

Big data en audit kwaliteit

Abstract

In deze thesis is onderzocht of het gebruik van big data en big data analytics de kwaliteit van een jaarrekeningcontrole kan verhogen. Dit is onderzocht door te kijken naar wat big data en big data analytics zijn en te kijken hoe een jaarrekeningcontrole wordt verricht zonder het gebruik van big data. Hierna is audit kwaliteit gedefinieerd als het niveau van assurance dat wordt bereikt en is gekeken naar de wijze waarop big data en big data analytics kunnen worden geïncorporeerd in de jaarrekeningcontrole. Big data en big data analytics kunnen helpen bij verkrijgen van audit bewijs, het voorspellen van verschillende gebeurtenissen en het identificeren van fraude. In deze thesis wordt dan ook geconcludeerd dat het gebruik van big data en big data analytics de audit kwaliteit kan verhogen, omdat het verkrijgen van audit bewijs, het voorspellen van verschillende gebeurtenissen en het identificeren van fraude zorgt voor een hogere mate van assurance.

Student: Jan van der Linden
Studentnummer: 434749
Inleverdatum: 30 augustus 2018
Universiteit: Erasmus Universiteit Rotterdam
Faculteit: Erasmus School of Economics
Begeleider: R. van der Wal

Inhoudsopgave

Hoofdstuk 1: Introductie	4
1.1 Inleiding	4
1.2 Probleemstelling	7
1.3 Relevantie	8
1.4 Methodologie.....	9
1.5 Structuur.....	9
Hoofdstuk 2: Big data en big data analytics	10
2.1 Inleiding	10
2.2 Big data	10
2.3 Big data analytics.....	13
2.4 Data mining	14
2.5 Artificial intelligence	17
2.6 Samenvatting	18
Hoofdstuk 3: De jaarrekeningcontrole	
3.1 Inleiding	19
3.2 Fase 1: Het plannen van de jaarrekeningcontrole en het identificeren van de risico's	20
3.3 Fase 2: Het bepalen van de strategie en het beoordelen van de risico's	20
3.4 Fase 3: Het uitvoeren van de controles	21
3.5 Fase 4: Concluderen en rapporteren	22
3.6 Samenvatting	22
Hoofdstuk 4: De kwaliteit van een jaarrekeningcontrole	23
4.1 Inleiding	23
4.2 Het niveau van assurance	23
4.3 De mate van naleving van standaarden	24
4.4 Verwachtingskloof	25
4.5 Kwaliteit volgens toezichthouders	25

4.6 Samenvatting	27
Hoofdstuk 5: Incorporatie van big data en big data analytics in de jaarrekeningcontrole....	28
5.1 Inleiding	28
5.2 Big data als audit bewijs	28
5.3 Voorspellingen op basis van big data	29
5.4 Outliers en fraudedetectie	30
5.5 Big data en de kwaliteit van audit	30
5.6 Samenvatting	31
Hoofdstuk 6: Conclusie en discussie	32
6.1 Conclusie	32
6.2 Discussie en vervolgonderzoek	33
Literatuurlijst	35

Hoofdstuk 1: Introductie

1.1 Inleiding

Informatie is al jarenlang cruciaal voor verschillende bedrijven. De informatierevolutie die vanaf het begin van de jaren '80 in gang is gezet, zorgt niet alleen voor competitieve voordelen voor bedrijven, maar verandert de structuur van een gehele industrie (Porter & Millar, 1985). De opgeslagen informatie groeit jaarlijks exponentieel met 23% (Hilbert & López, 2011). Deze extra opgeslagen data kan op vele manieren worden gebruikt door een bedrijf. De data dient echter nog wel geanalyseerd te worden. Door middel van 'data analytics' technieken zoals data mining kunnen bedrijven ontdekken wat voor gedrag hun consumenten vertonen, hierdoor kunnen effectieve marketing- en verkoopstrategieën worden geformuleerd (Berry & Linoff, 1997). De steeds maar groter wordende hoeveelheid data en formats van deze data, heeft ertoe geleid dat traditionele software deze data niet meer kan verwerken; om deze grote hoeveelheid data te analyseren zijn nieuwe software- en andere analysemethoden nodig (Sikos, 2015). Dit nieuwe fenomeen big data, is al succesvol gebruikt in de astronomie, detailhandel, politiek en bij zoekmachines (Murdoch & Detsky, 2013). Big data analytics wordt al op veel gebieden gebruikt, maar bij auditors is daar echter nog geen of bijna geen sprake van (Cao et al., 2015). Deze thesis zal zich dan ook voornamelijk richten op het gebruik van big data analytics door auditors.

Er zijn verschillende vormen van audit, allereerst dient er onderscheid te worden gemaakt tussen een interne en externe audit. Interne audits hebben als doel om waarde te creëren voor bedrijven door de effectiviteit en efficiëntie binnen het bedrijf te verhogen. Er zijn verschillende vormen van audits, namelijk: financial audits, information system audits, operational audits, compliance audits en investegative audits. Externe auditors geven een onafhankelijke beoordeling met betrekking tot de juistheid en volledigheid van de financiële verslagen. Hoewel er een duidelijk verschil zit tussen het werk van een interne auditor en externe auditor, komen veel concepten en technieken overeen (Romney & Steinbart, 2018). Hierdoor kunnen de resultaten van dit onderzoek voor beide partijen relevant zijn. Deze thesis richt zich echter enkel op financial audits door een externe partij, ook wel jaarrekeningcontroles genoemd.

Financial auditing is het proces van verifiëren van financiële verslagen van een bedrijf of een onderdeel daarvan, om zekerheid te kunnen bieden met betrekking tot de juistheid en volledigheid van het verslag. De juistheid van het verslag bestaat enerzijds uit de juiste financiële informatie in de verslagen en anderzijds uit een juiste toepassing van de accounting standaarden zoals International Financial Reporting Standards of United States Generally Accepted Accounting Principles.

Op basis van artikel 2:393 BW dient een rechtspersoon die een jaarverslag moet opstellen, deze te laten controleren door een registeraccountant. Niet elke rechtspersoon dient echter een jaarrekening op te stellen. Er wordt onderscheidt gemaakt tussen vier groottes van bedrijven. Om in een bepaalde categorie te komen dient een rechtspersoon twee opeenvolgende balansdata te voldoen aan twee van de drie onderstaande vereisten (artikelen 2:395a, 396 lid 1, 397 lid 1 BW):

Micro:

1. De waarde op de balans, op grond van de verkrijgings- en vervaardigingsprijs bedraagt maximaal €350.000.
2. De netto omzet bedraagt maximaal €700.000.
3. Het gemiddelde aantal werknemers over het boekjaar bedraagt maximaal 10.

