aankomstrijtmes, kansverdelingen horende bij beslissingspunten en doorlooptijden afzonderlijk ingebracht worden in een simulatiemodel van het bestudeerde proces. De output van deze simulatie zal de performantie en werking van het systeem beschrijven aan de hand van enkele KPI's zoals de gemiddelde doorlooptijd per entiteit en wachttijden. Er zal eveneens een simulatiemodel opgesteld en uitgevoerd worden waarin geen onderscheid gemaakt wordt tussen de verschillende entiteitentypes. Dit laat toe om in een vierde en laatste stap de resultaten van beide simulaties te vergelijken en vervolgens vast te stellen welke het beste aanleunt bij een simulatiemodel dat de realiteit weergeeft. Dit laatste model zal gebouwd

worden op basis van de informatie waarmee eveneens de oorspronkelijke log werd opgesteld. We trachten hiermee aan te tonen dat het maken van een onderscheid tussen entiteitentypes, die relevant zijn voor de uitvoering van een proces, leidt tot realistischere simulatiemodellen.

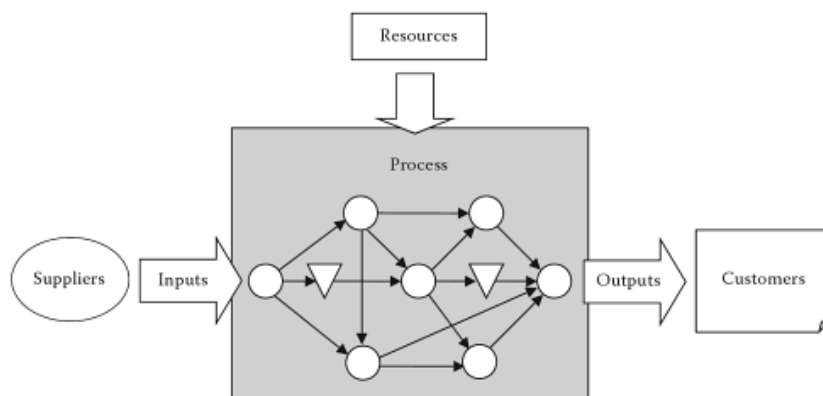
HOOFDSTUK 2: LITERATUURSTUDIE

1 BEDRIJFSPROCESSEN

De producten die bedrijven aanbieden aan de markt zijn steeds het resultaat van het uitvoeren van een aantal activiteiten. Deze activiteiten vormen samen een geheel van bedrijfsprocessen die geanalyseerd kunnen worden om de werking van de organisatie beter te begrijpen en te optimaliseren.

1.1 DEFINITIE

Een bedrijfsproces is een netwerk van logisch gerelateerde activiteiten met een duidelijk begin en einde en onderlinge afhankelijkheidsrelaties. In dergelijke processen worden resources zoals werknemers of machines gebruikt om inputs te transformeren in outputs om te voldoen aan wensen van klanten (Laguna & Marklund, 2013). Dit geheel wordt weergegeven in figuur 2.



Figuur 2: Overzicht definitie bedrijfsproces (Laguna & Marklund, 2013)

Elk bedrijfsproces wordt voltooid binnen één organisatie. Interacties met bedrijfsprocessen van andere organisaties zijn hierbij mogelijk.

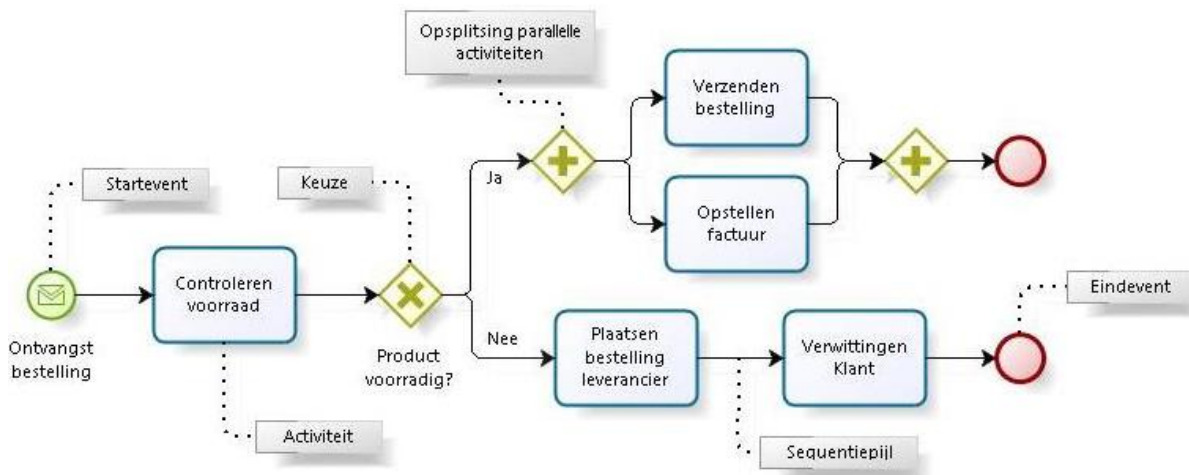
Om snel in te kunnen spelen op de veranderende noden van de markt en zo een competitief voordeel te bereiken en te behouden, is het identificeren en weergeven van activiteiten en hun onderlinge relaties belangrijk. Dit zorgt eveneens voor efficiënte en effectieve communicatie over processen tussen de stakeholders of belanghebbenden (Weske, 2012).

1.2 MODELLERINGSTAAL

Voor het weergeven van bedrijfsprocessen kunnen verschillende grafische notaties gebruikt worden waaronder de Business Process Modelling Notation (BPMN), de unified modeling language (UML) en Petri nets. In dit werk werd gekozen om gebruik te maken van BPMN, dat zich beperkt tot het modelleren van processen met behulp van grafische objecten. BPMN biedt hiermee ruim voldoende mogelijkheden voor de doelstellingen binnen deze masterproef. Het is bovendien een universele notatiestandaard die ontwikkeld werd met als doel processen begrijpelijk te maken voor iedereen die betrokken is bij het ontwerpen, beheren en monitoren van een proces. Op die manier wordt

getracht de kloof tussen het design en de implementatie van bedrijfsprocessen te verkleinen (White, 2004).

Bij het modelleren van processen in BPMN kunnen verschillende onderdelen of constructen gebruikt worden. De constructen die relevant zijn in het kader van deze masterproef zullen toegelicht worden aan de hand van het voorbeeld in figuur 3.



Figuur 3: Voorbeeld procesmodel BPMN

Het belangrijkste onderdeel van een procesmodel zijn activiteiten zoals bijvoorbeeld het controleren van de voorraad of het verzenden van een bestelling. Deze worden weergegeven door rechthoeken met afgeronde hoeken die de naam van de activiteiten bevatten. Activiteiten worden vervolgens verbonden met behulp van sequentiële pijlen die de onderlinge relatie hiertussen weergeven.

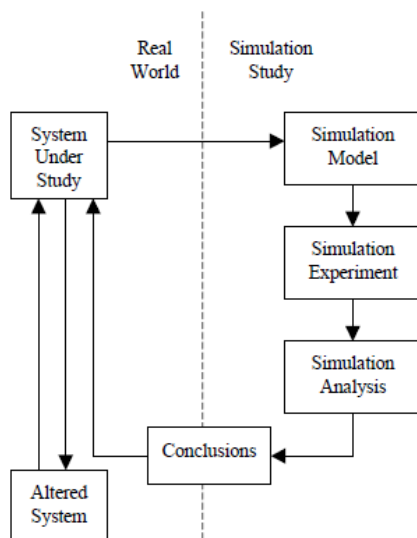
Daarnaast zijn in een procesmodel events aanwezig. Deze kunnen het begin of einde van een proces aangeven en worden weergegeven als cirkels. Deze cirkels kunnen symbolen bevatten die verwijzen naar een bepaald type event. Zo verwijst een enveloppe bijvoorbeeld naar het ontvangen of verzenden van een bericht. Het voorbeeldproces gaat van start na het ontvangen van een bestelling.

Binnen een procesmodel kunnen tenslotte opsplitsingen of beslissingspunten voorkomen. Deze beslissingspunten worden gekenmerkt door een ruit die eveneens voorzien kan zijn van een symbool om het type aan te geven. Zo kan er tussen verschillende activiteiten gekozen moeten worden (X) of kunnen meerdere activiteiten parallel plaatsvinden (+).

2 SIMULATIE

2.1 DEFINITIE

Om objectieve, geïnformeerde beslissingen te kunnen nemen in verband met reële bedrijfsprocessen is het belangrijk om hierover meer kennis op te doen. Het is echter niet altijd eenvoudig of mogelijk om deze processen direct te bestuderen. Daarom kan ervoor gekozen worden om processen indirect te bestuderen door het creëren van een simulatiemodel dat grote gelijkenissen vertoont met de werkelijkheid. Op die manier hoopt men dat analyses van een simulatiemodel eveneens betrouwbare informatie over de werkelijke bedrijfsprocessen opleveren. Dit is het achterliggende kernidee van simulaties (Paul, Hlupic, & Giaglis, 1998). De conclusies van simulatiestudies kunnen bovendien gebruikt worden om aanpassingen door te voeren in de werkelijke bedrijfsprocessen. Dit iteratieve proces wordt weergegeven in figuur 4.



Figuur 4: Schematische voorstelling simulatiestudie (Maria, 1997)

A. Maria omschreef modelleren als het produceren van een model (1997). Een model kan hierbij een representatie zijn van de opbouw en werking van het bedrijfsproces waarin men interesse heeft. Een model is gelijkaardig aan, maar eenvoudiger dan het proces dat het model representeert. Er moet hierbij dus een afweging gemaakt worden tussen realisme en eenvoud. Zo beweert S. Robinson dat het verhogen van de complexiteit van een model afnemende opbrengsten met zich meebrengt op het vlak van de accuraatheid van het model (2013). Dit is het gevolg van onterechte assumpties die gemaakt worden bij een gebrek aan informatie die nodig is om modellen met een hoge complexiteit te ondersteunen.

Simulatie verwijst naar een collectie van methodes en applicaties die toelaten om het gedrag van reële bedrijfsprocessen na te bootsen. Dit gebeurt gewoonlijk met behulp van computers en de daarvoor bestemde software (Kelton, Sadowski, & Swets, 2010). Softwaretools voor simulatie zijn in staat om de dynamiek van processen, zoals het ontstaan van wachtrijen, te modelleren en deze vervolgens visueel weer te geven. Dit kan het ontstaan van creatieve ideeën voor het verbeteren of herontwerpen van bedrijfsprocessen bevorderen (Paul, Hlupic, & Giaglis, 1998). Simulatietools tonen namelijk op een visuele manier hoe het proces werkt en waar problemen zich situeren.

Simulatietechnieken ontstonden reeds eind jaren '50 en hebben de voorbije 60 jaar een enorme evolutie doorgemaakt. De toegenomen toegankelijkheid hebben deze technieken voornamelijk te danken aan de verbetering van de performantie/prijs ratio's van hardware en aan de toegenemende flexibiliteit van softwareprogramma's die steeds eenvoudiger worden in gebruik (Kelton, Sadowski, & Swets, 2010).

Simulaties kunnen bovendien binnen diverse domeinen toegepast worden om systemen te modelleren en te analyseren. Voorbeelden hiervan zijn productie, transport, logistiek, commerciële netwerken, gezondheidszorg en de militaire industrie.

2.2 DOELSTELLINGEN EN PRAKTISCH NUT

Simulaties worden voornamelijk gebruikt om what-if analyses uit te voeren (van der Aalst, 2015). Hierbij kan geëxperimenteerd worden met het proces door bijvoorbeeld variërende waarden aan parameters toe te kennen of door aanpassingen te maken aan de structuur van het proces en te bestuderen hoe het proces reageert op deze veranderingen aan de hand van key performance indicators (KPI's). Enkele voorbeelden hiervan zijn de gemiddelde doorlooptijd, kost en de bezettingsgraad van resources.

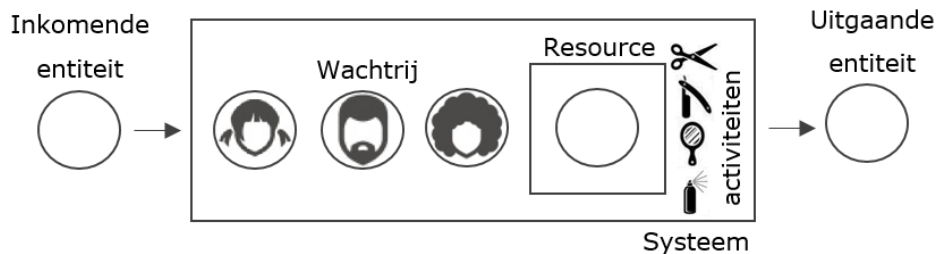
Het doel van een simulatie varieert van het meten van de performantie, het verbeteren van processen tot het optimaal ontwikkelen van onbestaande processen (Kelton, Sadowski, & Swets, 2010). Simulaties dienen met andere woorden gebruikt te worden vooraleer wijzigingen aan te brengen in een bestaand bedrijfsproces of voor het bouwen van een nieuw bedrijfsproces. Op die manier kan de faalkans om te voldoen aan de vereiste specificaties verkleind worden, kunnen onvoorziene bottlenecks weggewerkt worden en kan een te hoge of te lage bezettingsgraad van resources vermeden worden. Dit maakt simulaties een ideaal middel om de performantie van het proces te verbeteren (Maria, 1997).

Een ander toepassingsgebied voor simulatietechnieken is business process re-engineering (BPR). Dit wordt gedefinieerd als het fundamenteel herdenken en het radicaal herorganiseren van een bestaand bedrijfsproces (Hlupic & Robinson, 1998). Het gebruik van simulaties bij dergelijke projecten zou namelijk de risico's verlagen (Paul, Hlupic, & Giaglis, 1998).

De kracht van computersimulaties, gecombineerd met de eenvoud van flowcharts en spreadsheets bieden volgens Tumay het meest kosteneffectieve, accurate en snelle strategische wapen voor bedrijven om verscheidene alternatieven te evalueren vooraleer hier dure resources en tijd aan te alloceren (1996). Toch moet men zich ervan bewust zijn dat het creëren van een simulatiemodel een investering met zich meebrengt en een zekere expertise vereist. Wanneer een probleem opgelost kan worden met behulp van wiskundige methodes zoals lineair programmeren of wachtrijtheorie, dienen deze goedkopere en eenvoudigere methodes verkozen te worden. M. Aguilar, T. Rautert en A.J.G. Pater bevestigen dit en geven in hun studie aan dat een bepaalde graad van complexiteit vereist is om het gebruik van simulatiemethoden in plaats van eenvoudige analyses met spreadsheets te rechtvaardigen (1999).

2.3 ONDERDELEN SIMULATIEMODEL

Een simulatiemodel wordt steeds opgebouwd uit een aantal structurele componenten. De belangrijkste componenten in het kader van deze masterproef zijn entiteiten, resources, activiteiten, attributen, wachtrijen en beslissingspunten. Deze componenten zullen aan de hand van een beknopt voorbeeld (figuur 5) toegelicht worden.



Figuur 5: Voorbeeldschema simulatiemodel

Het voorbeeld bestaat uit een systeem waarin zich verschillende entiteiten en één resource bevinden. Stel dat dit simulatiemodel een kapsalon voorstelt. Entiteiten, weergegeven als cirkels, representeren dan de klanten die het kapsalon binnenkomen en een tijdje later vertrekken met een nieuw kapsel. In dit systeem is ook een resource aanwezig die de klanten behandelt, namelijk de kapper. De activiteiten die de resource uitvoert, waaronder haar knippen, wassen, kleuren enzoverder, hebben elk een bepaalde duurtijd. Wanneer het aankomstritme van de entiteiten hoger ligt dan de verwerkingssnelheid waaraan een resource entiteiten verwerkt, zal er een wachtrij ontstaan in het systeem.

2.3.1 Entiteiten

Entiteiten zijn de dynamische objecten in het model. Zij worden gecreëerd, bewegen zich voort in het model en verdwijnen uiteindelijk uit het systeem. Deze entiteiten veroorzaken veranderingen in de staat van de simulatie (Ingalls, 2008).

Aan verschillende entiteiten kan een verschillende prioriteit toegekend worden. Het toekennen van prioriteiten aan bepaalde entiteiten kan de verwerkingsvolgorde veranderen aangezien bepaalde entiteiten hierbij voorrang zullen krijgen op anderen. Zo zou het in ons voorbeeld mogelijk zijn om entiteiten die vooraf een afspraak maakten voorrang te geven.

Entiteiten hoeven niet altijd personen te zijn. In een productieomgeving kunnen onderdelen van een product entiteiten zijn die door een resource verwerkt of samengevoegd kunnen worden.

Tenslotte is ook het aankomstritme van entiteiten van belang bij het opstellen van een simulatiemodel. Dit hangt samen met de tussenaankomsttijd. Deze geeft aan hoeveel tijd gemiddeld verstrijkt tussen het arriveren van twee opeenvolgende entiteiten (Chung, 2003).

2.3.2 Resources

Resources stellen alles voor dat een beperkte capaciteit heeft binnen een simulatiemodel.

Voorbeelden hiervan zijn werknemers en machines. Gedurende de simulatie kunnen hieromtrent statistieken bewaard worden zoals de bezettingsgraad. Deze geeft aan hoeveel procent van de tijd

een resource actief is. In het gegeven voorbeeld is de kapper een resource. Wanneer er slechts één kapper aanwezig is in het systeem, zal deze alle klanten moeten verwerken met als gevolg een hoge bezettingsgraad. Het verhogen van de capaciteit door een extra kapper aan te nemen zal de bezettingsgraad van de resources verlagen.

Een resource kan zich in verschillende staten bevinden waaronder vrij, bezet, inactief of onbruikbaar (Chung, 2003). Een kapper is vrij wanneer er geen klanten aanwezig zijn in het kapsalon. Deze resource zal daarentegen bezet zijn zolang er zich klanten bevinden in het systeem. In complexere simulatiemodellen kunnen resources ook inactief zijn. Dit kan voorkomen wanneer er bijvoorbeeld een middagpauze genomen wordt. Daarnaast kan een resource onbruikbaar zijn. Hierbij denken we aan defecte machines die hersteld moeten worden alvorens deze opnieuw in staat zijn entiteiten te verwerken.

Tenslotte kunnen resources behoren tot een bepaalde resourceklasse. Deze resourceklassen groeperen gelijkaardige resources die bepaalde rollen kunnen vervullen. Voor elke rol moet bepaald worden welke taken of activiteiten deze resources wel of niet kunnen uitvoeren.

2.3.3 Activiteiten

Activiteiten zijn de taken die door een resource worden uitgevoerd. Binnen een bedrijfsproces is het mogelijk dat activiteiten sequentieel of parallel plaatsvinden. Activiteiten binnen het voorbeeld zouden het wassen, knippen en kleuren van haar kunnen zijn.

2.3.4 Attributen

Entiteiten, activiteiten en resources beschikken over bepaalde attributen. Voor elke entiteit, activiteit of resource nemen deze attributen een specifieke waarde aan. Op die manier geven deze attributen meer informatie over entiteiten, activiteiten of resources. Attributen horende bij entiteiten zouden in ons voorbeeld onder andere de leeftijd van de persoon kunnen zijn, alsook de gewenste behandeling, de haarkleur of het aantal keer dat deze klant al bij deze kapper is geweest. Daarnaast heeft elke activiteit een duurtijd en kan voor een resource bijgehouden worden of het een mens of machine is. Mogelijke attributen voor de resource in het voorbeeld zijn de leeftijd en ervaring.

2.3.5 Wachtrijen

Wachtrijen zijn verzamelingen van wachtende entiteiten die ontstaan door de wisselwerking tussen aankomsttijden van entiteiten, bedieningstijden van activiteiten en de beschikbaarheid van resources. Entiteiten moeten hierbij voor een onbepaalde tijd wachten.

Hoe de wachtrij verwerkt wordt is afhankelijk van de gekozen verwerkingsmethode. Vaak gaat het om first-in-first-out (FIFO), waarbij in tegenstelling tot last-in-first-out (LIFO) de entiteiten die het eerst toekwamen ook eerst behandeld worden door de desbetreffende resource (Chung, 2003). Bovendien dient bij de verwerkingsvolgorde rekening gehouden te worden met prioriteiten. Tenslotte zijn ook gemengde verwerkingsmethodes mogelijk.

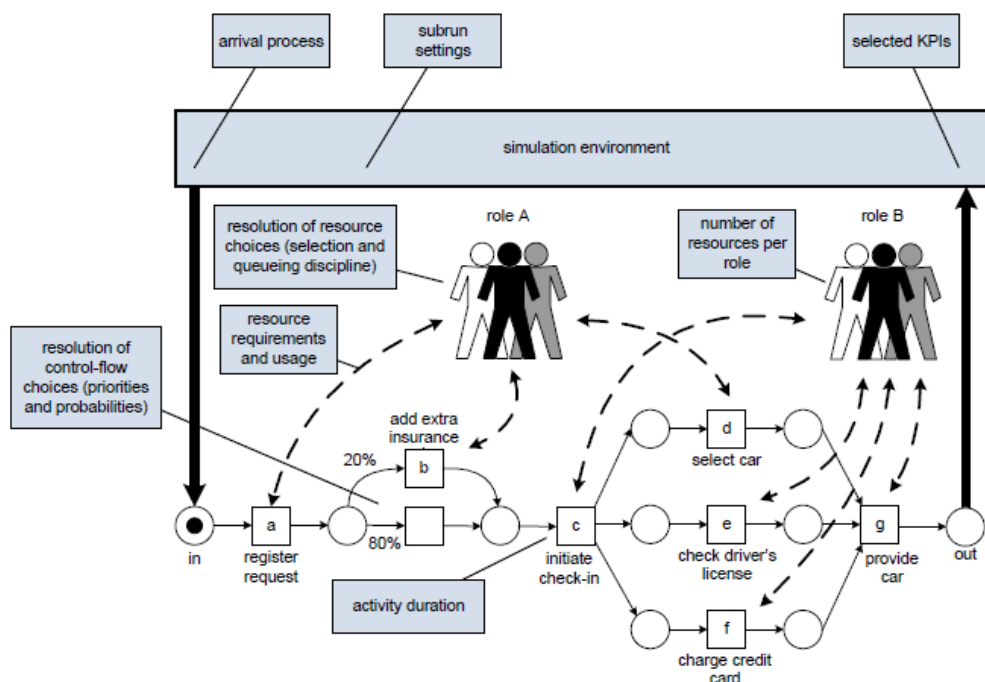
2.3.6 Beslissingspunten

Op beslissingspunten worden keuzes gemaakt worden die bepalend zijn voor de uitvoering van het proces. Zo kunnen er afhankelijk van het gekozen pad andere activiteiten uitgevoerd worden op een bepaalde entiteit. De beslissingen kunnen in een simulatiemodel gebaseerd zijn op kansverdelingen en/of afhankelijk zijn van bepaalde attribuutwaarden. Zo zou een vaste klant van de kapper, die al meer dan 20 keer teruggekomen is, een voorkeursbehandeling kunnen krijgen met extra activiteiten zoals een hoofdmassage.

2.4 INPUTS EN OUTPUTS

Het grootste verschil tussen simulaties en andere modellen is dat deze de dynamieken van een reëel systeem beter kunnen nabootsen door te werken met random inputs (Ingalls, 2008). Deze inputs behoren tot een bepaalde kansverdeling die de werkelijkheid zo goed mogelijk tracht te benaderen. Door te werken met willekeurige inputs is het nodig om per model of strategie meerdere iteraties uit te voeren om zo de willekeurigheid van de resultaten te beperken. Op die manier kunnen er betrouwbaarheidsintervallen opgesteld worden die de resulterende statistieken begrenzen.

De inputs en outputs die volgens W.M.P. van der Aalst nodig zijn voor het uitvoeren van een simulatie-experiment worden weergegeven in figuur 6 (2015). Naast het procesmodel moet eveneens het aankomstritme van nieuwe entiteiten gekend zijn, alsook de duurtijd van de activiteiten en de kansverdelingen in beslissingspunten en eventuele prioriteiten. Op deze plaatsen zou de aanwezigheid van relevante entiteitentypes een invloed kunnen hebben. Verder moet er ook informatie beschikbaar zijn over de gebruikte resources in het systeem, zoals de rollen die deze kunnen vervullen, het aantal resources per rol en hoe deze gebruikt worden.



Figuur 6: Noodzakelijke inputs en outputs simulatiemodel (van der Aalst, 2015)

De kwaliteit van de inputdata wordt door M. Aguilar, T. Rautert en A.J.G. Pater aangehaald als een kritische succesfactor voor het behalen van goede resultaten (1999). Inputdata moet betrouwbaar zijn, exact en statistisch relevant. Wanneer men hier niet genoeg aandacht aan besteedt is de kans groot dat de output van de simulatie niet betrouwbaar zal zijn ('Garbage-In-Garbage-Out' principe). Dit gegeven wordt bevestigd door R.G. Ingalls die beweert dat een model met ongepaste inputverdelingen tot ongeldige resultaten zal leiden (2008).

Na het bepalen van de inputs zal ook bepaald moeten worden hoe vaak de simulatie uitgevoerd moet worden, hoe lang de gesimuleerde periode moet zijn en hoe lang een eventuele opwarmperiode zal duren. Tenslotte dienen KPI's gedefinieerd te worden voor het bepalen van de output van het simulatiemodel (van der Aalst, 2015). In een artikel van K. Tumay werden de doorlooptijd, het aantal entiteiten in een proces, de bezettingsgraad van resources en de kost van activiteiten aangehaald als belangrijkste prestatie maatstaven (1996). De resulterende waarden voor de gedefinieerde KPI's kunnen vervolgens gebruikt worden om te beoordelen hoe realistisch een simulatiemodel is. Voorbeelden van maatstaven die door A. Rozinat, R.S. Mans, M. Song en W.M.P. van der Aalst reeds gebruikt werden om in hun onderzoek de originele logs te vergelijken met de gegenereerde logs zijn verwerkingstijden en wachttijden (2009).

2.5 PROBLEMEN

Ondanks de talrijke voordelen die simulatiemodellen kunnen bieden, worden deze technieken volgens W.M.P. van der Aalst slechts beperkt op een gestructureerde en effectieve manier gebruikt (2010). Dit zou het gevolg kunnen zijn van een gebrek aan trainingen of aan beperkingen van bestaande simulatietools. Daarnaast worden in de literatuur ook meer fundamentele problemen aangehaald die hier toegelicht zullen worden.

Een eerste fundamenteel probleem is dat simulaties niet in staat zouden zijn om operationele beslissingen te ondersteunen. Bij bestaande simulatietools zou de focus namelijk liggen op het maken van strategische beslissingen en dus onvoldoende op het ondersteunen van tactische en operationele beslissingen. Het doel van onderzoek hiernaar is om op korte termijn opties af te kunnen wegen om aanpassingen door te voeren in een werkend bedrijfsproces als reactie op contextuele veranderingen en onvoorziene omstandigheden. Om de implicaties op korte termijn in te kunnen schatten, dient de staat van het systeem en recente eventgeschiedenis verwerkt te worden in het simulatiemodel (Wynn, Dumas, Fidge, Ter Hofstede, & van der Aalst, 2008).

Een tweede probleem is dat de bestaande data onvoldoende wordt gebruikt bij het opstellen van een simulatiemodel. Organisaties van alle groottes worden heden ten dage ondersteund door Process Aware Information Systems (PAIS) die systematisch de werkelijke uitvoering van bedrijfsprocessen op een gestructureerde wijze documenteren in event logs (Nakatumba & van der Aalst, 2010). Deze documenten bevatten een verzameling van activiteiten, waaraan een tijdstip en eventueel een resource gekoppeld is. Dit laat toe om causale relaties tussen activiteiten te identificeren. Door gebruik te maken van event logs bij het opstellen van simulatiemodellen wordt getracht af te stappen van de manier waarop simulatiemodellen oorspronkelijk werden opgesteld. Traditioneel werden simulatiemodellen manueel opgesteld aan de hand van documentatie, interviews en observaties van het reële proces. Dit alles vergt veel tijd en geeft een grote kans op

fouten, aangezien het model dan eerder gebaseerd is op menselijke percepties dan op de realiteit zelf (Rozinat, Mans, Song, & van der Aalst, 2009). Om parameters van een simulatiemodel op een correcte manier te kunnen instellen, is het belangrijk om de beschikbare data uit event logs te exploreren.

