
Thesis: Bachelor

ERASMUS UNIVERSITEIT ROTTERDAM
Erasmus School of Economics

Bachelor scriptie

Begeleider: Bouwer, H.J.
Tweede beoordelaar: Aernoudts, R.H.R.M.

Naam student: Hamza Ou M'hamed
Student nummer: 509178

Datum definitieve versie:

E-mailadres: 509178ho@eur.nl
Studie: Economie & Bedrijfseconomie



‘Wat voor invloed heeft Big Data op het beleid voor externe fraudedetectie bij gemeentelijke organisaties’

Abstract

Deze paper behandelt het onderwerp Big Data en de verschillende voorkomende vormen van fraude bij gemeentelijke organisaties. De hoofdvraag die bij dit onderwerp beantwoord wordt, is: wat voor invloed heeft Big Data op het beleid voor externe fraudedetectie bij gemeentelijke organisaties. Om deze hoofdvraag te beantwoorden wordt eerst ingegaan op wat Big Data inhoudt gebaseerd op verschillende perspectieven en welke vormen van fraude voorkomen bij de gemeentelijke organisaties. Vervolgens zal toegelicht worden welke maatregelen gemeentes kunnen meenemen in hun beleid om deze vormen van fraude met behulp van Big Data tegen te werken. Daarnaast zal een empirisch onderzoek besproken worden waarin de websites, jaarrekeningen en accountantsverslagen van de drie gemeentes Bergen op Zoom, Roosendaal en Breda bekeken zijn. De paper wordt afgesloten met een conclusie waarin de deelvragen kort herhaald worden en er een duidelijk antwoord wordt gegeven op de onderzoeksvraag.

Inhoud

1. Introductie	3
2. Wat houdt Big Data in?	5
Huidige situatie.....	5
Technieken	8
Problemen	10
Uitdagingen	11
Ethische implicaties	13
3. Welke vormen van fraude komen voor bij gemeentelijke organisaties?.....	14
Definiëring van fraude.....	14
Sociale fraude	15
Fraude door bedrijven.....	17
4. Hoe wordt Big Data ingezet om deze vormen van fraude te detecteren?	19
Niet-eendimensionale manier.....	19
Uitwisselen van persoonsgegevens.....	21
Wisselwerking op orde brengen adressen en verschillende vormen van fraudedetectie.....	21
Gezamenlijk gebruik van databestanden voor fraudedetectie.....	22
Samenvatting.....	23
5. Opzet empirisch onderzoek.....	24
Hypotheses	24
Data	25
Methode.....	26
6. Resultaten.....	26
Eerste hypothese.....	26
Tweede hypothese	28
Derde hypothese	29
7. Conclusie	32
Bibliografie	33

1. Introductie

Big Data heeft een zeker magie in zich. Het bestaat uit een grote hoeveelheid gegevens waarover we kunnen beschikken en die we kunnen inzetten om het leven in verschillende aspecten te verbeteren (van der Voort & Cromptvoets, 2016). Big Data kan veel waarde opleveren voor onze economie, omdat met behulp van Big Data bedrijfsprocessen efficiënter ingericht kunnen worden. Ook kunnen door de nieuwe en betere informatie afkomstig uit Big Data beleidsevaluaties uitgevoerd worden. Verder bieden Big Datatools kansen sneller complexe patronen te vinden waardoor steeds meer data relevant wordt voor het vormen van beleid door bedrijven (Bongers, Veldkamp, & van de Vorst, 2015). Bovendien kan Big Data van belang zijn bij het uitvoeren van impact analyses, doelbereiking of effectmeting van beleidsinterventies. Hierdoor krijgen bedrijven toegang tot gegevens die meer real-time beschikbaar zijn. Door middel van de bovengenoemde voorbeelden kunnen beleidsdoelen gevonden worden die, als ze gerealiseerd zouden worden, maatschappelijk het meest effectief zullen zijn (Leeuw, 2019).

Gemeentes hebben veel verschillende taken, maar kunnen niet alles zelf doen. De gemeente heeft taken op verschillende gebieden, zoals: burgerzaken, openbare orde, zorg, integratie, onderwijs, verkeer, milieu en cultuur. Om alles te kunnen regelen hebben gemeentes de taken uitbesteed aan een gemeentelijke organisaties (ProDemos), die steeds meer gebruikmaken van Big Data, doordat ze zien wat voor mogelijkheden Big Data biedt. Het werken met Big Data kan leiden tot meer informatie en inzicht. Een van de doelen van gemeentes is momenteel het misbruik van voorzieningen krachtiger aan te pakken (ProDemos). Dat doel kan bereikt worden door middel van het gebruik van Big Data waardoor bijvoorbeeld de kans op misbruik wordt verkleind. Doordat organisaties door middel van Big Data meer fraude kunnen aanpakken, kan hun beleid aangepast worden naar een betere en bevorderde versie. De belangrijkste reden hiervoor is dat gemeentelijke organisaties meer data tot hun beschikking zullen hebben. Bedrijven kunnen door hun betere beleid meer waarde leveren aan de burgers, bijvoorbeeld door het analyseren van Big Data meer informatie te krijgen over de voorkeuren van de burger. Zo zal dit onderzoek een bijdrage leveren aan de maatschappij (Olsthoorn, 2016).

Dit onderzoek kijkt naar Big Data en wat voor invloed het heeft op het beleid voor fraudedetectie bij gemeentelijke organisaties. Dit leidt tot de volgende onderzoeksvraag:

‘Wat voor invloed heeft Big Data op het beleid voor fraudedetectie bij gemeentelijke organisaties?’

Om een helder antwoord te kunnen formuleren op de onderzoeksvraag zullen als eerst de volgende deelvragen worden beantwoord:

- Wat houdt Big Data in?
- Welke vormen van fraude komen voor bij gemeentelijke organisaties?
- Hoe wordt Big Data ingezet om deze vormen van fraude te detecteren?

De eerste deelvraag zal vanuit verschillende perspectieven bekeken worden. Vanuit het eerste perspectief zal een algemene definitie worden gegeven van Big Data en gekeken worden naar de huidige situatie met betrekking tot Big Data. Vervolgens zal gekeken worden naar de technieken die gebruikt worden en de problemen die gepaard gaan met Big Data. Bij het volgende perspectief zullen de uitdagingen besproken worden. Ook komt de ethische kant aan de orde waarop Big Data invloed heeft. Wanneer dit is uitgelegd, kan een overstap gemaakt worden van Big Data naar fraude. De verschillende vormen van externe fraude bij gemeentelijke bedrijven en het gebruik van Big Data om deze fraude te beperken zullen besproken worden in de tweede en derde deelvraag. Hierbij zal alleen gekeken worden naar de fraudes die extern plaatsvinden en niet naar de interne fraude door bijvoorbeeld ambtenaren. Vervolgens zal een onderzoek verricht worden door te kijken naar de websites, jaarrekeningen en accountantsverslagen van specifieke gemeentes. Het onderzoek zal op die manier een terugkoppeling maken naar de informatie in de eerder behandelde deelvragen. Het antwoord op de drie deelvragen wordt op die manier niet alleen geformuleerd door middel van onderzochte literatuur, echter wordt ook in kaart gebracht hoe dit is toegepast in de praktijk.

Met dit onderzoek kan er vastgesteld worden wat voor invloed Big Data heeft op het vormen van beleid voor fraudedetectie bij gemeentelijke bedrijven. Er is geen gebrek aan artikelen over Big Data met betrekking op de nadelen en voordelen hiervan. Bovendien is er veel informatie te vinden over wat Big Data als consequenties heeft voor de maatschappij. Alleen kan er geen eenduidig antwoord achterhaald worden op de vraag, in hoeverre gemeentelijke bedrijven Big Data toepassen op het gebied van fraude. Het onderzoek zal een relevante bijdrage leveren, omdat er gekeken zal worden naar de praktijk. Daarom zal dit onderzoek van belang zijn voor de literatuur en op die manier van wetenschappelijke relevantie zijn.

Om tot een antwoord te komen op de onderzoeksvraag wordt eerst het theoretisch kader geschetst. Hierin worden relevante concepten van het onderzoek toegelicht met belangrijke informatie over Big Data, verkregen uit onder andere de wetenschappelijke artikelen. Aansluitend wordt door middel van een literatuuronderzoek weergegeven wat voor invloed Big Data heeft op het vormen van één beleid voor fraudedetectie door gemeentelijke bedrijven. Tot slot volgt de conclusie waarin een antwoord wordt geformuleerd op de onderzoeksvraag door middel van de eerder beantwoorde deelvragen.

2. Wat houdt Big Data in?

Huidige situatie

Tot op de dag van vandaag is geen consensus over wat Big Data nou eigenlijk inhoudt. Eerst werd het bekeken vanuit het perspectief van bestandsgrootte. Big Data was data met een grootte dat niet verwerkt kon worden met het geheugen van pc's uit het verleden. Daarnaast wordt het vaak gedefinieerd als grote datasets die zijn gemaakt vanuit verschillende bronnen die zijn verkregen op verschillende momenten. Later werd er nauwer bekeken naar de karakteristieken van Big Data en werd de definitie verder uitgebreid. Daarbij werd ook gekeken vanuit het perspectief van de snelheid om data via traditionele en nieuwe bronnen te verzamelen. Naarmate de hoeveelheid data toeneemt, neemt het aantal onjuistheden toe en wordt de data minder accuraat. Voorstanders van het gebruik van Big Data vonden dat er niet gehecht moet worden aan accuratesse, omdat al hoewel de onjuistheden toe neemt op microniveau, het aantal nieuwe inzichten op macroniveau stijgt (Peters & Cheyney, 2018).

De focus op Big Data is in de loop van de tijd toegenomen en de vraag ernaar is exponentieel gegroeid (zie figuur 1). Steeds meer bedrijven, overheden en onderzoeksorganisaties gebruiken data-analyse om informatie te ontdekken en geld te verdienen met data om op die manier hun praktijken, diensten en producten verder te ontwikkelen (Falsafi & Grot, 2014).

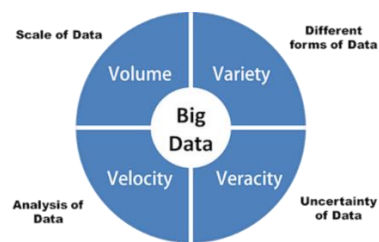


Figuur 1 Interesse in Big Data

De hoeveelheid informatie waarop computers tegenwoordig berekeningen op loslaat groeit steeds sneller. Elke dag creëren we 2,5 triljoen bytes aan nieuwe data die van verschillende bronnen komen, zoals: social media, Gps-signalen, medische zorg en gebruikersgegevens van telecommunicatiekanalen. Volgens de wet van Moore werkt de hardware steeds sneller, omdat de processorcapaciteit elke twee jaar verdubbelt. De enorme hoeveelheid gegevens wil men gebruiken ter ondersteuning van de kwaliteit van diverse besluitvorming. Daarentegen is de hoeveelheid

gegevens te groot geworden waardoor verschillende sets gegevens niet gekoppeld en geanalyseerd kunnen worden om bijvoorbeeld trends op het gebied van fraude te ontdekken (Stel, 2012).

Op die manier hebben de meeste bedrijven nog steeds de moeite de grote hoeveelheid gegevens over hun klanten, producten en partners effectief te gebruiken om tot een commerciële en strategische besluitvorming te komen. De gegevens zitten verstopt en verspreid over verschillende databases. Gartner heeft in 2012 gesteld dat Big Data *high-volume*, *high-velocity* en *high-variety* is (zie figuur 2). IBM voegde een vierde kenmerk toe, namelijk *veracity*. (Stel, 2012).



Figuur 2 4 v's van Big Data

De eerste eigenschap, *high-volume*, is het belangrijkste kenmerk van Big Data. De meeste mensen definiëren Big Data in termen van terabytes of petabytes. Daarnaast kan Big Data gedefinieerd worden in termen van aantal records, transacties, tabellen of bestanden (Russom, 2011). Dagelijks groeit het volume met hoge snelheid en wordt 2,3 triljoen gigabytes aan data gecreëerd. Door de snelle groei is er behoefte aan nieuwe databasemanagementsystemen om de Big Data op te kunnen vangen (Majster, 2013).