Klein:

1. De waarde op de balans, op grond van de verkrijgings- en vervaardigingsprijs bedraagt minimaal €350.000.
2. De netto omzet bedraagt minimaal €700.000.
3. Het gemiddelde aantal werknemers over het boekjaar bedraagt minimaal 10.

Middel:

1. De waarde op de balans, op grond van de verkrijgings- en vervaardigingsprijs bedraagt minimaal €6.000.000.
2. De netto omzet bedraagt minimaal €12.000.000.
3. Het gemiddelde aantal werknemers over het boekjaar bedraagt minimaal 50.

Groot:

1. De waarde op de balans, op grond van de verkrijgings- en vervaardigingsprijs bedraagt minimaal €20.000.000.
2. De netto omzet bedraagt minimaal €40.000.000.
3. Het gemiddelde aantal werknemers over het boekjaar bedraagt minimaal 250.

Op basis van artikel 2:396 lid 1 BW dient een rechtspersoon een jaarrekening op te stellen indien het om een rechtspersoon gaat dat in de categorie middel of groot zit.

De Nederlandse wet- en regelgeving is sterk gebaseerd op de vervaardigde richtlijnen van de Europese Unie (EU). De jaarrekeningcontrole is ook verplicht voor bedrijven gevestigd in de EU (richtlijn 2006/43/EG). Dit geldt net als bij de Nederlandse wetgeving enkel voor bedrijven die aan twee van de drie de hierboven genoemde vereisten voldoen (Richtlijn 2013/34/EU). In andere landen kan dit beleid echter sterk van het Nederlandse beleid afwijken. In de Verenigde Staten zijn er met betrekking tot de regelgeving omtrent audits sterke verschillen tussen beursgenoteerde en private ondernemingen (Engel et al., 2007). In de Verenigde Staten mag net als in Nederland, enkel een Certified Public Accountant (Amerikaanse benaming van een registeraccountant) een accountantsverklaring afgeven (Mayhew & Wilkins, 2003).

De jaarrekeningcontrole is een wettelijke verplichting. Het doen van een jaarrekeningcontrole dient een aantal doelen. Het primaire doel van een jaarrekeningcontrole is het vaststellen of de financiële verslagen juist en volledig zijn en in overeenstemming zijn met de algemene aanvaarde boekhoudprincipes (Elliott & Rogers, 1972). Het doel van de audit kan worden doorgetrokken naar het verbeteren van de betrouwbaarheid van informatie die gebruikt wordt voor investeringen (Elliott & Jacobson, 1998). Investeerders maken namelijk gebruik van de verstrekte informatie. Om investeerders vertrouwen te geven in de juistheid en volledigheid van de cijfers, dient er een audit te worden gedaan door een onafhankelijke accountant (Burke & Lee, 2015). Het doen van een onafhankelijke audit leidt tot kostenefficiëntie op vermogensmarkten (Elliott & Jacobsen, 1998). Het doen van een jaarrekeningcontrole is dus van groot belang voor investeerders. Niet enkel de grote financiële instellingen, maar ook 'kleinere' privé investeerders hebben baat bij een audit (ACCA, 2011). De voordelen van een audit beperken zich echter niet alleen tot investeerders, ook het bedrijf

dat wordt geaudit ervaart de audit vaak als toegevoegde waarde. Uit onderzoek blijkt dat het merendeel van besturen van ondernemingen ook een audit zou doen indien dit niet verplicht is, omdat het de kwaliteit van informatie verhoogd en interne documenten worden gecontroleerd (Collis et al., 2004). Ook de maatschappij als geheel heeft belang bij kwalitatief goede, onafhankelijke audits. Het Enron-schandaal is daar een voorbeeld van. Enron werd al vele jaren geaudit door hetzelfde bedrijf, Arthur Anderson. De relatie is door de jaren heen steeds informeler geworden, dit heeft de onafhankelijkheid van de auditor aangetast (Herrick & Barrioneuvo, 2002). De gevolgen van deze slechte audits bij Enron werden niet beperkt tot financiële schade bij investeerders, maar ook 5600 werknemers werden ontslagen (Ferrell & Ferrell, 2011; Cohen, 2006). Daarnaast zijn er miljarden dollars aan spaargeld en pensioenen verloren gegaan (Ferrell & Ferrell, 2011). Indien de kwaliteit van een audit niet voldoende is, kan dus de gehele samenleving daar de gevolgen van ondervinden.

1.2 Probleemstelling

De kans dat een auditor een onjuistheid in het jaarverslag vindt, is afhankelijk van de technologische mogelijkheden die de auditor heeft (DeAngelo, 1981). Het is daarom van belang dat de auditor de best mogelijke technologieën kan gebruiken bij de audit, het vinden van onjuistheden is namelijk één van de primaire doelen van een auditor (Richtlijn 2013/34/EU). Eén van de nieuwe technologieën die een auditor zou kunnen gebruiken is big data en de bijbehorende big data analytics. Uit eerder onderzoek blijkt dat auditors dat nog niet, of bijna niet doen (Cao et al., 2015). In deze thesis zal daarom de volgende hoofdvraag worden beantwoord:

Kan het gebruik van big data en big data analytics de kwaliteit van een jaarrekeningcontrole verbeteren?

Om deze hoofdvraag te beantwoorden zijn de volgende deelvragen opgesteld:

1. Wat zijn big data en big data analytics?
2. Hoe ziet een 'traditionele' jaarrekeningcontrole eruit zonder het gebruik van big data en big data analytics?
3. Wat is de definitie van kwaliteit met betrekking tot een jaarrekeningcontrole?

4. Op welke wijze kunnen big data en big data analytics worden geïncorporeerd in de jaarrekeningcontrole?

1.3 Relevantie

Deze thesis beschrijft op welke wijze big data en big data analytics gebruikt kunnen worden in een jaarrekeningcontrole, eerdere onderzoeken (Hoogduin et al., 2015; Thiprungsri & Vaserhelyi, 2011; Kogan et al., 2015; Alles, 2015) hebben zich ook al met deze vraag bezig gehouden. Het is echter nog onduidelijk in hoeverre big data en big data analytics de kwaliteit van een audit kunnen verbeteren. Deze thesis is dan ook wetenschappelijk relevant, omdat het bijdraagt aan de huidige situatie van wetenschappelijk onderzoek naar het gebruik van data analytics in auditing, en dan specifiek op het gebruik van big data in auditing.