Een laatste onderzoeksprobleem dat gesignaleerd wordt in de literatuur is het feit dat simulatiemodellen de realiteit vaak te sterk vereenvoudigen. Zo wordt het gedrag van resources soms te naïef gemodelleerd. In een studie van W.M.P. van der Aalst, J. Nakatumba, A. Rozinat en N. Russell werden reeds verschillende argumenten aangehaald waar het modelleren van human resources fout gaat (2008). Zo kunnen personen bij meerdere processen betrokken zijn en hebben human resources de neiging om werk in batches uit te voeren. Prioriteiten zijn eveneens moeilijk te modelleren omdat telkens opnieuw gekozen moet worden tussen taken. Daarnaast kunnen processen veranderlijk zijn, afhankelijk van de context. Tenslotte werken personen ook niet aan een constante snelheid.

Er werden reeds talrijke studies uitgevoerd die één of meerdere van deze problemen trachtten aan te pakken. Hierbij kan in de literatuur een algemene tendens opgemerkt worden waarbij getracht wordt simulatiemodellen nauwer te doen aansluiten bij de realiteit. Een deel van deze studies maken gebruik van process mining om een realistischer procesmodel te ontdekken dat gebruikt kan worden als vertrekpunt bij het opstellen van een simulatiemodel (Rozinat, Wynn, van der Aalst, ter Hofstede, & Fidge, 2009; Wynn, Dumas, Fidge, ter Hofstede, & van der Aalst, 2008; Rozinat, Mans, Song, & van der Aalst, 2009; Maruster & van Beest, 2009; Liu, Zhang, Li, & Jiao, 2012; Rozinat, Mans, Song, & van der Aalst, 2008). Het concept 'process mining' zal in deel drie van deze literatuurstudie nader toegelicht worden. Andere studies streven ernaar om onterechte veralgemeningen of assumpties weg te werken en proberen op die manier de realiteit beter te benaderen (Baines, Mason, Siebers, & Ladbroke, 2004; Nakatumba & van der Aalst, 2010; Nakatumba, Westergaard, & van der Aalst, 2012; van der Aalst, Nakatumba, Rozinat, & Russell, 2008). In een aantal van deze studies wordt process mining gebruikt in combinatie met simulatietechnieken.

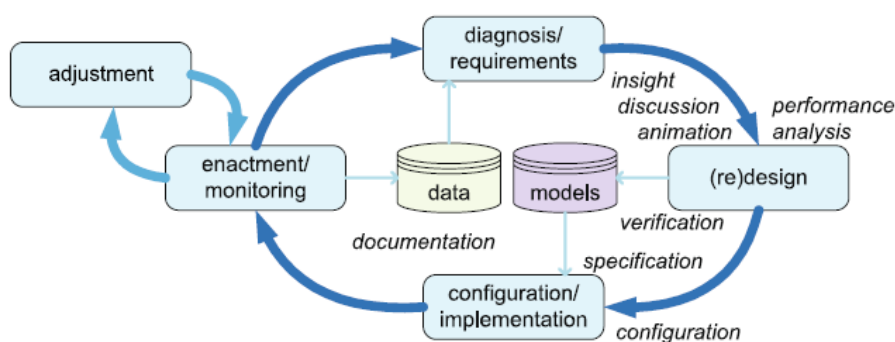
3 PROCESS MINING

3.1 DOELSTELLINGEN

Organisaties van alle groottes maken gebruik van Process Aware Information Systems (PAIS) om de uitvoering van bedrijfsprocessen te ondersteunen. Enkele voorbeelden hiervan zijn Workflow Management Systems (WMS), Customer Relationship Management (CRM) systems en Enterprise Resource Planning (ERP) systems. Deze informatiesystemen documenteren systematisch de uitgevoerde activiteiten op een gestructureerde wijze in event logs. Deze bevatten data in verband met de procesinstanties of cases, tijdstippen waarop taken werden uitgevoerd, personen die deze taken uitvoerden en eventueel aanvullende data. Event logs dienen als vertrekpunt voor process mining (van der Aalst et al., 2009).

Process mining is een vrij recente onderzoeksdiscipline die zich situeert tussen data mining en procesmodellering en -analyse. De voornaamste doelstelling is het ontdekken, monitoren en verbeteren van bedrijfsprocessen door het extraheren van kennis uit beschikbare event logs (Aguirre, Parra, & Alvarado, 2013).

Om dit onderzoeksdomein verder te situeren, toont figuur 7 de Business Process Modelling Life-cycle. Deze levenscyclus geeft de verschillende fases weer die aan bod komen bij het beheren van een bedrijfsproces. In de design-fase wordt een proces ontworpen. Het resulterende model wordt vervolgens in de configuratie- en implementatiefase omgezet in een werkend systeem. Daarop volgt de enactment- en monitoringfase waarin het management nagaat of er aanpassingen vereist zijn aan het huidige systeem. Een deel van deze gewenste veranderingen kunnen doorgevoerd worden door middel van aanpassingen aan de configuratie. Tenslotte zal het proces geëvalueerd worden in de diagnose- en requirementsfase en zal men hierin opkomende systeemvereisten monitoren. Slechte performantie of nieuwe eisen die opgelegd worden door de omgeving, kunnen een nieuwe iteratie van deze levenscyclus initiëren (van der Aalst, 2011).



Figuur 7: Business Process modelling life-cycle (van der Aalst, 2011)

In figuur 7 is duidelijk te zien dat modellen voornamelijk een rol spelen in de design-fase en configuratie- en implementatiefase. Data wordt daarentegen enkel verzameld en gebruikt bij de enactment- en monitoringfase en de diagnose- en requirementsfase. Tot voor kort werd de verzamelde data in beperkte mate en op onregelmatige wijze gebruikt bij het verbeteren van processen. Process mining biedt de mogelijkheid om de verzamelde data en de procesmodellen

optimaal op elkaar af te stemmen. Op die manier wordt het mogelijk om de kwaliteit van modellen en bijgevolg ook van de processen te verhogen (van der Aalst, 2011).

3.2 EVENT LOGS

Een event log, 'history', 'audit trail' of 'transaction file' bevat een collectie van traces. Dit zijn reeksen van events die gerelateerd zijn aan een bepaalde case. Events kunnen bestaan uit een aantal verschillende onderdelen:

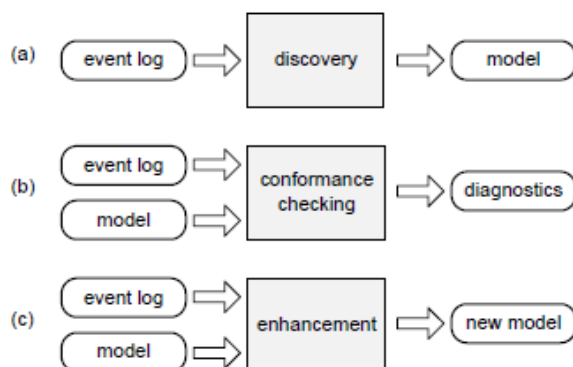
- **Een tijdstip**
- **Een instantie** - iets waarop een taak wordt uitgevoerd
- **Een performer** - de persoon die een activiteit initieert of afwerkt
- **Een activiteit** - de taak of operatie die uitgevoerd wordt
- **Een type** - START, STOP, GEANNULEERD,...

Events bevatten steeds een volgorde of tijdstip op basis waarvan deze gesorteerd kunnen worden. Dit laat toe om causale relaties tussen activiteiten te identificeren.

Ondanks het grote potentieel aan informatie dat deze event logs kunnen bevatten, zijn er enkele factoren die het gebruik in realistische situaties kunnen bemoeilijken. Zo zijn event logs niet altijd volledig. Zo hebben bepaalde paden in een procesmodel soms een kleinere kans op voorkomen, waardoor deze onopgemerkt kunnen blijven. Een event log is namelijk steeds een tijdsopname. Wanneer in de gelogde periode een bepaalde activiteit of situatie niet voorkomt, kan dit pad dat in werkelijkheid wel bestaat niet teruggevonden worden in de event log. Daarnaast kunnen er ook delen van de log foutief of incompleet zijn door een menselijke of technische fout. Tenslotte kunnen er zich ook uitzonderingen voordoen in de periode waarover de log zich uitstrekt (van der Aalst, Reijers, & Song, 2005). Het is dus belangrijk om de verzamelde event logs met voldoende aandacht te behandelen en te filteren waar nodig.

3.3 TYPES

In figuur 8 worden de drie types van process mining weergegeven die op dit moment gekend zijn in de literatuur met hun respectievelijke input en output. Process discovery, conformance checking en enhancement worden in de volgende paragrafen verder toegelicht.



Figuur 8: Basistypes van process mining met respectievelijke input/output (van der Aalst et al., 2012)

3.3.1 Discovery

Bij process discovery technieken wordt een event log genomen en wordt op basis hiervan een model geproduceerd zonder enige a priori informatie. Dit is de meest gebruikte process mining techniek (van der Aalst et al., 2012). Process discovery wordt vaak gebruikt als vertrekpunt voor andere analyses.

Met behulp van process discovery technieken is het mogelijk om drie verschillende perspectieven te ontdekken uit event logs (Nakatumba & van der Aalst, 2010):

- **Procesperspectief**

Dit perspectief focust zich op de volgorde waarin de verschillende activiteiten uitgevoerd kunnen worden. Het achterliggende doel hiervan is om een goed zicht te krijgen op alle mogelijke paden in het proces.

- **Organisatieperspectief**

Bij dit perspectief ligt de focus op de resources. Er wordt voornamelijk aandacht geschonken aan wie welke activiteiten uitvoert en hoe de verschillende actoren gerelateerd zijn aan elkaar. Het streefdoel binnen dit perspectief is om actoren onder te brengen in bepaalde rollen of organisatorische eenheden en om de relaties tussen deze actoren in kaart te brengen.

- **Case perspectief**

Hier ligt de focus eerder op de eigenschappen van de cases of instanties. Zo kunnen cases gekarakteriseerd worden door de paden die ze afleggen doorheen het proces of door waarden voor gerelateerde data-elementen.

3.3.2 Conformance checking

Conformance checking omvat technieken waarbij een bestaand procesmodel vergeleken wordt met een event log van hetzelfde proces. Dit kan gedaan worden om te controleren of de realiteit in overeenstemming is met het model en vice versa. Hiermee kan onder andere ontdekt worden of de regels en gehanteerde procedures steeds opgevolgd worden in de realiteit.

Conformance checking heeft tot doel het detecteren van afwijkingen. Deze kunnen gelokaliseerd worden en daarnaast kan ook de mate van afwijking bepaald worden. Dit domein tracht dus het procesmodel en de realiteit beter op elkaar af te stemmen (van der Aalst, 2011).

3.3.3 Enhancement

Het doel van process enhancement technieken is om het bestaande procesmodel uit te breiden of te verbeteren door gebruik te maken van informatie over het reële proces zoals opgenomen in de event logs. Dit domein heeft met andere woorden het veranderen of uitbreiden van het oorspronkelijke procesmodel als doel (van der Aalst, 2011). Zo kunnen bijvoorbeeld sociale interacties geanalyseerd worden met behulp van deze technieken (Aguirre, Parra, & Alvarado, 2013). Een andere mogelijkheid is het uitbreiden van het model door bottlenecks, service levels, doorlooptijden of frequenties weer te geven in het procesmodel (van der Aalst et al., 2012).

3.4 PROCESS MINING TOOLS

Process mining wordt gedaan met behulp van softwaretools. Enkele toonaangevende process mining tools zijn ProM en Disco.

Het ProM framework is een open-source tool die ontworpen werd om de ontwikkeling van process mining plug-ins te ondersteunen. Deze plug-ins kunnen onderverdeeld worden naargelang het type van de gebruikte process mining techniek: discovery, conformance en extension (van der Aalst et al., 2009).

Volgens S. Aguirre, C. Parra, en J. Alvarado is ProM goed in het ontdekken van processen, maar zijn de resulterende processchema's moeilijk te interpreteren door zakelijke gebruikers (2013). Een commercieel alternatief voor ProM is Disco. Dit pakket werd op de markt gebracht door Fluxicon en is momenteel in opkomst als een gebruiksvriendelijk pakket dat begrijpelijke visualisaties en animaties aanbiedt.

3.5 PROCESS MINING IN COMBINATIE MET SIMULATIE

In deze literatuurstudie kwamen reeds twee onderzoeksdomeinen aan bod, namelijk simulatie en process mining, waartussen raakvlakken aanwezig zijn. Zo kan process mining oplossingen bieden voor de problemen waar simulatiemodellen mee geconfronteerd worden. Bovendien ontdekte process mining een nieuw toepassingsgebied in simulaties. Deze twee domeinen vullen elkaar dus aan en leiden zo tot het ontdekken van innovatieve manieren om bedrijfsprocessen te analyseren en te verbeteren waar nodig.

3.5.1 Toekomstperspectief

De grootte van dit relatief nieuwe onderzoeksdomein waarin simulatie en process mining gecombineerd worden is momenteel beperkt. Desalniettemin geloven toonaangevende onderzoekers binnen deze domeinen in de potentiële voordelen van het combineren van beide technieken (Buijs, La Rosa, Reijers, van Dongen, & van der Aalst, 2013; Rozinat, Mans, Song, & van der Aalst, 2008). Bovendien werd het combineren van process mining met andere analysevormen zoals simulatie reeds aangehaald als een belangrijke uitdaging voor toekomstig onderzoek in het process mining manifesto (van der Aalst et al., 2012).

3.5.2 Onderzoek

3.5.2.1 Eerste onderzoekspogingen

Het idee om process mining toe te passen binnen workflowmanagement werd oorspronkelijk geïntroduceerd door Agrawal et al. (1998, in Rozinat, Mans, Song, & van der Aalst, 2009). Een eerste studie die ook effectief diverse data- en process mining resultaten integreerde om op die manier automatisch een compleet simulatiemodel te genereren was deze van A. Rozinat, R.S. Mans, M. Song, & W.M.P. van der Aalst (2009). In dit onderzoek werd reeds getoond hoe informatie omtrent vier perspectieven ontdekt en samengebracht kan worden.

Een eerste perspectief is het procesperspectief, waarbij causale relaties werden gezocht tussen uitgevoerde activiteiten om zo een procesmodel te bekomen. Een tweede onderzochte perspectief was de data. Hierbij werd een analyse van de beslissingspunten uitgevoerd om op die manier beslissingsregels te ontdekken. Het gaat hierbij om een vorm van decision mining waarbij gezocht wordt naar afhankelijkheden tussen de aanwezige attributen in de event log en de keuzes die gemaakt werden bij het uitvoeren van een proces. Indien mogelijk zullen de gemaakte keuzes gelinkt worden aan eigenschappen of attributen van individuele cases (Rozinat & van der Aalst, 2006). Algemene kwantitatieve en kwalitatieve informatie in een event log kan op die manier gebruikt worden om waardevolle inzichten op te doen over het proces. Een derde perspectief was de performantie. Hierbij werd het procesmodel vervolledigd door duurtijden en wachttijden van activiteiten toe te voegen. Bovendien werd ook aangegeven hoe groot de kans was dat bepaalde paden gekozen werden. Tenslotte werd ook informatie verzameld over de resources om zo de resourceklassen en de rollen die deze vervullen binnen het proces te achterhalen.

3.5.2.2 Organisatieperspectief

Hoewel het merendeel van de reeds uitgevoerde studies zich concentreert op het procesperspectief, wordt in een aantal studies dieper ingegaan op het organisatieperspectief. Zo werd in een onderzoek van M. Song en W.M.P. van der Aalst (2008) getracht een model van de organisatiestructuur, alsook de sociale netwerken en de informatiestroom tussen verschillende resources te ontdekken.

Andere studies gebruikten process mining om zo een realistischer beeld van resources te kunnen integreren in een simulatiemodel. Zo werden reeds twee studies uitgevoerd waarin getracht werd het effect van de werkdruk op de bedieningssnelheid te bepalen vanuit event logs (Nakatumba & van der Aalst, 2010; Nakatumba, Westergaard, & van der Aalst, 2012). Dit gaat in tegen het naïef modelleren van resources waarbij een constante werksnelheid verondersteld wordt.

3.5.2.3 Business process redesign

De aanpak waarbij process mining en simulatie gecombineerd worden, werd eveneens reeds toegepast binnen Business Process Redesign (BPR). Dit is het fundamenteel herdenken en het radicaal herorganiseren van een bestaand bedrijfssysteem (Hlupic & Robinson, 1998). Hierbij kunnen process mining technieken gebruikt worden om de 'As-is' situatie te bestuderen, waarna simulatietechnieken de mogelijkheid bieden verschillende 'To-be' situaties te beschouwen. Hierbij kunnen alternatieven voor het verbeteren van bedrijfsprocessen vergeleken worden. Bovendien kunnen de nodige parameters voor het opstellen van een simulatiemodel worden aangebracht door middel van process mining (Aguirre, Parra, & Alvarado, 2013).

Binnen dit domein werden door L. Mărușter en N.R.T.P. van Beest reeds drie case studies uitgevoerd in zeer diverse domeinen om zo de specifieke moeilijkheden en mogelijke oplossingen hiervoor in kaart te brengen (2007,2009). Hierbij valt op dat elk van de verwerkte event logs afhankelijk van hun respectievelijke inhoud een geheel andere aanpak vereiste. Dit kan een probleem vormen naar de toekomst toe aangezien er weinig uniformiteit is in het verwerken van event logs.

3.5.3 Conclusie

Er kan geconcludeerd worden dat de reeds uitgevoerde studies voornamelijk betrekking hebben op de activiteiten en hun duurtijd, het definiëren van resourceklassen en de beslissingslogica in een proces (Martin, Depaire, & Caris, 2014). Uit de beschikbare logs kan echter meer informatie geëxtraheerd worden, wat mogelijkheden biedt om simulatiemodellen nog realistischer te maken. Hier is dus nog ruimte voor verder onderzoek.

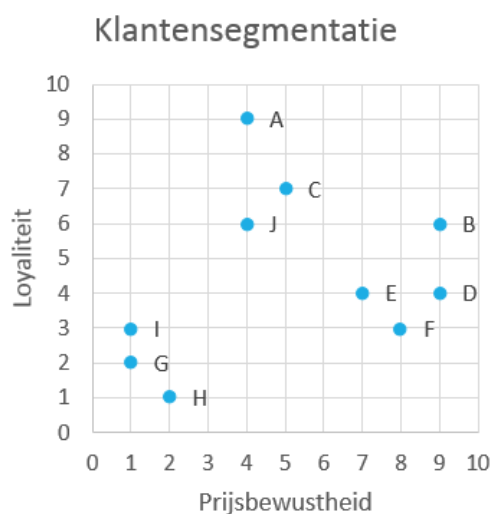
4 CLUSTERING

4.1 DEFINITIE

Clusteranalyse wordt door J.F. Hair, W.C. Black, B.J. Babin en R.E. Anderson omschreven als een groep van multivariate technieken met als doel het classificeren of groeperen van objecten op basis van bijhorende karakteristieken (2009). Bijgevolg ontstaan er deelgroepen met elementen die gelijkaardige kenmerken bezitten. Tussen de gevormde clusters onderling zijn idealiter duidelijke verschillen aanwezig. Hierdoor is het mogelijk om de clusters te beschrijven en de elementen hierin te profileren aan de hand van hun beschrijvende karakteristieken. Clustering wordt onder andere binnen marketing gebruikt om gelijkaardige klanten te groeperen en hiervoor een gepaste marketingstrategie op te stellen.

4.2 TYPES

Er zijn verschillende clusteringsmethoden waartussen gekozen kan worden, namelijk verschillende vormen van hiërarchische clustering en niet-hiërarchische clustering waaronder k-means. Binnen SPSS, een statistische analysesoftware, wordt eveneens een TwoStep component aangeboden voor clustering. Deze methoden zullen verder toegelicht worden aan de hand van het volgende voorbeeld (figuur 9). In dit voorbeeld stelt elk punt de score van een klant voor op twee dimensies, namelijk loyaliteit en prijsbewustheid.



Figuur 9: Voorbeeld clustering

4.2.1 Hiërarchische clustering

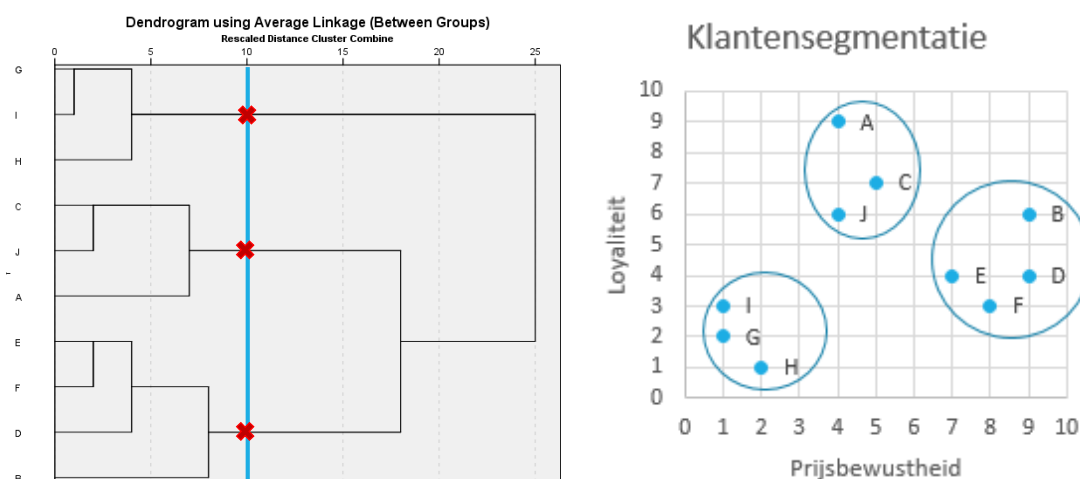
Bij hiërarchische clustering wordt allereerst de afstand tussen de aanwezige elementen (A,B,C,...) bepaald. Vervolgens worden de elementen achtereenvolgens samengevoegd op basis van deze afstanden. Deze afstanden worden verzameld in een proximity matrix (figuur 10). Elementen met de kleinste onderlinge afstand worden het eerst samengevoegd. Zo zal in dit voorbeeld punt I eerst samengevoegd worden met punt G aangezien de Euclidische afstand hiertussen kleiner is dan deze tussen andere puntencombinaties.

Proximity Matrix

Case	Euclidean Distance									
	A	B	C	D	E	F	G	H	I	J
A	,000	5,831	2,236	7,071	5,831	7,211	7,616	8,246	6,708	3,000
B	5,831	,000	4,123	2,000	2,828	3,162	8,944	8,602	8,544	5,000
C	2,236	4,123	,000	5,000	3,606	5,000	6,403	6,708	5,657	1,414
D	7,071	2,000	5,000	,000	2,000	1,414	8,246	7,616	8,062	5,385
E	5,831	2,828	3,606	2,000	,000	1,414	6,325	5,831	6,083	3,606
F	7,211	3,162	5,000	1,414	1,414	,000	7,071	6,325	7,000	5,000
G	7,616	8,944	6,403	8,246	6,325	7,071	,000	1,414	1,000	5,000
H	8,246	8,602	6,708	7,616	5,831	6,325	1,414	,000	2,236	5,385
I	6,708	8,544	5,657	8,062	6,083	7,000	1,000	2,236	,000	4,243
J	3,000	5,000	1,414	5,385	3,606	5,000	5,000	5,385	4,243	,000

Figuur 10: Hiërarchische clustering – Berekende afstanden

De afstanden waarop objecten worden samengevoegd kunnen vervolgens ook grafisch weergegeven worden in een dendrogram, zoals getoond wordt in figuur 11. Vervolgens wordt een maximale onderlinge afstand bepaald tussen de elementen. Deze afstand is bepalend voor het aantal gevormde clusters. Indien we in dit voorbeeld kiezen voor een maximale afstand van 10 zien we dat er 3 snijpunten zijn en bijgevolg 3 clusters, namelijk (G,I,H), (C,J,A) en (E,F,D,B). Hoe groter de maximaal toegelaten afstand tussen elementen in een cluster, hoe kleiner het aantal clusters zal zijn.



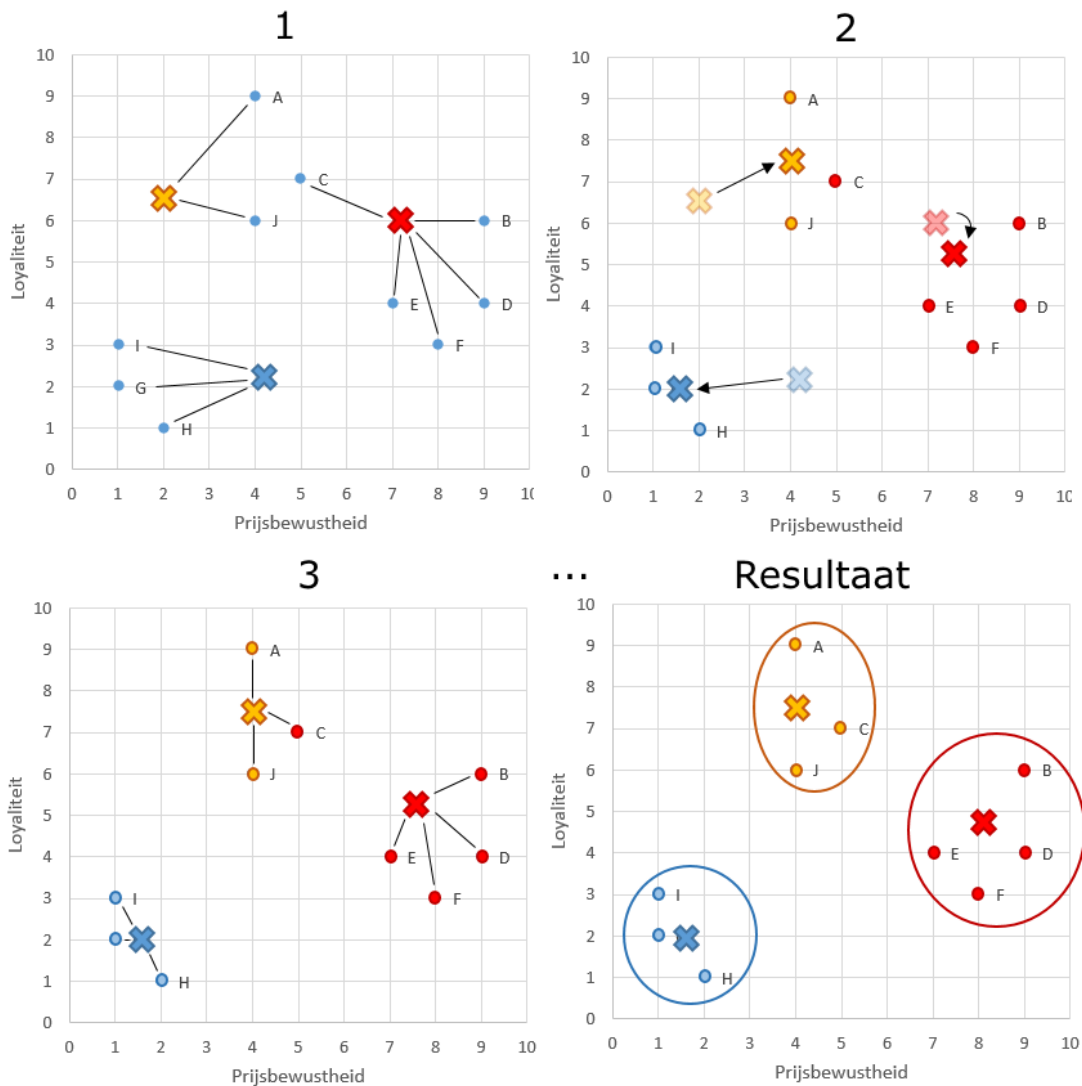
Figuur 11: Hiërarchische clustering – Dendrogram met maximale afstand 10

Voordelen van deze methode zijn de eenvoud en snelheid waarmee zeer veel clusteroplossingen aangeboden worden. Een belangrijk nadeel van deze methode is daarentegen dat eens besloten wordt om twee elementen samen in een cluster onder te brengen, niet teruggekomen kan worden op deze beslissing. Deze methode is bijgevolg gevoelig voor uitzonderingen of outliers. Indien twee outliers zeer dicht bij elkaar liggen zullen deze meteen samengevoegd worden tot een cluster. Op deze actie kan bij hiërarchische clustering niet teruggekomen worden.