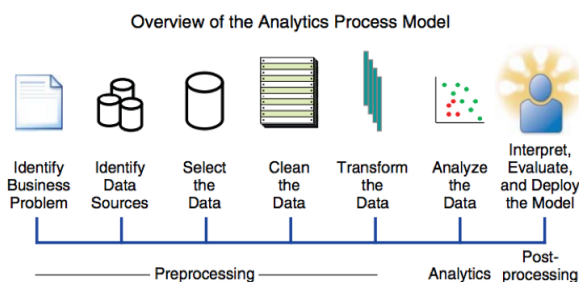
Ten tweede kan Big Data ook worden beschreven door het kenmerk *high-velocity*. De data worden met een grote snelheid gegenereerd en verwerkt. Vroeger duurde het even voor grote hoeveelheden gegevens verwerkt konden worden om op die manier specifieke data en de juiste informatie te verkrijgen. Daarentegen zijn data tegenwoordig *real-time* beschikbaar door de aanwezigheid van Big Data. Hierdoor kwamen nieuwe methodes aan het licht en kon meer informatie tegelijkertijd verwerkt worden (Majster, 2013). Het in *real-time* beschikbaar zijn van data is niet nieuw, omdat bedrijven al jarenlang gebruik maken van click-stream gegevens en web-gegevens waardoor datavolumes al snel groot worden. Daardoor wordt het steeds uitdagender om de hoeveelheid gegevens in *real-time* te begrijpen en eventueel actie te ondernemen (Russom, 2011).

Vervolgens heeft Big Data ook het kenmerk *Variety*. Het aanzienlijke volume en de grote snelheid van data gaat gepaard met de verscheidenheid aan vormen. Het volume en de variëteit zullen de komende jaren alleen maar toenemen (Majster, 2013). Big Data wordt groot gemaakt door de grote verscheidenheid aan bronnen waaruit de gegevens verzameld worden. Gebruikers van Big Data die

gegevens eerst alleen oppotten, gaan het nu meer analyseren op een complexer en geavanceerder niveau. Hoe groter het volume data is, hoe groter de variëteit in data zal zijn (Russom, 2011).

Tenslotte wordt Big Data verder gedefinieerd door het kenmerk *Veracity*. De data die gedeeld worden is niet altijd juist en kunnen ook oude informatie bevatten die niet meer relevant zal zijn. Datawetenschappers en IT-ers werken lange dagen om de juiste data te organiseren en te ontsluiten van de grote hoeveelheid gegevens. Als dat op een goede manier gebeurt, kan Big Data grote waarde opleveren voor bedrijven. Bedrijven kunnen bijvoorbeeld zakelijke trends voorspelen en fraudes tegengaan (Majster, 2013).

Om de juiste informatie te organiseren en te ontsluiten zijn een aantal stappen vereist. Het standaardmodel bevat 3 fasen met daarin verschillende activiteiten. Een bedrijf bepaalt zelf welke stappen noodzakelijk is om de juiste informatie te verkrijgen. Een voorbeeld van een stappenmodel is weergegeven in figuur 3. Om ervoor te zorgen dat de gegevens die het systeem binnenkomen schoon en bruikbaar zijn werd de creatiefase uitgevoerd. Het eerste proces, behoeftebepaling, is de belangrijkste stap. Dit levert een informatie-, rapportage- en gegevensbehoeften analyse van een bedrijf. Vervolgens volgt het verwerven van gegevens van interne afdelingen of externe tegenpartijen. Ten derde werden alle verworven gegevens geclassificeerd om het een duidelijke betekenis en context te geven. Daarna volgt het proces conversie waarbij de gegevens het juiste formaat kregen om te passen in de bestaande opslaginfrastructuur. De laatste stap is het opnemen van de gegevens in het informatiesysteem van het bedrijf (Coyne, Coyne, & Walker, 2018).



Figuur 3 Stappen Big Data systeem

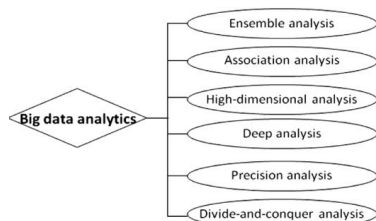
Toegang tot de gegevens moet voor lange termijn gegarandeerd zijn en de informatie die niet waardevol is, zal verwijderd moeten worden. Hiervoor is de tweede fase, onderhoud, in het model verantwoordelijk. De activiteiten die hierbij horen zijn opslag, indexering, vernieuwen, interpretatie en dispositie. Bij opslag zal er rekening gehouden worden met het type gegevens en de verwachte levensduur. Doordat de gegevens geïndexeerd worden door een locatie toe te wijzen, wordt het zoeken van de gegevens versneld. Met vernieuwen worden de data bijgewerkt om de inhoud van de gegevens te waarborgen. Bij het interpreteren worden de gegevens omgezet in een door mensen en

computer leesbare formaat. Tenslotte komen de gegevens in een lange-termijnopslag of worden ze vernietigd (Coyne, Coyne, & Walker, 2018).

De laatste fase, gebruik, zorgt ervoor dat besluitvormers toegang krijgen tot de informatie om het te analyseren. De eerste stap in deze fase is het zoeken van de informatie die wordt ondersteund door activiteiten in de andere fasen. Het snel kunnen vinden van informatie is eenvoudig als de informatie correct is geclassificeerd en opgeslagen. De tweede stap, analyseren, wordt gebruikt om nuttige informatie te ontfanen aan de gegevensopslag. Tenslotte vindt rapportage plaats waarbij informatie medegedeeld wordt met interne en externe besluitvormers (Coyne, Coyne, & Walker, 2018).

Technieken

Bij de vereiste stappen om de juiste informatie te organiseren en te ontsluiten zijn nieuwe technologieën en architecturen vereist voor het vastleggen en analyseren van Big Data om nuttige informatie te verkrijgen uit de grote hoeveelheid data. Met de bestaande traditionele technieken wordt het moeilijk de grote hoeveelheid gegevens effectief te analyseren. Door de eigenschappen van Big Data, volume, snelheid, variëteit, variabiliteit ontstaan veel uitdagingen bij het analyseren (Katal, Wazid, & Goudar, 2013). Om een beeld te krijgen van de mogelijke technieken om Big Data te analyseren worden zes technieken voorgesteld (zie figuur 4).



Figuur 4 zes technieken om Big Data te analyseren

Ensemble data-analyse, ook wel multi-dataset of multi-algoritme analyse genoemd, is gemaakt voor grote hoeveelheid gegevens. De hele set gegevens is vaak onderverdeeld in drie delen, namelijk trainingsgegevens, testgegevens en validatiegegevens. Om informatie uit het geheel van de drie delen te extraheren, zijn veel algoritmen geïntegreerd in het systeem. Doordat er meerdere leeralgoritmen worden geïntegreerd, kunnen er betere schattingsprestaties bereikt worden (Shu, 2016).

Big data wordt meestal verzameld zonder speciale steekproeven strategieën en verschillen de dataproducenten van de datagebruikers. Hierdoor wordt de oorzaak-gevolg relatie niet duidelijk voor de datagebruikers. Associatie-analyse is cruciaal voor multi-sourcing, multi-type en multi-domein data-analyse. Een associatie bestaat uit twee onderdelen, een antecedent (als) en een consequent (dan).

De twee onderdelen worden gemaakt door de gegevens te analyseren en frequente als/dan-patronen te identificeren. Hierbij wordt gebruik gemaakt van de twee criteria vertrouwen en ondersteuning. Bij ondersteuning wordt gekeken naar hoe vaak de items verschijnen en bij vertrouwen naar hoe vaak de als/dan-patroon juist blijkt. Bij het volgen van de patronen zijn meerdere sensoren ingeschakeld (Shu, 2016).

In de wiskunde wordt een dimensie van een object gedefinieerd als het minimumaantal coördinaten dat nodig is om een punt te specificeren. De afmeting van een object is onafhankelijk van de ruimte waarin het object is ingebed. Daarvoor bestaat de derde techniek, high-dimensional analysis, bij steekproeven van meerdere willekeurige variabelen. De vloek van dimensionaliteit komt in het groot naar voren bij data-analyse in hoog-dimensionale ruimtes, omdat het volume van de ruimte te snel toe neemt waardoor de beschikbare gegevens schaars worden. Deze schaarste zal leiden tot de statistische significantiefout en de ongelijkheid tussen de objecten. Om de variabiliteit van originele variabelen zo hoog mogelijk te houden wordt gebruik gemaakt van afmetingsreductie. Dat is het proces waarbij het aantal willekeurige variabelen verminderd worden en een subset wordt gevonden van de oorspronkelijke variabelen (Shu, 2016).

Daarnaast is Big Data groot genoeg voor complexe kunstmatige neurale netwerktrainingen. Diep analysis zorgt voor het kunnen verkennen van complexe structuur eigenschappen van Big Data. Door middel van diepgaande analyse kunnen onwaarneembare variabelen, verborgen parameters, verborgen hiërarchieën, lokale correlaties en complexe verdeling van willekeurige variabelen gevonden worden. Het algemene kunstmatige neurale netwerk heeft veel niet-lineaire verwerkingseenheden van activeringseenheden (neuronen) die gezien worden als de variabelen. Diepgaande analyse wordt gezien als een transformatiefunctie om input om te zetten in output. Het kunstmatige neurale netwerk is krachtig in complexe datamodellering door de meerdere lagen, meerdere variabelen en niet-lineaire activeringseenheden (Shu, 2016).

Om de juistheid van data te evalueren vanuit het perspectief van datagebruik en datakwaliteit wordt er gebruik gemaakt van precisie-analyse. Daarbij wordt de Big Data verzameld uit het systeem van mens-machine (computer) en omgeving (dingen). Ook gebruikersvoorkeuren zijn een vorm van de speciale soorten gegevens. Die gebruikersgegevens zijn nodig om voor de analyse een integrale model en optimalisatiecriteria te vinden. Om de modellen van gebruikersvoorkeuren en gebruikersvereisten te bouwen, onderscheiden gebruikers zich van elkaar en zijn ze gekoppeld aan een sociaal netwerk. Door middel van gepersonaliseerde marketing, één-op-één marketing, vragen marketeers rechtstreeks persoonlijke voorkeuren om op die manier een uniek productaanbod te maken voor elke klant (Shu, 2016).

Divide-and-conquer analysis is een algemeen computationele strategie om de efficiëntie van het oplossen van problemen en snelheid van de berekening van Big Data te verbeteren. Door middel van deze analyse wordt een probleem opgesplitst in twee of meer deelproblemen totdat deze eenvoudig genoeg worden om direct op te lossen. Vervolgens worden de oplossingen samengevoegd tot één oplossing voor het oorspronkelijke probleem. Eerst zal gedistribueerd computergebruik, zoals cloud computing en gedistribueerde intelligentie, gebruikt kunnen worden voor de berekening van de analyse in de ruimte. Daarna zal parallel computing, cluster of multi-core computing, gebruikt kunnen worden voor de berekening van de analyse op het perspectief van tijd (Shu, 2016).

Problemen

Al hoewel verschillende technieken aanwezig zijn om Big Data te kunnen verwerken, komen problemen nog regelmatig voor. Een van de problemen met Big Data is dat het gevoelig kan zijn en een conceptueel, technisch en juridische kant heeft. Big Data kan leiden tot nieuwe inzichten over een persoon waarvan de persoon niet had gewild dat de gegevenseigenaar dit weet. Ook wordt informatie verzameld over personen waarvan ze zich niet bewust zijn en vervolgens gebruikt wordt om waarde toe te voegen aan het bedrijf. Als data gebruikt wordt om beslissingen te nemen, zal het nauwkeurig, volledig en op tijd moeten zijn. Door het toevoegen van de noodzaak om data op te stellen en beschikbaar te maken voor overheidsinstellingen, wordt de data beheer- en bestuursproces complexer. Bedrijven zullen de gegevens niet willen delen om de cultuur van geheimhouding en concurrentievermogen niet te bedreigen (Katal, Wazid, & Goudar, 2013).

Daarnaast kan de grote hoeveelheid gegevens niet opgeslagen worden door de beschikbare ruimtes. Het uitbesteden van de data aan Cloud leek in eerste instantie een optie, maar het uploaden van de gegevens duurt te lang en de gegevens veranderen te snel waardoor het niet in realtime verkrijgbaar is. Ook die methodes kosten veel tijd, omdat het vinden van geschikte elementen gepaard gaat met het scannen van de hele dataset. Vervolgens brengt Big Data veel analytische uitdagingen met zich mee. Het type analyse is afhankelijk van de te behalen resultaten oftewel de besluitvorming. Dit kan met behulp van twee technieken: het opnemen van enorme hoeveelheid data in de analyse of vooraf bekijken welke data relevant is (Katal, Wazid, & Goudar, 2013).