Deze thesis beperkt zich niet enkel tot de wetenschappelijke relevantie, maar heeft ook een praktische relevantie. Deze thesis zal duidelijk maken aan auditors en accountantskantoren of het gebruik van big data en big data analytics de kwaliteit van een jaarrekeningcontrole zal verbeteren. De beantwoording van deze vraag is relevant, want accountantskantoren concurreren onderling, en proberen een jaarrekeningcontrole van hoge kwaliteit af te leveren aan hun klanten. Deze thesis is ook relevant voor bedrijven waarbij een jaarrekeningcontrole dient te worden gedaan. Bij het beoordelen welk accountantskantoor de jaarrekeningcontrole met de hoogste kwaliteit kan uitvoeren, kan het relevant zijn of er wel of geen gebruik wordt gemaakt van big data analytics.

Daarnaast zijn er recent rapporten verschenen van verschillende toezichthouders dat de accountantscontrole momenteel kwalitatief onvoldoende is. De Autoriteit Financiële Markten (AFM) geeft aan na een onderzoek bij acht grote accountantskantoren dat de kwaliteit van accountantscontrole onvoldoende is (AFM, 2017). De Monitoring Commissie Accountancy (MCA) bracht recent een rapport naar buiten waarin werd aangegeven dat er verbetering is met betrekking tot de accountantscontrole, maar dat de accountantscontrole structureel kwalitatief onvoldoende is (MCA, 2018). De toezichthouders geven aan dat de kwaliteit van de accountantscontrole dient te worden verbeterd. In deze thesis zal worden gekeken of big data en big data analytics daar een rol bij kunnen spelen.

1.4 Methodologie

De hoofdvraag in deze thesis zal worden beantwoord aan de hand van een literatuuronderzoek. Waarbij de hoofdvraag wordt beantwoord aan de hand van verschillende deelvragen. Bij elke deelvraag zal relevante literatuur worden gezocht en zal de deelvraag worden beantwoord aan de hand van deze literatuur.

1.5 Structuur

In het tweede hoofdstuk van deze thesis zal de eerste deelvraag worden behandeld. Deze deelvraag luidt: Wat zijn big data en big data analytics? In dit hoofdstuk zal worden gekeken welke kenmerken big data precies bezit en wat er wordt verstaan onder big data analytics. Ook zullen enkele big data analytics methodes worden behandeld. Het derde hoofdstuk zal antwoorden geven op de tweede deelvraag. Deze deelvraag luidt: Hoe ziet een ‘traditionele’ jaarrekeningcontrole eruit zonder het gebruik van big data en big data analytics? Hierbij zal de jaarrekeningcontrole in verschillende fases worden uitgesplitst om zo een overzichtelijker beeld te krijgen van de jaarrekeningcontrole. In het vierde hoofdstuk zal antwoord worden gegeven op de derde deelvraag. Deze deelvraag luidt: Wat is de definitie van kwaliteit met betrekking tot een jaarrekeningcontrole? Om deze vraag te beantwoorden zal worden gekeken naar welke stromingen er bestaan met betrekking tot audit kwaliteit en zal worden gekozen welke definitie binnen deze thesis gebruikt zal worden. Het vijfde hoofdstuk geeft antwoord op de vierde deelvraag. Deze deelvraag luidt: Op welke wijze kunnen big data en big data analytics worden geïncorporeerd in de jaarrekeningcontrole? Hierbij zal worden gekeken naar relevante wetenschappelijke literatuur. Daarnaast zal in dit hoofdstuk de link worden gelegd tussen audit kwaliteit en de toepassing van big data en big data analytics in de jaarrekeningcontrole. In het zesde hoofdstuk zal een conclusie worden getrokken met betrekking tot de hoofdvraag en zullen na de discussie aanbevelingen worden gegeven voor vervolgonderzoek.

Hoofdstuk 2: Big data en big data analytics

2.1 Inleiding

Dat het gebruik van data van belang is binnen verschillende organisaties, is al jaren bekend. De afgelopen jaren zijn er door verschillende nieuwe technologieën, zoals: the internet of things en cloud computing verschillende nieuwe mogelijkheden om data te genereren en data op te slaan (Zhou et al., 2016). Hierdoor wordt er meer en meer data opgeslagen die gebruikt kan worden om waarde voor een organisatie te creëren (Villars et al., 2011). In dit hoofdstuk zal worden gekeken naar wat big data en big data analytics zijn.

2.2 Big data

Big data is de term die wordt gebruikt om data aan te duiden die niet door traditionele databases kan worden opgeslagen en geanalyseerd (Hashem et al., 2015). Big data wordt gekenmerkt door een aantal verschillende aspecten, deze worden ook wel de 3 V's genoemd die zijn geïntroduceerd door IT onderzoeks- en adviesbureau Gartner (Gandomi & Haider, 2015; Hashem et al., 2015). Dit zijn:

- **Volume:** Dit refereert aan de hoeveelheid data (Hashem et al., 2015). Bij big data gaat het om datasets met een grote hoeveelheid data. Er is echter geen grens waarbij een bepaalde dataset groot genoeg is om als big data gedefinieerd te kunnen worden, dit maakt dat volume relatief moet worden bekeken. Uit een onderzoek komt wel naar voren dat ongeveer de helft van alle respondenten datasets met een grote van één terabyte zien als big data (Schroek et al., 2012). Wat nu als big data wordt gezien, zal over een aantal jaar mogelijk niet meer zijn dan een normale dataset. Dat komt door de steeds maar groeiende hoeveelheid data, waardoor de definitie van een grote dataset steeds groter zal worden (Gandomi & Haider, 2015).
- **Variety:** Dit refereert aan de verscheidenheid aan typen data die kunnen worden verzameld. Dit wordt ook wel aangeduid als de heterogeniteit van data. Er dient onderscheid gemaakt te worden in verschillende typen data, namelijk: gestructureerde data, niet-gestructureerde data en semi-gestructureerde data. Gestructureerde data bestaat uit data in relationele datasets, waarin de data wordt weergegeven in rijen en kolommen die op een bepaalde wijze met elkaar in verband

staan (Gandomi & Haider, 2015). Gestructureerde data wordt door de eenvoudige structuur vaak gebruikt voor data analyse. Bedrijven die enkel dit type data gebruiken lopen echter veel informatie mis, gestructureerde data beslaat namelijk slechts 5% van alle data (Cukier, 2010). Ongestructureerde data bestaat uit de data die niet kan worden geanalyseerd door traditionele methoden waarin de data in rijen en kolommen staat. Voorbeelden van ongestructureerde data zijn: Teksten, video's, audio-opnamen en afbeeldingen. De grootste hoeveelheid opgeslagen data bestaat dan ook uit ongestructureerde data (Cukier, 2010). Semi-gestructureerde data zijn datasets die kenmerken hebben van gestructureerde data en ongestructureerde data (Abitboul, 1997). Een voorbeeld van semi-gestructureerde data is een XML-bestand. Big data heeft dus een hoge verscheidenheid aan type data die kunnen worden gebruikt voor de analyse.