4.2.2 Niet-hiërarchische clustering

Naast hiërarchische clusteringstechnieken waarbij elementen achtereenvolgens worden samengevoegd tot wanneer er slechts één cluster resteert en waarbij achteraf een optimaal aantal clusters bepaald wordt, zijn er ook niet-hiërarchische clusteringstechnieken. K-means is een voorbeeld van een populaire niet-hiërarchische techniek waarbij het aantal clusters (parameter k) reeds vooraf vastgelegd dient te worden. De stappen waaruit deze techniek bestaat kunnen

gevolgd worden aan de hand van het uitgewerkte voorbeeld in figuur 12. Allereerst worden k punten gekozen als clustercentra. Alle elementen worden hierna toegewezen aan het dichtstbijzijnde clustercentrum (1) door de afstanden tot deze centra te berekenen. Vervolgens wordt het nieuwe gemiddelde van alle elementen die behoren tot een bepaalde cluster berekend om zo nieuwe en realistischere clustercentra te bekomen (2) en de instanties opnieuw te verdelen (3). Dit proces wordt herhaald tot opeenvolgende iteraties eenzelfde toewijzing van punten aan de clusters oplevert (Witten, Frank, & Hall, 2011).



Figuur 12: K-means clustering toegepast op voorbeeld

Voordelen van deze methode zijn het gebruik van iteraties waardoor beter omgegaan kan worden met uitzonderingen en onbelangrijke clusters. Bovendien vraagt deze methode minder rekenkracht dan hiërarchische methoden waardoor deze methode ook gebruikt kan worden voor grotere datasets. Een nadeel is dat vooraf reeds het aantal gewenste clusters bepaald moet worden. Daarnaast kan het zo zijn dat de optimale oplossing niet gevonden wordt, aangezien de oorspronkelijke clustercentra vaak willekeurig bepaald worden. Een optimale oplossing is een oplossing waarbij de clusters op zo'n manier gevormd werden, dat de verschillen tussen elementen binnen eenzelfde cluster minimaal zijn en tussen de clusters onderling maximaal.

4.2.3 TwoStep clustering

De TwoStep clustercomponent is een nieuwere methode die gecreëerd werd om een antwoord te bieden op de nadelen die samenhangen met het gebruik van hiërarchische of niet-hiërarchische clusteringsmethoden. Zo kan deze methode goed om met grotere datasets, bepaalt deze automatisch het optimale aantal clusters en laat deze gebruikers toe om zowel categorische als continue variabelen op te nemen in de clustering. Bij andere methoden wordt de keuze van het aantal clusters overgelaten aan de uitvoerder van de analyse. Daarnaast is ook het standaardiseren van de waarden van de variabelen die gebruikt worden noodzakelijk om te voorkomen dat bepaalde variabelen meer doorwegen bij de clustering. Ook dit is bij de andere technieken een taak voor de uitvoerder van de analyse.

Deze methode is gebaseerd op een aanpak die uit twee fases bestaat (SPSS Inc., 2001). In de eerste fase wordt een sequentiële voorverwerking gedaan waarbij gelijkaardige elementen reeds samengevoegd worden in een groot aantal kleine subclusters. Alle elementen worden hierin één voor één verwerkt. Op basis van een afstandscriterium wordt voor elk element bepaald of dit toegevoegd kan worden aan een reeds bestaande cluster ofdat er een nieuwe cluster gecreëerd moet worden voor dit element. De resultaten die hieruit voortkomen worden gebruikt in de tweede fase. Hierin worden de reeds gevormde subclusters verder samengevoegd totdat het gewenste aantal clusters bereikt wordt. Deze tweede stap vertoont dus gelijkenissen met de hiërarchische clusteringsmethode. In hoofdstuk drie zullen de nodige inputs en outputs van TwoStep clustering in meer detail besproken worden.

4.3 TOEPASSINGEN

4.3.1 Combinatie van clustering en simulatie

In de literatuur wordt doorgaans relatief weinig aandacht besteed aan de entiteiten binnen simulatiemodellen. Toch zouden er tussen entiteiten onderling verschillen aanwezig kunnen zijn die de uitvoering van het proces voor sommige entiteiten kunnen beïnvloeden. Het kan bijgevolg belangrijk zijn deze extra informatie te gebruiken bij het opstellen van een simulatiemodel.

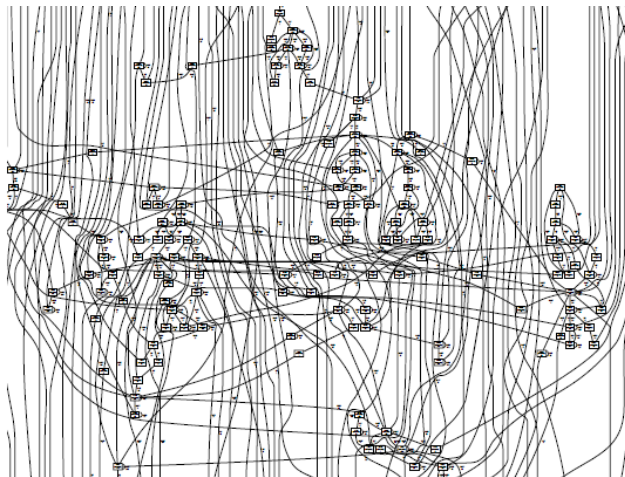
In het onderzoek van M.W. Isken en B. Rajagopalan (2002) ging de aandacht, in tegenstelling tot vele andere studies, uit naar de entiteiten binnen het systeem. De onderzoekers ondernamen hierbij reeds een poging om een type of klasse toe te wijzen aan de entiteiten in het systeem. Deze informatie moest daarna dienen als input bij het opstellen van een simulatiemodel. Het onderzoek vond plaats in een ziekenhuisomgeving waar de schaarse middelen optimaal benut dienen te worden. Simulatiemodellen worden binnen deze omgeving reeds gebruikt om allocatieproblemen omtrent resources aan te pakken. Het gaat hier voornamelijk om het bepalen van de optimale omvang van de infrastructuur en de personeelsbezetting. De entiteiten binnen het voorgestelde systeem zijn patiënten. Een patiëntentype of -klasse werd in dit onderzoek gedefinieerd als een groep patiënten die een gelijkaardige hoeveelheid resources consumeren. De grootste uitdaging was om een gepast aantal entiteitenklassen te bepalen. Te weinig klassen zouden het namelijk moeilijk maken om een realistisch model op te stellen dat de geobserveerde variatie tussen de entiteiten kan verklaren. Te veel patiëntentypes daarentegen kunnen vanuit een

modeleerstandpunt niet haalbaar zijn. De onderzoekers gebruikten data mining technieken en meer bepaald de K-means clusteringmethode om patiënten te groeperen naargelang de benodigde behandeling. Bij het bepalen van het aantal clusters speelde domeinkennis een grote rol. In dit onderzoek werd namelijk geen gebruik gemaakt van process mining om de mogelijke paden doorheen het systeem in kaart te brengen. In dit onderzoek werd aangetoond dat clusteranalyse gebruikt kan worden ter ondersteuning bij het bouwen van een simulatiemodel. Bovendien komt het belang van het onderscheiden van entiteitentypes in dit artikel duidelijk naar voren.

4.3.2 Combinatie van clustering en process mining

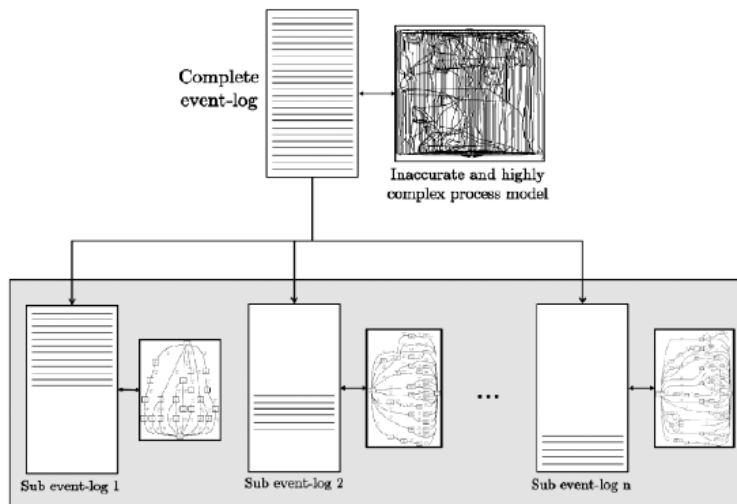
Terwijl process mining technieken er steeds beter in slagen om processen in kaart te brengen, gaan gaat het vaak fout wanneer het om complexe, ongestructureerde processen gaat. Deze processen kunnen voornamelijk teruggevonden worden in flexibele omgevingen zoals bijvoorbeeld gezondheidszorg, klantrelatiebeheer en productontwikkeling (De Weerd, vanden Broucke, Vanthienen, & Baesens, 2013). Informatiesystemen in dergelijke omgevingen bieden vaak meer vrijheid aan de gebruiker, wat zorgt voor een hoge variëteit aan procesuitvoeringen.

Wanneer de reeds ontwikkelde methodologieën toepast worden op reële, complexe data ontstaat er vaak een 'spaghettiproces' (figuur 13). Dit is voornamelijk het gevolg van de schending van de gemaakte assumpties omtrent de event logs. Zo wordt er bij het gebruik van process mining technieken steeds van uitgegaan dat event logs betrouwbaar zijn en dat er één exact proces bestaat dat gereflecteerd wordt in de log. Een andere oorzaak die door R.P.J.C. Bose en W.M.P. van der Aalst wordt aangehaald, is het toepassen van process discovery algoritmes zonder de ruwe data vooraf te bewerken (Bose & van der Aalst, 2009). De overvloed aan connecties in de resulterende procesmodellen die hierdoor ontstaan, maken het zo goed als onmogelijk om bruikbare informatie uit het model te halen.



Figuur 13: Onderdeel 'spaghettiproces' (van der Aalst & Günther, 2007)

Een techniek die gebruikt kan worden om met dergelijke ongestructureerde processen om te gaan is trace clustering. Hierbij worden traces gegroepeerd in clusters, zodat elke resulterende cluster een coherente set van traces bevat. Voor elke cluster kan er vervolgens een procesmodel opgesteld worden (Bose & van der Aalst, 2009). Deze techniek wordt weergegeven in figuur 14.



Figuur 14: Werking trace clustering techniek (De Weerd, vanden Broucke, Vanthienen, & Baesens, 2013)

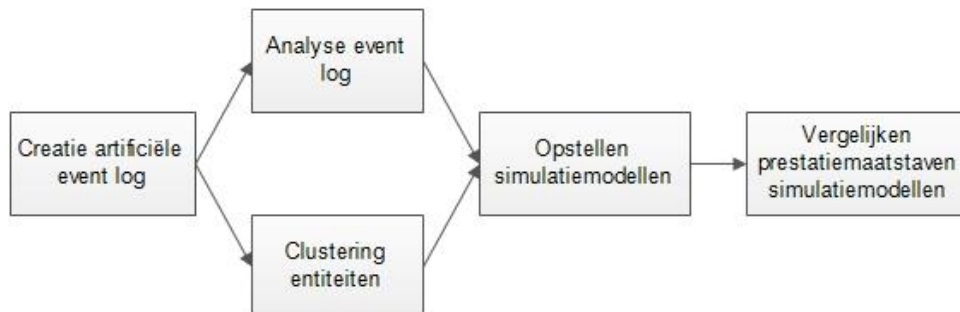
Het basisprincipe van trace clustering is dat er similariteiten en niet-similariteiten tussen traces worden gedefinieerd. Op basis hiervan wordt de event log onderverdeeld in een aantal clusters zodat alle traces in een bepaalde cluster gelijkenissen vertonen met elkaar en verschillen met traces uit andere clusters. De gekozen maatstaf voor gelijkenissen is bepalend voor de kwaliteit van de clusters. Goede clusters zullen leiden tot modellen met een hoge fitnesswaarde en een zo laag mogelijke complexiteit (Bose & van der Aalst, 2009). Hoe hoger de fitnesswaarde, hoe meer traces, of opeenvolgingen van events die voorkomen in de event log, ook mogelijk zijn in het model. Dit is één van de kwaliteitsmaatstaven die gebruikt kunnen worden om de kwaliteit van ontdekte procesmodellen te bepalen.

Een voordeel van trace clustering is dat de grote variëteit van het gedrag in de event log van een ongestructureerd proces op een adequate wijze weergegeven kan worden door gebruik te maken van meerdere procesmodellen. Process discovery technieken worden hier dan ook toegepast op subsets van gedrag om op die manier de accuraatheid van de modellen te verbeteren en deze verstaanbaarder te maken (De Weerd, vanden Broucke, Vanthienen, & Baesens, 2013). Door gebruik te maken van trace clustering kan eveneens het overmatige gebruik van veralgemeningen beperkt worden. Wanneer men tracht om één model te maken voor een proces met zeer grote onderlinge verschillen tussen cases, leidt dit vaak tot modellen die te veel gedrag toelaten. In dergelijke modellen worden veel meer procesuitvoeringen toegelaten dan in de log voorkomen (de Medeiros et al., 2008).

Bij onderzoek naar trace clustering werden reeds pogingen ondernomen om de methodes te automatiseren en te optimaliseren door een zo goed mogelijke maatstaf voor gelijkenissen te bepalen. Zo werd door R.P.J.C. Bose en W.M.P. van der Aalst een algoritme uitgewerkt op basis van generieke beweringsafstanden (2009). Deze geeft aan hoeveel bewerkingen nodig zijn om een bepaalde sequentie te veranderen in een andere. Dit leidde tot betere resultaten dan clustering op basis van Euclidische afstanden in een vectorruimte. In een andere studie werd een event log iteratief geclusterd totdat het mogelijk werd om een precies model op te stellen voor elke cluster (de Medeiros et al., 2008). Deze techniek leidde ook tot stabielere resultaten dan voorgaande algoritmes en werd geïmplementeerd in ProM.

HOOFDSTUK 3: METHODOLOGIE

In dit onderdeel zullen de verschillende stappen in de methodologie toegelicht worden. Deze stappen worden eveneens schematisch weergegeven in figuur 15.



Figuur 15: Overzicht methodologie

1 CREATIE ARTIFICIËLE EVENT LOG

Allereerst zal er een artificiële event log gecreëerd worden. Deze wordt aangemaakt op basis van een procesmodel in combinatie met gedefinieerde formules en kansverdelingen. Met oog op de doelstelling, namelijk het onderscheiden van entiteitentypes waarvoor de procesuitvoeringen verschillen, is het belangrijk dat de entiteiten binnen dit procesmodel een aantal keuzes kunnen maken. Deze keuzemogelijkheden zorgen ervoor dat entiteiten verschillende paden kunnen afleggen doorheen het model. Daarnaast zal ook de capaciteit van de resources en het aankomst ritme van de entiteiten binnen het procesmodel bepaald worden. Door bovendien een voldoende groot aantal entiteiten op te nemen in de event log, worden eventuele problemen bij het toepassen van clusteringsmethoden voorkomen. Zo kunnen bijvoorbeeld minderheidsgroepen vaak niet ontdekt worden wanneer er onvoldoende data beschikbaar is. Daarnaast is het aantal observaties ook bepalend voor het aantal variabelen die opgenomen kunnen worden in de clustering.

In dit model zullen ook attributen worden toegekend aan de aanwezige entiteiten. De attribuutwaarden zullen meer informatie verschaffen over de entiteiten en zullen bovendien een invloed hebben op de procesuitvoering voor een bepaalde entiteit. Hierbij wordt getracht de verschillende manieren waarop het entiteitentype invloed kan hebben op het verloop van het proces te integreren in de event log. Zo zullen het aankomst ritme, gemaakte keuzes en de duurtijd van activiteiten in het proces afhankelijk zijn van attribuutwaarden of entiteitentypes. Deze relaties worden beschreven en in formules gegoten.

Tenslotte zal ervoor gezorgd worden dat er verschillende entiteitentypes aanwezig zijn in de event log. Het procentuele voorkomen van deze types zal bepaald worden. Het moet bovendien mogelijk zijn om binnen dit proces voldoende verschillen te integreren tussen de aanwezige entiteiten. Hiervoor kan gebruik gemaakt worden van attributen. De gekozen types zullen verschillende kenmerken hebben, waardoor ook het procesverloop verschillend zal zijn. Het doel hierna is het terugvinden van de aanwezige types in de event log zonder gebruik te maken van enige voorkennis.

In een volgende fase bestaat de doelstelling enerzijds uit het analyseren van de event log en anderzijds uit het clusteren van de entiteiten om zo aanwezige entiteitentypes te kunnen onderscheiden. De kennis aangaande het opgestelde model wordt vanaf nu enkel nog gebruikt ter controle. Zowel na het clusteren van de entiteiten als na het opstellen van de simulatiemodellen zal een controle plaatsvinden.

2 BESCHRIJVENDE ANALYSE EVENT LOG

Het analyseren van de log zal allereerst gebeuren met behulp van eenvoudige datamanipulaties, filtering en subprocedures. Hiermee zal de data geordend en verrijkt worden voor een eerste analyse. Vervolgens zal gebruik gemaakt worden van process mining tools om het achterliggende proces van de event log duidelijker in kaart te brengen. Tenslotte zullen ook beschrijvende statistieken van de attribuutwaarden van de entiteiten beschouwd worden om zo een beter beeld te krijgen van de waarden die deze attributen kunnen aannemen.

Bij het clusteren van entiteiten zal gebruik gemaakt worden van de attribuutwaarden die hieraan gekoppeld zijn, maar ook van enkele andere eigenschappen die afgeleid kunnen worden uit de log. Het gaat hier om eigenschappen die beschrijven hoe de entiteit door het proces is gegaan en deze kunnen met behulp van datamanipulaties bekomen worden. We denken hier aan de totale doorlooptijd, wachttijden en de keuzes die gemaakt werden.

3 CLUSTERING ENTITEITEN

3.1 VOORBEREIDEND WERK

Een eerste stap bij het uitvoeren van een clusteranalyse is het controleren of de beschikbare dataset voldoende groot is. Dit is nodig opdat ook de kleinere groepen in de populatie ontdekt kunnen worden. Een vuistregel hierbij is dat het aantal observaties groter moet zijn dan 2^n , waarbij n staat voor het aantal gebruikte attributen bij de clustering (Mooi & Sarstedt, 2011). Met behulp van deze formule kan bepaald worden hoeveel attributen maximaal opgenomen kunnen worden bij een bepaald aantal observaties.

Een volgende stap is het controleren van de correlaties tussen de variabelen die we wensen te gebruiken bij het clusteren. Correlatiecoëfficiënten hoger dan 0,90 zouden namelijk de resultaten kunnen vertekenen (Mooi & Sarstedt, 2011). Deze worden opgespoord door een correlatiematrix op te stellen op basis van de Pearson correlatiecoëfficiënt.

Na een inhoudelijke controle van de dataset, volgt de effectieve analyse. Hierbij zal de TwoStep clustermethode van SPSS gebruikt worden. Deze keuze werd gemaakt door het afwegen van de reeds vermelde voor- en nadelen van de verschillende beschikbare methodes.

3.2 INPUT TWOSTEP CLUSTERING

Naast de gekozen variabelen of attributen vereist deze methode als input ook een maximumgrens voor het aantal clusters die gevormd mogen worden. Het maximum aantal clusters dat gebruikt zal

worden als input in de TwoStep clustering, werd bepaald door allereerst een hiërarchische clustering uit te voeren. Met de resultaten hiervan kan een scree plot opgesteld worden waarop de variabiliteit binnen de groepen wordt weergegeven in functie van het aantal clusters. Hierop is te zien dat deze variabiliteit steeds minder sterk afneemt naarmate er meer clusters gecreëerd worden. Er is met andere woorden een dalende opbrengst verbonden aan het toevoegen van meer clusters. In deze scree plot moet er gezocht worden naar een knikpunt om zo een beeld te krijgen van het optimale aantal clusters.

3.3 OUTPUT TWOSTEP CLUSTERING

De output van een TwoStep clustering bevat een figuur die de kwaliteit van de clusteroplossing weergeeft. Deze is gebaseerd op het werk van L. Kaufman en P. J. Rousseeuw uit 1990 omtrent de interpretatie van clusterstructuren. 'Good', 'Fair' en 'Poor' verwijzen respectievelijk naar sterke, beperkte en een gebrek aan significante bewijzen van een aanwezige clusterstructuur (IBM Knowledge Center, 1989, 2011). De maatstaf van samenhang en scheiding die in deze figuur wordt weergegeven, kan berekend worden aan de hand van de volgende formule:

$$\frac{(B - A)}{\text{Max}(A, B)}$$

Hierbij stelt A de afstand voor van een bepaald element tot het clustercentrum en B de afstand tot het dichtstbijzijnde clustercentrum van een cluster waartoe het element niet behoort. Een waarde van één zou hier betekenen dat alle elementen zich op het clustercentrum bevinden.

Een andere figuur die beschouwd zal worden, geeft aan in welke mate de verschillende attributen gebruikt werden bij het clusteren. Het is namelijk belangrijk dat de clustering niet volledig gebaseerd is op één kenmerk, wanneer meerdere kenmerken opgenomen worden in de clustering.

Na het identificeren van een aantal clusters moet eveneens nagegaan worden of deze ook voldoende verschillen vertonen op het vlak van procesuitvoeringen. Hierbij zal gecontroleerd moeten worden of het aankomstritme verschillend is, of er verschillen merkbaar zijn in de genomen beslissingen en tenslotte of de duurtijd van de activiteiten verschilt voor de gevonden clusters. Indien dit het geval is, dan werden er clusters ontdekt die waarschijnlijk kunnen bijdragen aan het opstellen van een realistischer simulatiemodel. De nodige informatie kan hiervoor eveneens teruggevonden worden in een overzichtelijke outputtabel.

3.4 BESCHRIJVING GEVORMDE CLUSTERS

Een volgende stap in de clustering is het omschrijven van de populatie in elk van de ontdekte clusters. Hiervoor kan eveneens gebruik gemaakt worden van de outputtabel die alle gemiddelde of meest voorkomende attribuutwaarden voor de verschillende clusters bevat.

3.5 CONTROLE

Aangezien gewerkt wordt met een artificiële setting, kan vervolgens de vergelijking gemaakt worden tussen de werkelijke groepen en de clusters. Dit om na te gaan hoeveel entiteiten juist geclassificeerd werden en of er al dan niet grote classificatiefouten gemaakt werden. Indien het resultaat goed is, kan de gevonden clusteroplossing gebruikt worden voor verdere analyse. Dit zal niet mogelijk zijn bij reële data, maar er zijn andere methoden beschikbaar om de kwaliteit van een clusteroplossing te beoordelen.

4 OPSTELLEN SIMULATIEMODELLEN

4.1 INPUT SIMULATIEMODEL

In deze analyse zullen de aankomsttimes, de gemaakte keuzes en de duurtijden van activiteiten gezocht worden voor de verschillende clusters. Deze kunnen bekomen worden met behulp van filters. Vervolgens zal de verzamelde data verwerkt worden met behulp van een distribution fitting tool. Dit zijn softwaretools die verschillende statistische verdelingen toepassen op een reeks gegevens om zo de best passende verdeling en bijhorende parameters te ontdekken. Deze kunnen vervolgens gebruikt worden bij het opstellen van de simulatiemodellen. Zo kunnen alle statistische verdelingen ingevoerd worden, alsook de keuzeverdelingen en de algemene structuur van het proces.

4.2 SCENARIO'S

Naast het model met entiteitentypes of clusters zal er ook een model opgesteld worden zonder entiteitentypes. Hierbij zullen alle klanten als één type klant aanzien worden, met één aankomsttime, één keuzepatroon en eenzelfde gemiddelde duurtijd per activiteit.

Tenslotte wordt ook de realiteit, zoals omschreven bij het opstellen van het procesmodel, in een simulatiemodel gegoten. Hierbij worden alle vooraf gedefinieerde kansverdelingen, attribuutwaarden en formules ingevoerd. Dit model zal de werkelijkheid representeren en dient bijgevolg als basis voor de vergelijking.

5 VERGELIJKING SIMULATIEMODELLEN VOLGENS PRESTATIEMAATSTAVEN

Om in te schatten of het model met entiteitentypes dichter aansluit bij de realiteit zullen enkele veelgebruikte prestatie maatstaven waaronder de totale doorlooptijd, alsook het gemiddeld aantal entiteiten in het systeem, de wachttijden bij verschillende resources en hun respectievelijke bezettingsgraad vergeleken worden. Idealiter kan er hier besloten worden dat het model met entiteitentypes dichter aansluit bij het model dat de realiteit representeert. Dit zou aantonen dat het onderscheiden van aanwezige entiteitentypes en het opnemen van informatie hieromtrent bij het opstellen van een simulatiemodel een waardevolle bijdrage kan leveren.

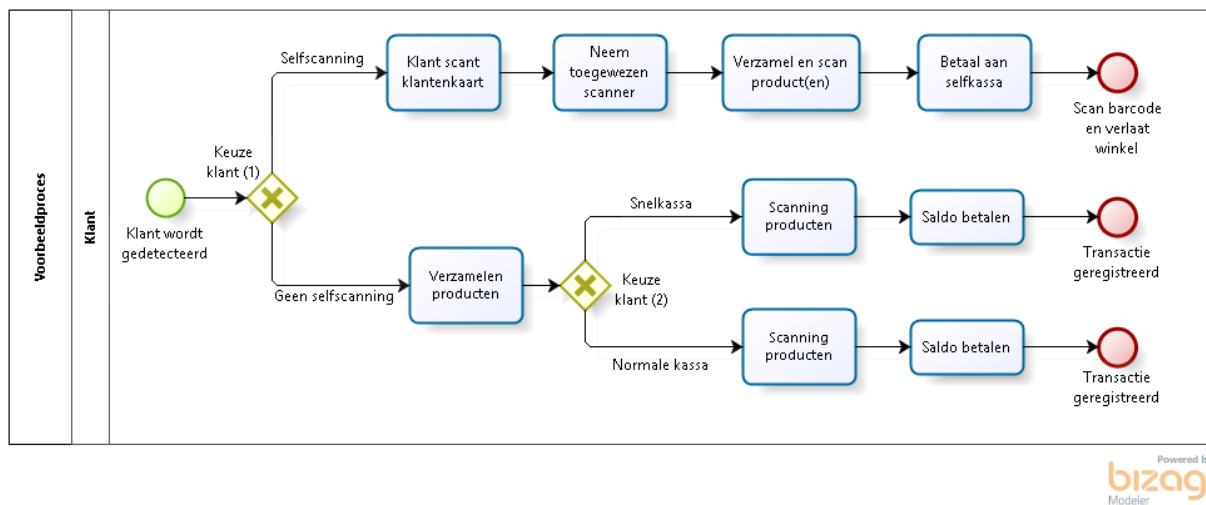
HOOFDSTUK 4: TOEPASSING METHODOLOGIE

De beschreven methode uit hoofdstuk 3 zal in dit onderdeel uitgewerkt worden aan de hand van een artificieel voorbeeld.

1 CREATIE ARTIFICIËLE EVENT LOG

1.1 HET PROCESMODEL

Er werd gekozen voor een winkelproces aangezien dit een herkenbare setting is. Het gecreëerde proces (figuur 16) werd relatief eenvoudig gehouden om op die manier de transparantie van het onderzoek te verhogen. Het systeem in dit proces is een supermarkt.



Figuur 16: Procesmodel winkel

Door de tijd heen zullen er verschillende klanten of entiteiten het systeem betreden. Een eerste keuze die een klant zal maken, is of deze al dan niet gebruik zal maken van de aanwezige self-scanningsfaciliteiten. Indien de klant beslist om hier gebruik van te maken, zal deze een klantenkaart inscannen en de toegewezen scanner nemen. Hiermee zal de klant zich doorheen de winkel verplaatsen en de gewenste producten verzamelen en inscannen. Wanneer de klant tenslotte bij de selfkassa komt, moet deze klant enkel betalen. Daarna verlaat de entiteit het systeem.

Indien de klant geen self-scanning wenst te gebruiken, zal deze zich eerst door de winkel bewegen en producten verzamelen. Wanneer de klant arriveert bij de kassa's, moet er een keuze gemaakt worden tussen de snelkassa en de normale kassa. Deze keuze zal afhankelijk zijn van het aantal producten dat de klant verzameld heeft en van de omvang van de wachtrijen bij de respectievelijke kassa's. Klanten die meer dan tien producten wensen te kopen zullen aan de normale kassa moeten betalen. In beide gevallen worden de verzamelde producten pas aan de kassa door een kassierster gescand. Tenslotte zal elke klant betalen en de winkel verlaten.

1.2 ASSUMPTIES

1.2.1 Opvolging winkelproces

Een assumptie die gemaakt wordt, is dat alle klanten beschikken over een klantenkaart. Deze is nodig om aan self-scanning te doen en eveneens om te kunnen betalen. Een mogelijke reden hiervoor is dat de desbetreffende winkel op deze manier informatie kan verzamelen over de klanten en hun winkelgedrag om zo bijvoorbeeld gepersonaliseerde aanbiedingen te doen.