Ook zal Big Data komende jaren nog veel groeien waardoor meer bekwame werknemers hiervoor nodig zullen zijn. De vaardigheden van deze werknemers moeten zich niet beperken op technische vaardigheden, maar ook uitstrekken tot analytische, interpretatieve en creatieve vaardigheden. Wanneer een storing optreedt moet de aangerichte schade binnen een acceptabele grens blijven in plaats van de hele taak opnieuw te moeten beginnen. De belangrijkste taak is om de faalkans tot op

een laag mogelijk niveau te brengen. Hoe meer gewerkt wordt aan een lagere faalkans, hoe hoger de kosten worden (Katal, Wazid, & Goudar, 2013).

Cloud computing is voortgekomen uit het schaalbaarheidsprobleem van Big Data en aggregeert meerdere ongelijke werklasten met verschillende prestatiedoelen in grote clusters. Daarentegen kost het veel geld om de middelen te delen en brengt het uitdagingen met zich mee zoals het uitvoeren van verschillende taken en het efficiënt omgaan met de systeemstoringen in de grote clusters. Het gebruik van meer gegevens zal leiden tot betere resultaten voor besluitvorming of voorspellende analyses. Ondernemers zullen steeds meer gegevens willen en de IT-afdeling zal de technische aspecten in gedachte nemen hierbij. Zo zal Big Data zich richten op kwaliteit gegevensopslag in plaats van zeer grote hoeveelheden irrelevante gegevens om tot betere conclusies te komen. Dit kan leiden tot verschillende vragen, zoals: welke hoeveelheid gegevens is voldoende voor besluitvorming en welke gegevens zijn relevant (Katal, Wazid, & Goudar, 2013).

Bij Big Data is sprake van heterogene gegevens. Aan de ene kant vertegenwoordigen ongestructureerde gegevens bijna alle gegevens die worden geproduceerd, zoals social media, e-mails en opgenomen vergaderingen. Aan de andere kant is sprake van gestructureerde data waarbij het georganiseerd is op een gemechaniseerd en beheersbaar manier (zie figuur 5). Het omzetten van alle ongestructureerde data in gestructureerde data is niet mogelijk. Het werken met ongestructureerde data is omslachtig en kostbaar (Katal, Wazid, & Goudar, 2013).



Figuur 5 gestructureerde data versus ongestructureerde data

Uitdagingen

Naast de problemen die ontstaan bij het werken met Big Data, zijn er ook uitdagingen waarmee bedrijven te maken krijgen. Big Data heeft impact gehad op de manier waarop bedrijven hun data verzamelen, analyseren en hier inzichten aan ontleend hebben. De data staan verspreid over meerdere computers wereldwijd en vanwege de omvang is het niet mogelijk om de data naar één computer te transporteren alvorens de analyse uit te voeren. De algoritmes zullen vaak aangepast moeten worden op de plaats waar de data zijn opgeslagen en zullen de deelresultaten van deze aanpassingen gestuurd moeten worden naar een centrale unit. De deelresultaten zullen samengevoegd worden tot een

eindresultaat. Dit is niet mogelijk voor alle analyses, waardoor niet altijd een precies beeld gevormd kan worden van de samenhang (Wieringa, 2015).

Bedrijven zijn continu aan het leren en nieuwe methodes aan het implementeren om met Big Data om te kunnen gaan. Het blijft echter een uitdaging op een reële oplossing te vinden voor de grote en complexe hoeveelheid data. Een van de grootste uitdagingen hierbij is het om kunnen gaan met de hoge kosten. Hardware apparatuur is voor de meeste bedrijven te duur. Een Big Datasysteem moet veel kunnen verwerken, stabiel zijn en ingewikkelde netwerkconfiguraties bevatten die gemaakt zijn door specialisten. Bedrijven kunnen kiezen voor een open netwerk, maar ze hebben nog steeds specialisten nodig hiervoor. Een keerzijde van een open netwerk is dat de onderhoud en ondersteuning hiervan niet wordt bijgeleverd zoals bij een betaald netwerk. Om een Big Data-oplossing te kunnen onderhouden heeft een bedrijf een onderhoudsteam nodig (Tole, 2013).

Doordat de data realtime beschikbaar is, weten de klanten dat nieuwe inzichten sneller kenbaar worden gemaakt. Een gevolg dat samenhangt hiermee is dat klanten verwachten dat ook daadwerkelijk snel zal worden gereageerd op de nieuwe inzichten. Ook zijn tegenwoordig meer databronnen toegankelijk geworden. De grootste waarde voor een onderneming wordt geleverd door het samenbrengen en analyseren van de vele databronnen en het genereren van meer en betere inzichten. Dit levert soms veel problemen op doordat niet altijd een koppeling gemaakt kan worden tussen de databronnen. Dit kan komen omdat de databronnen niet volledig zijn en gegevens ontbreken (Wieringa, 2015).

Daarnaast worden de softwareoplossingen door de mogelijkheden van een hardware, omdat de hardware niet harder kan werken dan de huidige technologieën bieden. Doordat de gegevens worden gesorteerd en besteld kunnen processen worden geoptimaliseerd. Hierdoor kan de beste prestatie bereikt worden en probeert de software het gebrek aan verwerkingssnelheid te compenseren. Een computerprogramma kan niet leren of zich aanpassen aan nieuwe soorten informatie tenzij het geprogrammeerd is om het te kunnen behandelen. Daarom zijn menselijke capaciteiten nodig om de gegevens te sorteren en het proces op die manier te versnellen. Bedrijven zullen menselijke analysevaardigheden nodig hebben om de waardevolle informatie te extraheren voor verder gebruik (Tole, 2013).

Sommige bedrijven kunnen meer profiteren van de voordelen van Big Data, omdat ze meer meegeëvolueerd zijn dan hun concurrenten. Bedrijven die een achterstand gelopen hebben proberen steeds meer om ook Big Data om te zetten in betere inzichten. Verschillende bedreigingen zijn aanwezig bij het willen profiteren van de Big Data successen, bijvoorbeeld het gebrek het analytisch talent en privacy-schending. Big Data zal aangepakt moeten worden met het gebruik van theorieën en

methoden uit verschillende vakgebieden zoals wiskunde, statistiek, patroonherkenning en machine learning. Een Data Scientist kan goed werken met verschillende technieken en methoden uit de wiskunde, statistiek en computer science. Een bedrijf kan veel data verzamelen uit verschillende bronnen, maar zonder een Data Scientist kunnen ze hier geen nuttige informatie uit halen (Wieringa, 2015).

Ethische implicaties

De mogelijkheden van Big Data brengt een groot hoeveelheid aan ethische bezwaren voor individuen, organisaties en de samenleving. Voorbeelden van ethische kwesties zijn privacy, besluitvorming, profilering, toezicht op individuen en gebrek aan transparantie in Big Data-waardeketen. Big Data-analyse vertrouwt op algoritmen om menselijke beslissingen te ondersteunen met het doel de smaak en gedrag van individu te voorspellen op basis van eerder getoond gedrag. Een tweede onethische sociale impact kan komen doordat individuen geprofileerd worden door middel van discriminatie. Profilering kan plaatsvinden doordat de mensen opzettelijk in bepaalde groepen in te delen op basis van ras, geslacht, sociale en economische status (Asadi Someh, Breidbach, Davern, & Shanks, 2016).

Individueen hebben het recht volgens de Internationale Telecommunicatie-unie om de informatie die over hen bekendgemaakt is te controleren. De informatie over individuen is mogelijk verzameld van hun online profielen. De bedrijven gebruiken dergelijke informatie om marketingstrategieën te ontwikkelen en diensten uit te breiden. Bedrijven hebben verschillende keuzes met betrekking tot waar de informatie opgeslagen kan worden. Het kan worden opgeslagen op Cloud systemen, in-house systemen of een hybride oplossing. Door de gegevens op te slaan op een Cloud systeem bespaart een bedrijf veel kosten, maar de gegevensbeveiliging blijft hierbij een grote kwestie (Tole, 2013).

Doordat de gegevensbeveiliging een uitdaging voor bedrijven blijft, kiezen sommige bedrijven ervoor hun eigen infrastructuur te bouwen voor de opslag en manipulatie van de data. Daarentegen zijn voor kleinere bedrijven de kosten om eigen infrastructuur te implementeren te hoog. Ook zullen er meer opgeleide personeel nodig zijn om het systeem te kunnen onderhouden. Het enige voordeel dat behaald wordt, is het oplossen van het probleem wat betreft privacy. Een ander oplossing voor dit probleem is het versleutelen van de informatie met een persoonlijke sleutel. Hierdoor wordt het onleesbaar voor personen die geen toestemming hebben om de informatie te zien. Een keerzijde van het gebruik van versleuteling is dat je hetzelfde software die de informatie heeft versleuteld moet gebruiken om de informatie te kunnen lezen en analyseren (Tole, 2013).

Daarnaast zijn individuen niet bewust van wat er met de verzamelde gegevens gebeurt. De gegevens gaan door een informatiewaardeketen, van volgbedrijven naar bedrijven voor aggregatie van

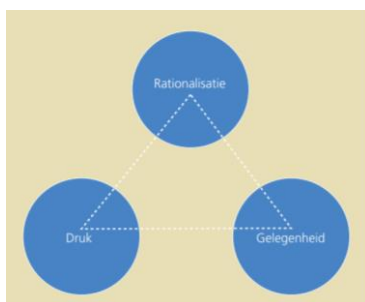
gegevens, naar retailers. De data worden uit meerdere bronnen gecombineerd en vormt een nieuw beeld van individuen op basis van de data. Ethische kwesties kunnen in elke segment van de waardeketen voorkomen, waarbij de uiteindelijke eigenaar van de gegevens het gebruikt voor doeleinden die verschillen van oorspronkelijke doeleinden (Asadi Someh, Breidbach, Davern, & Shanks, 2016).

3. Welke vormen van fraude komen voor bij gemeentelijke organisaties?

Definiëring van fraude

Gemeentes hebben op veel gebieden een monopolistische positie en hebben bevoegdheden zoals het opleggen van boetes en dergelijke. Wanneer sprake is van fraude bij gemeentes, zal dit resulteren in negatieve publiciteit, bijvoorbeeld een negatief nieuwsartikel over deze fraude. De gemeente zal vervolgens het vertrouwen dat geschaad is moeten herstellen en maatregelen moeten nemen om toekomstige incidenten te voorkomen. Een gemeente zal zich moeten realiseren welke frauderisico's kunnen ontstaan en hoe zij hier mee om kunnen gaan. Het is belangrijk de fraudedetectie intern en extern zichtbaar te maken vanwege de maatschappelijke functie. Hierbij wordt aandacht gezet op externe fraude, de fraude die gepleegd wordt door burgers en externe partijen. Om de fraudedetectie transparant te maken naar de buitenwereld is expliciete aandacht voor en beheersing nodig van frauderisico's (de Groot, van Vliet, Vriend, & van Vugt, 2007).

Fraude is een verzamelnaam voor onder ander verwijtbaar, niet-toegestaan of niet-gewenst gedrag. De kans op fraude wordt gedefinieerd door middel van het begrip frauderisico. Een frauderisico bestaat uit drie onderdelen (zie figuur 6). Vanuit een organisatie kan een druk worden veroorzaakt die leidt tot fraude, oftewel een prikkel. Om het voorkomen van frauderisico te bevorderen, dient er te worden geconcentreerd op het punt 'gelegenheid'. De gelegenheid is de mogelijkheid die een organisatie heeft om te kunnen frauderen en hoe eenvoudig het is om überhaupt fraude te kunnen plegen. Het derde onderdeel, rationalisatie, houdt rekening met het gedrag van mensen op de manier van hoe bereid mensen zijn om fraude te plegen. (de Groot, van Vliet, Vriend, & van Vugt, 2007).

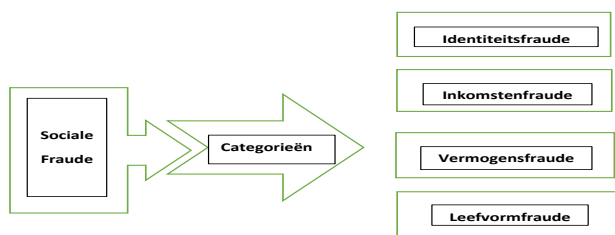


Figuur 6 Fraudedriehoek

Ook zullen de gemeentes een proactieve houding moeten aannemen ten aanzien van het voorkomen van fraudes. Er dient een analyse van de frauderisico te worden opgenomen als onderdeel van het risicomanagement van de gemeente. Door middel van de analyse van de frauderisico kunnen de opgemerkte frauderisico's beperkt worden met adequate beheersmaatregelen. Hierbij zullen grenzen gesteld moeten worden om op die manier niet elke persoon te bestempelen als fraudeur. Door het maken van een frauderisicoanalyse worden gemeentes bewust van de voorkomende fraudes. Als gevolg hierop kunnen de gemeentes eerder maatregelen nemen en op die manier een voorsprong hebben op de fraudeur (de Groot, van Vliet, Vriend, & van Vugt, 2007).