- **Velocity:** Dit refereert aan de snelheid waarmee data wordt gegenereerd en de snelheid waarin de data geanalyseerd dient te worden voor gebruik (Gandomi & Haider, 2015). Een sneller gebruik van verschillende data wordt in de huidige maatschappij verwacht. Een voorbeeld daarvan zijn restaurant-beoordeelplatformen zoals: Yelp en TripAdvisor. Waar restaurantbezoekers hun keuze waar te gaan eten in het verleden baseerden op de jaarlijks uitkomende restaurantgids, maken zij nu de keuze door verschillende reviews van de afgelopen dagen te lezen (Walker, 2015). Niet enkel consumenten verwachten advies op basis van recente gegevens. Ook bedrijven hechten waarde aan de snelheid waarmee data kan worden geanalyseerd. Zo willen snel kunnen ingrijpen indien er iets fout gaat. 'Continuous auditing' is hier een voorbeeld van. Bij 'continuous auditing' wordt voortdurend zekerheid geboden met betrekking tot de kwaliteit en geloofwaardigheid van de informatie die wordt gepresenteerd (Elam et al., 2002). 'Continuous auditing' is enkel mogelijk indien de snelheid waarmee de data kan worden geanalyseerd hoog genoeg is. 'Continuous auditing' vindt namelijk plaats op 'real-time basis' (Elam et al., 2002).

Deze drie kenmerken kunnen worden aangeduid als de drie klassieke V's. Deze zijn in de wetenschappelijke literatuur aangevuld met drie nieuwe V's, namelijk:

- **Value:** Eén van de belangrijkste aspecten van big data is dat het waarde creëert voor de gebruiker (Chen & Zhang, 2014). Big data is betekenisloos indien het geen waarde

creëert (Zhou et al., 2016). Waardecreatie is het belangrijkste aspect van big data en refereert aan het proces waarin verborgen waarde wordt gecreëerd uit grote datasets (Hashem et al., 2015). Softwareontwikkelaar Oracle introduceerde het kenmerk 'value' voor big data. Big data wordt namelijk gekenmerkt door data die een lage waarde hebben relatief aan het volume van de data indien deze niet juist geanalyseerd worden, maar door het analyseren van deze data kan veel waarde worden gecreëerd doordat het volume zo groot is (Gandomi & Haider, 2015).

- **Veracity:** Dit refereert aan de waarheidsgetrouwheid van big data. Waarheidsgetrouwheid van big data is de laatste jaren een steeds belangrijkere rol gaan spelen, doordat er meer aandacht wordt besteed aan de kwaliteit van data (Lukoianova & Rubin, 2014). Doordat big data veel onzekerheden bevat, is 'veracity' één van de belangrijkste waarde creërende aspecten van big data (Schroeck et al., 2012). De onzekerheden van big data komen voort uit de inconsistentie en onvolledigheid van de data, dubbelzinnigheden in de data, wachttijden en doordat de data in benaderd worden door modellen (Lukoianova & Rubin, 2014). International Business Machines Corporation (IBM) introduceerde 'veracity' als vierde V, omdat sommige bronnen van informatie samengaan met onbetrouwbaarheid van data, de noodzaak om deze onnauwkeurige en onzekere gegevens te verwerken is een belangrijk aspect van big data (Gandomi & Haider, 2015; Lukoianova & Rubin, 2014).
- **Variability:** Dit aspect is geïntroduceerd door softwareontwikkelaar SAS en refereert aan de variabiliteit omtrent de hoeveelheid data. De snelheid waarmee big data wordt gecreëerd is namelijk niet constant over tijd, maar heeft pieken en dalen (Gandomi & Haider, 2015).

Zoals alle verschillende hierboven besproken aspecten van big data aantonen, is big data een veelomvattend begrip. Of bepaalde data als big data wordt omschreven of als een reguliere dataset is context afhankelijk en dient per industrie afzonderlijk te worden bekeken (Gandomi & Haider, 2015). De definitie van big data die in deze thesis zal worden gebruikt is: 'Een verzameling van grote datasets met een grote diversiteit aan type data met bepaalde kenmerken waardoor het lastig is deze te verwerken door middel van traditionele dataverwerkingssoftware' (Chen & Zhang, 2014).

2.3 Big data analytics

Big data op zichzelf staand is waardeloos (Gandomi & Haider, 2015). Om waarde te creëren dient de data te worden geanalyseerd. Het proces om van ruwe data naar waardevolle informatie te gaan bestaat uit de volgende vijf stappen (Labrinidis & Jagadish, 2012), hierbij bestaan de eerste 3 stappen uit data management en de laatste 2 stappen uit data analyse (Gandomi & Haider, 2015):

- 1. Verkrijgen en opnemen:** Allereerst dient data te worden verkregen. Verschillende bronnen kunnen enorme hoeveelheden data verkrijgen, niet alle data is daarentegen belangrijk. Er dienen filters te worden aangebracht op welke data dient te worden opgenomen en welke data niet relevant is en direct kan verdwijnen (Labrinidis & Jagadish, 2012).
- 2. Extractie en 'cleaning':** Uit de opgeslagen data dient enkel de relevante data uit de grote dataset te worden gehaald. En deze relevante data dient 'schoongemaakt' te worden. Dat wil zeggen, de data dient op een bepaalde wijze gestructureerd te worden zodat het geschikt is voor analyse (Labrinidis & Jagadish, 2012).
- 3. Integratie en aggregatie:** De data uit verschillende databronnen en formats dient te worden geïntegreerd en geaggregeerd voordat er kan worden overgegaan naar een analyse waarin alle relevante data wordt meegenomen (Gandomi & Haider, 2015).
- 4. Analyse:** De data dient op een bepaalde manier geanalyseerd te worden (Labrinidis & Jagadish, 2012). Hiervoor bestaan verschillende methodes, daar zal later in dit hoofdstuk aandacht aan worden besteed.
- 5. Interpretatie:** De laatste stap is dat de geanalyseerde data dient te worden geïnterpreteerd. Het hele proces staat in het teken dat er bepaalde acties worden ondernomen op basis van de data. Indien de personen die beslissingen moeten maken de geanalyseerde data niet begrijpen, zullen er ook geen acties worden ondernomen. Interpretatie vormt dus een belangrijke schakel in het proces (Labrinidis & Jagadish, 2012).