Een tweede assumptie is dat klanten automatisch gedetecteerd worden bij het arriveren in de winkel, wat bijvoorbeeld gerealiseerd kan worden met behulp van RFID-technologie. Dit laat het bedrijf toe om na te gaan hoe lang een bepaalde klant winkelt. Door deze technologie zou het ook mogelijk zijn om voor een bepaalde klant het gekozen pad binnen dit procesmodel in kaart te brengen.

1.2.2 Resources

De kassiersters vormen samen met de kassa's, de handscanners en het apparaat voor het inscannen van de klantenkaart de resources die in dit proces aangegrepen kunnen worden. We veronderstellen dat deze winkel een gemiddelde grootte heeft met beperkte resources. Er zal één selfkassa aanwezig zijn, één snelkassa en twee gewone kassa's. Tevens werden 40 handscanners voorzien voor self-scanning. Deze resources worden met hun respectievelijke capaciteiten weergegeven in tabel 1. Daarnaast werd de assumptie gemaakt dat deze resources continu beschikbaar zijn. Zo zullen de kassiersters bijvoorbeeld geen pauzes nemen.

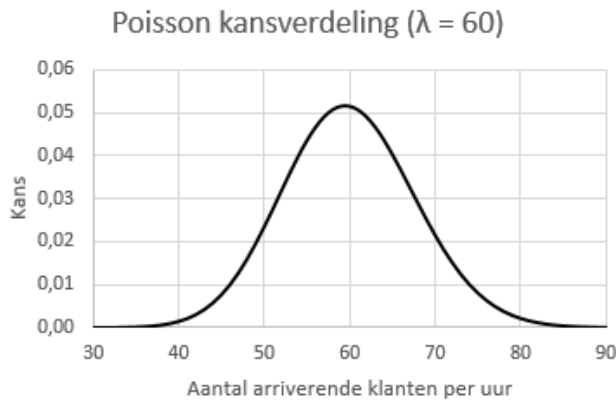
Tabel 1: Capaciteiten resources

Nr	Resourcetype	Capaciteit
r1	Apparaat inscannen klantenkaart	∞
r2	Handscanners	40
r3	Selfkassa	1
r4	Snelkassa	1
r5	Reguliere kassa	2

1.2.3 Entiteiten

1.2.3.1 Aankomst ritme

In de event log zullen er 800 klanten of entiteiten aanwezig zijn. Het aankomst ritme van deze entiteiten is Poisson verdeeld. Dit is een discrete kansverdeling die de kans weergeeft op een bepaald aantal willekeurige voorkomens van een bepaalde gebeurtenis binnen een zeker tijdsinterval. Het aantal voorkomens in een periode is hierbij bovendien onafhankelijk van het aantal voorkomens in elke andere periode. De gebeurtenis waarvoor het aantal voorkomens hier wordt ingeschat, is het arriveren van een klant. Er worden gemiddeld 60 klanten ($\lambda=60$) verwacht op een tijdsinterval van één uur. Deze kansverdeling wordt weergegeven in figuur 17. Hierop is te zien dat de kans dat er per uur minder dan 40 of meer dan 80 klanten arriveren miniem is. Dit gekozen aankomst ritme komt overeen met een gemiddelde tussenaankomsttijd van één minuut.



Figuur 17: Poisson kansverdeling met gemiddelde 60

Een assumptie die hier gemaakt wordt, is dat het gemiddeld aantal klanten per uur stabiel is doorheen de dag en doorheen de week.

1.2.3.2 Attributen

Om onderlinge verschillen te kunnen ontdekken, is het nodig om voor elke entiteit of klant informatie over bepaalde kenmerken te verzamelen onder de vorm van attributen. Dergelijke attribuutwaarden kunnen verzameld worden met behulp van een enquête. Verzamelde informatie is relevant wanneer deze de keuzes die een bepaalde klant maakt op beslissingspunten in het proces kan verklaren of wanneer deze gerelateerd is aan de doorlooptijd. Dit is de totale tijd die een bepaalde entiteit in het systeem doorbrengt. De relevante attributen en de mogelijke waarden die deze kunnen aannemen voor de verschillende klanten worden vermeld in tabel 2.

Tabel 2: Relevante variabelen in verband met entiteiten en hun mogelijke waarden

Variabele	Mogelijke waarden	Type
Gezinsgrootte	1 / 2 / 3 / 4 / 5 / 6	Categorisch
Winkelfrequentie	< 1x per week (0,5x) / 1x / 2x / 3x / elke dag (7x)	Categorisch
Aantal producten	Formule	Continu
	Range	3 - 240
Leeftijd	18 - 80	Continu
Ervaring met self-scanning	1 / 2 / 3 of meer	Categorisch
Nieuwe klant	0 / 1	Categorisch

Een eerste attribuut is de gezinsgrootte. Er wordt van uitgegaan dat grotere gezinnen een hoger aantal producten zal aankopen, wat resulteert in een hogere doorlooptijd. Een assumptie die we hierbij maken is dat elk gezinslid wekelijks 20 producten nodig zal hebben. Het aantal producten is dus lineair verdeeld in functie van de gezinsgrootte.

Een ander attribuut dat een invloed zal uitoefenen op het aantal gekochte producten is de winkelfrequentie. Wanneer meerdere keren per week gewinkeld wordt, zal het aantal gekochte producten per winkelbezoek afnemen. De winkelfrequentie en het aantal producten zijn dus omgekeerd evenredig.

Om het aantal producten te berekenen wordt een combinatie van de twee voorgaande variabelen gemaakt. De formule waarmee deze waarde berekend kan worden, werd opgenomen in tabel 2. Deze variabele kan 22 verschillende waarden aannemen, liggende tussen drie en 240, afhankelijk van de gezinsgrootte en de winkelfrequentie.

Een vierde attribuut dat van belang is, is de leeftijd van de klant. Zo gaan we ervan uit dat oudere klanten meer tijd nodig zullen hebben om de nodige producten te verzamelen. Daarnaast zullen jongere klanten ook sneller geneigd zijn om voor self-scanning te kiezen. De leeftijd van de klanten zal uniform verdeeld zijn tussen 18 en 80 jaar.

Het voorlaatste attribuut is de ervaring die de klant heeft met self-scanning. Dit geeft aan of het de eerste, tweede of derde keer en meer is dat hij gebruik heeft gemaakt of zal maken van de self-scanningsfaciliteiten. We gaan ervan uit dat de klant vanaf drie keer voldoende ervaring heeft opgebouwd met de scanner om efficiënt te werk te gaan en weinig tijd te verliezen aan het handmatig inscannen van producten. Het gaat hier om een leereffect waarbij het inscannen van producten bij self-scanning steeds minder lang zal duren afhankelijk van de reeds opgedane ervaring. Deze variabele zal daarnaast ook een invloed hebben op de keuze voor self-scanning. Zo zal de kans dat iemand voor self-scanning kiest groter zijn wanneer deze persoon hier reeds ervaring mee heeft ten opzichte van iemand die hier nog nooit mee heeft gewerkt.

Tenslotte is het ook belangrijk om te weten of het om een nieuwe klant gaat, aangezien deze meer tijd nodig zal hebben om de benodigde producten te kunnen terugvinden in de winkel. De schikking van de winkel zal namelijk onbekend zijn voor deze klanten. Er worden vooral nieuwe klanten verwacht in de lagere leeftijdscategorie.

1.2.3.3 Types

De drie entiteitentypes die we wensen te integreren binnen dit procesmodel zijn jongeren, gezinnen en ouderen met een respectievelijk voorkomen van 20, 50 en 30%. We geloven namelijk dat er duidelijke verschillen aanwezig zijn in de kenmerken en winkelgewoontes van deze drie klantentypes, die een weerslag zullen hebben op de manier waarop het proces doorlopen wordt. De verschillen tussen de attribuutwaardes voor deze types worden aan de hand van kansverdelingen weergegeven in tabel 3.

De kansverdelingen van de variabelen 'Gezinsgrootte', 'Winkelfrequentie', 'Ervaring met selfscanning' en 'Nieuwe klant' zijn discreet. Dit betekent dat aan elk van de mogelijke attribuutwaarden van deze categorische variabelen een bepaalde kans gekoppeld is. De kansverdeling van de variabele 'Leeftijd' is daarentegen continu aangezien de attribuutwaarden uniform verdeeld zijn. Het aantal producten is tenslotte gebaseerd op een formule die twee categorische variabelen omvat, namelijk 'Gezinsgrootte' en 'Winkelfrequentie'. In tabel 3 worden zowel de range als de gemiddelde waarden van het aantal producten voor elk entiteitentype weergegeven.

Tabel 3: Koppeling attributen aan entiteitentypes

			Entiteitentypes		
			Jongeren	Gezinnen	Ouderen
Variabelen	Gezinsgrootte	1	0,75	0,05	0,47
		2	0,22	0,10	0,47
		3	0,02	0,30	0,06
		4	0,01	0,35	0
		5	0	0,17	0
		6	0	0,03	0
	Winkelfrequentie	< 1x per week	0,20	0,10	0,10
		1x	0,20	0,40	0,85
		2x	0,20	0,25	0,05
		3x	0,20	0,10	0
		Elke dag	0,20	0,15	0
	Aantal producten	Gemiddelde	9	55	24
		Range	3 - 160	3 - 240	3 - 120
	Leeftijd	Range	18-25	26-64	65-80
	Ervaring met self-scanning	1	0,2	0,7	1
2		0,1	0,1	0	
3 of meer		0,7	0,2	0	
Nieuwe klant	Ja	0,15	0,05	0,01	
	Nee	0,85	0,95	0,99	

1.2.4 Procesuitvoering

1.2.4.1 Aankomstrijtme

Met behulp van het procentuele voorkomen van elk van de aanwezige entiteitentypes kan het aankomstrijtme en de tussenaankomsttijd per type berekend worden (tabel 4). De poissonverdeling geeft het gemiddelde aantal aankomsten weer op een tijdsinterval van één uur of 3600 seconden.

Tabel 4: Aankomstrijtme en tussenaankomsttijd per entiteitentype

Entiteitentype	Procentueel voorkomen	Aankomstrijtme	Tussenaankomsttijd
Jongere	20%	Poisson ($\lambda=12$)	300 seconden
Gezin	50%	Poisson ($\lambda=30$)	120 seconden
Oudere	30%	Poisson ($\lambda=18$)	200 seconden
Totaal	100%	Poisson ($\lambda=60$)	60 seconden

1.2.4.2 Keuzes beslissingspunten

De gehanteerde logica in de twee beslissingspunten van dit proces wordt weergegeven in tabel 5.

Tabel 5: Logica beslissingspunten

Beslissingspunt	Logica
Keuze 1: Self-scanning	Kans op self-scanning = $65\% - \text{leeftijd} + 5\% * \text{ervaring self-scanning}$ $= 0 (\geq 65 \text{ jaar})$
Keuze 2: Snelkassa ⇔ Normale kassa	Als Aantal producten ≤ 10 en Grootte wachtrij snelkassa $\leq (\text{Grootte wachtrij normale kassa}/2)$ Dan Snelkassa Anders Normale kassa

De eerste formule zorgt ervoor dat oudere personen in mindere mate geneigd zullen zijn om de self-scanningsfaciliteiten te gebruiken. Daarnaast levert de opgedane ervaring met self-scanning een hogere kans op om hiervoor te kiezen. Belangrijk is dat het steeds een kans blijft die zal variëren tussen 0 en 62%, afhankelijk van de leeftijd en de reeds opgedane ervaring met self-scanning. Het is bijvoorbeeld niet zo dat jongeren met veel ervaring hier steeds voor zullen kiezen.

Een tweede keuze is deze tussen een normale kassa en een snelkassa. Deze keuze is afhankelijk van het aantal producten dat een klant wil kopen en de lengte van de respectievelijke wachtrijen. Indien een klant meer dan tien producten wenst te kopen zal deze automatisch bij een normale kassa terechtkomen. Klanten met minder dan tien producten zullen een afweging maken tussen beide kassa's. Als de wachtrij bij de snelkassa korter of even lang is als deze bij de twee normale kassa's, zal voor de snelkassa gekozen worden. In het andere geval zal de klant kiezen voor de normale kassa.

1.2.4.3 Duurtijd activiteiten

De duurtijd van de activiteiten in het proces wordt bepaald door middel van formules of kansverdelingen. De formules werden zoals aangegeven in tabel 6 opgebouwd met behulp van de reeds besproken attributen.

Het scannen van de klantenkaart in geval van self-scanning heeft een vaste duurtijd van 1 seconde. De periode tussen het scannen van de klantenkaart en het nemen van de toegewezen scanner duurt vervolgens gemiddeld 20 seconden. Er wordt hier gekozen voor een gecensureerde normale verdeling met een standaardafwijking van 15 seconden. Hierbij worden negatieve waarden omgezet naar 1 seconde.

Daarna zal een klant overgaan tot het verzamelen van producten. Deze activiteit zal doorgaans het meeste tijd in beslag nemen. De duurtijd is vooral afhankelijk van het aantal gekochte producten en bestaat zowel uit een vaste als een variabele component. Hierdoor zal een klant onafhankelijk van het aantal gekochte producten minstens vijf minuten in de winkel doorbrengen. Factoren die een winkelbezoek kunnen verlengen zijn of het al dan niet om een nieuwe klant of een oudere persoon gaat. Hierbij wordt elke klant met een leeftijd hoger dan 65 beschouwd als een oudere

persoon. Self-scanning zorgt eveneens voor een langere duurtijd voor het verzamelen van producten aangezien de klant hierbij zelf de producten scant, wat tijd zal besparen aan de kassa. Bij klanten met ervaring in self-scanning zal dit echter sneller gaan dan bij onervaren klanten.

Een volgende activiteit is het scannen van producten aan de kassa bij klanten die niet kozen voor self-scanning. We gaan ervan uit dat de kassierster bij oudere personen, met een leeftijd hoger dan 65, haar snelheid zal aanpassen. Voor het betalen aan de normale kassa en de snelkassa werd gekozen voor een uniforme verdeling tussen de 15 en 30 seconden. Variaties kunnen ontstaan door de gekozen betaalwijze. Hierover wordt geen data bijgehouden en daarom werd hier geopteerd voor een uniforme verdeling.

Om de duurtijd van het betalen aan de selfkassa te beschrijven zal tenslotte een triangulaire verdeling gebruikt worden die afhankelijk is van de besproken entiteitentypes. Dit is bovendien de enige duurtijd die afhankelijk is van het entiteitentype en niet van de bijhorende attributen. Deze activiteit duurt doorgaans iets langer dan het betalen aan de andere kassa's omdat hier een andere aanpak gehanteerd wordt zonder kassierster.

Tabel 6: Bepaling duurtijd activiteiten

Activiteit	Duurtijd
Scannen klantenkaart	1 seconde
Toegewezen scanner nemen	Gecensureerde normale verdeling met $\mu = 20$ sec, $\sigma = 15$ sec
Verzamelen producten	5 min + 20 sec * aantal producten + 10 sec * aantal producten * nieuwe klant (ja/nee) + 10 sec * aantal producten * (leeftijd \geq 65) (ja/nee) + (10 sec / ervaring self-scanning (1/2/3 of meer)) * aantal producten * self-scanning (ja/nee)
Scannen producten (normale kassa/ snelkassa)	2 sec * aantal producten + 1 sec * aantal producten * (leeftijd \geq 65) (ja/nee)
Betalen (normale kassa/ snelkassa)	Uniforme verdeling (15-30 sec)
Betalen (selfkassa)	Triangulaire verdeling met (minimum, gemiddelde, maximum): <ul style="list-style-type: none"> • Jongeren (25, 30, 60) • Gezinnen (30, 40, 70) • Ouderen (40, 60, 90)

2 BESCHRIJVENDE ANALYSE EVENT LOG

Met behulp van de informatie en de formules die in het eerste deel besproken werden, werd een event log en een controlebestand gemaakt. Vanaf dit moment zal er hierop verder gewerkt worden om het achterliggende proces te achterhalen en relevante entiteitentypes te onderscheiden. Deze zullen vervolgens de inputs zijn voor een simulatiemodel met entiteitentypes. De formules, het procesmodel en de kennis over de entiteitentypes uit het vorige onderdeel (Creatie procesmodel) zullen met andere woorden niet gebruikt worden bij het opstellen van een simulatiemodel met en zonder entiteitentypes.

2.1 MANIPULATIES LOG EN EERSTE ANALYSE

We beschikken over drie bestanden waaronder één event log, één bestand met de attribuutwaarden van de entiteiten en één bestand dat de toewijzing van een type aan elke entiteit bevat. Dit laatste document zal uitsluitend gebruikt worden ter controle, na het clusteren van de aanwezige entiteiten. Onderstaande figuren tonen een onderdeel van de logbestanden om een beter beeld te geven van de effectieve inhoud.

	A	B	C	D	E
1	Entity	Activity	Start_timestamp	End_timestamp	Resource_class
2	3	scannen klantenkaart	103,34	104,34	r1
3	4	scannen klantenkaart	122,73	123,73	r1
4	3	toewijzing scanner	104,34	139,47	r2
5	4	toewijzing scanner	123,73	154,59	r2
6	9	scannen klantenkaart	385,25	386,25	r1
7	9	toewijzing scanner	386,25	399,37	r2
8	4	verzamel producten	154,59	592,59	r2
9	3	verzamel producten	139,47	619,47	r2
10	4	betaal aan selfkassa	592,59	626,36	r3

Figuur 18: Document 1 – De event log

	A	B	C	D	E	F	G
1	Case_id	Aantal_producten	Gezinsgrootte	Winkelfrequentie	Ervaring_selfscan	Nieuwe_klant	Leeftijd
2	1	30.0	3	2	1	0	55
3	2	20.0	1	1	1	0	71
4	3	6.0	2	7	1	0	45
5	4	6.0	2	7	3	0	24
6	5	40.0	2	1	1	0	75

Figuur 19: Document 2 – De attribuutwaarden per entiteit

Allereerst werden de drie bestanden verzameld in één Excel-document om het uitvoeren van datamanipulaties te vergemakkelijken. Daarna werd de event log in het eerste tabblad gesorteerd per entiteit en vervolgens ook per tijdstip. Dit geeft een beter beeld van welke activiteiten er voor een entiteit uitgevoerd worden en hoe lang een entiteit zich ongeveer in het systeem bevindt. Vervolgens werden hieraan twee kolommen toegevoegd, waaronder één kolom toe die de duurtijd van elke activiteit en een andere kolom die de wachttijden weergeeft. Duurtijden van de activiteiten worden gedefinieerd als het verschil tussen de eindtijd en de starttijd. Wachttijden komen voor wanneer de eindtijd van één activiteit en de starttijd van de daarop volgende activiteit niet samenvallen. Na deze eerste aanpassingen ziet de event log er als volgt uit:

	A	B	C	D	E	F	G
1	Entity	Activity	Start_timestamp	End_timestamp	Resource_class	Duurtijd activiteiten	Wachttijden
2	1	verzamel producten	0,00	900,00		900,00	0,00
3	1	scan producten reguliere kassa	900,00	960,00	r5	60,00	0,00
4	1	betaal producten reguliere kassa	960,00	979,26	r5	19,26	0,00
5	2	verzamel producten	56,97	956,97		900,00	0,00
6	2	scan producten reguliere kassa	956,97	1016,97	r5	60,00	0,00
7	2	betaal producten reguliere kassa	1016,97	1035,09	r5	18,13	0,00
8	3	scannen klantenkaart	103,34	104,34	r1	1,00	0,00
9	3	toewijzing scanner	104,34	139,47	r2	35,13	0,00
10	3	verzamel producten	139,47	619,47	r2	480,00	6,89
11	3	betaal aan selfkassa	626,36	682,68	r3	56,32	0,00

Figuur 20: Tabblad 1 – Event log inclusief duurtijden en wachttijden

Uit een eerste visuele analyse van de event log blijkt dat we beschikken over een log met 800 entiteiten en 2614 gelogde uitvoeringen van activiteiten of events. Er komen in het proces negen verschillende activiteiten voor waaronder 'verzamel producten', 'scan producten reguliere kassa', 'betaal producten reguliere kassa', 'scan klantenkaart', 'toewijzing scanner', 'betaal aan selfkassa', 'scan producten snelkassa' en 'betaal producten snelkassa'. Uit de benaming van deze activiteiten kunnen we afleiden dat het hier gaat om een winkelproces met verschillende kassa's. Daarnaast zien we ook dat er vijf resourceklassen voorkomen (r1 – r5). Bovendien valt op dat de activiteit 'verzamel producten' doorgaans ruimschoots meer tijd in beslag neemt dan de andere activiteiten. Bovendien vindt deze activiteit plaats voor elke entiteit. Hieruit leiden we af dat dit een belangrijke activiteit is binnen het proces.

Vervolgens worden de totale duurtijd en eventuele wachttijden toegevoegd aan de gegevens omtrent de entiteiten. Dit kan nuttig zijn voor de clustering aangezien deze waarden meer informatie geven over het procesverloop voor een bepaalde entiteit. Om dit te realiseren werd een subprocedure geschreven in Visual basic for Excel (Bijlage 1.1). Deze telt de verschillende duurtijden per case op en vult deze aan in het tweede tabblad. Hetzelfde gebeurt voor de wachttijden (Bijlage 1.2). Dit leidt tot het volgende resultaat:

	A	B	C	D	E	F	G	H	I
1	Case_id	Aantal_producten	Gezinsgrootte	Winkelfrequentie	Ervaring_selfscan	Nieuwe_klant	Leeftijd	Duurtijd activiteiten	Wachttijd
2	1	30	3	2	1	0	55	979,26	0,00
3	2	20	1	1	1	0	71	978,13	0,00
4	3	6	2	7	1	0	45	572,45	6,89

Figuur 21: Tabblad 2 – De eigenschappen per entiteit inclusief duurtijden en wachttijden

Tenslotte wordt eveneens een opsplitsing gemaakt van de totale duurtijd van de activiteiten. Elke activiteit krijgt hierbij een eigen kolom. Aan elke activiteit werd bovendien een nummer gekoppeld om de leesbaarheid van het document te verhogen (figuur 22).

- Act1 (Scannen klantenkaart)
- Act2 (Toewijzing scanner)
- Act3 (Verzamelen producten)
- Act4 (Betaal aan selfkassa)
- Act5 (Scannen producten reguliere kassa)
- Act6 (Betalen producten reguliere kassa)
- Act7 (Scannen producten snelkassa)
- Act8 (Betalen producten snelkassa)

Figuur 22: Koppeling activiteitennummers

Indien een bepaalde activiteit niet plaatsvond voor een bepaalde entiteit zal deze cel leeg zijn. Dit werd eveneens gerealiseerd met behulp van een subprocedure (bijlage 1.3). Deze procedure zoekt voor elke entiteit de duurtijden van de voltooide activiteiten op in het eerste tabblad en plaatst deze vervolgens in de correcte kolom op het tweede tabblad. Het resultaat wordt weergegeven in figuur 23.

H	I	J	K	L	M	N	O	P
Act1	Act2	Act3	Act4	Act5	Act6	Act7	Act8	Totale duurtijd activiteiten
		900		60	19			979,2642219
		900		60	18			978,1250702
1	35	480	56					572,4457701
1	31	438	34					503,6318977
		1500		120	24			1644,124879

Figuur 23: Onderdeel tabblad 2 – Opsplitsing duurtijden activiteiten

2.2 PROCESS MINING

Om een beter beeld te kunnen vormen van het achterliggende proces van deze event log zal hierop process mining toegepast worden. Hiervoor zullen we gebruik maken van het commerciële pakket Disco. We kiezen voor dit pakket omdat dit voldoende ondersteuning biedt voor onze doelstelling, namelijk het in kaart brengen van het proces voor het ondersteunen van de clustering en het bouwen van een gepast simulatiemodel.

Na het inladen van de log in deze tool krijgen we een schematisch overzicht te zien van het proces zoals het zich in werkelijkheid voordoet. In Disco is het zowel mogelijk om de frequentie (figuur 24) te bestuderen waarmee activiteiten uitgevoerd worden, alsook de performantie (figuur 25) hiervan. Op die manier kan er meer informatie verkregen worden over de wachttijden en de duurtijden van de verschillende activiteiten binnen het systeem.

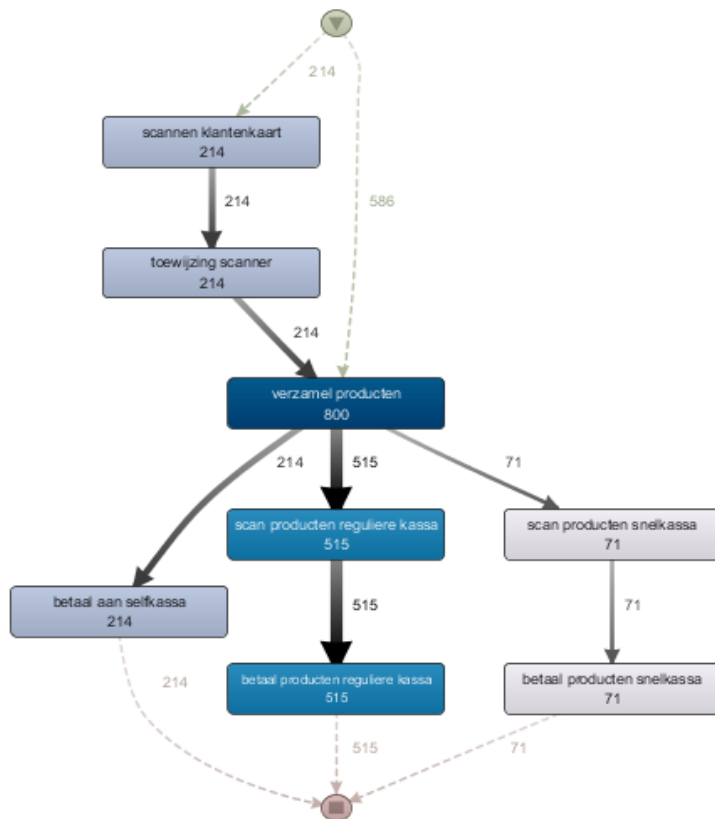
Uit deze figuren kunnen we afleiden dat het proces steeds van start gaat door het scannen van de klantenkaart of met het verzamelen van producten. Na het verzamelen van de gewenste producten, zijn er vervolgens drie mogelijke paden die leiden tot het beëindigen van het proces. Zo kan er betaald worden aan een selfkassa of kunnen de producten gescand en betaald worden aan een reguliere of een snelkassa. Deze drie mogelijkheden worden opgesomd in tabel 7.

Tabel 7: Geobserveerde procesuitvoeringen in event log

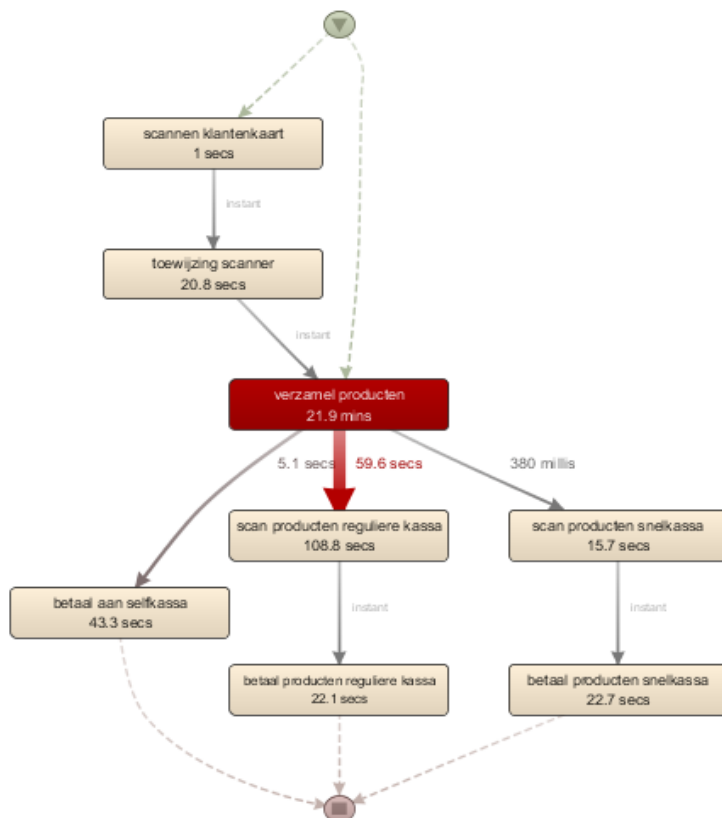
Doorlopen pad	Aantal entiteiten	Gemaakte keuze
Verzamel producten → Scan producten reguliere kassa → Betaal producten reguliere kassa	515	Reguliere kassa (1)
Scannen klantenkaart → Toewijzing scanner → Verzamel producten → Betaal aan selfkassa	214	Selfkassa (2)
Verzamel producten → Scan producten snelkassa → Betaal producten snelkassa	71	Snelkassa (3)

Tenslotte merken we op dat de activiteit 'verzamel producten' duidelijk het meeste tijd in beslag neemt. Dit bevestigt wat we reeds ontdekten bij de visuele analyse van de event log. Voorts

merken we op dat de wachttijden gemiddeld langer zijn bij de reguliere kassa. Enkel voor de kassa's komen er wachttijden voor, dit zijn bijgevolg de kritische resources binnen het systeem.