Sociale fraude

De voornaamste hoofdvorm van fraude door burgers die voorkomt bij gemeentelijke organisatie is sociale fraude. Sociale fraude kan vervolgens onderverdeeld worden in onder ander de subcategorieën: identiteitsfraude, inkomstenfraude, vermogensfraude en leefvormfraude (zie figuur 7). Fraude bij gemeentelijke organisaties kan naast de sociale fraude door burgers ook gepleegd worden door bedrijven. Dit zijn bedrijven die gecontracteerd zijn door de gemeentes en niet het geen leveren wat van tevoren afgesproken is met de aanbieders (Sociaalweb).



Figuur 7 Onderverdeling sociale fraude bij gemeentelijke organisaties

Om te beginnen is sociale fraude het opzettelijk verzwijgen van informatie of het opzettelijk geven van onjuiste informatie. Dit met doel om onterecht aanspraak te maken op een sociale uitkering of om deze te behouden. Hiermee wordt het sociale stelsel aangetast en gaat het gepaard met hoge kosten voor de gemeentes. Nederland zorgt ervoor dat ieder burger de minimale middelen heeft om voor zichzelf te kunnen zorgen. Doordat sociale fraude gepleegd wordt, wordt het maatschappelijke draagvlak hiervan aangetast en lijden de goeden onder de kwaden. Een fraudeur kan door middel van sociale fraude uitkeringen aanvragen en ontvangen op andermans naam. De complexiteit van het systeem zorgt voor kwetsbaarheden en vooraf controleren is tijdrovend. Er ontstaat een trade-off tussen de noodzakelijke controle vooraf het proces en de wens van burgers om het proces snel af te handelen (van Geldrop & de Vries, 2015).

Identiteitsfraude

Een manier om sociale fraude te plegen is door het toepassen van identiteitsfraude om in aanmerking te komen voor een uitkering. Identiteitsfraude is een nationale en wereldwijde bedreiging voor de veiligheid van naties en hun burgers, de economie en wereldwijde handel, omdat het een breed scala aan misdaden mogelijk maakt en terrorisme. Het staat op de voorgrond als maatschappelijk probleem voor meerdere jaren en het publiek is bewust van de gevaren die hiermee gepaard gaan wat betreft hun persoonlijke en financiële veiligheid (Gordon & Willox, Jr., 2003). Dit betekent dat iemand met oneerlijke bedoelingen zich opzettelijk uitgeeft onder een identiteit van een ander bestaande of fictieve persoon. De fraudeur kan gebruik maken van persoonsnummers, acties of foto's om geloofwaardig te maken dat de tegenpartij met een desbetreffende persoon te maken heeft (Grijpink, 2004).

In het verleden vond identiteitsfraude plaats door middel van *dumpster diving*. Hackers vertrouwen op het feit dat mensen niet de waarde kennen van de informatie die ze weggooien. Ze vinden gevoelige en zelfs geclassificeerde informatie in de prullenbak waar de hackers een organisatie gemakkelijker kan aanvallen. Een aanval kan zorgen voor reputatieschade, verlies van vertrouwen van het publiek en financiële schade. Hackers omzeilen bij zulke aanvallen de technische controles en focussen op de menselijke zwaktes (Scott D. Applegate, 2009). Tegenwoordig gebruiken fraudeurs uitsluitend digitale middelen om persoonlijke gegevens van derden te verkrijgen. Het is slechts een kwestie voor de fraudeur van zoeken naar het meeste kwetsbare systeem. Voorbeelden van digitale middelen zijn: websites, telefoons, virussen en valse e-mails. Met de gevonden gegevens kan een fraudeur weer verder zoeken om nieuwe informatie te verkrijgen van het slachtoffer. De organisaties houden zich bezig met het verbeteren van de maatregelen om dit tegen te gaan. Wanneer identiteitsfraude toch plaatsvindt, kan het effect hiervan doorgewerkt worden naar andere systemen en is het nauwelijks om te keren (van Geldrop & de Vries, 2015).

Met identiteitsfraude kan ten onrechte aanspraak gemaakt worden op uitkeringen. De mensen krijgen dan een uitkering waar ze eigenlijk geen recht op hebben. De gemeente mag niet alleen uitgaan van vermoeden, maar moet daarnaast daadwerkelijk kijken of er sprake is van fraude. Om nog een voorbeeld te geven waarbij sprake is van identiteitsfraude wordt er gekeken naar de kinderopvang. De wetgever vindt dat het voor ouders zo eenvoudig mogelijk moet worden gemaakt om toeslagen te kunnen ontvangen voor onder andere kinderopvang. Het werd zo eenvoudig gemaakt dat fraudeurs DigiD nummers van moeders, die geen gebruik maakten van kinderopvang, gebruikten om grote sommen toeslagen te ontvangen. De fraudeurs vertelden de alleenstaande moeders dat ze nog aanspraak konden doen op een ander toeslag naast de kinderbijslag door hun DigiD af te geven. Ook

was in een aantal gevallen sprake van alleenstaande moeders die doelbewust meewerkten aan de fraude. Er werd bijvoorbeeld voor drie kinderen uit een gezin voor een maximaal hoeveelheid aan toeslag aangevraagd. Dit is gewoonlijk ongebruikelijk dat meerdere kinderen vijf dagen per week voor het maximaal aantal uren bij een kinderopvang zitten. Doordat er geen controle plaatsvond, keerde de belastingdienst het opgevraagde bedrag toch uit (Sociaalweb).

Inkomstenfraude & Vermogensfraude

Daarnaast kan met inkomstenfraude ook ten onrechte een uitkering of subsidie worden ontvangen. Hierbij wordt inkomsten uit arbeid of een ander uitkering niet opgegeven waardoor een persoon ten onrechte een (te hoge) uitkering heeft ontvangen. Inkomsten spelen een rol bij het bepalen van de hoogte van de uitkering. Dit kan ook met behulp van vermogensfraude, waarbij een persoon niet opgeeft dat hij over vermogen beschikt. Vermogensfraude kan zich voordoen wanneer een persoon een bankrekeningsaldo, auto of onroerend goed niet opgeeft. Dit vermogen kan zich ook in het buitenland bevinden.

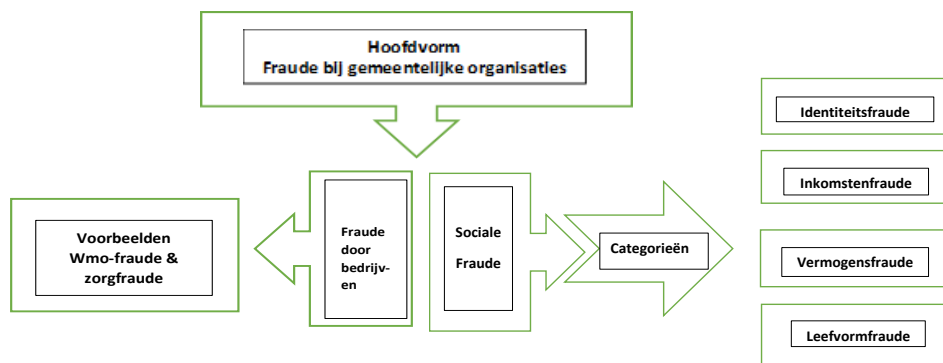
Leefvormfraude

Als laatste kan ten onrechte een uitkering aangevraagd worden door middel van leefvormfraude. Een persoon moet in de gemeente verblijven waar de uitkering is aangevraagd en de woonsituatie is medebepalend voor de hoogte van de uitkering (Alexa Crawls, 2007). Het kost veel werk om fraude te kunnen vaststellen, omdat bijvoorbeeld bij een vermogensfraude gekeken moeten worden naar waar het geld zit en bij woonfraude moet gecheckt worden hoeveel mensen ongeoorloofd samenwonen (Sociaalweb).

Fraude door bedrijven

Naast de sociale fraude die gepleegd wordt door burgers is er ook fraude die gepleegd wordt door bedrijven (zie figuur 8). De gemeentelijke organisaties hebben de taken om burgers begeleiding, dagbesteding en beschermd wonen te bieden, namelijk de wet maatschappelijke ondersteuning (Wmo). Het doel van de gemeente is mensen langer thuis te kunnen laten wonen en deel te laten nemen aan het dagelijkse leven. De gemeente kan hier hulp voor aanbieden wanneer een persoon hier niet in staat voor is door bijvoorbeeld een woonvoorziening of vervoersvoorziening aan te bieden. De kwaliteit en rechtmatigheid van deze ondersteuning valt onder hun verantwoordelijkheid. Op het terrein van begeleiding en beschermd wonen gaat veel geld om, van wel duizenden tot tienduizenden euro's per cliënt per jaar. Het vormt voor malafide aanbieders een interessante mogelijkheid geld te verdienen door de combinatie van kwetsbare mensen en de hoge geldbedragen. De organisaties

bieden zo geld aan dat bedoeld is voor jongeren met bijvoorbeeld een beperking, maar wordt daar uiteindelijk niet voor gebruikt (Sociaalweb).



Figuur 8 onderverdeling voorkomende fraudes bij gemeentelijke organisaties

Dit komt doordat informatiestromen van gemeenten vaak niet op orde zijn. Gemeentes hebben de verantwoordelijkheid op grond van de jeugdwet ervoor te zorgen dat de jeugdigen en hun ouders bij een aangewezen organisatie terecht kunnen met vragen en zorgen over de opvoeding, ontwikkeling, geestelijke gezondheid of veiligheid van een kind. Een jeugdige met hulpaanvraag zich kan aanmelden bij verschillende partijen, bijvoorbeeld het jeugdteam, huisarts of medische specialist. Hierdoor is er weinig zicht op de toegang, omvang en soort van de zorg van jeugdigen en kunnen er geen realistisch begrotingen gemaakt worden. De gemeentelijke organisatie kunnen moeilijk de organisatie van de jeugdhulp bijsturen en strategieën ontwikkelen om fraudes te voorkomen of op te sporen. De zorgaanbieders geven weinig informatie af aan de gemeentes door gebrek aan bereidwilligheid en omwille van privacy van patiënten. Hierdoor kunnen verschillende vormen van fraude gepleegd worden zoals: de jeugdige kan een ernstigere voorstelling maken van zijn klachten om zorg te maximaliseren, zorgaanbieders kunnen minder zorg aanbieden dan gefactureerd is of jeugdigen kunnen frauderen met het persoonsgebonden budget (Fenger, Chin-A-Fat, Frankowski, & van der Torre, 2016).

Bij deze vorm van fraude is er een onderscheid tussen fraude door cliënten en fraude door aanbieders. De aanbieders worden meestal gecontracteerd door de gemeentes. Hierbij komt het voor dat de door de gemeente ingeschakelde aanbieders niet hetgeen leveren aan de cliënt wat door gemeente is toegekend. Een voorbeeld hierbij is dat de gemeente individuele begeleiding toekent aan een cliënt, maar in de praktijk een goedkoper groepsbegeleiding aangeboden krijgt. Achteraf brengt de aanbieder een hoger tarief voor individuele begeleiding in rekening aan de gemeente. In de gemeentelijke Wmo-verordening moeten regels opgesteld worden om oneigenlijk gebruik van de wet of ten onrechte ontvangen van een voorziening te detecteren. De cliënten zullen hierbij relevante informatie moeten

verstrekken, informatieplicht, en meewerken aan het onderzoek, medewerkingsplicht, zodat de gemeente zelf regels kan maken om deze vorm van fraude te detecteren (Sociaalweb).

4. Hoe wordt Big Data ingezet om deze vormen van fraude te detecteren?

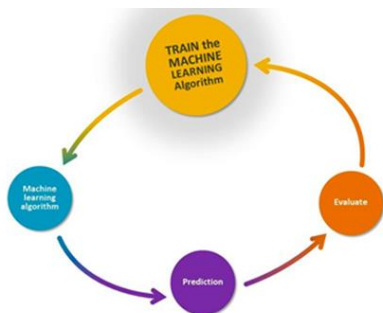
Big Data biedt steeds meer mogelijkheden bij het opsporen en het in beeld brengen van fraudeurs. Bij het frauderen wordt misbruik gemaakt van het vertrouwen. Hoe meer een gemeentelijke organisatie kwetsbaar is voor fraudes, hoe meer beheersmaatregelen de organisatie zal moeten treffen. De vier grootste vormen van sociale fraude zijn: identiteitsfraude, inkomstenfraude, vermogensfraude en leefvormfraude. Het grootste knelpunt voor gemeentes is de bewijslast. Gemeentelijke organisaties hebben snel een vermoeden en krijgen een signaal binnen van fraude, maar het kost veel werk te kijken of daadwerkelijk sprake is van fraude. Doordat gemeentes werken aan het verbeteren van de mogelijkheid aan de burgers bijstandsfraude te melden en op die manier een groter hoeveelheid gegevens, Big Data, te creëren wordt een beleid gevormd van het verbeteren van de administratie. Hierdoor wordt de bewijslast voor de gemeentes eenvoudiger gemaakt (Gemeente Bergen op Zoom).