Uit de vijf stappen blijkt dat big data analytics slechts een sub-proces is in het proces van waardecreatie van big data (Gandomi & Haider, 2015). Big data analytics is het toepassen van geavanceerde analysetechnieken op big data (Russom, 2011). Big data analytics bestaat dus

enerzijds uit big data en anderzijds uit geavanceerde analyse technieken (advanced analytics). Wat onder big data wordt verstaan is eerder in dit hoofdstuk beschreven, het vervolg van dit hoofdstuk zal zich dan ook richten op wat onder geavanceerde analyse technieken wordt verstaan.

Geavanceerde analysetechnieken zijn analysetechnieken die in staat zijn big data te analyseren, waar traditionele analysetechnieken dat niet hebben (Russom, 2011). Gegeven het brede spectrum aan geavanceerde analyse technieken (Gandomi & Haider, 2015), gaat het geven van een complete lijst van deze technieken buiten het doel van deze thesis. Wel zullen enkele belangrijke geavanceerde analyse technieken worden besproken die relevant worden geacht voor deze thesis.

2.4 Data Mining

Data mining is geavanceerde analysetechniek die als doel heeft kleine ‘goudklompjes’ aan kennis te vinden in de grote hoop ruwe data (Han et al., 2011). Een betere term voor data mining zou dan ook zijn ‘knowledge mining from data’ (Han et al., 2011). De verschillende data miningtechnieken kunnen onderverdeeld worden in de zes primaire taken of doelen van data mining (Kantardzic, 2011; Fayyad et al., 1996):

- **Classification:** Dit is een taak van data mining die data in verschillende vooraf vastgestelde groepen of klassen classificeert (Kantardzic, 2011; Fayyad et al., 1996). Het classificatieproces bestaat uit twee stappen, een trainingstap en classificatiestap (Han et al., 2011). In de trainingstap wordt het algoritme aangeleerd welke data uit de trainings-dataset bij welke groep hoort, waarna in de classificatiestap de nieuwe data wordt geclassificeerd in vooraf vastgestelde groepen (Han et al., 2011). Voorbeelden van classificaties zijn: het classificeren van risicovolle aanvragers van leningen bij een bank (Han et al., 2011; Fayyad et al., 1996), maar ook ingewikkeldere classificatieprocessen zoals het automatisch classificeren van bepaalde objecten uit databases met een grote hoeveelheid afbeeldingen (Fayyad et al., 1996; Djorgovski et al., 1996) en het classificeren van verschillende kosten binnen ‘cost accounting’. De belangrijkste data mining technieken die gebruikt kunnen worden voor het classificeren van data zijn: decision trees, naïve bayes, neural networks (Dangare & Apte, 2012) en k-nearest neighbor (Sumathi & Sivanandam, 2006).

- **Regression:** Dit is een vorm van 'machine learning' waarin een bepaalde continue waarde wordt voorspeld aan de hand van enkele variabelen (Mehta, 2017; Fayyad 1996; Kantardzic, 2011). Er kan sprake zijn van een enkelvoudige regressie, in dat geval wordt de waarde van de afhankelijke variabele geschat op basis van één andere variabele (Mehta, 2017). Een voorbeeld hiervan is het vaststellen van de audit kwaliteit op basis van de grootte van het accountantskantoor dat de audit uitvoert. De audit kwaliteit wordt echter niet enkel door de grootte van het accountantskantoor beïnvloed (Choi et al., 2010), er bestaat dan ook een meervoudige regressie waarin de waarde wordt voorspeld aan de hand van meerdere onafhankelijke variabelen (Mehta, 2017). Een voorbeeld hiervan is de audit kwaliteit, op basis van verschillende factoren, zoals: grootte van het accountantskantoor, het aantal uur dat is besteed aan de audit door de auditor en de onafhankelijkheid van de auditor.
- **Clustering:** Dit is een taak van data mining waarin het doel is om categoriën of clusters te identificeren om de data te beschrijven (Fayyad et al., 1996; Kantardzic, 2011). Clustering is net als classification een techniek die door middel van 'machine learning', data in verschillende groepen classificeert. Het verschil tussen classification en clustering is dat classification bestaat uit 'supervised learning' en clustering uit 'unsupervised learning' (Jain, 2010). Dat wil zeggen dat bij classification de data wordt verdeeld in vooraf vastgestelde groepen, terwijl bij clustering de groepen worden geïdentificeerd door de software. Clustering wordt daarom ook wel beschreven als een speciaal type van classification (Jain & Dubbes, 1988). De clusters kunnen verschillende eigenschappen hebben. Er dient onderscheid gemaakt te worden in niet-exclusieve en exclusieve clusters. Niet-exclusieve clusters, ook wel overlappende cluster genoemd, kunnen data bevatten die bij meerdere clusters horen, terwijl exclusieve clusters data bevatten die enkel in één cluster zitten (Jain & Dubbes, 1988; Tan et al., 2017). Binnen 'human resource accounting' zou groeperen op basis van geslacht een voorbeeld van een exclusief cluster zijn, terwijl het groeperen op basis van welke taak iemand heeft binnen het bedrijf een voorbeeld is van een niet-exclusief cluster. Een persoon kan namelijk meerdere taken hebben binnen een bedrijf. De belangrijkste clustering technieken zijn: k-means, hierarchical clustering, density-based clustering (Sumathi & Sivanandam, 2006) en fuzzy clustering (Tan et al., 2017).

Hierbij dient de kanttekening te worden geplaatst dat de traditionele k-means techniek niet geschikt is om big data te analyseren. Enkele nieuwe k-means technieken zoals multi-view k-means zijn wel geschikt om big data te analyseren (Cai et al., 2013).