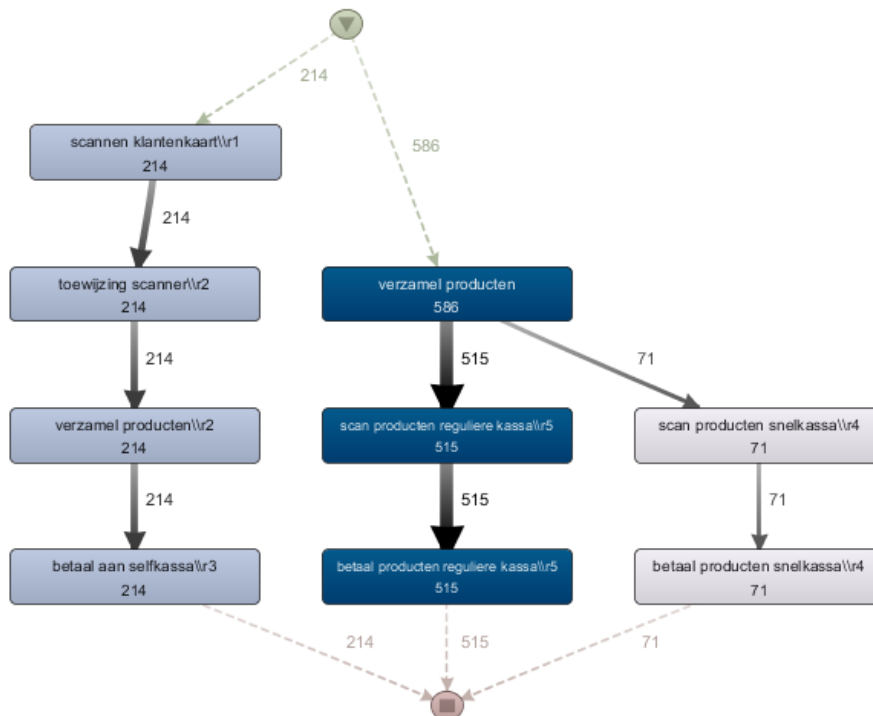


Figuur 24: Procesmodel Disco (Frequentie)



Figuur 25: Procesmodel Disco (Performantie – Gemiddelde duurtijd)

In een volgende stap kunnen de resources gekoppeld worden aan de activiteiten (figuur 26). Hierbij wordt duidelijk welke activiteiten gebruik maken van de resources uit bepaalde resourceklassen. Bovendien is het onderscheid tussen de verschillende procesuitvoeringen, die in tabel 7 reeds vergeleken werden, hierop duidelijk zichtbaar.



Figuur 26: Procesmodel Disco (Activiteit \ Resource)

Het gekozen kassatype kan vervolgens opgenomen worden in de tabel met gegevens over de entiteiten. Dit attribuut ('Keuze') geeft meer informatie over het procesverloop voor een entiteit. Om deze gegevens te bekomen werd gebruik gemaakt van filtering. De gefilterde gegevens werden samengevoegd in een hulptabel en vervolgens gesorteerd. Dit leidt tot het volgende resultaat:

	A	B	C	D	E	F	G	H	I	J
1	Case_id	Aantal_producten	Gezinsgrootte	Winkelfrequentie	Ervaring_selfscan	Nieuwe_klant	Leeftijd	Duurtijd activiteiten	Wachttijd	Keuze
2	1	30	3	2	1	0	55	979,26	0,00	reguliere kassa
3	2	20	1	1	1	0	71	978,13	0,00	reguliere kassa
4	3	6	2	7	1	0	45	572,45	6,89	selfkassa
5	4	6	2	7	3	0	24	503,63	0,00	selfkassa
6	5	40	2	1	1	0	75	1644,12	0,00	reguliere kassa

Figuur 27: Tabblad 2 – De eigenschappen per entiteit inclusief keuze

Omdat deze tabel volledig bestaat uit numerieke gegevens werd ervoor gekozen om 'reguliere kassa', 'selfkassa' en 'snelkassa' te vervangen door de respectievelijke waarden 1,2 en 3.

2.3 BESCHRIJVENDE STATISTIEKEN

Een laatste stap in deze beschrijvende analyse van de event log is het nader analyseren van de attribuutwaarden. Hierbij beschouwen we enkele beschrijvende statistieken voor de attributen die de entiteiten beschrijven. Opmerkelijk is hier dat er zowel categorische als continue variabelen voorkomen in het desbetreffende logbestand. Daarnaast zijn er zowel attributen die de entiteit beschrijven als attributen die betrekking hebben op de procesuitvoering voor elke entiteit.

2.3.1 Continue attributen

Voor de continue variabelen worden het aantal instanties, de minimum- en maximumwaarde, het gemiddelde en de standaardafwijking weergegeven in figuur 28.

	N	Minimum	Maximum	Mean	Std. Deviation
Aantal_producten	800	3	240	40,53	34,57
Leeftijd	800	18	79	48,00	19,88
Act1 (Scannen klantenkaart)	214	1	1	1,00	,00
Act2 (Toewijzing scanner)	214	0	66	20,77	13,46
Act3 (Verzamelen producten)	800	360	6300	1311,36	823,91
Act4 (Betaal aan selfkassa)	214	27	68	43,35	8,89
Act5 (Scannen producten reguliere kassa)	515	22	480	108,76	70,59
Act6 (Betalen producten reguliere kassa)	515	15	30	22,15	4,42
Act7 (Scannen producten snelkassa)	71	6	30	15,75	6,64
Act8 (Betalen producten snelkassa)	71	15	30	22,58	4,71
Totale duurtijd activiteiten	800	381,01	6725,41	1416,43	873,72
Wachttijd	800	,00	428,79	39,74	74,18

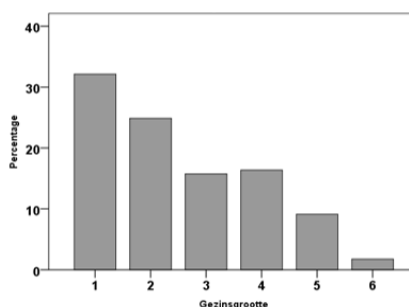
Figuur 28: Beschrijvende statistieken continue variabelen

2.3.2 Categorische attributen

Vervolgens zullen ook de frequenties en overeenkomstige percentages van de categorische variabelen besproken worden aan de hand van onderstaande figuren. Het gaat hier om de gezinsgrootte, winkelfrequentie, ervaring met self-scanning, nieuwe klanten en de gemaakte keuzes binnen het proces.

Allereerst is er de gezinsgrootte. Hierbij valt op dat gezinnen met een kleine omvang (één of twee personen) het meest voorkomen in de event log.

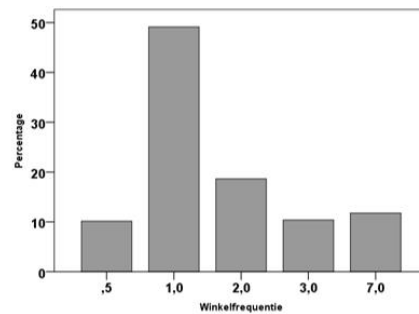
	Frequency	Percent	Cumulative Percent
1	257	32,1	32,1
2	199	24,9	57,0
3	126	15,8	72,8
4	131	16,4	89,1
5	73	9,1	98,3
6	14	1,8	100,0
Total	800	100,0	



Figuur 29: Beschrijvende statistieken categorische variabelen - Gezinsgrootte

De meest courante winkelfrequentie is één keer (49,1%). De overige opties komen ongeveer in gelijke mate voor (10,1% - 18,6%).

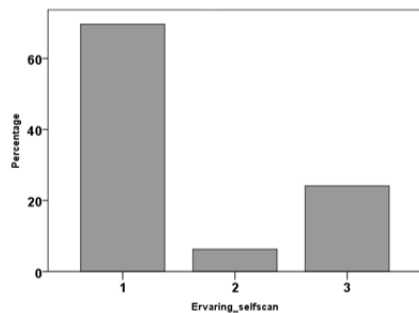
Winkelfrequentie			
	Frequency	Percent	Cumulative Percent
,5	81	10,1	10,1
1,0	393	49,1	59,3
2,0	149	18,6	77,9
3,0	83	10,4	88,3
7,0	94	11,8	100,0
Total	800	100,0	



Figuur 30: Beschrijvende statistieken categorische variabelen – Winkelfrequentie

Uit figuur 31 blijkt dat maar liefst 69,6% van de entiteiten in de event log beperkte ervaring heeft met self-scanning. 24,1% van de entiteiten heeft wel reeds voldoende ervaring hiermee.

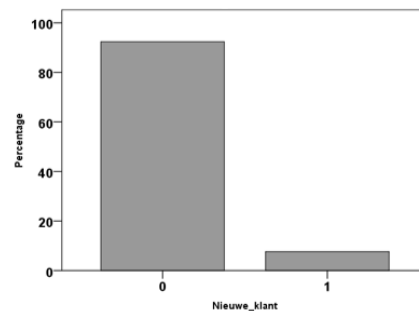
Ervaring self-scanning			
	Frequency	Percent	Cumulative Percent
1	557	69,6	69,6
2	50	6,3	75,9
3	193	24,1	100,0
Total	800	100,0	



Figuur 31: Beschrijvende statistieken categorische variabelen – Ervaring self-scanning

Slechts 7,6% van de entiteiten zijn nieuwe klanten.

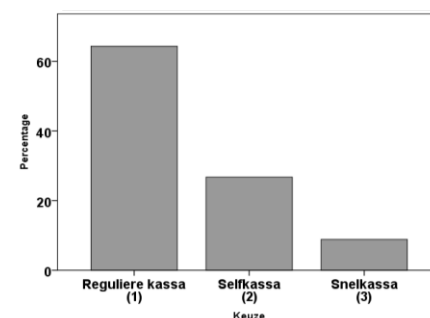
Nieuwe klant			
	Frequency	Percent	Cumulative Percent
0	739	92,4	92,4
1	61	7,6	100,0
Total	800	100,0	



Figuur 32: Beschrijvende statistieken categorische variabelen – Nieuwe klant

Figuur 33 geeft de verdeling van de entiteiten over de drie verschillende kassatypes weer.

Keuze			
	Frequency	Percent	Cumulative Percent
1	515	64,4	64,4
2	214	26,8	91,1
3	71	8,9	100,0
Total	800	100,0	



Figuur 33: Beschrijvende statistieken categorische variabelen – Keuze

3 CLUSTERING ENTITEITEN

Om de entiteiten uit de event log in te delen in een aantal types, wordt op zoek gegaan naar groepen van entiteiten waartussen grote gelijkenissen aanwezig zijn. Tussen de groepen onderling hopen we voldoende grote verschillen waar te nemen. Hiervoor zal gebruik gemaakt worden van de aanwezige clusteringsmethoden in SPSS. Meer bepaald wordt hier gekozen voor TwoStep clustering om de reeds vernoemde voordelen die deze methode biedt. Om relevant te zijn voor het opstellen van een simulatiemodel, is het belangrijk dat de ontdekte groepen of clusters verschillen vertonen op het vlak van aankomstrijtme, genomen beslissingen en doorlooptijd.

3.1 ONDERSCHIEDEN ENTITEITENTYPES

3.1.1 Voorbereidend werk

De gebruikte dataset bevat informatie over 800 observaties. We kunnen bijgevolg maximaal negen attributen opnemen in de clustering wegens de reeds vermeldde vuistregel.

Uit de opgestelde correlatiematrix in figuur 34 blijkt dat er hoge correlaties zijn tussen het aantal producten, de duurtijd van activiteit 3 ('verzamel producten'), 5 ('scan producten reguliere kassa') en 7 ('scan producten snelkassa') en de totale doorlooptijd. Om collineariteitsproblemen te voorkomen kiezen we ervoor om van deze variabelen enkel het aantal producten te gebruiken in de verdere analyse. Hierna blijven er nog 13 variabelen over die gebruikt kunnen worden bij de clustering.

Correlations																	
	Aantal producten	Gezins-grootte	Winkel-frequentie	Ervaring selfscan	Nieuwe klant	Leeftijd	Act1	Act2	Act3	Act4	Act5	Act6	Act7	Act8	Totale duurtijd activiteiten	Wachttijd	Keuze
Aantal producten	1	,577**	-,490**	-,157**	-,046	,058	^b	-,085	,950**	,251**	,957**	,000	,900**	,050	,957**	,075	-,298**
Gezinsgrootte		1	,090 [*]	-,121**	-,043	-,072 [*]	^b	-,044	,511**	,329**	,379**	-,021	,079	,088	,507**	,049	-,162**
Winkelfrequentie			1	,223**	,071	-,313**	^b	,121	-,525**	-,043	-,444**	-,003	-,564**	-,020	-,529**	-,146**	,482**
Ervaring selfscan				1	,142**	-,595**	^b	,071	-,227**	-,144 [*]	-,070	,016	-,449**	,086	-,234**	-,152**	,393**
Nieuwe klant					1	-,206**	^b	-,074	,047	-,053	-,031	,032	-,113	,089	,038	-,005	,129**
Leeftijd						1	^b	-,045	,144**	,421**	-,006	,013	,719**	-,269 [*]	,159**	,193**	-,506**
Act1							^b	^b	^b	^b	^b	^b	^b	^b	^b	^b	^b
Act2								^b	^b	^b	^b	^b	^b	^b	^b	^b	^b
Act3								1	-,100	-,031	^b	^b	^b	^b	-,086	-,136 [*]	^b
Act4									1	,249**	,980**	,007	,933**	-,041	,999**	,064	-,262**
Act5										1	^b	^b	^b	^b	,258 [*]	,013	^b
Act6											1	-,003	^b	^b	,983**	-,006	^b
Act7												1	^b	^b	,011	,014	^b
Act8													1	-,100	,939**	-,013	^b
Totale duurtijd activiteiten															1	,075	-,286**
Wachttijd																1	-,336**
Keuze																	1

** . Correlation is significant at the 0.01 level (2-tailed).
^{*} . Correlation is significant at the 0.05 level (2-tailed).
^b . Cannot be computed because at least one of the variables is constant.

Figuur 34: Correlatiematrix met Pearson correlatiecoëfficiënt

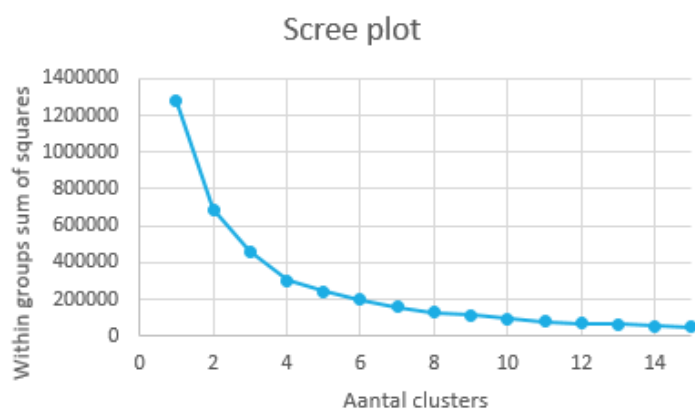
Aangezien maximaal slechts negen variabelen gebruikt mogen worden door de omvang van de dataset, wordt ervoor gekozen om de informatie omtrent duurtijden van individuele activiteiten niet te gebruiken. Daarnaast zal ook de wachttijd niet opgenomen worden in de analyse omdat deze variabele niet beschrijvend is voor de entiteit. Deze variabele is louter het gevolg van de

aanwezigheid van een groter aantal entiteiten in het systeem op een bepaald moment. Tenslotte blijven er nog zeven attributen over die gebruikt zullen worden in de clustering, namelijk 'Aantal producten', 'Gezinsgrootte', 'Winkelfrequentie', 'Ervaring selfscan', 'Nieuwe klant', 'Leeftijd' en 'Keuze'.

3.1.2 Input TwoStep clustering

Allereerst werd reeds besloten om de beschikbare attributen in de log en een keuze-attribuut als input te gebruiken voor de clustering. Dit keuze-attribuut verwijst naar de gekozen procesuitvoering voor een bepaalde entiteit. Voor de clustering wensen we bijgevolg zowel twee continue als vijf categorische variabelen te gebruiken. TwoStep clustering laat toe om deze op een eenvoudige manier in te brengen in de clustering en normaliseert hierbij de continue variabelen. De overige attributen werden opgenomen als evaluatievariabelen. Een evaluatievariabele zal niet gebruikt worden bij de vorming van de clusters, maar kan achteraf inzichten bieden in de verschillen tussen de gevormde clusters.

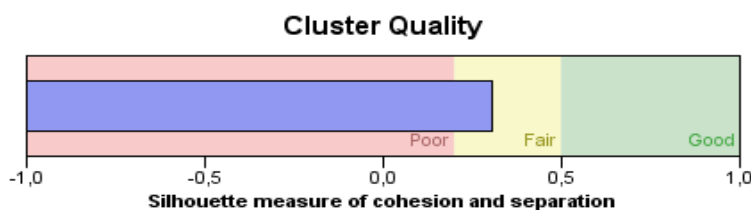
Uit de grafiek in figuur 35 kunnen we afleiden dat een oplossing met twee à vier clusters optimaal is. Op deze punten komt er namelijk een lichte knik voor in de grafiek. Een bovengrens van zes clusters lijkt hier dan ook een gepaste keuze.



Figuur 35: Scree plot voor bepaling aantal clusters

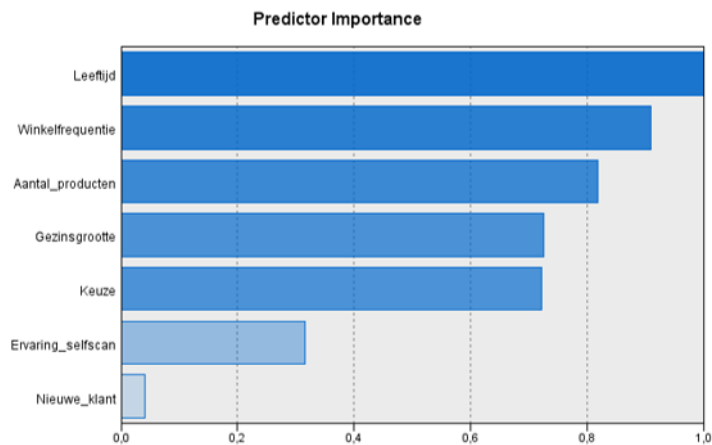
3.1.3 Output TwoStep clustering

Door middel van de TwoStep clustermethode bekomen we vervolgens een oplossing bestaande uit vier clusters. Zoals te zien is in figuur 36, is deze oplossing niet perfect, maar wel voldoende om verder mee aan de slag te gaan.



Figuur 36: Kwaliteit clusteroplossing

Bovendien worden de verschillende attributen in grote mate gebruikt bij het vormen van de clusters (figuur 37). Dit geeft aan dat de gebruikte attributen allen een deel uitmaken van de clustering. Enkel de variabele 'nieuwe klant' heeft een beperkte invloed. Dit kan verklaard worden doordat het een dummyvariabele is waarbij 92,4% van de observaties de waarde '0' heeft.



Figuur 37: Invloed gekozen attributen voor clusteroplossing

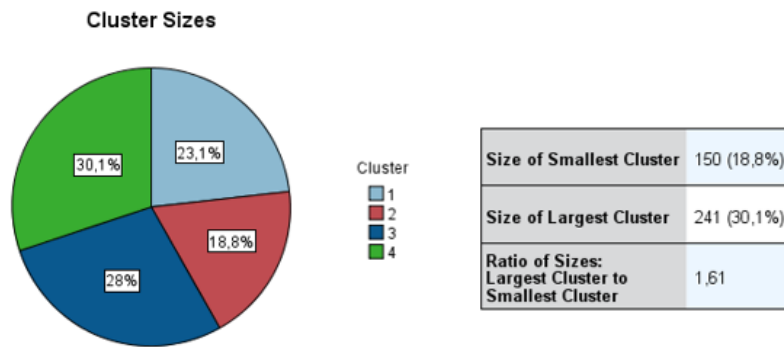
In figuur 38 wordt tenslotte een overzichtstabel weergegeven met de gemiddelde of vaakst voorkomende waarden van de gebruikte clustervariabelen en evaluatievariabelen voor de gevormde clusters. De kleur geeft hierbij aan hoe belangrijk de desbetreffende variabele was voor het vormen van de clusters. Aan de hand van deze outputtabel zullen de gevormde clusters vergeleken worden in de volgende sectie. Na een eerste blik op deze tabel kunnen we besluiten dat er tussen de clusters onderling voldoende verschillen zijn.

Cluster	1	2	3	4
Label				
Description				
Size	23,1% (185)	18,8% (150)	28,0% (224)	30,1% (241)
Inputs	Leeftijd 29,29	Leeftijd 44,58	Leeftijd 41,97	Leeftijd 70,08
	Winkelfrequentie 7,0 (34,1%)	Winkelfrequentie 2,0 (54,7%)	Winkelfrequentie 1,0 (78,6%)	Winkelfrequentie 1,0 (83,4%)
	Aantal_producten 11,76	Aantal_producten 30,07	Aantal_producten 78,93	Aantal_producten 33,44
	Gezinsgrootte 1 (61,6%)	Gezinsgrootte 4 (41,3%)	Gezinsgrootte 4 (30,4%)	Gezinsgrootte 2 (54,8%)
	Keuze 2 (61,1%)	Keuze 1 (78,7%)	Keuze 1 (69,2%)	Keuze 1 (100,0%)
	Ervaring_selfscan 3 (60,0%)	Ervaring_selfscan 1 (72,7%)	Ervaring_selfscan 1 (66,5%)	Ervaring_selfscan 1 (100,0%)
	Nieuwe_klant 0 (85,9%)	Nieuwe_klant 0 (92,7%)	Nieuwe_klant 0 (89,3%)	Nieuwe_klant 0 (100,0%)
Evaluation Fields	Act1 1,00	Act1 1,00	Act1 1,00	Act1 .
	Act2 21,65	Act2 23,19	Act2 18,19	Act2 .
	Act3 595,35	Act3 981,53	Act3 2.167,05	Act3 1.270,95
	Act4 40,32	Act4 45,38	Act4 47,36	Act4 .
	Act5 26,00	Act5 58,17	Act5 165,94	Act5 97,10
	Act6 24,00	Act6 22,26	Act6 21,81	Act6 22,30
	Act7 15,75	Act7 .	Act7 .	Act7 .
	Act8 22,58	Act8 .	Act8 .	Act8 .
	Totale duurtijd activiteiten 648,73	Totale duurtijd activiteiten 1.059,67	Totale duurtijd activiteiten 2.317,45	Totale duurtijd activiteiten 1.390,35
	Wachttijd 3,24	Wachttijd 52,22	Wachttijd 41,05	Wachttijd 58,79

Input (Predictor) Importance

Figuur 38: Output two-step clustering SPSS - overzichtstabel

3.1.4 Beschrijving gevormde clusters



Figuur 39: Omvang gevormde clusters

3.1.4.1 Cluster 1

Deze cluster omvat entiteiten met een gemiddelde leeftijd van 29 jaar. De winkelfrequentie ligt voor deze klanten vaak zeer hoog, waardoor het aantal gekochte producten eerder beperkt is. Het gaat hier om kleine gezinnen die meestal bestaan uit één persoon. Deze klanten hebben vaak reeds veel ervaring opgedaan met self-scanning en kiezen ook vaker voor deze manier van winkelen. Daarnaast valt op dat deze klanten het minste tijd doorbrengen in de winkel. In dit segment komen relatief meer nieuwe klanten voor dan in andere segmenten.

3.1.4.2 Cluster 2

Deze tweede cluster omvat entiteiten met een gemiddelde leeftijd van 45 jaar. De gemiddelde winkelfrequentie ligt voor deze klanten op twee keer, waardoor het aantal gekochte producten per keer lager is. Het gaat hier om gezinnen die meestal bestaan uit vier personen. Deze klanten hebben meestal weinig met self-scanning en kiezen hier ook minder vaak voor. Deze klanten brengen meer tijd door in de winkel dan de klanten uit de eerste cluster, maar minder dan andere klanten. In dit segment komt slechts een beperkt aantal nieuwe klanten voor.

3.1.4.3 Cluster 3

Cluster drie omvat entiteiten met een gemiddelde leeftijd van 42 jaar. De gemiddelde winkelfrequentie ligt voor deze klanten op één keer, wat samenhangt met een hoger aantal gekochte producten. Dit is bovendien het hoogste aantal van alle clusters. Het gaat hier eveneens om gezinnen die meestal bestaan uit vier personen. Deze klanten hebben meestal weinig ervaring met self-scanning en kiezen hier ook minder vaak voor. Deze klanten brengen het meeste tijd door in de winkel. In dit segment komt ook slechts een beperkt aantal nieuwe klanten voor.

3.1.4.4 Cluster 4

De laatste cluster omvat voornamelijk oudere personen met een gemiddelde leeftijd van 70 jaar. Deze winkelen over het algemeen één keer en kopen een vergelijkbaar aantal producten als de entiteiten in cluster twee. Toch doen zij hier langer over en brengen zij dus meer tijd door in de winkel. Meestal bestaat het gezin van een klant in deze cluster uit twee personen. Deze klanten hebben absoluut geen ervaring met self-scanning en maken hier geen gebruik van. In dit segment komen bovendien geen nieuwe klanten voor.

3.2 CONTROLE

Om de controle uit te kunnen voeren, zullen er twee groepen samengevoegd moeten worden. In de oorspronkelijke log werden namelijk slechts drie types ingebracht. De drie oorspronkelijke entiteitentypes waren: jongeren, gezinnen en ouderen. Wanneer we kijken naar de omschrijving van de inhoud van de ontdekte clusters zien we de volgende overeenkomsten:

Tabel 8: Overeenkomsten clusters voor controledoeleinden

Entiteitentype	Oude clusters (3)	Nieuwe clusters (4)
Jongere	1	1
Gezin	2	2&3
Oudere	3	4

Met behulp van Excel worden de nodige transformaties uitgevoerd en wordt geteld hoeveel entiteiten in een juiste groep geplaatst werden. Hieruit blijkt dat 664 van de 800 eenheden correct geclassificeerd werden, wat overeen komt met 83% van de totale populatie. Onderstaande confusion matrix toont op welke manier de fout geclassificeerde entiteiten werden ingeschat. Hierin valt op dat de grote fouten beperkt zijn. Zo werden er geen jongeren geclassificeerd als oudere en werden er slechts acht ouderen als jongere. De meest voorkomende fout is het classificeren van gezinnen als jongeren. Dit gebeurde voor 54 entiteiten, wat overeenkomt met 39,7% van alle foutieve classificaties.

Tabel 9: Confusion matrix

Confusion matrix		Voorspelde groep		
		1	2	3
Werkelijke groep	1	123	37	0
	2	54	318	18
	3	8	19	223

4 OPSTELLEN SIMULATIEMODELLEN

Een volgende stap in de methodologie is het opstellen van de nodige simulatiemodellen. Hierbij werd gekozen om drie verschillende simulatiemodellen op te stellen, namelijk één model dat de realiteit representeert en simulatiemodellen met en zonder entiteitentypes. Dit is nodig om in een volgende stap de vergelijking te kunnen maken tussen de realiteit en het simulatiemodel met en zonder entiteitentypes. Deze vergelijking zal gebaseerd worden op een aantal prestatie maatstaven die reeds besproken werden in de literatuurstudie.

4.1 INPUT SIMULATIEMODEL

In dit onderdeel wordt het aankomst ritme, de gemaakte keuzes en de duurtijd van de activiteiten voor de verschillende clusters beschreven. Hierbij kan verwezen worden naar de originele verdelingen in tabel 4,5 en 6. Bij het maken van een vergelijking hiertussen kan opgemerkt worden dat de originele verdelingen zelden exact teruggevonden worden. Daarnaast wordt tenslotte ook de capaciteit van de verschillende resources in het systeem bestudeerd.

4.1.1 Aankomst ritme

Een eerste parameter die vereist is bij het opstellen van een simulatiemodel is de tussenaankomsttijd. Deze kan verschillend zijn voor elk van de gevonden clusters. De kansverdeling die deze parameter het best omschrijft wordt bepaald door de startmomenten van de eerste activiteiten voor de entiteiten binnen een bepaalde cluster samen te brengen, te sorteren, het verschil hiertussen te berekenen en deze verschillen in te brengen in de input analyzer van Arena. De bekomen kansverdelingen worden weergegeven in tabel 10.