Niet-eendimensionale manier

Mensen kunnen beperkt aantal waarnemingen aan door onder ander te kijken naar het uiterlijk van een persoon, oftewel etnisch profileren. Met behulp van Big Data is het niet eendimensionaal, omdat daarmee het gedrag in tijd van mensen met veel verschillende kenmerken in beeld wordt gebracht en niet alleen gekeken wordt naar zichtbare kenmerken zoals het uiterlijk van een desbetreffende persoon. Het bevat alles wat bekend is over een bepaald persoon door onder andere het grotere volume (een van de 4 v's van Big Data) en gaat veel verder dan alleen waarnemingen door onder ander de grotere variabiliteit (Sociaalweb). Doordat met Big Data verder gegaan kan worden dan alleen waarnemingen, wordt het steeds belangrijker voor gemeentes bij fraudedetectie. Met behulp van de combinatie van Big Data en Machine Learning, wordt het frauderen met een bijstandsuitkering voorspeld. Bij Machine Learning wordt fraudedetectie alleen ingezet bij de gevallen waar de kans op fraude het grootst is. Op deze manier worden minder vooroordelen gevormd over de burgers en worden minder mensen onterecht lastiggevallen. Dit komt doordat met Machine Learning aandachtig gekeken wordt door de computer naar de variabelen en wordt de kans berekend op fraude. Er zal hierbij geen sprake zijn van profiling en de methode zal niet eendimensionaal zijn. Gemeentelijke organisaties zullen hiermee hun beleid moeten aanvullen en aanpassen. Het eindoordeel ligt wel altijd bij de mens door het uitvoeren van een rechtmatigheidsonderzoek. Een uitkering mag niet gestopt worden, omdat de computer aangeeft dat een hoge kans op fraude bestaat bij een desbetreffende persoon (Sociaalweb).

Door middel van Big Data kan een set data gevormd worden waarvan bekend is wie de fraudeurs zijn. Vervolgens gebruikt de computer Machine Learning om te zoeken naar variabelen in de data die horen bij een fraudeur. Ook ontwikkelt Machine Learning een algoritme dat voorspellingen kan maken met de data. Het algoritme onderzoekt cases en variabelen en bekijkt in hoeverre de combinatie van de variabelen van één case lijkt op die van de gevonden fraudeurs. Hierna wordt een kans op fraude gegeven door de computer (Sociaalweb).

Zowel de correcte als de foutieve voorspellingen kunnen teruggegeven worden aan de computer zodat de computer leert van de patronen die zitten in de nieuwe data (zie figuur 7). Na meerdere iteraties kan de nauwkeurigheid hierdoor flink verbeteren. Dit is de kracht van Machine Learning, omdat naarmate de computer meer voorspellingen terugkrijgt wordt het voorspellend model steeds sterker. Gemeentes hebben onderzoeken gedaan waarbij wel fraude is geconstateerd en waarbij geen fraude geconstateerd is. De gemeentes weten bij de gevallen waarbij geen fraude gepleegd is niet zeker of er ook daadwerkelijk geen fraude gepleegd is. Het percentage fraudeergevallen is erg klein ten opzichte van het geheel. Hierdoor zal een kleine set met data aanwezig zijn waarop de algoritmes toegepast kunnen worden (Sociaalweb).



Figuur 7 Machine Learning proces

Een gemeentelijke organisatie moet voldoende cases hebben om algoritmes te bouwen en de data zal goed geregistreerd moeten zijn. Wanneer dit niet het geval is, zal de data eerst gerepareerd moeten worden voordat voorspellingen gemaakt kunnen worden. De organisatie zal geen onderzoek kunnen doen in het ene systeem en fraude constateren in een ander systeem. Hierdoor was geen koppelsleutel aanwezig tussen de systemen en bevat de constatering onzuiverheden. Als de gemeentelijke organisatie data bevat waarmee niet voorspeld kan worden, zal de organisatie moeten investeren in de eigen data. Er zal een businesscase opgesteld worden waarin wordt aangegeven wat Machine Learning zal kosten en wat het zal opleveren. Op die manier zal er bepaald worden of het gebruik van Machine Learning geschikt is voor een specifieke gemeente (Sociaalweb).

Uitwisselen van persoonsgegevens

De gemeentelijke organisaties zullen voorzichtig om moeten gaan met de data. Ze krijgen te maken met privacyregels en regels met betrekking tot welke data waarvoor gebruikt mogen worden. Er wordt voornamelijk data gebruikt uit de uitkeringsadministratie. De vraag die constant gesteld zal moeten worden is: is er verenigbaar gebruik van die gegevens voor het doeleinde dat je wilt bereiken? Eerst was alles vanuit de Wet bescherming persoonsgegevens (WPB) geregeld. Daarna werd het vanuit de Algemene verordening gegevensbescherming (AVG) geregeld waarin meer regels staan dan de WPB. Voor elk onderzoek zal nu een privacy impact assessment nodig zijn. Daarin staat welke gegevens gebruikt gaan worden, hoe die gegevens beveiligd zijn en of er verenigbaar gebruik is. De gemeentelijke organisaties krijgen zo meer administratieve handelingen voordat onderzoeken uitgevoerd kunnen worden om op die manier risico's beter in te kunnen schatten (Sociaalweb).

Het uitwisselen van persoonsgegevens tussen twee verschillende organisatie is verboden. De ene gemeentes kent andere waardes toe aan variabelen dan de andere gemeente. Zo is het huurprijsniveau in Bergen op Zoom veel lager dan bijvoorbeeld Rotterdam. Hierdoor kunnen de bestanden niet gecombineerd worden en zijn het losstaande trajecten in verschillende gemeentes. Wanneer de bestanden gecombineerd konden worden, werden algoritmes hierdoor sterker gemaakt. Er zouden mogelijkheden liggen waarbij het algoritme van het ene project op het andere wordt gezet zonder het meenemen van persoonlijke gegevens. Hier wordt nog naar gekeken doordat Nederland een van de eerste is die sociale fraude voorspellen met behulp van Machine Learning (Sociaalweb).

Wisselwerking op orde brengen adressen en verschillende vormen van fraudedetectie

Gemeentes proberen een van de vier vormen van sociale fraude, identiteitsfraude, nadrukkelijk tegen te gaan door een wisselwerking plaats te laten vinden tussen het op orde brengen van onder ander adressen en verschillende vormen van fraudedetectie. Verkeerde woonadressen opgeven en verkeerde identiteiten zijn de grootste vormen van sociale fraude om op die manier uitkeringen, studiefinanciering en toeslagen te kunnen ontvangen. Ook speelt sociale fraude een grote rol bij zwaardere criminaliteiten. Het heeft nut om risico gestuurd adresonderzoek op te stellen om een hoger percentage onjuiste adressen te traceren en daarmee een beleid te kunnen vormen. Data-analyse is nodig om patronen te herkennen en op basis van die patronen gericht te zoeken naar fraude. Dit zal leiden tot risicoprofielen: bij welke kenmerken is er grote kans op sociale fraude. Ook werden na een paar jaar mechanismes ingesteld om de toeslagen te controleren. Door middel van data te koppelen werd het maximaal aantal uren dat ouders toeslag kunnen ontvangen gekoppeld aan het aantal werkuren van de minst werkende ouder (Olsthoorn, 2016).

De toename in data leidt tot de vraag naar nieuwe geavanceerde methodes voor het handelen en analyseren van de data. Door middel van de geavanceerde analyses kunnen inzichten worden verkregen die anders moeilijk te realiseren zijn (van Geldrop & de Vries, 2015). Zo gaat fraudedetectie meer en meer hand in hand met het verbeteren van adreskwaliteit. Opsporingsdiensten kunnen met een goed werkend stelsel van basisregistraties sneller relevantere informatie verzamelen. Die verzamelende informatie zal vervolgens gekoppeld worden om op die wijze een fraudegeval te kunnen ontdekken of te ondersteunen. Door middel van de Big Data kunnen analyses uitgevoerd worden om nieuwe risicofactoren te ontdekken zonder dat er sprake is van een concrete verdenking. Er zal een centrale informatiedatabase system opgebouwd moeten worden voor incidenten met identiteitsfraude. Er is behoefte aan het meten van de omvang, reikwijdte en trends van identiteitsfraude te meten. De informatie-uitwisselingsnetwerk zal minder beperkt en gemakkelijker beschikbaar moeten zijn, omdat meer toegang nodig zal zijn voor globale gegevens om te bepalen of een aangeboden identiteit geldig is (Gordon & Willox, Jr., 2003). Hierbij spelen vragen ten aanzien van privacy een rol, zoals de principes van subsidiariteit. Zonder voorwaarden hiervoor zouden veel valse verdenkingen gemaakt kunnen worden. Dit kan leiden tot grote gevolgen voor een burger wanneer deze informatie verspreid zal worden naar verschillende instanties (Olsthoorn, 2016).

Om dit te voorkomen zal beveiliging onderdeel worden van de bedrijfscultuur. Een goede verdedigingsstrategie tegen social engineering bevat het beveiligingsbewustzijn, training, technische controles in combinatie met een beleid dat met een effectieve informatiemanagementstrategie waarmee informatie wordt geïdentificeerd en beschermd. Werknemers zullen opgeleid moeten worden tegen sociale technische aanvallen door het kennen van de soorten aanvallen, de technieken om deze aanvallen tegen te gaan en de schade die het kan opleveren. Met onder andere prestatiebeoordelingen die periodiek plaatsvinden zullen werknemers hun dagelijkse routine hierop aanpassen. Ze worden aangemoedigd om te voldoen aan dit beleid (Scott D. Applegate, 2009).

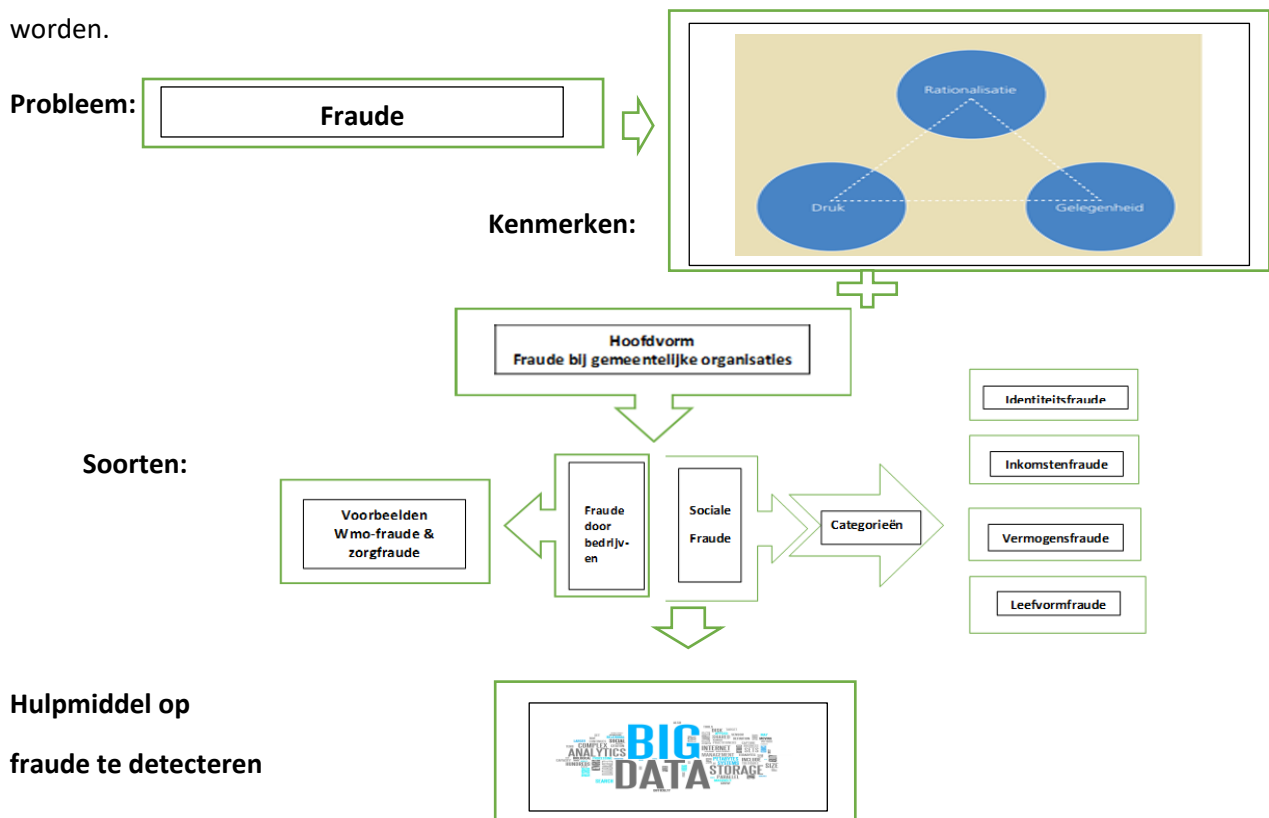
Gezamenlijk gebruik van databestanden voor fraudedetectie