- **Summarization:** Dit omvat de technieken die als doel hebben een compacte beschrijving te geven van een bepaalde dataset (Fayyad et al., 1996; Kantardzic, 2011). Door middel van summarization wordt de data geabstraheerd en gegeneraliseerd. Hierdoor wordt de dataset verkleind, waardoor er een algemeen overzicht van de data ontstaat (Sumathi & Sivanandam, 2006). Deze summarization technieken bestaan zowel uit eenvoudige beschrijvende statistiek, zoals het gemiddelde en de standaard deviatie, als complexere methoden zoals multivariate visualisation (Fayyad et al., 1996). Een belangrijke toepassing van summarization is 'automated report generation' (Fayyad et al., 1996). 'Automated report generation' wordt over het algemeen gebruikt als beginpunt om de data beter te begrijpen (Vesanto & Hollmén, 2004).
- **Dependency Modeling:** Dit is een taak van data mining waarin modellen worden ontdekt op basis van een dataset (Kantardzic, 2011). Deze modellen bestaan uit twee aspecten. Het eerste aspect is de structuur van het model, de structuur geeft aan welke variabelen van elkaar afhankelijk zijn. Daarnaast kan het model ook uit een kwantitatief aspect bestaan, in dat geval wordt de sterkte van de relaties tussen de variabelen weergegeven op een numerieke schaal (Fayyad et al., 1996). Dependency modeling werkt goed bij hele grote datasets. Bij grote, complexe datasets is het echter niet altijd mogelijk om modellen te maken voor de gehele dataset. Dependency modeling wordt dan ook vaak gebruikt om gedeelten van een dataset te modeleren (Kantardzic, 2011).
- **Change and Deviation Detection:** Deze taak van data mining is erop gericht de significante veranderingen of verschillen in de data ten opzichte van eerder gemeten of normale waardes (Fayyad et al., 1996). De veranderingen en verschillen worden waargenomen door algoritmes die outliers ontdekken (Kantardzic, 2011). Door het vinden van deze outliers kunnen fouten in de data worden ontdekt (Arning et al., 1996). Het vinden van de fouten in de data is van belang bij data miningtechnieken waar gebruik wordt gemaakt van machine learning. Door fouten in de trainings-dataset worden de bevindingen de van verschillende machine learning technieken, zoals

decision trees, onnauwkeuriger (Quinlan, 1986). Outliers zijn niet enkel van belang om fouten uit de dataset te halen, er zijn ook situaties waarin niet de algemene patronen van een dataset interessant zijn, maar de outliers (Breunig et al., 2000). Een voorbeeld van een situatie waarin outliers interessant kunnen zijn is fraude detectie (Bolton & Hand, 2001).

2.5 Artificial intelligence

Kunstmatige intelligentie is een belangrijke rol gaan spelen binnen verschillende wetenschapsgebieden. Onder wetenschappers bestaat onenigheid over de precieze definitie van kunstmatige intelligentie. De verschillende definities zijn onder te verdelen in vier categorieën (Russell & Norvig, 1995), namelijk:

- Systemen die denken als mensen (Haugeland, 1985; Bellman, 1978).
- Systemen die zich gedragen als mensen (Kurzweil, 1990; Rich & Knight, 1991).
- Systemen die rationeel denken (Charniak & McDermott, 1985; Winston, 1992).
- Systemen die zich rationeel gedragen (Schalkoff, 1990; Luger & Stubblefield, 1993).

Kunstmatige intelligentie is een overkoepelende term voor systemen die aan één van de hierboven staande definities voldoen. De overkoepelende term kunstmatige intelligentie omhult een aantal technologieën. In het vervolg van dit hoofdstuk zullen een aantal van deze technologieën worden besproken. De verschillende technologieën waaruit kunstmatige intelligentie bestaat, zijn (Russell & Norvig, 1995):

- **Natural language processing:** Dit zijn technologieën om 'natuurlijke taal' te begrijpen of te gebruiken voor verschillende nuttige doeleinden. De 'natuurlijke taal' kan uit tekst bestaan, maar ook uit audio (Chowdhury, 2003).
- **Knowledge representation:** Dit onderdeel van kunstmatige intelligentie houdt zich bezig met het opslaan (Russell & Norvig, 1995) van informatie en de wijze waarop de informatie wordt vertoond (Brachman et al., 1992).
- **Automated reasoning:** Dit zijn technologieën die erop gericht zijn om systemen logische gevolgtrekkingen te laten maken (Harrison, 2009). Door automated reasoning kunnen kunstmatige intelligentie systemen vragen beantwoorden en nieuwe conclusies trekken (Russell & Norvig, 1995).

- **Machine learning:** Hiermee worden zelflerende algoritmen bedoeld, die zich op basis van eerdere data kan aanpassen aan omstandigheden en patronen kan ontdekken (Russel & Norvig, 1995)

De verschillende technologieën die onder kunstmatige intelligentie worden geschaard, spelen ook een belangrijke rol bij andere data analyse methoden zoals data mining (Quinlan, 1986).

2.6 Samenvatting

In dit hoofdstuk is de eerste deelvraag beantwoord, deze vraag luidt: Wat zijn big data en big data analytics? Allereerst is er onderzocht wat er onder big data wordt verstaan aan de hand van een aantal aspecten van big data. Deze aspecten bestonden uit de drie traditionele V's en de drie 'nieuwe' V's. Big data wordt gedefinieerd als: 'Een verzameling van grote datasets met een grote diversiteit aan type data met bepaalde kenmerken waardoor het lastig is deze te verwerken door middel van traditionele dataverwerkingssoftware' (Chen & Zhang, 2014).

Hierna is er onderzocht wat er onder big data analytics wordt verstaan. Onder big data analytics wordt het toepassen van geavanceerde analysetechnieken op big data verstaan (Russom, 2011). Tot slot zijn in dit hoofdstuk twee geavanceerde analysetechnieken besproken, namelijk: data mining en artificial intelligence.

Hoofdstuk 3: De jaarrekeningcontrole

3.1 Inleiding

In de introductie van deze thesis is uitgelegd wat een jaarrekeningcontrole is en wat het belang van de jaarrekeningcontrole is. In dit hoofdstuk zal worden uitgelegd hoe een traditionele jaarrekeningcontrole eruit ziet, onder traditioneel wordt verstaan: een jaarrekeningcontrole waarbij geen gebruik wordt gemaakt van big data en big data analytics.

De precieze wijze waarop een jaarrekeningcontrole plaatsvindt verschilt tussen de accountantskantoren. De meeste accountantskantoren hanteren wel soortgelijke methodes. Deze methode moet echter wel voldoen aan International Standard on Auditing (ISA) 315. In Nederland stelt het bestuur van de Koninklijke Nederlandse Beroepsorganisatie van Accountants (NBA) de nadere voorschriften controle- en overige standaarden (NV COS) op. De standaarden van de NBA dienen te worden nageleefd doordat de NBA een bij wet opgerichte organisatie is (artikel 2 Wet op het accountantsberoep). Op basis van artikel 3 sub a van de Wet op het accountantsberoep is één van de taken van de NBA het bevorderen van een goede beroepsuitoefening door accountants, onder meer door het vaststellen van beroepsreglementering. De controle van de naleving van deze standaarden is neergelegd bij de AFM op basis van de Wet toezicht accountantsorganisaties (Wta).