Tabel 10: Gepaste kansverdelingen tussenaankomsttijd

Entiteiten	Gepaste kansverdeling
Cluster 1	GAMM(282, 0.92)
Cluster 2	WEIB(330, 1.1)
Cluster 3	GAMM(200, 1.05)
Cluster 4	GAMM(177, 1.13)
Alle entiteiten	WEIB(59.3, 0.951)

4.1.2 Keuzes beslissingspunten

Een tweede parameter die gekend moet zijn om het simulatiemodel te kunnen opstellen zijn de verhoudingen waarmee de verschillende keuzemogelijkheden voorkomen. Deze zijn nodig om de kansverdelingen te bepalen in de beslissingspunten van het model. Tabel 11 geeft het aantal entiteiten weer dat per cluster een bepaalde keuze maakt en het procentuele voorkomen hiervan. Ook de totaalpercentages per cluster en per keuze kunnen hierin teruggevonden worden.

Tabel 11: Gemaakte keuzes beslissingspunten

Keuze	Clusters				Alle entiteiten
	1	2	3	4	
Reguliere kassa	1	118	155	241	515
	0,54%	78,67%	69,20%	100%	64,375%
Selfkassa	113	32	69	0	214
	61,08%	21,33%	30,80%	0%	26,75%
Snelkassa	71	0	0	0	71
	38,38%	0%	0%	0%	8,875%
Totaal	185	150	224	241	800
	23,1%	18,8%	28%	30,1%	100%

4.1.3 Duurtijd activiteiten

Daarnaast is het ook noodzakelijk om te weten hoe lang de verschillende activiteiten voor de entiteiten uit verschillende clusters zullen duren. Hierbij wordt een gelijkaardige procedure gevolgd waarbij de duurtijden van elke activiteit verzameld worden en verwerkt worden door de input analyzer om op die manier de best aansluitende kansverdeling met bijhorende parameters te bekomen. De resultaten hiervan kunnen teruggevonden worden in tabel 12. In deze tabel kan eveneens gezien worden dat bepaalde activiteiten voor bepaalde clusters niet of zelden plaatsvonden.

Tabel 12: Gepaste kansverdelingen duurtijd activiteiten

Activiteiten	Entiteiten	Gepaste kansverdeling	
Scannen klantenkaart	Cluster 1	1	
	Cluster 2	1	
	Cluster 3	1	
	Cluster 4	n.v.t.	
	Alle entiteiten	1	
Toewijzing scanner	Cluster 1	$-0.5 + 56 * \text{BETA}(1.1, 1.75)$	
	Cluster 2	$0.999 + 52 * \text{BETA}(0.901, 1.2)$	
	Cluster 3	$66 * \text{BETA}(1.14, 3.01)$	
	Cluster 4	n.v.t.	
	Alle entiteiten	$66 * \text{BETA}(1.03, 2.3)$	
Verzamel producten	Cluster 1	Self-scanning	$369 + \text{EXPO}(306)$
		Geen self-scanning	$360 + 340 * \text{BETA}(1.05, 2.18)$
	Cluster 2	Self-scanning	$575 + 1.73e+003 * \text{BETA}(0.895, 1.27)$
		Geen self-scanning	$520 + 980 * \text{BETA}(0.859, 1.36)$
	Cluster 3	Self-scanning	$\text{NORM}(2.47e+003, 786)$
		Geen self-scanning	$700 + 5.6e+003 * \text{BETA}(1.26, 4.04)$
	Cluster 4	Self-scanning	n.v.t.
		Geen self-scanning	$700 + \text{WEIB}(582, 1.07)$
Alle entiteiten	$360 + 5.94e+003 * \text{BETA}(0.96, 5.03)$		
Scan producten reguliere kassa	Cluster 1 (slechts 1 waarde)	26	
	Cluster 2	$21 + 99 * \text{BETA}(0.909, 1.51)$	
	Cluster 3	$\text{NORM}(166, 86.7)$	
	Cluster 4	$40 + \text{WEIB}(59.1, 1.13)$	
	Alle entiteiten	$21 + \text{LOGN}(107, 158)$	
Betaal producten reguliere kassa	Cluster 1 (slechts 1 waarde)	23.54	
	Cluster 2	$15 + 15 * \text{BETA}(0.894, 0.955)$	
	Cluster 3	$15 + 15 * \text{BETA}(0.846, 0.957)$	
	Cluster 4	$15 + 15 * \text{BETA}(0.958, 1.01)$	
	Alle entiteiten	$15 + 15 * \text{BETA}(0.893, 0.982)$	
Scan producten snelkassa	Cluster 1	$\text{POIS}(15.7)$	
	Alle entiteiten		
	Cluster 2,3 en 4	n.v.t.	
Betaal producten snelkassa	Cluster 1	$\text{UNIF}(14.5, 30.5)$	
	Alle entiteiten		
	Cluster 2,3 en 4	n.v.t.	
Betaal aan selfkassa	Cluster 1	$26.5 + 35 * \text{BETA}(1.16, 1.74)$	
	Cluster 2	$\text{NORM}(45.5, 5.93)$	
	Cluster 3	$34 + \text{WEIB}(14.7, 1.54)$	
	Cluster 4	n.v.t.	
	Alle entiteiten	$26 + 42 * \text{BETA}(1.82, 2.59)$	

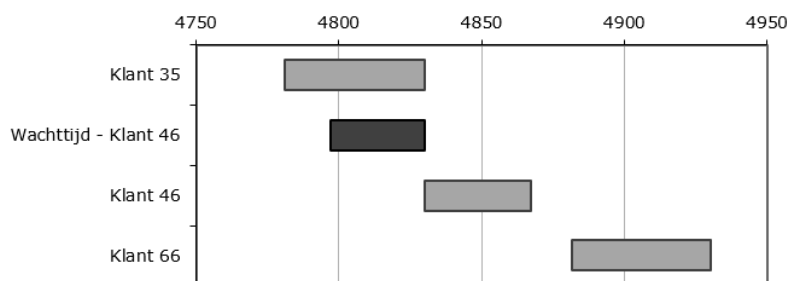
4.1.4 Resources

Als laatste factor moeten we ook de nodige informatie verzamelen over de gebruikte resources. Zo is het belangrijk om te weten welke resources waar in het proces gebruikt worden en hoe groot de capaciteit van deze resourceklassen is. Op die manier weten we hoeveel entiteiten er gelijktijdig verwerkt kunnen worden.

Om deze capaciteit te ontdekken zullen we op zoek gaan naar een entiteit die moest wachten aangezien op dat moment de volledige capaciteit van de desbetreffende resource bereikt was. Vervolgens kijken we naar de situatie van de resources op dat ogenblik om te weten te komen hoeveel entiteiten er behandeld worden. Een assumptie die we hierbij maken is dat de capaciteit van de resources constant is.

Bij zowel resource één als twee komen geen wachttijden voor in de event log. We kunnen er dus vanuit gaan dat er geen tekorten zullen zijn in een gelijkaardige situatie. Om concrete informatie te verkrijgen over de limieten van deze capaciteiten, kan in reële situaties gebruik gemaakt worden van documentatie of interviews.

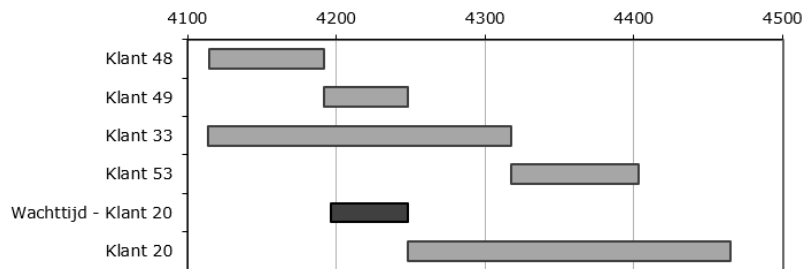
Voor resourceklasse drie komen er wel wachttijden voor. Zo is in figuur 40 te zien dat op het tijdstip waarop klant 46 arriveerde aan resource drie (4797,10), deze nog bezig was met het verwerken van klant 35. Klant 46 werd vervolgens in een wachtrij geplaatst totdat klant 35 deze resource niet meer nodig had (4830,27). Hieruit kunnen we afleiden dat de capaciteit van deze resource één bedraagt. Een gelijkaardig onderzoek doen we voor resources vier en vijf.



Figuur 40: Capaciteitsbepaling resourceklasse 3

Voor resource vier zien we een gelijkaardig patroon verschijnen. Deze resourceklasse beschikt eveneens over een capaciteit van één klant.

Voor resource vijf kan in figuur 41 opgemerkt worden dat deze zowel bezig was met de verwerking van klant 49 (4191,72 – 4248,29) en 33 (4113,65 – 4317,82) op het moment dat klant 20 arriveerde aan de reguliere kassa (4195,99). Van zodra de resource klaar is met 49 zal deze starten met klant 20 (4248,29 – 4464,96). Hieruit kunnen we afleiden dat de capaciteit van deze resource twee bedraagt. Verder kunnen we hieruit afleiden dat wie eerst aankomt eerst bediend zal worden in dit systeem. Dit zou verder nagegaan moeten worden voor alle cases om zeker te kunnen zijn dat er nergens prioriteiten toegepast worden. Aangezien er geen informatie gekend is over eventuele prioriteiten, veronderstellen we dat deze niet gebruikt worden binnen dit systeem.



Figuur 41: Capaciteitsbepaling resourceklasse 5

Een korte samenvatting van de capaciteit van de respectievelijke resources in dit systeem, wordt gegeven in tabel 13.

Tabel 13: Overzicht capaciteit resourceklassen

Resource	Capaciteit
R1 – Scanner klantenkaart	Onbeperkt
R2 – Handscanners	Onbeperkt
R3 – Selfkassa	1
R4 – Snelkassa	1
R5 – Reguliere kassa	2

4.2 SCENARIO'S

4.2.1 Scenario 1: De realiteit

In dit eerste scenario zal alle informatie die gebruikt werd bij het opstellen van de event log verwerkt worden. Het gaat hier zowel om entiteitentypes en bijhorende kansverdelingen van attribuutwaarden, als om de formules die de gemaakte keuzes en duurtijden van de activiteiten bepalen. Dit model zal beschouwd worden als de realiteit waarmee de andere modellen vergeleken zullen worden. In bijlage twee zal de inhoud van dit model aan de hand van enkele figuren weergegeven worden.

4.2.2 Scenario 2: Model zonder entiteitentypes

In dit model zal geen onderscheid gemaakt worden tussen de verschillende entiteitentypes die aanwezig zijn in de event log. Hierbij worden de kansverdelingen die ontdekt werden voor alle entiteiten gebruikt. De inhoud van dit model kan teruggevonden worden in bijlage drie.

4.2.3 Scenario 3: Model met entiteitentypes

In het laatste model zullen alle opgedane inzichten omtrent de aanwezige entiteitentypes geïntegreerd worden in het model. Hiermee bedoelen we het verschil in aankomstrijme, beslissingen en duurtijden van activiteiten voor de ontdekte clusters. Dit simulatiemodel wordt weergegeven in bijlage vier.

5 VERGELIJKING SIMULATIEMODELLEN VOLGENS PRESTATIEMAATSTAVEN

In deze laatste stap zullen de resulterende statistieken van de drie modellen voor de gekozen prestatemaatstaven vergeleken worden met elkaar. Zo zullen de gemiddelde totale doorlooptijd, het gemiddeld aantal entiteiten in het systeem, de gemiddelde wachttijden en de bezettingsgraad per kassa met elkaar vergeleken worden. Dit zijn namelijk universele maatstaven die in de literatuur meermaals aangehaald of toegepast worden. We hopen hierbij aan te kunnen tonen dat deze statistieken voor het simulatiemodel "met entiteitentypes" ten opzichte van het simulatiemodel "zonder entiteitentypes" beter aansluiten bij deze van de realiteit.

Om deze gemiddeldes en bijhorende betrouwbaarheidsintervallen te bekomen werden de drie opgestelde simulatiemodellen 100 keer uitgevoerd. Indien er meerdere entiteitentypes aanwezig waren in het model werd een gewogen gemiddelde gemaakt en werden de betrouwbaarheidsintervallen hiervan herberekend aan de hand van de volgende formules:

$$\text{Gewogen gemiddelde} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i} \text{ met } w_i \text{ gewichten en } x_i \text{ gemiddeldes}$$

$$BI_{95\%} = \bar{X} \pm \text{Half width}$$

$$\text{Half width} = 1,96 \sigma \text{ waaruit volgt dat } \sigma = \frac{\text{Half width}}{1,96}$$

$$\text{Var}(X) = \sigma^2$$

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y) \text{ met } \text{Cov}(X, Y) = 0$$

Deze laatste formule is bovendien toepasbaar op meer dan twee variabelen en kan uitgebreid worden als volgt:

$$\text{Var}\left(\sum_i^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} a_i a_j \text{Cov}(X_i, X_j) \text{ met } \text{Cov}(X, Y) = 0$$

Hierbij wordt de assumptie gemaakt dat de observaties onafhankelijk zijn.

Hoe deze formules concreet gebruikt worden om het nieuwe betrouwbaarheidsinterval te berekenen wordt weergegeven in figuur 42.

	A	B	C	D	E	F	G	H
1	SCENARIO 1	Bijlage 5.5				Bijlage 5.1		
2			gewicht	gewicht ²	gemiddelde	half width	standaardafwijking	standaardafwijking ²
3	Jongere	160,99	0,201238	0,040497	875,1	9,36	4,775510204	22,80549771
4	Gezin	401,24	0,50155	0,251552	1734,7	11,87	6,056122449	36,67661912
5	Gepensioneerde	237,77	0,297213	0,088335	1524,48	8,11	4,137755102	17,12101728
6	TOTAAL	800			1499,236233		3,414970702	
7								
8		=B3/B6*E3+B4/B6*E4+B5/B6*E5			lower limit	1492,5429	=E6-1,96*G6	=SQRT(D3*H3+D4*H4+D5*H5)
9					upper limit	1505,9296	=E6+1,96*G6	

Figuur 42: Berekening betrouwbaarheidsinterval: Totale doorlooptijd - Scenario 1

Voor het totale aantal entiteiten in het systeem werden de gemiddelden van de verschillende deelgroepen opgeteld volgens de volgende formule:

$$E(X + Y) = E(X) + E(Y)$$

Vervolgens werden de betrouwbaarheidsintervallen hier berekend met de volgende formule:

$$Var(X + Y) = Var(X) + Var(Y) + Cov(X, Y) \text{ met } Cov(X, Y) = 0$$

Hierbij werd eveneens onafhankelijkheid verondersteld. De concrete toepassing van de formules voor het berekenen van het gemiddeld aantal klanten in het systeem en het bijhorende betrouwbaarheidsinterval, wordt getoond in figuur 43.

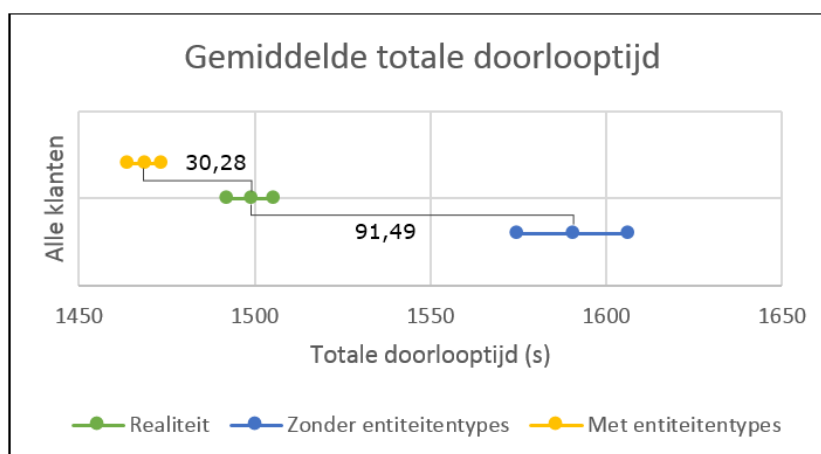
	A	B	C	D	E
17	SCENARIO 3		Bijlage 5.2		
18		gemiddelde	half width	standaardafwijking	standaardafwijking ²
19	cluster 1	2,2884	0,02	0,010204082	0,000104123
20	cluster 2	3,2344	0,04	0,020408163	0,000416493
21	cluster 3	10,1033	0,12	0,06122449	0,003748438
22	cluster 4	6,7416	0,07	0,035714286	0,00127551
23	TOTAAL	22,3677	0,1459452	0,074461834	0,005544565
24		=B19+B20+B21+B22	=1,96*D23	=SQRT(E23)	=E19+E20+E21+E22
25		lower limit	22,22	=B23-C23	
26		upper limit	22,51	=B23+C23	

Figuur 43: Berekening betrouwbaarheidsinterval: Aantal entiteiten in het systeem - Scenario 3

In bijlage vijf kunnen de relevante onderdelen van de outputrapporten die gebruikt werden voor de berekeningen teruggevonden worden. Daarnaast werd ook de resulterende tabel met alle gemiddelden en betrouwbaarheidsintervallen die in dit onderdeel grafisch weergegeven zullen worden, opgenomen in bijlage zes.

5.1 TOTALE DOORLOOPTIJD

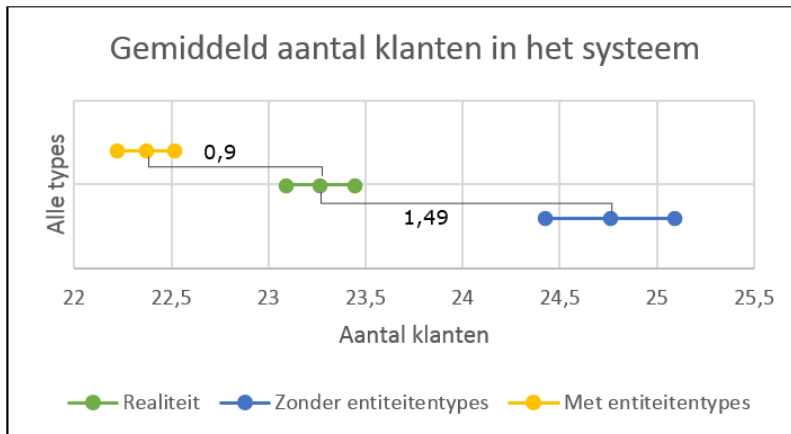
Allereerst wordt de gemiddelde totale doorlooptijd voor alle klanten in het systeem vergeleken. In deze eerste grafiek is te zien dat de waarden voor het model met entiteitentypes dichter bij de realiteit liggen dan deze van het model zonder entiteitentypes. Het gaat hier om een verschil van respectievelijk 30,28 seconden ten opzichte van 91,49 seconden tussen deze gemiddeldes.



Figuur 44: Vergelijking resultaten – Gemiddelde totale doorlooptijd

5.2 TOTAAL GEMIDDELD AANTAL ENTITEITEN IN HET SYSTEEM

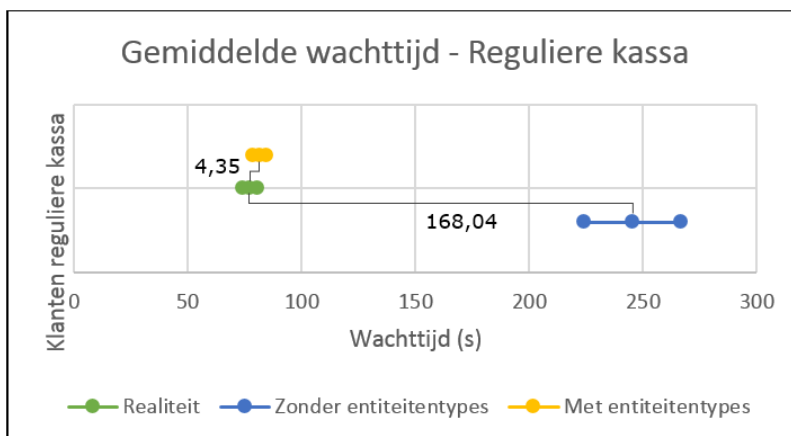
Een volgende prestatemaatstaf die besproken zal worden is het gemiddeld aantal entiteiten in het systeem. Hierbij kan eveneens opgemerkt worden dat de resultaten van het model met entiteitentypes dichter aansluiten bij de realiteit. De verschillen tussen de gemiddeldes bedragen respectievelijk 0,9 en 1,49 klanten.



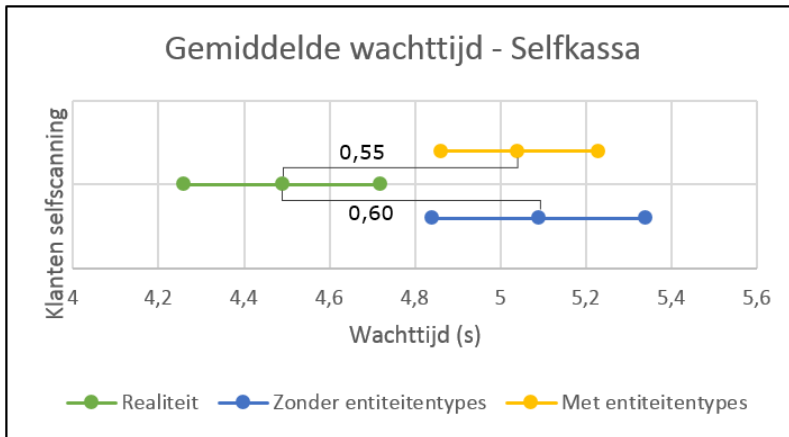
Figuur 45: Vergelijking resultaten – Totaal gemiddeld aantal entiteiten in het systeem

5.3 GEMIDDELTE WACHTTIJDEN PER TYPE KASSA

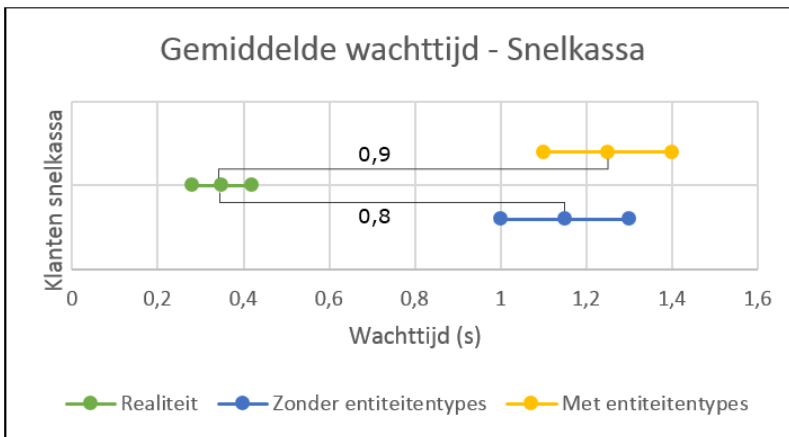
Vervolgens zullen de gemiddelde wachttijden per type kassa besproken worden in de drie opgestelde modellen. Hierbij valt op dat hoewel het model met entiteitentypes zeer goed scoort voor de wachttijden aan de reguliere kassa, de verschillen bij de selfkassa en snelkassa zeer beperkt zijn. Het gaat hierbij in verhouding ook over verwaarloosbare wachttijden van slechts enkele seconden of zelfs maar een fractie hiervan. De respectievelijke verschillen tussen de gemiddeldes kunnen teruggevonden worden in onderstaande grafieken.



Figuur 46: Vergelijking resultaten – Gemiddelde wachttijd (Reguliere kassa)



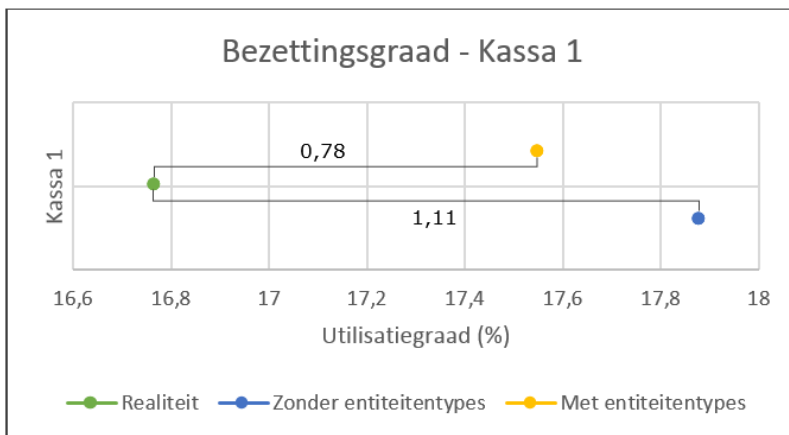
Figuur 47: Vergelijking resultaten – Gemiddelde wachttijd (Selfkassa)



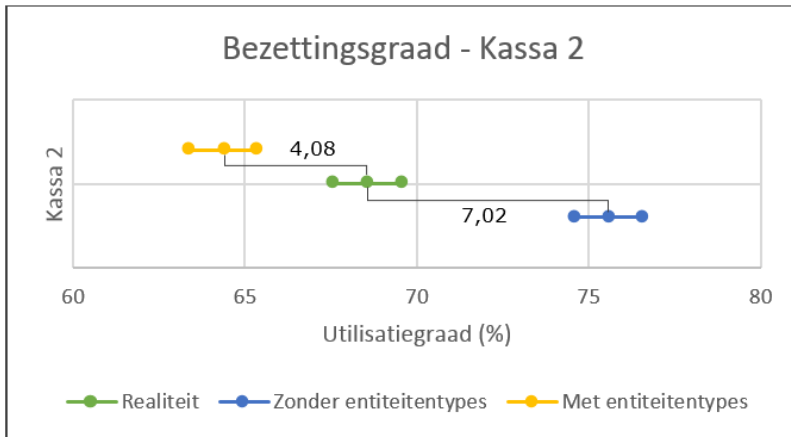
Figuur 48: Vergelijking resultaten – Gemiddelde wachttijd (Snelkassa)

5.4 BEZETTINGSGRAAD KASSA'S

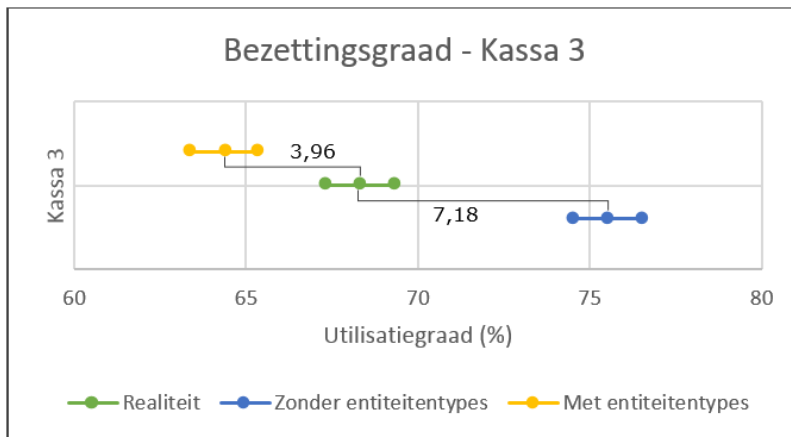
Tenslotte zal ook de bezettingsgraad per kassa besproken worden voor de drie opgestelde modellen. Hierbij valt op dat de resulterende prestatie maatstaven van het model met entiteitentypes steeds iets beter aansluiten bij de realiteit. Toch blijven de verschillen hier eerder beperkt. De exacte waarden kunnen zowel teruggevonden worden in onderstaande grafieken als in de overzichtstabel in bijlage zes.