Ten tweede kan Big Data ingezet worden om de tweede soort fraude, fraude door bedrijven, te detecteren. Bij deze fraude wordt niet hetgeen wat door gemeente is toegekend geleverd aan de client door de gemeente ingeschakelde aanbieders. Jaarlijks wordt miljarden aan onregelmatigheden van het medisch declareren gevonden door individuele verzekeraars. Als oplossing hiervoor willen de verzekeraars eenvoudiger inzage wat tegenstrijdig is met privacybelangen. De Tweede kamer eiste intensieve samenwerking op tussen alle partijen die betrokken waren bij de financiering van de zorg en een samenwerking voor een werkprogramma voor aanpak van fraude in de zorg. Er worden mogelijke projectgroepen of werkgroepen opgericht voor bijvoorbeeld specifieke risicoanalyses. Dit

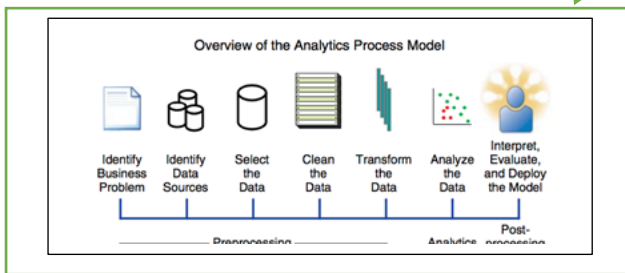
vereist gezamenlijk gebruik van databestanden voor fraudedetectie en vanuit dat punt analyses van grote hoeveelheden data. Daarbij wordt informatie gedeeld over ernstige voorkomende fraudes en de daarbij benodigde persoonsgegevens zodat een beter beleid gevormd kan worden om deze vorm van fraude door bedrijven tot het minimum te kunnen brengen (Olsthoorn, 2016).

Samenvatting

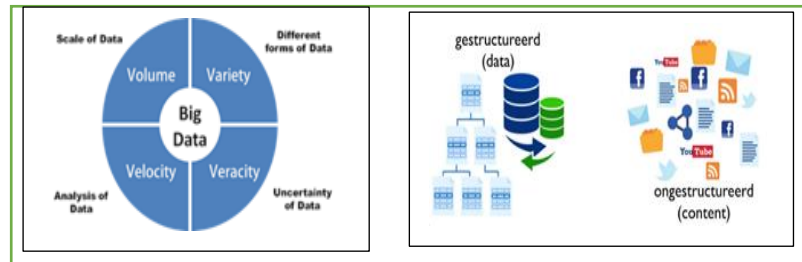
Hieronder wordt schematisch herhaald hoe Big Data invloed heeft op het beleid voor externe fraudedetectie bij gemeentelijke organisaties door middel van eerder besproken informatie. Gemeentelijke organisaties willen het probleem fraude oplossen door meer fraudes te gaan detecteren en te bestrijden. De kans op fraude wordt gedefinieerd door middel van het begrip frauderisico dat bestaat uit drie onderdelen: rationalisatie, druk en gelegenheid. Er kunnen verschillende vormen van fraude voorkomen, namelijk: sociale fraude door burgers dat vervolgens onderverdeeld kan worden in onder andere de subcategorieën: identiteitsfraude, inkomstenfraude, vermogensfraude en leefvormfraude en daarnaast fraude bij gemeentelijke organisaties door bedrijven. Fraude kan sneller en beter gedetecteerd worden door middel van Big Data dat *high-volume*, *high-velocity*, *high-variety* en *veracity* is. Bij Big Data is sprake van heterogene gegevens en het bevat alle gegevens die worden geproduceerd, zoals social media, e-mails en opgenomen vergaderingen. Door middel van Big Data kunnen bijvoorbeeld databestanden voor fraudedetectie gezamenlijk gebruikt worden en vanuit dat punt analyses van grote hoeveelheden data uitgevoerd worden.



Werking



Kenmerken



Maatregelen gebruik big data

- Niet-eendimensionale manier gebruiken om machine learning toe te passen zodat meer terechte fraudes opgespoord kunnen worden
- Uitwisselen van bestanden onderling zonder het uitwisselen van persoonsgegevens waardoor sterkere algoritmes gemaakt kunnen worden om fraude op te sporen
- Wisselwerking plaats laten vinden tussen het op orde brengen van onder ander adressen en verschillende vormen van fraudedetectie
- Gezamenlijk gebruik van databestanden voor fraudedetectie en vanuit dat punt analyses van grote hoeveelheden data.

5. Opzet empirisch onderzoek

Hypotheses

Met betrekking tot Big Data kunnen gemeentelijke organisaties beleidsmaatregelen treffen, waaronder het uitwisselen van persoonsgegevens. Hiertoe worden algoritmes sterker gemaakt waardoor meer terechte fraudes, sociale fraude en fraudes door bedrijven, opgespoord kunnen worden met bijvoorbeeld machine learning. Het uitwisselen van persoonsgegevens tussen twee verschillende organisatie was in eerste instantie verboden. Gemeentes kennen onderling verschillende waardes toe aan variabelen waardoor de bestanden niet gecombineerd kunnen worden. Dit komt omdat de gemeentes verschillen van elkaar op het gebied van omgeving, werkgelegenheid en grootte. Wanneer de bestanden gecombineerd konden worden zonder dat persoonsgegevens meegenomen worden, werden algoritmes hierdoor sterker gemaakt, omdat meer verbanden gelegd kunnen worden tussen gemeentes en eerder fraudes herkend kunnen worden (Sociaalweb).

Hierbij kan de volgende hypothese bekeken worden:

- Gemeentes wisselen bestanden uit aan elkaar zonder dat persoonsgegevens meegenomen worden zodat algoritmes sterker gemaakt kunnen worden waardoor meer fraudes, sociale fraude en fraude door bedrijven, opgespoord kunnen worden

Gemeentes kunnen bij het vormen van beleid ook een andere maatregel treffen, namelijk: wisselwerking tussen het op orde brengen van onder ander adressen en verschillende vormen van fraudedetectie. De pakkans is kleiner wanneer iedere Nederland gecontroleerd moet worden op fraude door middel van een willekeurige streekproef. Een oplossing hiervoor is dat gemeentes een beleid vormen door risico gestuurd adresonderzoek op te stellen om een hoger percentage onjuiste adressen te traceren. Gemeentes werken aan het verbeteren van de mogelijkheid aan de burgers bijstandsfraude te melden op hun website. Op die manier wordt een groter hoeveelheid gegevens gecreëerd, Big Data, en daarmee beleid gevormd van het verbeteren van de administratie. Om patronen te herkennen in de resultaten van dit onderzoek is data-analyse nodig en op basis daarvan gericht te zoeken naar fraude. Met deze patronen kunnen risicoprofielen opgesteld worden.

Hierbij kan de volgende hypothese bekeken worden:

- *Door risico gestuurd adresonderzoek kunnen gemeenten een hoger percentage onjuiste adressen traceren, waardoor het gemeentelijk beleidsproces geoptimaliseerd wordt.*

Daarnaast kunnen gemeentes bij het vormen van beleid de maatregel nemen om gezamenlijk binnen de gemeentes gebruik te maken van databestanden voor fraudedetectie. Door het gezamenlijk gebruik kunnen analyses uitgevoerd worden op grotere hoeveelheden data. Zo eiste de Tweede Kamer dat er intensief samengewerkt moest worden door de betrokken partijen zodat de data onderling gedeeld kan worden, bijvoorbeeld een samenwerking voor aanpak van fraude in de zorg. Ook worden mogelijke project- en werkgroepen opgericht voor bijvoorbeeld specifieke risicoanalyses. Hierbij wordt informatie gedeeld over ernstige fraudegevallen en de daarbij benodigde persoonsgegevens zodat een beter beleid gevormd kan worden om fraude door bedrijven tegen te gaan (Olsthoorn, 2016).

Hierbij kan de volgende hypothese bekeken worden:

- *Gemeentes werken in projectengroepen of werkgroepen om specifieke risicoanalyses uit te voeren met gezamenlijk gebruik van databestanden voor fraudedetectie en vanuit dat punt analyses van grote hoeveelheden data.*

Data

Om te kijken of drie hypotheses met zekere mate als geldig gezien kunnen worden, wordt gebruik gemaakt van de volgende drie gemeentes: Bergen op Zoom (66811 inwoners), Roosendaal (77031 inwoners) en Breda (180937 inwoners). Er is gekozen voor deze drie gemeentes op basis van verschillende criteria. Ten eerste is gekozen voor deze drie gemeentes vanwege hun ligging. Doordat ze alle drie in de buurt van elkaar liggen bestaat een grotere kans dat er samen wordt gewerkt en

overeenkomsten zijn tussen de gemeentes zijn. Ten tweede moet het een middelgrote organisatie zijn, omdat zulke organisaties vaak over een ICT-afdeling beschikken en in staat zijn om op die manier onderzoeken uit te voeren. Dat is bij kleine organisaties niet het geval. De websites, jaarrekeningen en accountantsverslagen van deze drie gemeentes zullen nader bekeken worden.

Methode

Om te kijken of de hypotheses geldig zijn is onderzoek uitgevoerd. In dit onderzoek is deskresearch uitgevoerd en zal gekeken worden naar de jaarrekeningen, accountantsverslagen en uit de jaren 2019 en 2020 om op die manier meest recentelijke informatie te vinden. De jaarrekening zal bekeken worden op de trefwoorden *fraude*, *samenwerking* en *risico's*. Op die manier zullen de jaarrekeningen informatie geven over wat er gedaan wordt aan fraude, of er samengewerkt wordt met gemeentes en wat gedaan wordt aan risico's van bijvoorbeeld een bijstandsuitkering. De accountantsverslagen zullen alleen doorgrond worden door middel van het trefwoord *fraude* om te kijken hoe fraudes in jaarrekening gecontroleerd en gedetecteerd worden. Daarnaast zal gekeken worden naar de raadstukken van 2021 om te onderzoeken of gemeentes daadwerkelijk bezig zijn met fraudes met behulp van de trefwoorden *fraudes*, *samenwerking* en *werkgroepen*.

Ook zal gekeken worden naar de websites van de desbetreffende gemeentes. Op de website zal gezocht worden op de trefwoorden: *fraude*, *samenwerking*, *werkgroepen*. Op die manier zal gekeken worden of er daadwerkelijk een mogelijkheid is om fraude te melden op de websites, of gemeentes samenwerken en of gemeentes werkgroepen hebben opgericht om specifieke risicoanalyses uit te voeren. Het onderzoek is valide en betrouwbaar. Er is hierbij onderzocht wat in eerste instantie onderzocht moest worden en betrouwbaar omdat de onderzoeksresultaten feiten zijn en hetzelfde zouden zijn wanneer dit onderzoek opnieuw gedaan zou worden.