Bij een audit kan onderscheid gemaakt worden tussen de volgende vier fases (Norgaard & Kühn, 2013):

Fase 1: Het plannen van de jaarrekeningcontrole en het identificeren van de risico's.

Fase 2: Het bepalen van de strategie en het beoordelen van de risico's.

Fase 3: Het uitvoeren van de controles.

Fase 4: Concluderen en rapporteren.

De risico's binnen de jaarrekeningcontrole worden aangeduid als audit risico. Het audit risico bestaat uit het inherente risico, het controlerisico en het detectie risico. Het inherente risico bestaat uit risicofactoren die voor onjuistheden in financiële verslagen kunnen zorgen die los staan van fouten in controlemechanismen. Het controlerisico bestaat uit het risico dat er fouten ontstaan in de controlemechanismen. Het detectie risico bestaat uit het risico dat het

de auditor niet lukt om 'material misstatements' te ontdekken in de jaarrekening (Arens et al., 2010).

3.2 Fase 1: Het plannen van de jaarrekeningcontrole en het identificeren van de risico's

De eerste fase bestaat voornamelijk uit het leren kennen van het bedrijf waarbij de jaarrekeningcontrole plaatsvindt. Onder beter leren kennen wordt verstaan: de wijze van bedrijfsuitvoering, de eigenaar of eigendomsstructuren, investeringen die het bedrijf heeft gemaakt dan wel plant te gaan maken en de wijze waarop het bedrijf is gestructureerd en gefinancierd (ISA 315). Bij het leren kennen van het bedrijf wordt niet enkel gekeken naar de interne situatie binnen het bedrijf, maar ook de industrie waarin het bedrijf actief is. De auditor dient namelijk ook goed geïnformeerd te zijn over de concurrenten, klanten en leveranciers (PwC, 2013).

Er wordt gekeken naar de administratieve organisatie en interne controle (AO/IC) binnen het bedrijf. Het duidelijk maken van de AO/IC helpt de auditor in het bepalen van risico's die van materieel belang zijn omtrent fraude en fouten. In deze fase wordt ook bepaald welke materialiteit, zoals bedoeld in NV COS 320, er wordt gehandhaafd (ISA 315; Norgaard & Kühn, 2013).

Tot slot wordt in deze fase gekeken wat de significante rekeningen zijn. Bij het overwegen of er sprake is van een significante rekening wordt gekeken naar de kwantitatieve en kwalitatieve kenmerken van de rekeningen. Kwantitatieve overwegingen bestaan zijn gericht op de hoogte van het saldo bij een bepaalde rekening. Des te hoger het saldo, des te groter de kans op materiële fouten. De kwalitatieve overwegingen zijn gericht op het type rekening en de gevoeligheid van de informatie die de rekening bevat (Norgaard & Kühn, 2013).

3.3 Fase 2: Het bepalen van de strategie en het beoordelen van de risico's

In deze fase worden de 'significant classes of transactions' (SCOTs) geïdentificeerd, dit zijn transacties die een significante invloed kunnen hebben op de materialiteit van het jaarverslag. Nadat de relevante SCOTs zijn geïdentificeerd wordt gekeken welke IT applicaties relevant zijn voor de jaarrekeningcontrole. Wanneer het duidelijk is welke IT applicaties relevant zijn voor

de jaarrekeningcontrole wordt er gekeken hoe deze IT applicaties werken. Er wordt hierbij gekeken of er sprake is van een autorisatieprocedure en hoe deze in zijn werking gaat, of er sprake is van scheiding van taken, of er nog ander relevant IT beleid is en of er nog andere controleprocedures zijn (Norgaard & Kühn, 2013).

Door het bekijken van de hierboven genoemde zaken kan de auditor enerzijds vaststellen welke AO/IC worden gebruikt en anderzijds vaststellen wat de zaken zijn die mis kunnen gaan. Niet alle zaken die mis kunnen gaan worden in kaart gebracht. Enkel de zaken waar de waarschijnlijkheid voldoende hoog is dat er een onjuistheid ontstaat in de jaarrekening en waarbij de onjuistheid een omvang behelst die materieel is worden relevant geacht. In deze fase van de audit is de auditor op de hoogte van (Norgaard & Kühn, 2013):

- De wijze waarop transacties worden geïnitieerd, gecorrigeerd, verwerkt en gerapporteerd.
- Welke fouten kunnen ontstaan in het proces.
- Welke controlemechanismen bestaan die de kans op fouten verkleinen.

Nu de auditor hiervan op de hoogte is zal er worden beoordeeld welke AO/IC dienen te worden gecontroleerd en zal er een risicobeoordeling worden opgemaakt van de risico's die tezamen het audit risico behelzen (Norgaard & Kühn, 2013).

Het bepalen van het audit risico is van belang bij het bepalen van de wijze waarop de AO/IC gecontroleerd worden. Er wordt beoordeeld welke controletechniek zal worden gebruikt, welke duur de controle zal hebben en wat de omvang van de controle zal zijn. Dit zal op een wijze worden vastgesteld zodat er voldoende audit bewijs wordt verkregen voor een succesvolle jaarrekeningcontrole. Ook zal worden vastgesteld hoe de journaalposten zullen worden gecontroleerd en op welke wijze andere audit procedures zullen worden uitgevoerd (Norgaard & Kühn, 2013).

3.4 Fase 3: Het uitvoeren van de controles

In de derde fase worden de verschillende testen die in de tweede fase zijn gepland uitgevoerd. De auditor zal beginnen met het controleren van de interne controle binnen het bedrijf, om te bepalen of deze betrouwbaar genoeg zijn om erop te kunnen vertrouwen. Hierna zal er een

inhoudelijke controle plaatsvinden. Bij deze controle wordt geverifieerd of de jaarrekening juist en volledig is. De samenstelling van de controleprocedures is sterk afhankelijk van de situatie binnen het bedrijf. In een situatie waarin de interne controle sterk is, zal de auditor over het algemeen meer gebruik maken van 'Substantive Analytical Procedures'. Dat zijn procedures die erop gericht zijn om te controleren of de financiële cijfers in overeenstemming zijn met andere cijfers, anders gezegd zijn bepaalde ratio's logisch, en of de financiële cijfers in overeenstemming zijn met niet-financiële informatie. Deze procedures zorgen dan ook voor indirecte audit bewijs doordat enkel de verhoudingen worden gecontroleerd. In het geval dat blijkt dat de interne controles binnen een bedrijf zwak zijn, zal de auditor over het algemeen meer gebruik maken van 'Substantive Tests of Detail of Balance'. Hierbij wordt direct audit bewijs gezocht voor een bepaalde transacties en bepaalde grootboekrekeningen. Dat bewijs kan worden verkregen door het opzoeken van facturen of het opvragen van bankgegevens (Arens et al., 2010).