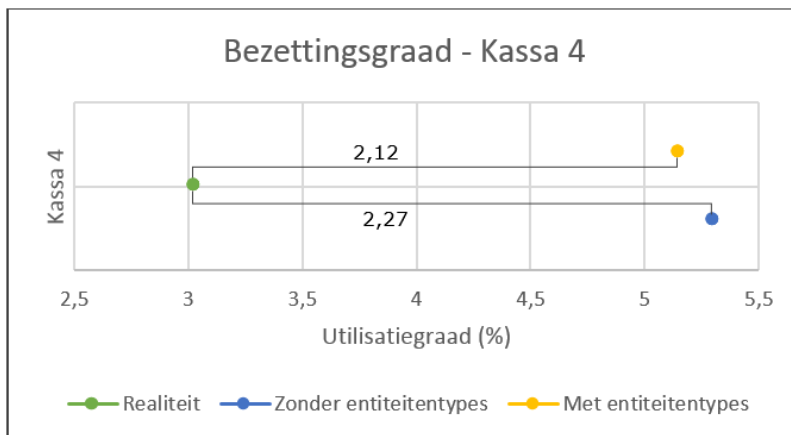
Figuur 49: Vergelijking resultaten – Bezettingsgraad (Kassa 1)



Figuur 50: Vergelijking resultaten – Bezettingsgraad (Kassa 2)



Figuur 51: Vergelijking resultaten – Bezettingsgraad (Kassa 3)



Figuur 52: Vergelijking resultaten – Bezettingsgraad (Kassa 4)

HOOFDSTUK 5: CONCLUSIES

Een relatief nieuw onderzoeksdomein combineert process mining en simulatietechnieken. Op die manier wordt getracht een antwoord te bieden aan de kritieken die momenteel gerelateerd zijn aan simulatietechnieken. Zo wordt beweerd dat deze onvoldoende ondersteuning bieden voor operationele beslissingen, onvoldoende gebruik maken van beschikbare data en dat resources soms te naïef gemodelleerd worden. Binnen dit domein werden reeds enkele onderzoekspogingen ondernomen, maar er werd nog geen universele, geïntegreerde aanpak gecreëerd om op basis van event logs simulatiemodellen te ontwikkelen. Daarnaast valt ook op dat de ondernomen pogingen zich voorlopig beperkten tot activiteiten en hun duurtijd, het definiëren van resourceklassen en de beslissingslogica in een proces. Onderzoekers geven bovendien aan grote voordelen te zien in het combineren van process mining en simulatietechnieken.

In deze masterproef werd getracht om een simulatiemodel dichter te laten aanleunen bij de realiteit door hierin entiteitentypes te integreren die geleerd werden uit een event log. De gebruikte methode voor het extraheren van relevante entiteitentypes en het ontwikkelen van een realistischer simulatiemodel bestond uit de volgende stappen:

Allereerst werd een beschrijvende analyse uitgevoerd op de beschikbare event log. Hierbij werden zowel vergelijkende statistieken gebruikt om de attribuutwaarden te bestuderen als process mining om de structuur van het proces in kaart te brengen. Andere methoden die gebruikt werden om meer inzicht te verkrijgen in de event log waren onder andere het filteren en sorteren van data in Excel.

Vervolgens werd een clusteranalyse van de entiteiten uitgevoerd met de TwoStep clustercomponent in SPSS om op die manier relevante entiteitentypes te ontdekken. Relevante entiteitentypes werden hierbij gedefinieerd als types waarvoor de uitvoering van het proces verschilt op het vlak van aankomstrijtme, gemaakte beslissingen en duurtijd van de activiteiten. Voor de clustering werden zowel attributen van de entiteiten opgenomen alsook attributen die betrekking hadden op de procesuitvoering. Een controle wees uit dat 83% van de entiteiten correct geclassificeerd was.

Tenslotte werd informatie verzameld over het aankomstrijtme, de duurtijd van de activiteiten en de gemaakte beslissingen voor de verschillende clusters. Deze informatie werd geïntegreerd bij het opstellen van het simulatiemodel met entiteitentypes. Daarnaast werd ook een simulatiemodel opgesteld zonder entiteitentypes en een model dat de realiteit weerspiegelt. De prestatie maatstaven die vervolgens gebruikt werden om de resultaten van de drie gecreëerde scenario's te vergelijken waren de gemiddelde totale doorlooptijd, het totaal gemiddeld aantal klanten in het systeem, de gemiddelde wachttijden voor de drie aanwezige kassatypes en de bezettingsgraad van elke kassa.

Na uitvoering van het empirisch onderzoek kan besloten worden dat het opnemen van kennis omtrent relevante entiteitentypes er effectief voor kan zorgen dat het resulterende simulatiemodel beter aansluit bij de realiteit. Een uitzondering op deze conclusie vormde de gemiddelde wachttijd

bij de snelkassa. Hier bleek het model zonder entiteitentypes iets beter aan te sluiten bij de realiteit. De desbetreffende gemiddelde wachttijd in realiteit was zeer klein (0,35 seconden) en het verschil tussen beide scenario's was verwaarloosbaar (1,15s \Leftrightarrow 1,25s). Een artificiële setting met een hoger aankomst ritme, bij eenzelfde verwerkingscapaciteit en bijgevolg iets langere wachttijden, had hier eventueel meer duidelijkheid kunnen scheppen.

Toch merken we op dat er nog steeds een verschil aanwezig is tussen de realiteit en het model met entiteitentypes. Dit kan mogelijk deels verklaarbaar zijn door het feit dat de genomen beslissingen bij het model met entiteitentypes gebaseerd zijn op kansen en niet zozeer op de attribuutwaarden van deze entiteiten. Decision mining zou hiervoor in verder uitgediept en toegepast kunnen worden in toekomstig onderzoek. Bovendien werden in deze masterproef enkele assumpties en simplificaties gemaakt. Zo zijn resources binnen dit systeem bijvoorbeeld continu beschikbaar en werd er niet gewerkt met prioriteiten.

Binnen de beperkte tijdspanne waarin deze masterproef tot stand kwam, is het eveneens niet gelukt om de ontwikkelde methode toe te passen op reële data. Er werd in dit onderzoek bijgevolg enkel gewerkt met een geheel artificiële setting. Verder onderzoek zou de bruikbaarheid van de ontwikkelde methodologie kunnen onderzoeken in een reële omgeving.

LIJST VAN GERAADPLEEGDE WERKEN

- Aguilar, M., Rautert, T., & Pater, A. J. G. (1999). *Business process simulation: a fundamental step supporting process centered management*. Paper presented in the 31st conference on Winter simulation: Simulation—a bridge to the future-Volume 2, Phoenix, Arizona, USA. New York, NY: ACM. doi:10.1145/324898.325282
- Aguirre, S., Parra, C., & Alvarado, J. (2013). Combination of Process Mining and Simulation Techniques for Business Process Redesign: A Methodological Approach. *Data-Driven Process Discovery and Analysis*, 162, 24–43.
- Baines, T., Mason, S., Siebers, P.-O., & Ladbrook, J. (2004). Humans: the missing link in manufacturing simulation? *Simulation Modelling Practice and Theory*, 12(7-8), 515–526. doi:10.1016/S1569-190X(03)00094-7
- Bose, R. P. J. C., & van der Aalst, W. M. P. (2009). *Context Aware Trace Clustering: Towards Improving Process Mining Results*. Paper presented in the 2009 SIAM International Conference on Data Mining. Philadelphia, USA: SIAM. doi: <http://dx.doi.org/10.1137/1.9781611972795.35>
- Chung, C. A. (2004). *Simulation Modeling Handbook: A Practical Approach*. Boca Raton, FL: CRC Press.
- De Medeiros, A. K. A., Guzzo, A., Greco, G., van der Aalst, W. M. P., Weijters, A. J. M. M., Van Dongen, B. F., & Saccà, D. (2008). Process mining based on clustering: A quest for precision. *Business Process Management Workshops*, 4928, 17–29. doi: 10.1007/978-3-540-78238-4_4
- De Weerd, J., vanden Broucke, S., Vanthienen, J., & Baesens, B. (2013). Active trace clustering for improved process discovery. *Knowledge and Data Engineering, IEEE Transactions on*, 25(12), 2708–2720. doi: 10.1109/TKDE.2013.64
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th edition). Upper Saddle River: Pearson Prentice Hall
- Hlupic, V., & Robinson, S. (1998). *Business process modelling and analysis using discrete-event simulation*. Paper presented in the 30th conference on Winter simulation. Washington, DC, USA: IEEE Computer Society Press.
- IBM Knowledge Center (1989, 2011). *Model Summary View*. Geraadpleegd via http://www-01.ibm.com/support/knowledgecenter/SSLVMB_20.0.0/com.ibm.spss.statistics.help/clusterviewer_modelsummary_panel.htm
- Ingalls, R. G. (2008). Introduction to simulation. *Proceedings of the 40th Conference on Winter Simulation*, 17–26.

- Isken, M. W., & Rajagopalan, B. (2002). Data Mining to Support Simulation Modeling of Patient Flow in Hospitals. *Journal of Medical Systems*, 26(2), 179–197.
doi:10.1023/A:1014814111524
- Kelton, W.D., Sadowski, R.P. & Swets, N.B. (2010). *Simulation with Arena* (5th edition). New York, NY: McGraw-Hill.
- Laguna, M., & Marklund, J. (2013). *Business Process Modeling, Simulation and Design*. (2nd edition). Boca Raton, FL: CRC Press.
- Liu, Y., Zhang, H., Li, C., & Jiao, R. J. (2012). Workflow simulation for operational decision support using event graph through process mining. *Decision Support Systems*, 52(3), 685–697.
- Maria, A. (1997). *Introduction to modeling and simulation*. Paper presented in the 29th conference on Winter simulation. Washington, DC: IEEE Computer Society. doi: 10.1145/268437.268440
- Martin, N., Depaire, B., & Caris, A. (2014). *The use of process mining in a business process simulation context: Overview and challenges*. Paper presented in the 2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM). Orlando, FL: IEEE. doi: 10.1109/CIDM.2014.7008693
- Märuster, L., & van Beest, N. R. T. P. (2009). Redesigning business processes: a methodology based on simulation and process mining techniques. *Knowledge and Information Systems*, 21(3), 267–297.
- Mooi, E., & Sarstedt, M. (2011). *A Concise Guide to Market Research*. Berlin Heidelberg: Springer-Verlag. doi: 10.1007/978-3-642-12541-6_9
- Nakatumba, J., & van der Aalst, W. M. P. (2010). Analyzing resource behavior using process mining. *Business Process Management Workshops*, 43, 69–80.
- Nakatumba, J., Westergaard, M., & van der Aalst, W. M. P. (2012). Generating Event Logs with Workload-Dependent Speeds from Simulation Models. *Advanced Information Systems Engineering Workshops*, 112, 383–397.
- Paul, R. J., Hlupic, V., & Giaglis, G. M. (1998). *Simulation modeling of business processes*. Paper presented in the 3rd UK Academy of Information Systems Conference. Lincoln, UK: McGraw-Hill.
- Robinson, S. (2013). *Conceptual modeling for simulation*. Paper presented in the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World. Piscataway, NJ: IEEE Press.
- Rozinat, A., Mans, R. S., Song, M., & van der Aalst, W. M. P. (2008). Discovering colored Petri nets from event logs. *International Journal on Software Tools for Technology Transfer*, 10(1), 57–74. doi: 10.1007/s10009-007-0051-0

- Rozinat, A., Mans, R. S., Song, M., & van der Aalst, W. M. P. (2009). Discovering simulation models. *Information Systems*, 34(3), 305–327. doi: 10.1016/j.is.2008.09.002
- Rozinat, A., & van der Aalst, W. M. P. (2006). Decision mining in business processes. *Beta, Research School for Operations Management and Logistics*. Geraadpleegd via http://www.processmining.org/_media/publications/beta_164.pdf
- Rozinat, A., Wynn, M. T., van der Aalst, W. M. P., ter Hofstede, A. H. M., & Fidge, C. J. (2009). Workflow simulation for operational decision support. *Data & Knowledge Engineering*, 68(9), 834–850.
- Song, M., & Van der Aalst, W. M. P. (2008). Towards comprehensive support for organizational mining. *Decision Support Systems*, 46(1), 300–317. doi:10.1016/j.dss.2008.07.002
- SPSS Inc. (2001). *The SPSS TwoStep Cluster Component - A scalable component to segment your customers more effectively*. Geraadpleegd via http://www.spss.ch/upload/1122644952_The%20SPSS%20TwoStep%20Cluster%20Component.pdf
- Tumay, K. (1996). *Business process simulation*. Paper presented in the 28th conference on Winter simulation. Washington, DC: IEEE Computer Society. doi: 10.1145/256562.256581
- Van Beest, N. R. T. P., & Mărușter, L. (2007). A Process Mining Approach to Redesign Business Processes-a case study in gas industry. Paper presented in the International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC). Timisoara: IEEE. doi: 10.1109/SYNASC.2007.50
- Van der Aalst, W. M. P. (2010). Business process simulation revisited. *Enterprise and Organizational Modeling and Simulation*, 63, 1–14. doi: 10.1007/978-3-642-15723-3_1
- Van der Aalst, W. M. P. (2011). *Process Mining: Discovery, conformance and enhancement of business processes*. Berlin Heidelberg: Springer-Verlag. doi: 10.1007/978-3-642-19345-3
- Van der Aalst, W. M. P. (2015). Business process simulation survival guide. *Handbook on Business Process Management 1*, 337–370.
- Van der Aalst, W. M. P., Adriansyah, A., de Medeiros, A. K. A., Arcieri, F., Baier, T., Blickle, T., ... Wynn, M. (2012). Process mining manifesto. *Business process management workshops*, 99, 169–194. doi: 10.1007/978-3-642-28108-2_19
- Van der Aalst, W. M. P., & Günther, C. W. (2007). *Finding structure in unstructured processes: The case for process mining*. Paper presented in the Seventh International Conference on Application of Concurrency to System Design (ACSD). Bratislava: IEEE. doi:10.1109/ACSD.2007.50
- Van der Aalst, W. M. P., Nakatumba, J., Rozinat, A., & Russell, N. (2008). Business Process Simulation: How to get it right. *BPM Center Report BPM-08-07, BPMcenter.Org*. Geraadpleegd via <http://www.win.tue.nl/~jnakatum/publications/SimulationPaper.pdf>

- Van der Aalst, W. M. P., van Dongen, B. F., Günther, C. W., Rozinat, A., Verbeek, E., & Weijters, A. J. M. M. (2009). *ProM: The Process Mining Toolkit*. Paper presented in the Business Process Management Demonstration Track (BPM Demos 2009). Ulm, Germany: DBLP.
- Van der Aalst, W. M. P., Reijers, H. A., & Song, M. (2005). Discovering social networks from event logs. *Computer Supported Cooperative Work (CSCW)*, 14(6), 549–593. doi: 10.1007/s10606-005-9005-9
- Weske, M. (2012). *Business Process Management: Concepts, Languages, Architectures* (2nd edition) Berlin Heidelberg: Springer-Verlag. doi: 10.1007/978-3-642-28616-2
- White, S. A. (2004). *Introduction to BPMN*. IBM Cooperation, Geraadpleegd via <https://www.ibm.com/developerworks/community/files/basic/anonymous/api/library/7624eb5a-089a-41bf-9b71-b3c33739e18d/document/e908d328-7b50-40e3-8107-70af4e6bb48f/media>
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3rd edition) San Francisco, CA: Morgan Kaufmann.
- Wynn, M. T., Dumas, M., Fidge, C. J., Ter Hofstede, A. H. M., & van der Aalst, W. M. P. (2008). Business process simulation for operational decision support. *Business Process Management Workshops*, 4928, 66–77.

BIJLAGEN

BIJLAGE 1: SUBPROCEDURES

1.1 SUBPROCEDURE OPTELLEN DUURTIJDEN

```
Sub totale_duurtijd()  
Dim HuidigeKlant, VolgendeKlant As Integer  
Dim Doorlooptijd As Double  
Dim Waarde, Som As Double  
Dim i, j As Integer  
  
Worksheets("Sheet1").Activate  
Worksheets("Sheet1").Range("A2").Activate  
Som = 0  
i = 2  
j = 8  
  
Do While ActiveCell.Value <> Empty  
    HuidigeKlant = ActiveCell.Value  
    VolgendeKlant = ActiveCell.Offset(1, 0).Value  
  
    If HuidigeKlant = VolgendeKlant Then  
        Waarde = ActiveCell.Offset(0, 5).Value  
        Som = Som + Waarde  
    Else  
        Waarde = ActiveCell.Offset(0, 5).Value  
        Doorlooptijd = Som + Waarde  
        Worksheets("Sheet2").Activate  
        Worksheets("Sheet2").Cells(i, j) = Doorlooptijd  
        i = i + 1  
        Som = 0  
    End If  
  
    Worksheets("Sheet1").Activate  
    ActiveCell.Offset(1, 0).Activate  
Loop  
  
End Sub
```

1.2 SUBPROCEDURE OPTELLEN WACHTTIJDEN

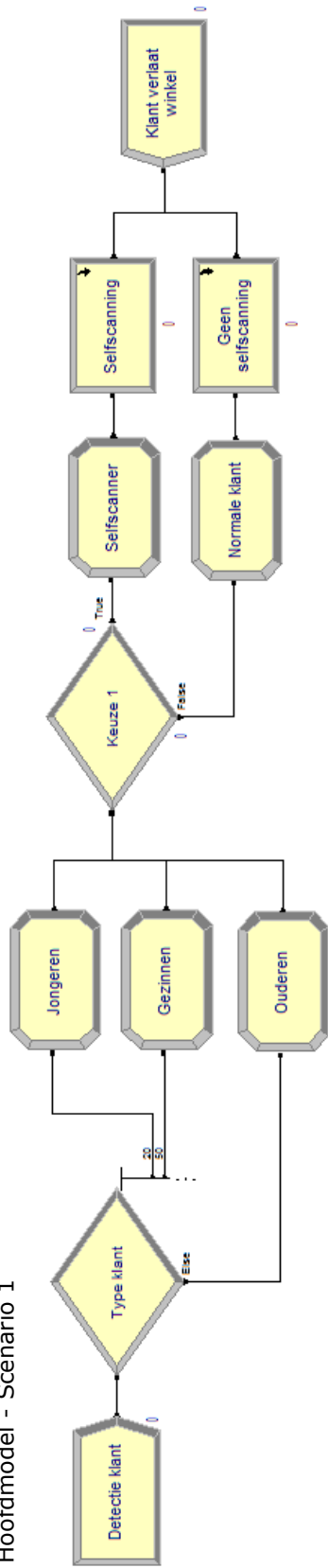
```
Sub totale_wachttijd()  
Dim HuidigeKlant, VolgendeKlant As Integer  
Dim Wachttijd As Double  
Dim Waarde, Som As Double  
Dim i, j As Integer  
  
Worksheets("Sheet1").Activate  
Worksheets("Sheet1").Range("A2").Activate  
Som = 0  
i = 2  
j = 9  
  
Do While ActiveCell.Value <> Empty  
    HuidigeKlant = ActiveCell.Value  
    VolgendeKlant = ActiveCell.Offset(1, 0).Value  
  
    If HuidigeKlant = VolgendeKlant Then  
        Waarde = ActiveCell.Offset(0, 6).Value  
        Som = Som + Waarde  
    Else  
        Waarde = ActiveCell.Offset(0, 6).Value  
        Wachttijd = Som + Waarde  
        Worksheets("Sheet2").Activate  
        Worksheets("Sheet2").Cells(i, j) = Wachttijd  
        i = i + 1  
        Som = 0  
    End If  
  
    Worksheets("Sheet1").Activate  
    ActiveCell.Offset(1, 0).Activate  
Loop  
  
End Sub
```

1.3 SUBPROCEDURE INDIVIDUELE DUURTIJD ACTIVITEITEN

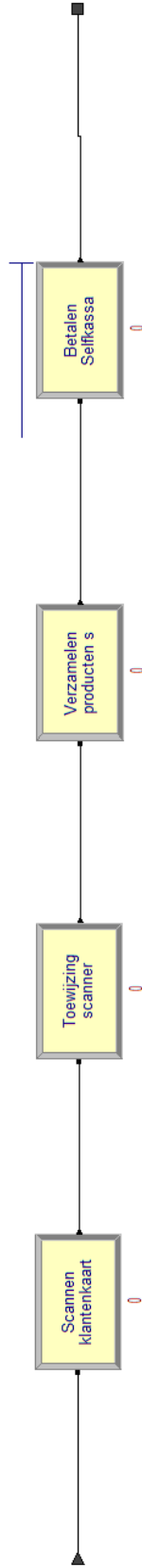
```
Sub activiteitenduurtijd()  
  
Dim Activiteit, Duurtijd As Integer  
Dim i, j As Integer  
  
Worksheets("Sheet1").Activate  
Worksheets("Sheet1").Range("A2").Activate  
i = 2  
Do While ActiveCell.Value <> Empty  
    i = ActiveCell.Value  
    Activiteit = ActiveCell.Offset(0, 2).Value  
    Duurtijd = ActiveCell.Offset(0, 6).Value  
  
    Worksheets("Sheet2").Activate  
    Select Case Activiteit  
        Case 1  
            Worksheets("Sheet2").Cells(i + 1, 8) = Duurtijd  
        Case 2  
            Worksheets("Sheet2").Cells(i + 1, 9) = Duurtijd  
        Case 3  
            Worksheets("Sheet2").Cells(i + 1, 10) = Duurtijd  
        Case 4  
            Worksheets("Sheet2").Cells(i + 1, 11) = Duurtijd  
        Case 5  
            Worksheets("Sheet2").Cells(i + 1, 12) = Duurtijd  
        Case 6  
            Worksheets("Sheet2").Cells(i + 1, 13) = Duurtijd  
        Case 7  
            Worksheets("Sheet2").Cells(i + 1, 14) = Duurtijd  
        Case 8  
            Worksheets("Sheet2").Cells(i + 1, 15) = Duurtijd  
    End Select  
  
    Worksheets("Sheet1").Activate  
    ActiveCell.Offset(1, 0).Activate  
Loop  
  
End Sub
```

BIJLAGE 2: SIMULATIEMODEL REALITEIT

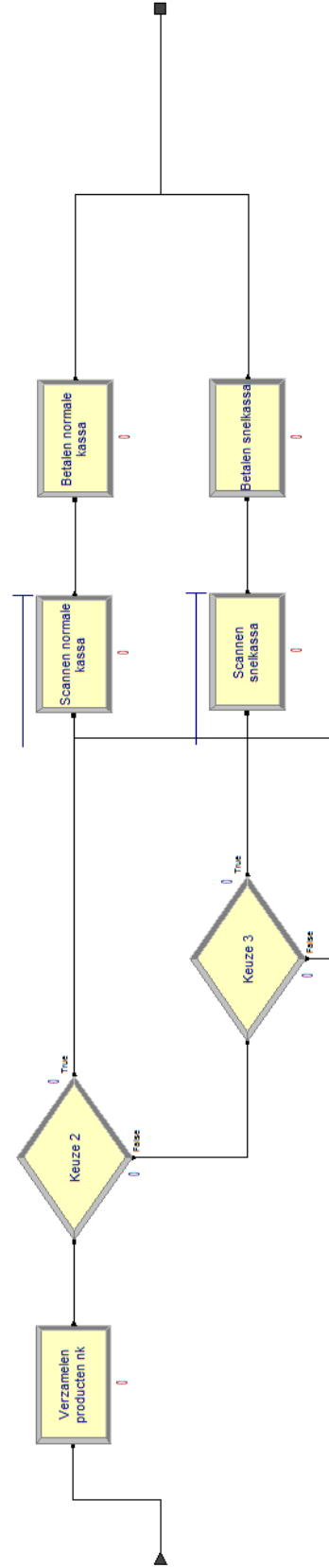
Hoofdmodel - Scenario 1



Submodel - Selfscanning



Submodel - Geen selfscanning



Submodellen

Process - Basic Process			
	Name	Type	Report Statistics
1	Selfscanning	Submodel	<input checked="" type="checkbox"/>
2	Geen selfscanning	Submodel	<input checked="" type="checkbox"/>

Entiteiten

Entity - Basic Process									
	Entity Type	Initial Picture	Holding Cost / Hour	Initial VA Cost	Initial NVA Cost	Initial Waiting Cost	Initial Tran Cost	Initial Other Cost	Report Statistics
1	Klant	Picture.Report	0.0	0.0	0.0	0.0	0.0	0.0	<input checked="" type="checkbox"/>
2	Jongere	Picture.Report	0.0	0.0	0.0	0.0	0.0	0.0	<input checked="" type="checkbox"/>
3	Gezin	Picture.Report	0.0	0.0	0.0	0.0	0.0	0.0	<input checked="" type="checkbox"/>
4	Gepensioneerde	Picture.Report	0.0	0.0	0.0	0.0	0.0	0.0	<input checked="" type="checkbox"/>

Resources

Resource - Basic Process									
	Name	Type	Capacity	Busy / Hour	Idle / Hour	Per Use	StateSet Name	Failures	Report Statistics
1	Kassa 1	Fixed Capacity	1	0.0	0.0	0.0		0 rows	<input checked="" type="checkbox"/>
2	Kassa 2	Fixed Capacity	1	0.0	0.0	0.0		0 rows	<input checked="" type="checkbox"/>
3	Kassa 4	Fixed Capacity	1	0.0	0.0	0.0		0 rows	<input checked="" type="checkbox"/>
4	Kassa 3	Fixed Capacity	1	0.0	0.0	0.0		0 rows	<input checked="" type="checkbox"/>

Sets van resources

Set - Basic Process			
	Name	Type	Members
1	Kassas	Resource	2 rows

Members	
	Resource Name
1	Kassa 2
2	Kassa 3

Wachtrijen

Queue - Basic Process				
	Name	Type	Shared	Report Statistics
1	Scannen normale kassa.Queue	First In First Out	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2	Scannen snelkassa.Queue	First In First Out	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3	Betalen Selfkassa.Queue	First In First Out	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Voorbeeld toewijzing eigenschappen - Jongeren

Assignments				
	Type	Attribute Name	Entity Type	New Value
1	Attribute	Jongere persoon	Entity 1	1
2	Attribute	Gezinslid	Entity 1	0
3	Attribute	Oudere persoon	Entity 1	0
4	Entity Type	Attribute 9	Jongere	1
5	Attribute	Gezinsgrootte	Entity 1	DISC(0.75, 1, 0.97, 2, 0.99, 3, 1, 4)
6	Attribute	Winkelfrequentie	Entity 1	DISC(0.20, 0.5, 0.40, 1, 0.60, 2, 0.80, 3, 1, 7)
7	Attribute	Nieuwe klant	Entity 1	DISC(0.15,1,1,0)
8	Attribute	Ervaring	Entity 1	DISC(0.2, 1, 0.3, 2, 1, 3)
9	Attribute	Leeftijd	Entity 1	UNIF(18,25)
10	Attribute	Kans op selfscanning	Entity 1	65-Leeftijd + 5 * ervaring
11	Attribute	Aantal producten	Entity 1	(Gezinsgrootte * 20)/Winkelfrequentie

Keuzes beslissingspunten hoofdmodel

Decide - Basic Process			
	Name	Type	Percent True
1	Type klant	N-way by Chance	50
2	Keuze 1	2-way by Chance	Kans op selfscanning

Conditions	
	Percent True
1	20
2	50

Aankomstritme entiteiten

Create - Basic Process						
Name	Entity Type	Type	Expression	Units	Entities per Arrival	First Creation
1	Detectie klant	Klant	Expression	Seconds	1	0.0
			POIS(60)		800	

Duurtijd activiteiten: Submodel – Selfscanning

Process - Basic Process						
Name	Action	Delay Type	Units	Value	Std Dev	Expression
1	Scannen klantenkaart	Delay	Seconds	1	2	1
2	Verzamelen producten	Delay	Seconds	1	2	$300 + 20 * \text{Aantal producten} + 10 * \text{Aantal producten} * \text{Nieuwe klant} + 10 * \text{Aantal producten} * \text{Oudere persoon} + (10 / \text{Ervaring}) * \text{Aantal producten} * \text{Keuze selfscanning}$
3	Betalen Selfkassa	Seize Delay Release	Seconds	1	2	$\text{Jongere persoon} * \text{TRIA}(25 , 30 , 60) + \text{Gezinslid} * \text{TRIA}(30 , 40 , 70) + \text{Oudere persoon} * \text{TRIA}(40 , 60 , 90)$
4	Toewijzing scanner	Delay	Seconds	20	15	1