6. Resultaten

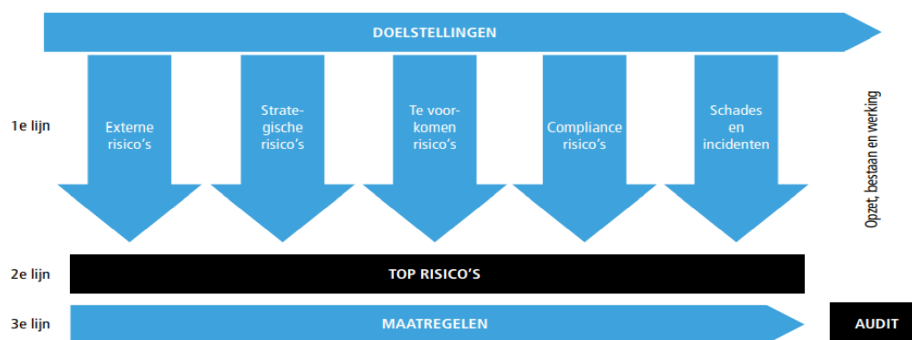
Eerste hypothese

Bij alle drie de gemeentes was vermeld dat er wordt samengewerkt en bestanden uitgewisseld worden. Ook zal aandacht geschonken worden aan het opsporen en tegengaan van fraudes. Op de website van Bergen op Zoom, bij het zoeken op het trefwoord *samenwerking*, staat dat Bergen op Zoom en Roosendaal werken voor een sterke regio. Zij zullen samen optrekken en elkaar aanvullen waar nodig, omdat er kansen zijn om de huidige samenwerking verder te intensiveren. Ze zullen gezamenlijk een plan neerleggen om de gemeentes in West-Brabant-West verder te kunnen

versterken en op de kaart te kunnen zetten (Gemeente Bergen op Zoom). Door deze samenwerking zullen hoogstwaarschijnlijk meer informatie met elkaar gedeeld worden en zullen regels gevormd worden om minder fraude voor te laten komen in de desbetreffende gemeentes.

In de jaarrekening van Roosendaal is te vinden op het trefwoord *fraude* onder het kopje *programma 3 Werken & Ondernemen* dat er bewust lager ingezet wordt op fraudepreventie. Dit komt omdat de aanpak van het voorkomen en detecteren van fraudes niet langer het gewenste rendement opleverde. Het geld werd zo besteed aan Werk en Participatie waarmee een voordeel behaald werd van dertigduizend euro. Ook blijkt na het zoeken op het trefwoord *samenwerking* en *risico's* dat er intensief samengewerkt met de gemeenten in West-Brabant West en West-Brabant Oost om te kunnen komen tot gecontroleerde zorgoverdracht gekeken wordt naar de ontwikkelingen wat betreft de risico's van een groeiende bijstand (Gemeente Roosendaal, 2019).

Verder valt in de jaarrekening van Breda bij het zoeken op het trefwoord *fraude* onder het kopje *weerstandvermogen en risicobeheersing* te zien dat fors wordt ingezet op het opsporen en voorkomen van fraude met adressen door het gebruiken van risicoprofielen. Hierdoor volgt Breda een van de maatregelen die zijn besproken in hoofdstuk 4, namelijk: wisselwerking plaats laten vinden tussen het op orde brengen van onder ander adressen en verschillende vormen van fraudedetectie. De gemeente Breda wil met zo min mogelijk tegenslagen doelen realiseren. Dit gebeurt in drie lijnen (zie figuur 11), waarbij het centrale uitgangspunt bewustwording is. Hierbij vallen fraudes onder schades en incidenten en valt het onder de 1^e lijn risico's. Die vormen van schades en incidenten komen dagelijks voor (Gemeente Breda, 2019).



Figuur 11 beleid Breda: uitgangspunten en proces

Ook in de raadstukken van gemeente Bergen op Zoom van juni 2021 is te zien dat er wordt samengewerkt met andere gemeentes in West-Brabant, begrotingen worden gemaakt en jaarrekeningen met elkaar worden uitgewisseld (Gemeente Bergen op Zoom, 2021). Hetzelfde geldt

voor de raadstukken van gemeente Roosendaal (Gemeente Roosendaal, 2020) en Breda (Gemeente Breda, 2021) waarin ook staat dat er wordt samengewerkt met de gemeentes in West-Brabant.

Gemeentes	Websites	Jaarrekeningen	Accountantsverslagen	Raadstukken
Bergen op Zoom	✓	X	X	✓
Roosendaal	✓	✓	X	✓
Breda	X	✓	X	✓

- Inde bovenstaande tabel is te zien dat elk gemeente bij minimaal 2 van de 4 bronnen wat weergegeven wat betreft het samenwerken en delen van bestanden met elkaar. Dit komt overeen met de eerste hypothese: *Gemeentes wisselen bestanden uit aan elkaar zonder dat persoonsgegevens meegenomen worden zodat logaritmen sterker gemaakt kunnen worden waardoor meer fraudes, sociale fraude en fraude door bedrijven, opgespoord kunnen worden.* De hypothese kan op basis hiervan met zekere mate als geldig gezien worden.

Tweede hypothese

Bij alle drie de gemeentes kan geconstateerd worden dat fraude gemeld kan worden en meer focus wordt gelegd op het risico gestuurd adres onderzoek. Wanneer er wordt gekeken naar de websites van de gemeentes Bergen op Zoom, Roosendaal en Breda op het trefwoord *fraude* is te zien dat een mogelijkheid geboden wordt aan de burgers bijstandsfraude te melden. Bij alle drie de gemeentes is de webpagina op dezelfde manier opgesteld en kunnen de burgers eenvoudig en overzichtelijke fraudes melden. Dit kan gedaan worden door het opgeven van duidelijke feiten die gebruikt worden als signaal dat kan leiden tot een onderzoek als gevolg (zie figuur 8 met als voorbeeld Bergen op Zoom). Het eventuele onderzoek kan leiden tot een uitkomst wat weer gebruikt kan worden als bewijs. Met dit bewijs kunnen maatregelen genomen worden. Met het melden van fraude moeten de volgende gegevens verzameld worden door de melder: naam, adres, vervoersmiddel, kenteken, adres, werktijden, inkomsten en samenwoonadres. Ook kan bijstandsfraude gemeld worden in de vorm van dat mensen gescheiden zijn en dat niet hebben gemeld, zwart werken, niet daadwerkelijk wonen op een adres en niet opgeven van vermogen (binnen- en/of buitenland) (Gemeente Bergen op Zoom).

Home > Melden aan de gemeente > Bijstandsfraude melden

Bijstandsfraude melden

Bijstandsfraude is het opzettelijk geven van verkeerde informatie of het verzwaren van informatie om onrecht een (hogere) uitkering te krijgen.

[Bijstandsfraude melden](#)

Als u melding maakt van een vermoeden van bijstandsfraude, is het belangrijk dat u concrete, duidelijke feiten meldt. Uw tip wordt beschouwd als een signaal wat een onderzoek tot gevolg kan hebben. De uitkomst van het eventuele onderzoek vormt dan het bewijs en kan leiden tot maatregelen. Het is daarom belangrijk dat u de volgende zaken meldt.

Gegevens van de persoon waarover u melding maakt:

- * Naam;
- * Adres;
- * Vervoermiddel (auto, fiets, brommer, openbaar vervoer, etc.);
- * Eventueel kenteken;
- * Werkadres (indien van toepassing);
- * Werktijden;
- * Inkomsten;
- * Samenwoonadres (indien van toepassing).

Hoe kunt u fraude melden? +

Voorbeelden bijstandsfraude -

- * Melden dat men gescheiden is gaan wonen, maar dat niet doet;
- * Niet opgeven van (extra) inkomsten uit bijvoorbeeld zwart werk;
- * Inschrijven op een adres, maar daar niet daadwerkelijk wonen;
- * Melden dat men alleen woont, maar in feite samenwoont;
- * Niet opgeven van vermogen (in binnenland en/of buitenland).

Figuur 8 Melden van fraude bij gemeentelijke organisaties

Daarnaast kost het veel werk voor gemeentelijke organisaties om te kijken of daadwerkelijk sprake is van fraude. Door het laten melden van fraude wordt het bewijzen van fraude vergemakkelijkt en kan er risico gestuurd te werk gaan. Door de binnengekomen fraudemeldingen kunnen gemeentes adressen beter op orde brengen en analyses uitvoeren op grotere hoeveelheid gegevens. Bergen op Zoom zoals weergegeven in de raadsstukken van juni 2021 grote bedragen aan fraudebestrijding om misbruik van sociale voorzieningen tegen te gaan. In de aankomende jaren zal Bergen op Zoom geld blijven uitgeven om deze vorm van fraude te beperken (Geneente Bergen op Zoom , 2021).

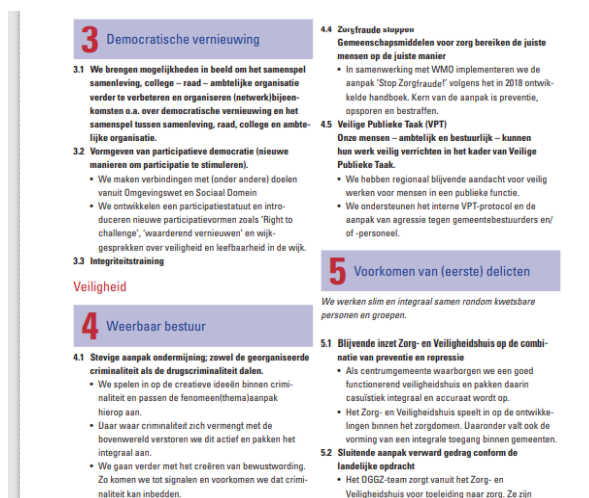
Gemeentes	Websites	Jaarrekeningen	Accountantsverslagen	Raadsstukken
Bergen op Zoom	✓	X	X	✓
Roosendaal	✓	X	X	X
Breda	✓	X	X	X

- In de bovenstaande tabel is af te lezen dat bij alle drie de gemeentes het fraude melden mogelijk is gemaakt op de websites en gesteld kan worden dat de andere bronnen niet relevant hoeven te zijn voor de tweede hypothese. De tweede hypothese: *Door risico gestuurd adresonderzoek kunnen gemeenten een hoger percentage onjuiste adressen traceren, waardoor het gemeentelijk beleidsproces geoptimaliseerd wordt*, kan met zekere mate als geldig gezien worden.

Derde hypothese

Bij alle drie de gemeentes kan geconstateerd worden dat fraude een steeds belangrijker onderdeel van de bedrijfsvoering wordt en fraudes eenvoudiger gedetecteerd en opgespoord kunnen worden. Zo gebruikt Bergen op Zoom zoals weergegeven in de jaarrekening bij het zoeken op het trefwoord

fraude onder het kopje *bestuur en maatregelen* middelen om gezamenlijk gebruik te gaan maken van databestanden en analyses te voeren op grotere hoeveelheid data. Op die manieren willen ze er bijvoorbeeld voor zorgen dat de gemeenschapsmiddelen voor zog de juiste mensen bereiken op de juiste manier. Hiervoor wordt er samengewerkt met WMO (Wet maatschappelijke ondersteuning) en wordt de aanpak ‘Stop Zorgfraude!’ (zie figuur 9) geïmplementeerd. De kern van deze aanpak is het preventie, opsporen en bestraffen van deze vorm van fraude. Om fraude op te sporen hebben toezichthouders, opsporingsdiensten, gemeenten en uitvoeringsorganisaties hun krachten gebundeld. Hierbij komen alle signalen over fraude en misbruik, zowel van meldingen van burgers als van aanbieders, bij het informatieknooppunt terecht. Op die manier worden gegevens uitgewisseld bij uitvoering, preventie en bestrijden van vormen van fraude (Gemeente Bergen op Zoom, 2019).



Figuur 9 Jaartukken 2020 wat betreft het stoppen van fraude

In de jaarrekening van Breda bij het trefwoord *projectgroepen* onder het kopje *duurzaam wonen in Breda* is te zien dat de gemeentes, corporaties en zorgaanbieder al jarenlang werken in verschillende vormen van samenwerkingsverbanden. Er zijn verschillende werkgroepen en projectgroepen aanwezig voor verschillende thema's die het wonen en zorg raken. Hier zal fraude ook een van de besproken thema's zijn oor de verschillende thema's die het wonen en zorg raken, zijn diverse werkgroepen en projectgroepen actief. De ambities op alle niveaus zullen verboden worden russen de gemeente, corporatie en de zorgaanbieders (Gemeente Breda, 2019).

Ook vormt IT zoals weergegeven in de accountantsverslagen van de drie gemeentes op het trefwoord *fraude* een steeds belangrijker onderdeel van de bedrijfsvoering en daarmee de jaarrekeningcontrole. Door middel van IT zijn er steeds meer mogelijkheden en wordt het analyseren en detecteren gemakkelijker gemaakt door de grotere hoeveelheid real-time beschikbare informatie. Bij het voeren

van de controle van de jaarrekening wordt door de gemeentes gebruik gemaakt van een mix van systeem-en gegevensgerichte werkzaamheden en een mix van vaste teamleden en ingeschakelde experts. Door samen te werken, de maatregel gezamenlijk gebruik, kunnen de verschillende partijen focussen op de posten waar zij belang en kennis over hebben. Dit leidt tot een kwalitatief goede controle waarmee gericht wordt op het in kaart kunnen brengen van de voor de jaarrekening van belang zijnde risico's (zie figuur 10). Door het gebruik van externe deskundigen en verbonden partijen kan achterhaald worden of de jaarrekening volledig is en er vormen van fraude vastgesteld kunnen worden (Gemeente Breda, 2020). Daarnaast worden frauderisicoanalyses (zie figuur 12) uitgevoerd om ervoor te zorgen dat de jaarrekening geen afwijkingen bevat als gevolg van fraude en fouten. Het is mogelijk dat niet alle fraudes ontdekt zijn door de hoge maar geen absolute mate van zekerheid. Hierbij wordt gefocust op het valse informatie verstrekken in geschrifte, het opzettelijk verkeerde voorstellen van zaken en het opzettelijk nalaten van transacties (Gemeente Roosendaal, 2020).



Figuur 10 Controle aanpak jaarrekeningen

Gemeentes	Websites	Jaarrekeningen	Accountantsverslagen	Raadstukken
Bergen op Zoom	X	✓	✓	X
Roosendaal	X	✓	✓	X
Breda	X	✓	✓	X

- In de bovenstaande tabel is te zien dat bij alle drie de gemeentes relevante informatie vermeld staat in de jaarrekeningen en accountantsverslagen. Dit komt overeen met de derde hypothese: *Gemeentes werken in projectengroepen of werkgroepen om specifieke risicoanalyses uit te voeren met gezamenlijk gebruik van databestanden voor fraudedetectie*

en vanuit dat punt analyses van grote hoeveelheden data. De derde hypothese kan met zekere mate als geldig gezien worden.

7. Conclusie

Door middel van Big Data kan meer fraude worden aangepakt door gemeentes en hun beleid aangepast worden naar een betere en bevorderde versie. De belangrijkste reden hiervoor is dat gemeentelijke organisaties meer data tot hun beschikking zullen hebben. De onderzoeksvraag die hierbij centraal stond is: wat voor invloed heeft Big Data op het beleid voor externe fraudedetectie bij gemeentelijke organisaties? Alle drie de hypothesen kunnen met zekere mate als geldig gezien worden aan de hand van de informatie die door middel van grootste deel van de deskresearch (websites, jaarrekeningen en accountantsverslagen) gevonden. Dit kan worden teruggekoppeld naar de eerder besproken maatregelen die gemeentes kunnen nemen met behulp van Big Data om het aantal fraudegevallen te laten afnemen.

Doordat Big Data steeds meer mogelijkheden biedt om fraudeurs in beeld te brengen en mogelijkheden bij het opsporen ervan gebruiken gemeentes wel degelijk de maatregelen. Dit voorkomt het voorkomen van fraudes waardoor minder misbruik gemaakt wordt van het vertrouwen dat gemeentes biedt aan hun burgers. Daarentegen wordt bij de gemeenteraad weinig raadsvoorstellen gegeven over de voorkomende fraudes en de maatregelen die genomen worden om deze fraudes te beperken. Er wordt alleen gesproken over de samenwerking tussen de verschillende gemeentes. Concluderend heeft Big Data wel degelijk invloed op het beleid voor externe fraudedetectie door middel van de genoemde maatregelen die genomen kunnen worden hiervoor.

Dit onderzoek bevat alleen informatie door middel van het bekijken van informatie dat gepubliceerd is door gemeentes en kan niet teruggekoppeld worden naar wat daadwerkelijk in de praktijk gebeurt. Er kan op basis van dit onderzoek niet geconcludeerd worden of de informatie die gemeentes publiceren over fraudedetectie ook daadwerkelijk in de praktijk wordt toegepast. Ook is gekeken naar drie gemeentes in West-Brabant waardoor de conclusie wellicht anders geweest kan zijn wanneer er gekeken zou zijn naar drie gemeentes in het noorden van het land. Specifieke aanbevelingen voor vervolgonderzoek zouden zijn het gebruiken van grotere sample van gemeentes, verdeeld over heel Nederland. Dit zou een beter beeld moeten geven van de verschillende soorten maatregelen die genomen worden. Daarnaast zou dit onderzoek zich moeten richten op de praktijk door middel van het afnemen van interviews of enquêtes om te onderzoeken of de gevonden maatregelen in de bronnen overeenkomen met wat in de praktijk gedaan wordt.

Bibliografie

- Bakermans, C. (2015, Mei). *Big Data een modekreet of the next best thing?* Opgehaald van deloitte: <https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/legal/deloitte-nl-big-data-een-modekreet-of-the-next-best-thing.pdf>
- Bongers, F., Veldkamp, J., & van de Vorst, T. (2015, februari 11). *Big Data: een zoektocht voor beleid*. Opgehaald van pixelspark: <https://pixelspark.nl/wp-content/uploads/2015/04/Big-data.pdf>
- Brans, M., & Craeghs, J. (2000). *Praktisch handboek voor gemeentebestuur*. Brugge: Die Keure.
- Leeuw, F. L. (2019, juni). *Evaluatieonderzoek, Big Data en Artificiële Intelligentie: een verkenning*. Opgehaald van beleidsonderzoekonline: http://www.beleidsonderzoekonline.nl/tijdschrift/bso/2019/6/BO_2213-3550_2019_000_006_001.pdf
- Olsthoorn, P. (2016). *Big Data voor Fraudebestrijding*. Opgehaald van fortoranjenieuws: <https://fortoranjenieuws.nl/wp-content/uploads/2019/10/72887-big-data-voor-fraudebestrijding.pdf>
- Van der Voort, H., & Cromptvoets, J. (2016). *Big data: een zoektocht naar instituties*. Opgehaald van researchgate: https://www.researchgate.net/profile/Haiko-Voort/publication/297726611_Big_data_een_zoektocht_naar_instituties/links/57dc01a908ae5292a379a481/Big-data-een-zoektocht-naar-instituties.pdf
- Alexa Crawls. (2007, Februari 14). *Fraudebestrijding: algemene informatie*. Opgehaald van SZW: https://web.archive.org/web/20070214041026/http://home.szw.nl/navigatie/dossier/dsp_dossier.cfm?set_id=982&link_id=23148#1485120
- Asadi Someh, I., Breidbach, C., Davern, M., & Shanks, G. (2016). Ethical implications of big data analytics. *Twenty-Fourth European Conference on Information Systems*.
- Coyne, E., Coyne, J., & Walker, K. (2018). Big Data information governance by accountants. *International Journal of Accounting and Information Management*, pp. 153-170. doi:<http://dx.doi.org.eur.idm.oclc.org/10.1108/IJAIM-01-2017-0006>.
- de Groot, A., van Vliet, J., Vriend, P., & van Vugt, R.-A. (2007, Juni). *Gemeente Governance Fraude*. Opgehaald van vng: https://vng.nl/files/vng/vng/Documenten/Extranet/Marz/CVA/Integriteit/2007handboek_gemeente_governance_fraude.pdf
- Falsafi, B., & Grot, B. (2014, July-Aug). Big Data [Guest editors' introduction]. *IEEE Micro*, pp. 4-5. doi:10.1109/MM.2014.65.
- Fenger, M., Chin-A-Fat, N., Frankowski, A., & van der Torre, L. (2016). *Naar rechtmatige zorg in het gemeentelijke sociale domein*. Faculteit der Sociale Wetenschappen, Erasmus University Rotterdam.
- Gemeente Bergen op Zoom. (2019, september 17). *Begroting 2020 Meerjarenraming 2021-2023*. Opgehaald van findo: https://findo.nl/jaarstukken/Begroting_2020_GEMEENTE20_0748.pdf
- Gemeente Bergen op Zoom. (sd). *Bijstandsfraude melden*. Opgehaald van bergenopzoom: <https://www.bergenopzoom.nl/bijstandsfraude-melden>
- Gemeente Breda. (2019). *Jaarstukken 2019*. Opgehaald van Breda jaarverslag 2019: <https://breda.jaarverslag-2019.nl/assets/docs/boekwerk%20jaarstukken%202019.pdf>

- Gemeente Breda. (2020, juni 3). *accountantsverslag*. Opgehaald van accountantsrapport: [file:///C:/Users/Gebruiker/Downloads/4.1.B%20Accountantsrapport%20\(3\).pdf](file:///C:/Users/Gebruiker/Downloads/4.1.B%20Accountantsrapport%20(3).pdf)
- Gemeente Breda. (2021, juni 18). *Gemeente Breda*. Opgehaald van breda.raadsinformatie: https://breda.raadsinformatie.nl/document/10296015/1/Raadsbrief_18062021_RES%20Wes%20Brabant%20incl%20bijlage
- Gemeente Roosendaal. (2019). *Jaarstukken 2019*. Opgehaald van raad Roosendaal: <https://raad.roosendaal.nl/Vergaderingen/Commissie/2020/02-juli/19:30/Bijlage-Jaarstukken-Roosendaal-2019.pdf>
- Gemeente Roosendaal. (2020, juni 30). *Raadsmededeling*. Opgehaald van raad.roosendaal: <https://raad.roosendaal.nl/documenten/Raadsmededeling/39-2020-Accountantsverslag-2019-en-reactie-van-het-college-1.pdf>
- Gemeente Roosendaal. (2020, juni 25). *Samenvatting besluiten en toezeggingen raadsvergadering*. Opgehaald van raad.roosendaal: <https://raad.roosendaal.nl/documenten/Besluitenlijst/03-Samenvatting-besluiten-en-toezeggingen-Raad-25-juni-2020.pdf>
- Geneente Bergen op Zoom . (2021, juni 24). *Raadsvergadering*. Opgehaald van ris2.ibabs.eu: <https://ris2.ibabs.eu/Agenda/Details/BergenOpZoom/49ae7665-5a44-43df-835c-0bf5d69ddea6>
- Gordon, G., & Willox, Jr., N. (2003, october 28). *Identity Fraud: A Critical National and Global Threat*. Opgehaald van veracity lexis-nexis: <http://veracity.lexis-nexis.com/presscenter/hottopics/ECIRReportFINAL.pdf>
- Grijpink, J. (2004, Januari 1). Identity fraud as a challenge to the constitutional state1. *Computer Law & Security Review*, pp. 29-36. doi:10.1016/S0267-3649(04)00006-8
- Katal, A., Wazid, M., & Goudar, R. (2013, augustus). Big Data: Issues, Challenges, Tools and Good Practices. *Sixth international conference on contemporary computing IEEE*, pp. 404-409.
- Majster, D. (2013, november 5). *de vier V's van Big Data*. Opgehaald van intersystems: <https://www.intersystems.com/impuls-blog/de-vier-v-s-van-big-data/#:~:text=Om%20meer%20inzicht%20te%20krijgen,%2C%20Velocity%2C%20Variety%20en%20Veracity.>
- Peters, L., & Cheyney, M. (2018). *wat is "Big Data"?* Opgehaald van research vumc: https://research.vumc.nl/ws/files/8832229/Peters_Cheyney_VvE_epistel_2018_Wat_is_Big_Data.pdf
- ProDemos. (sd). *Wat doet de gemeente?* Opgehaald van ProDemos: <https://prodemos.nl/kennis-en-debat/publicaties/informatie-over-politiek/de-gemeente/wat-doet-de-gemeente/>
- Russom, P. (2011). *Big Data Analytics*. Opgehaald van vivomente: <https://vivomente.com/wp-content/uploads/2016/04/big-data-analytics-white-paper.pdf>
- Scott D. Applegate , M. (2009). Social Engineering: Hacking the Wetware. *Information Security Journal: A Global Perspective*, 40-46. doi:10.1080/19393550802623214
- Shu, H. (2016). Big Data analytics: six techniques. *Geo-spatial Information Science*, pp. 119-128. doi: 10.1080/10095020.2016.1182307
- Sociaalweb. (sd). *Bijstandsfraude voorspellen met big data*. Opgehaald van magazine.sociaalweb: <http://magazines.sociaalweb.nl/fraude#!/magazine-fraude>

- Stel, P. (2012, oktober). *Wat betekent 'Big Data' eigenlijk?* Opgehaald van executivefinance: <https://executivefinance.nl/wp-content/uploads/2015/02/FC201205071.pdf>
- Tole, A. (2013). Big Data challenges. *Database Systems Journal*, 31-40.
- van Geldrop, A., & de Vries, T. (2015). *Fraude Loont... nog steeds*. Den Haag: Stichting Toekomstbeeld der Techniek. Opgehaald van file:///C:/Users/Gebruiker/Downloads/Fraude-loont...nog-steeds%20(1).pdf
- Wieringa, J. (2015). *Ontwikkelingen in het marktonderzoek*. Opgehaald van researchgate: https://www.researchgate.net/profile/Paul-Dekker-3/publication/273195765_Ontwikkelingen_in_het_marktonderzoek_2015/links/55227e380cf29dcabb0d6b32/Ontwikkelingen-in-het-marktonderzoek-2015.pdf#page=9