Duurtijd activiteiten: Submodel – Geen selfscanning

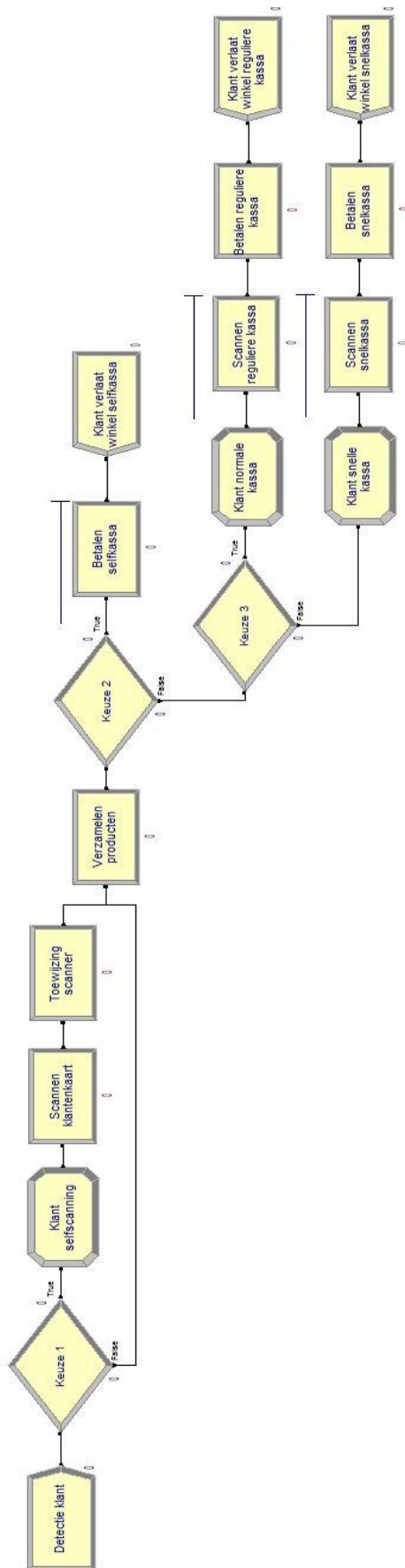
Process - Basic Process						
Name	Action	Delay Type	Units	Mi...	M...	Expression
1	Verzamelen producten	Delay	Seconds	5	1.5	$300 + 20 * \text{Aantal producten} + 10 * \text{Aantal producten} * \text{Nieuwe klant} + 10 * \text{Aantal producten} * \text{Oudere persoon} + (10 / \text{Ervaring}) * \text{Aantal producten} * \text{Keuze selfscanning}$
2	Scannen normale kassa	Seize Delay	Seconds	5	1.5	$2 * \text{Aantal producten} + 1 * \text{Aantal producten} * \text{Oudere persoon}$
3	Scannen snelkassa	Seize Delay	Seconds	5	1.5	1
4	Betalen normale kassa	Delay Release	Seconds	15	30	1
5	Betalen snelkassa	Delay Release	Seconds	15	30	1

Keuzes beslissingspunten: Submodel – Geen selfscanning

Decide - Basic Process				
Name	Type	If	Attribute Name	Value
1	Keuze 2	2-way by Condition	Aantal producten	>
2	Keuze 3	2-way by Condition	Attribute 1	>=
				$\text{NQ}(\text{Scannen normale kassa.Queue}) \geq \text{NQ}(\text{Scannen snelkassa.Queue})$

BIJLAGE 3: SIMULATIEMODEL ZONDER ENTITEITENTYPES

Hoofdmodel - Scenario 2



Entiteiten

Entity - Basic Process									
	Entity Type	Initial Picture	Holding Cost / Hour	Initial VA Cost	Initial NVA Cost	Initial Waiting Cost	Initial Tran Cost	Initial Other Cost	Report Statistics
1	Klant	Picture.Green Ball	0.0	0.0	0.0	0.0	0.0	0.0	<input checked="" type="checkbox"/>

Resources

Resource - Basic Process									
	Name	Type	Capacity	Busy / Hour	Idle / Hour	Per Use	StateSet Name	Failures	Report Statistics
1	Kassa 1	Fixed Capacity	1	0.0	0.0	0.0		0 rows	<input checked="" type="checkbox"/>
2	Kassa 2	Fixed Capacity	1	0.0	0.0	0.0		0 rows	<input checked="" type="checkbox"/>
3	Kassa 3	Fixed Capacity	1	0.0	0.0	0.0		0 rows	<input checked="" type="checkbox"/>
4	Kassa 4	Fixed Capacity	1	0.0	0.0	0.0		0 rows	<input checked="" type="checkbox"/>

Sets van resources

Set - Basic Process			
	Name	Type	Members
1	Reguliere kassas	Resource	2 rows

Members	
	Resource Name
1	Kassa 2
2	Kassa 3

Wachtrijen

Queue - Basic Process				
	Name	Type	Shared	Report Statistics
1	Betalen selfkassa.Queue	First In First Out	<input type="checkbox"/>	<input checked="" type="checkbox"/>
2	Scannen reguliere kassa.Queue	First In First Out	<input type="checkbox"/>	<input checked="" type="checkbox"/>
3	Scannen snelkassa.Queue	First In First Out	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Aankomstrijtme entiteiten

Create - Basic Process								
	Name	Entity Type	Type	Expression	Units	Entities per Arrival	Max Arrivals	First Creation
1	Detectie klant	Klant	Expression	WEIB(59.3 , 0.951)	Seconds	1	800	0.0

Duurtijden activiteiten

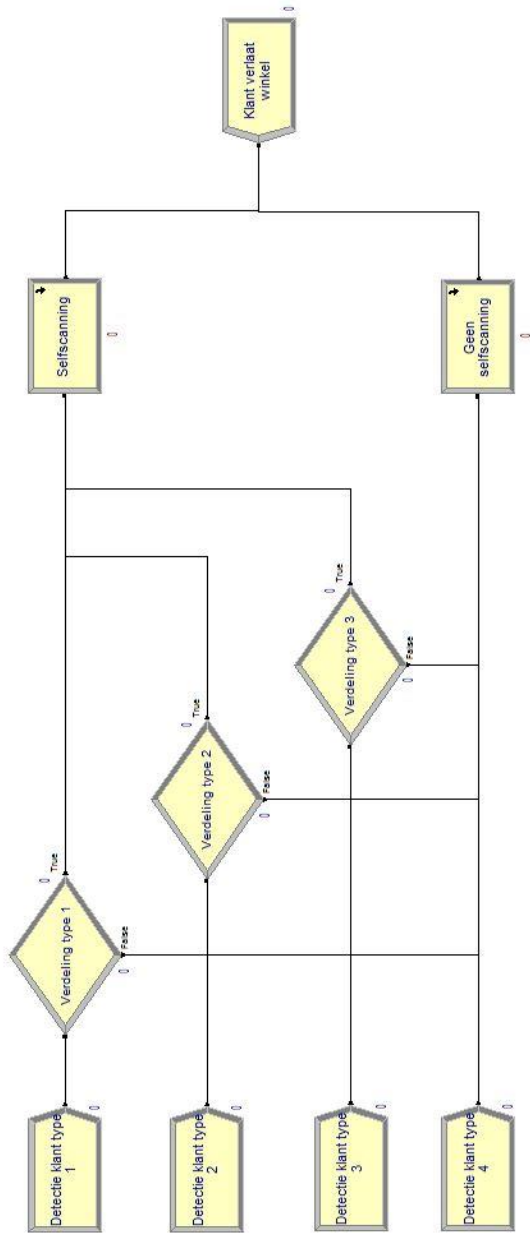
Process - Basic Process										
	Name	Type	Action	Priority	Resources	Delay Type	Units	Allocation	Value	Expression
1	Scannen klantenkaart	Standard	Delay	Medium(2)	0 rows	Constant	Seconds	Non-Value Added	1	1
2	Toewijzing scanner	Standard	Delay	Medium(2)	0 rows	Expression	Seconds	Non-Value Added	1	66 * BETA(1.03 , 2.3)
3	Verzamelen producten	Standard	Delay	Medium(2)	0 rows	Expression	Seconds	Value Added	1	360 + 5.94e+003 * BETA(0.96 , 5.03)
4	Betalen selfkassa	Standard	Seize Delay Release	Medium(2)	1 rows	Expression	Seconds	Value Added	1	26 + 42 * BETA(1.82 , 2.59)
5	Scannen reguliere kassa	Standard	Seize Delay	Medium(2)	1 rows	Expression	Seconds	Value Added	1	21 + LOGN(107 , 158)
6	Betalen reguliere kassa	Standard	Delay Release	Medium(2)	1 rows	Expression	Seconds	Value Added	1	15 + 15 * BETA(0.893 , 0.982)
7	Scannen snelkassa	Standard	Seize Delay	Medium(2)	1 rows	Expression	Seconds	Value Added	1	POIS(15.7)
8	Betalen snelkassa	Standard	Delay Release	Medium(2)	1 rows	Expression	Seconds	Value Added	1	UNIF(14.5, 30.5)

Keuzes beslissingspunten

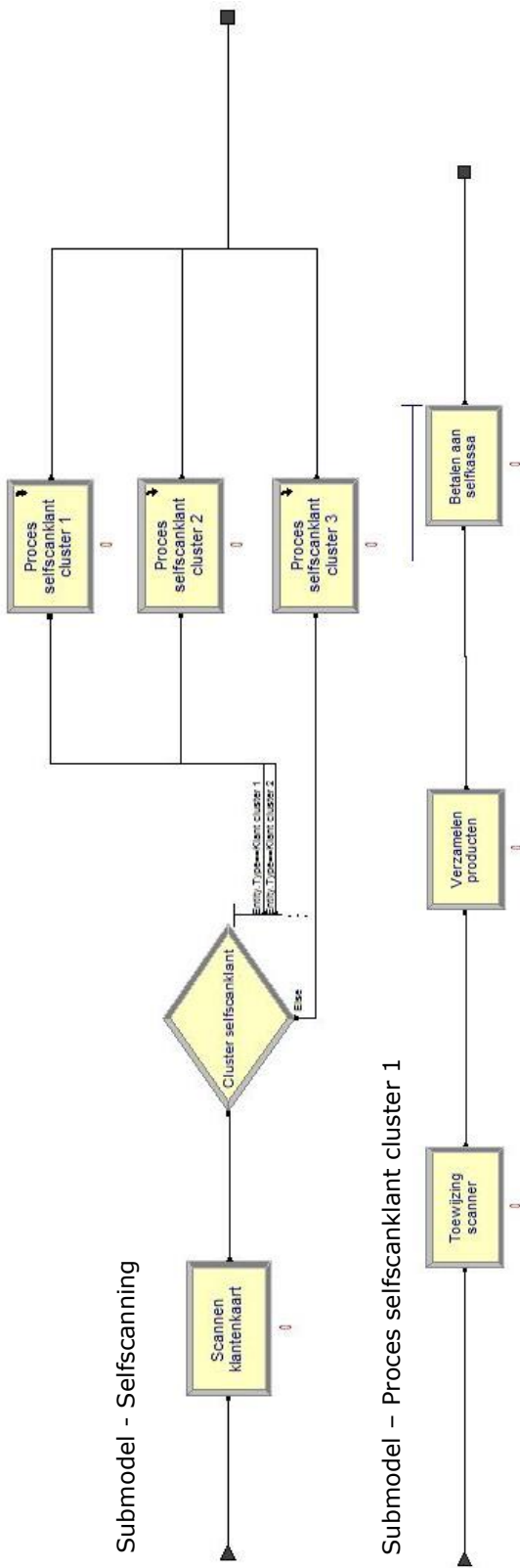
Decide - Basic Process								
	Name	Type	Percent True	If	Attribute Name	Is	Value	
1	Keuze 1	2-way by Chance	26.75	Entity Type	Attribute 1	>=	1	
2	Keuze 2	2-way by Condition	50	Attribute	Type	==	1	
3	Keuze 3	2-way by Chance	87.88	Entity Type	Attribute 1	>=	1	

BIJLAGE 4: SIMULATIEMODEL MET ENTITEITENTYPES

Hoofdmodel - Scenario 3

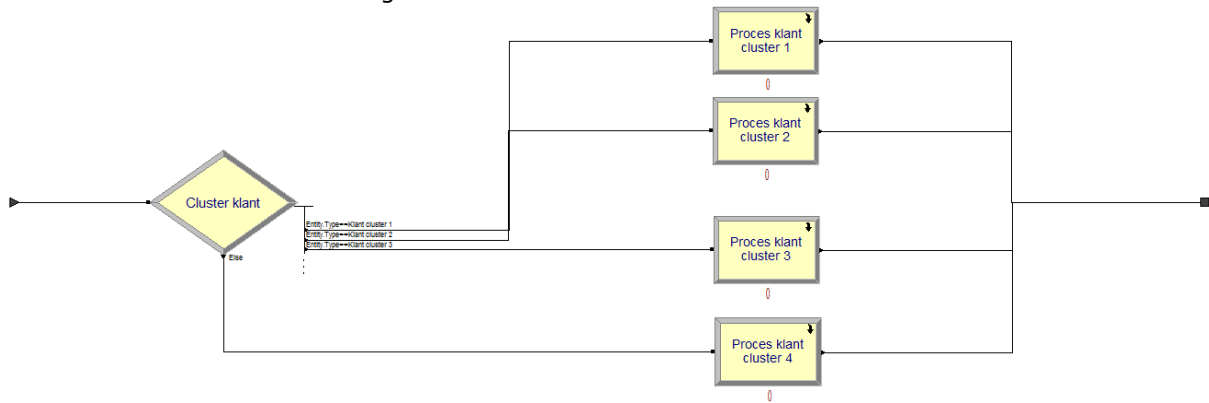


Submodel - Selfscanning

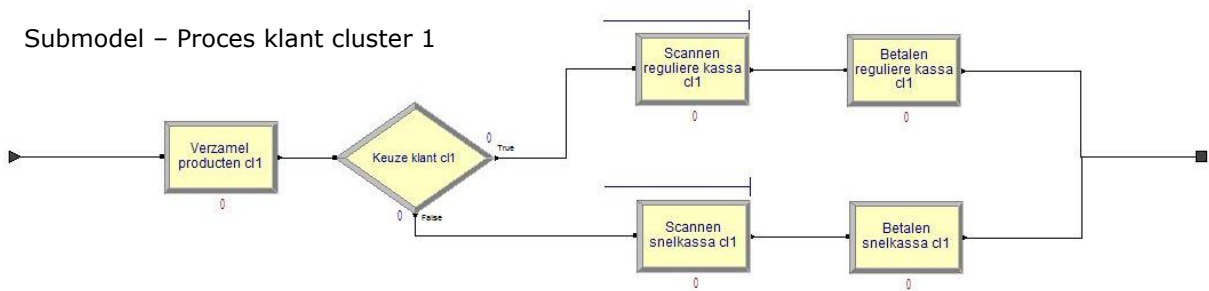


Submodel - Proces selfscanklant cluster 1

Submodel – Geen selfscanning



Submodel – Proces klant cluster 1



Entiteiten

Entity - Basic Process									
	Entity Type	Initial Picture	Holding Cost / Hour	Initial VA Cost	Initial NVA Cost	Initial Waiting Cost	Initial Tran Cost	Initial Other Cost	Report Statistics
1	Klant cluster 1	Picture.Report	0.0	0.0	0.0	0.0	0.0	0.0	✓
2	Klant cluster 2	Picture.Report	0.0	0.0	0.0	0.0	0.0	0.0	✓
3	Klant cluster 3	Picture.Report	0.0	0.0	0.0	0.0	0.0	0.0	✓
4	Klant cluster 4	Picture.Report	0.0	0.0	0.0	0.0	0.0	0.0	✓

Resources

Resource - Basic Process									
	Name	Type	Capacity	Busy / Hour	Idle / Hour	Per Use	StateSet Name	Failures	Report Statistics
1	Kassa 1	Fixed Capacity	1	0.0	0.0	0.0		0 rows	✓
2	Kassa 2	Fixed Capacity	1	0.0	0.0	0.0		0 rows	✓
3	Kassa 3	Fixed Capacity	1	0.0	0.0	0.0		0 rows	✓
4	Kassa 4	Fixed Capacity	1	0.0	0.0	0.0		0 rows	✓

Sets van resources

Set - Basic Process			
	Name	Type	Members
1	Set 1	Resource	2 rows

Members	
	Resource Name
1	Kassa 2
2	Kassa 3

Wachtrijen

Queue - Basic Process				
	Name	Type	Shared	Report Statistics
1	Scannen reguliere kassa c1.Queue	First In First Out	<input type="checkbox"/>	✓
2	Scannen snelkassa c1.Queue	First In First Out	<input type="checkbox"/>	✓
3	Scannen reguliere kassa c2.Queue	First In First Out	<input type="checkbox"/>	✓
4	Scannen reguliere kassa c3.Queue	First In First Out	<input type="checkbox"/>	✓
5	Scannen reguliere kassa c4.Queue	First In First Out	<input type="checkbox"/>	✓
6	Betalen aan selfkassa 2.Queue	First In First Out	<input type="checkbox"/>	✓
7	Betalen aan selfkassa 3.Queue	First In First Out	<input type="checkbox"/>	✓
8	Betalen aan selfkassa.Queue	First In First Out	<input type="checkbox"/>	✓

Aankomstrijtme entiteiten

Create - Basic Process								
	Name	Entity Type	Type	Expression	Units	Entities per Arrival	Max Arrivals	First Creation
1	Detectie klant type 1	Klant cluster 1	Expression	GAMM(282 , 0.92)	Seconds	1	185	0.0
2	Detectie klant type 2	Klant cluster 2	Expression	WEIB(330 , 1.1)	Seconds	1	150	0.0
3	Detectie klant type 3	Klant cluster 3	Expression	GAMM(200, 1.05)	Seconds	1	224	0.0
4	Detectie klant type 4	Klant cluster 4	Expression	GAMM(177, 1.13)	Seconds	1	241	0.0

Keuzes beslissingspunten hoofdmodel

Decide - Basic Process			
	Name	Type	Percent True
1	Verdeling type 1	2-way by Chance	61.08
2	Verdeling type 2	2-way by Chance	21.33
3	Verdeling type 3	2-way by Chance	30.80

Submodellen

Process - Basic Process			
	Name	Type	Report Statistics
1	Selfscanning	Submodel	<input checked="" type="checkbox"/>
2	Geen selfscanning	Submodel	<input checked="" type="checkbox"/>

Submodellen - Selfscanning

Process - Basic Process									
	Name	Type	Action	Delay Type	Units	Allocation	Value	Report Statistics	
1	Proces selfscanklant cluster 1	Submodel	Delay	Triangular	Hours	Value Added	1	<input checked="" type="checkbox"/>	
2	Proces selfscanklant cluster 2	Submodel	Delay	Triangular	Hours	Value Added	1	<input checked="" type="checkbox"/>	
3	Proces selfscanklant cluster 3	Submodel	Delay	Triangular	Hours	Value Added	1	<input checked="" type="checkbox"/>	
4	Scannen klantenkaart	Standard	Delay	Constant	Seconds	Value Added	1	<input checked="" type="checkbox"/>	

Duurtijd activiteiten: Submodel – Proces selfscanklant cluster 1

Process - Basic Process										
	Name	Type	Action	Resources	Delay Type	Units	Allocation	Expression	Report Statistics	
1	Toewijzing scanner	Standard	Delay	0 rows	Expression	Seconds	Value Added	$-0.5 + 56 * BETA(1.1, 1.75)$	<input checked="" type="checkbox"/>	
2	Verzamelen producten	Standard	Delay	0 rows	Expression	Seconds	Value Added	$369 + EXPO(306)$	<input checked="" type="checkbox"/>	
3	Betalen aan selfkassa	Standard	Seize Delay Release	1 rows	Expression	Seconds	Value Added	$26.5 + 35 * BETA(1.16, 1.74)$	<input checked="" type="checkbox"/>	

Keuzes beslissingspunten: Submodel – Proces selfscanklant cluster 1

Decide - Basic Process			
	Name	Type	
1	Cluster selfscanklant	N-way by Condition	2 rows

Conditions		
	If	Entity Type
1	Entity Type	Klant cluster 1
2	Entity Type	Klant cluster 2

Submodellen – Geen selfscanning

Process - Basic Process			
	Name	Type	Re...
1	Proces klant cluster 1	Submodel	<input checked="" type="checkbox"/>
2	Proces klant cluster 2	Submodel	<input checked="" type="checkbox"/>
3	Proces klant cluster 3	Submodel	<input checked="" type="checkbox"/>
4	Proces klant cluster 4	Submodel	<input checked="" type="checkbox"/>

Duurtijd activiteiten: Submodel – Proces klant cluster 1

Process - Basic Process										
	Name	Type	Action	Resources	Delay Type	Units	Allocation	Expression	Re...	
1	Verzamel producten cl1	Standard	Delay	0 rows	Expression	Seconds	Value Added	$360 + 340 * BETA(1.05, 2.18)$	<input checked="" type="checkbox"/>	
2	Scannen reguliere kassa cl1	Standard	Seize Delay	1 rows	Expression	Seconds	Value Added	26	<input checked="" type="checkbox"/>	
3	Scannen snelkassa cl1	Standard	Seize Delay	1 rows	Expression	Seconds	Value Added	POIS(15.7)	<input checked="" type="checkbox"/>	
4	Betalen reguliere kassa cl1	Standard	Delay Release	1 rows	Expression	Seconds	Value Added	23.54	<input checked="" type="checkbox"/>	
5	Betalen snelkassa cl1	Standard	Delay Release	1 rows	Expression	Seconds	Value Added	UNIF(14.5, 30.5)	<input checked="" type="checkbox"/>	

Keuzes beslissingspunten: Submodel – Proces klant cluster 1

Decide - Basic Process			
	Name	Type	
1	Cluster klant	N-way by Condition	3 rows

Conditions		
	If	Entity Type
1	Entity Type	Klant cluster 1
2	Entity Type	Klant cluster 2
3	Entity Type	Klant cluster 3

BIJLAGE 5: ONDERDEEL OUTPUTRAPPORTEN ARENA

5.1 TOTALE DOORLOOPTIJD

5.1.1 Scenario 1: De realiteit

Total Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Gepensioneerde	1524.48	8,11	1424.50	1638.44	616.01	4891.60
Gezin	1734.70	11,87	1607.66	1881.45	373.18	8403.64
Jongere	875.10	9,36	763.14	996.42	373.16	6776.60

5.1.2 Scenario 2: Model zonder entiteitentypes

Total Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Klant	1590.73	15,70	1458.53	1820.17	388.12	7214.78

5.1.3 Scenario 3: Model met entiteitentypes

Total Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Klant cluster 1	650.01	3,86	600.18	708.19	382.91	3895.95
Klant cluster 2	1132.61	6,82	1052.11	1202.61	560.76	2410.98
Klant cluster 3	2369.74	14,74	2203.70	2528.86	47.7237	5983.47
Klant cluster 4	1469.72	7,86	1357.70	1569.17	757.81	6191.34

5.2 TOTAAL GEMIDDELD AANTAL ENTITEITEN IN HET SYSTEEM

5.2.1 Scenario 1: De realiteit

WIP	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Gepensioneerde	7.0307	0,08	6.1572	8.2339	0.00	19.0000
Gezin	13.5019	0,15	11.8317	15.3550	0.00	28.0000
Jongere	2.7326	0,05	1.9834	3.3663	0.00	11.0000
Klant	0.00	0,00	0.00	0.00	0.00	1.0000

5.2.2 Scenario 2: Model zonder entiteitentypes

WIP	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Klant	24.7539	0,33	20.9304	28.9140	0.00	58.0000

5.2.3 Scenario 3: Model met entiteitentypes

WIP	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Klant cluster 1	2.2884	0,02	1.9534	2.7171	0.00	13.0000
Klant cluster 2	3.2344	0,04	2.7069	3.8310	0.00	12.0000
Klant cluster 3	10.1033	0,12	8.8016	11.7750	0.00	27.0000
Klant cluster 4	6.7416	0,07	5.7445	7.8652	0.00	21.0000

5.3 GEMIDDELDE WACHTTIJDEN PER TYPE KASSA

5.3.1 Scenario 1: De realiteit

Waiting Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Betalen Selfkassa.Queue	4.4916	0,23	2.2995	9.0296	0.00	157.92
Scannen normale kassa.Queue	77.2636	3,40	44.4024	115.87	0.00	893.13
Scannen snelkassa.Queue	0.3542	0,07	0.00	1.5825	0.00	47.9554

5.3.2 Scenario 2: Model zonder entiteitentypes

Waiting Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Betalen selfkassa.Queue	5.0905	0,25	2.5969	8.0056	0.00	165.66
Scannen reguliere kassa.Queue	245.30	21,31	82.8314	625.45	0.00	2588.62
Scannen snelkassa.Queue	1.1505	0,15	0.00	3.3053	0.00	71.9523

5.3.3 Scenario 3: Model met entiteitentypes

Waiting Time	Average	Half Width	Minimum Average	Maximum Average	Minimum Value	Maximum Value
Betalen aan selfkassa 2.Queue	4.8891	0,46	0.5162	11.7961	0.00	127.49
Betalen aan selfkassa 3.Queue	4.9144	0,31	1.6220	9.5234	0.00	108.04
Betalen aan selfkassa.Queue	5.1642	0,27	1.9740	8.3951	0.00	114.57
Scannen reguliere kassa cl1.Queue	53.7014	22,35	0.00	813.44	0.00	813.44
Scannen reguliere kassa cl2.Queue	80.8540	4,75	39.8908	154.44	0.00	1002.40
Scannen reguliere kassa cl3.Queue	81.8844	5,64	0.00	181.04	0.00	1145.72
Scannen reguliere kassa cl4.Queue	81.8977	4,55	46.4674	174.31	0.00	966.30
Scannen snelkassa cl1.Queue	1.2545	0,15	0.00	3.4204	0.00	72.5356

5.4 BEZETTINGSGRAAD KASSA'S

5.4.1 Scenario 1: De realiteit

Scheduled Utilization	Average	Half Width	Minimum Average	Maximum Average
Kassa 1	0.1677	0,00	0.1414	0.1936
Kassa 2	0.6860	0,01	0.6055	0.7443
Kassa 3	0.6836	0,01	0.6047	0.7525
Kassa 4	0.03028014	0,00	0.02227257	0.04236419

5.4.2 Scenario 2: Model zonder entiteitentypes

Scheduled Utilization	Average	Half Width	Minimum Average	Maximum Average
Kassa 1	0.1788	0,00	0.1490	0.2036
Kassa 2	0.7562	0,01	0.6588	0.8808
Kassa 3	0.7554	0,01	0.6417	0.8826
Kassa 4	0.05298431	0,00	0.03787083	0.07758551

5.4.3 Scenario 3: Model met entiteitentypes

Scheduled Utilization	Average	Half Width	Minimum Average	Maximum Average
Kassa 1	0.1755	0,00	0.1493	0.2117
Kassa 2	0.6452	0,01	0.5348	0.7153
Kassa 3	0.6440	0,01	0.5312	0.7173
Kassa 4	0.05151887	0,00	0.04059437	0.06327722

5.5 AANTAL ENTITEITEN PER TYPE

5.5.1 Scenario 1: De realiteit

Number In	Average	Half Width	Minimum Average	Maximum Average
Gepensioneerde	237.77	2,29	218.00	272.00
Gezin	401.24	2,76	367.00	430.00
Jongere	160.99	2,38	135.00	188.00
Klant	800.00	0,00	800.00	800.00

5.5.2 Scenario 2: Model zonder entiteitentypes

Number In	Average	Half Width	Minimum Average	Maximum Average
Klant	800.00	0,00	800.00	800.00

5.5.3 Scenario 3: Model met entiteitentypes

Number In	Average	Half Width	Minimum Average	Maximum Average
Klant cluster 1	185.00	0,00	185.00	185.00
Klant cluster 2	150.00	0,00	150.00	150.00
Klant cluster 3	224.00	0,00	224.00	224.00
Klant cluster 4	241.00	0,00	241.00	241.00

Vergelijkende tabel: Scenario 1, 2 en 3						
	Scenario 1		Scenario 2		Scenario 3	
	De realiteit		Model zonder entiteitentypes		Model met entiteitentypes	
	Gemiddelde	BI 95%	Gemiddelde	BI 95%	Gemiddelde	BI 95%
Totale doorlooptijd	1499.24	[1492.54, 1505.93]	1590.73	[1575.03, 1606.43]	1468.96	[1463.95, 1473.97]
Totaal aantal entiteiten	23.27	[23.09, 23.44]	24.75	[24.42, 25.08]	22.37	[22.22, 22.51]
Wachttijd						
Reguliere kassa	77.26	[73.86, 80.66]	245.30	[223.99, 266.61]	81.61	[78.68, 84.55]
Snelkassa	0.35	[0.28, 0.42]	1.15	[1, 1.3]	1.25	[1.10, 1.40]
Selfkassa	4.49	[4.26, 4.72]	5.09	[4.84, 5.34]	5.04	[4.86, 5.23]
Bezettingsgraad						
Kassa 1	16.77	[16.77, 16.77]	17.88	[17.88, 17.88]	17.55	[17.55, 17.55]
Kassa 2	68.60	[67.60, 69.60]	75.62	[74.62, 76.62]	64.52	[63.52, 65.52]
Kassa 3	68.36	[67.36, 69.36]	75.54	[74.54, 76.54]	64.40	[63.40, 65.40]
Kassa 4	3.03	[3.03, 3.03]	5.30	[5.30, 5.30]	5.15	[5.15, 5.15]

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:

Het ontwikkelen van realistische simulatiemodellen door het onderscheiden van entiteitentypes met behulp van process mining

Richting: **master in de toegepaste economische wetenschappen: handelsingenieur in de beleidsinformatica**

Jaar: **2015**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Verstrepen, Brighid

Datum: **1/06/2015**