3.5 Fase 4: Concluderen en rapporteren

In de laatste fase wordt al het audit bewijs samengevoegd en wordt er een algemene conclusie met betrekking tot het jaarverslag geformuleerd (Arens et al., 2010). Aan het einde van de jaarrekeningcontrole stelt de accountant een accountantsverklaring op waarin staat vermeld dat de accountant heeft vastgesteld dat de jaarrekening voldoet aan de vereisten die op grond van de relevante regelgeving verplicht zijn.

3.6 Samenvatting

In dit hoofdstuk is onderzocht op welke wijze een traditionele jaarrekeningcontrole plaatsvindt. De jaarrekeningcontrole is hierbij opgedeeld in vier fases welke ieder afzonderlijk zijn behandeld. Ook is in dit hoofdstuk besproken dat het audit risico een grote rol speelt binnen de jaarrekeningcontrole.

Hoofdstuk 4: De kwaliteit van een jaarrekeningcontrole

4.1 Inleiding

De meest geciteerde definitie van audit kwaliteit is van DeAngelo (1981) (Tritschler, 2014). DeAngelo (1981) definieert audit kwaliteit als: ‘De waarschijnlijkheid dat een auditor een onjuistheid (material misstatement) detecteert en deze onjuistheid rapporteert’. Over de definitie van kwaliteit met betrekking tot een jaarrekeningcontrole bestaat echter veel discussie. Er zijn verschillende definities, toch zijn deze definities over het algemeen in te delen in twee grote stromingen, namelijk (Tritschler, 2014):

1. De kwaliteit van een jaarrekeningcontrole is gebaseerd op het niveau van assurance dat wordt bereikt.
2. De kwaliteit van een jaarrekeningcontrole is gebaseerd op de mate waarin de geldende standaarden worden nageleefd.

In dit hoofdstuk zullen de twee stromingen uiteen worden gezet om te bepalen welke definitie van kwaliteit er in deze thesis voor de jaarrekeningcontrole wordt gebruikt. Daarnaast zal er ook worden gekeken naar de wijze waarop de AFM en de MCA audit kwaliteit zien. Tot slot zal worden gekeken naar de ‘expectation gap’ van audit kwaliteit.

4.2 Het niveau van assurance

De definitie van DeAngelo (1981) bestaat uit twee aspecten. Enerzijds dient de auditor in staat te zijn om de onjuistheden te detecteren. De waarschijnlijkheid dat een auditor een onjuistheid detecteert is afhankelijk van verschillende factoren, zoals de technologische mogelijkheden, de auditprocedure die wordt gebruikt bij een bepaalde audit en de hoeveelheid gegevens die daadwerkelijk wordt gecontroleerd in de steekproef. Anderzijds dient de auditor onafhankelijk te zijn, omdat hij de onjuistheden ook moet rapporteren (DeAngelo, 1981; Tritschler, 2014). De onafhankelijkheid van de auditor wordt als één van de belangrijkste, als niet de belangrijkste, pijler(s) van een audit gezien. Voor investeerders is de onafhankelijkheid van de audit dan ook belangrijk voor het verlagen van het informatie-risico dat zij bij elke investering lopen. Door de onafhankelijkheid van de auditor kunnen zij meer

vertrouwen op de informatie in de jaarrekening en dus beter geïnformeerde beslissingen maken (Elliott & Jacobson, 1998).

De definitie van DeAngelo (1981) kan worden ingedeeld in de stroming waarbij de definitie van audit kwaliteit het bereikte niveau van assurance is. Door het detecteren en rapporteren van material misstatements zal het niveau van assurance namelijk toenemen. Na de publicatie van DeAngelo (1981) gingen vele wetenschappers mee met deze definitie van audit kwaliteit. Enkele met een identieke definitie zoals Palmrose (1988), maar ook definities die sterk gebaseerd zijn op de definitie van DeAngelo (1981). Titman en Trueman (1986) definieerden audit kwaliteit als de nauwkeurigheid waarmee auditors informatie verstrekken aan investeerders. Knechel (2009) vat deze stroming dan ook heel goed samen en definieert audit kwaliteit als de bereikte mate van assurance (Tritschler, 2014).

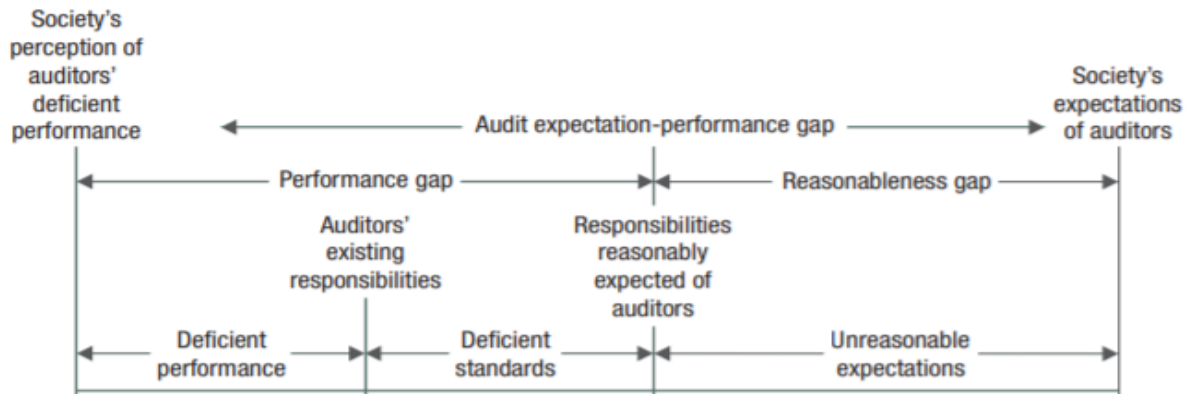
4.3 De mate van naleving van standaarden

De andere stroming met betrekking tot audit kwaliteit richt zich niet op de mate van assurance die te audit verschaft, maar op de mate waarin de relevante standaarden worden nageleefd (Krishnan & Schauer, 2001; Tritschler, 2014). De aanname die hierbij wordt gemaakt is dat indien een auditor niet voldoet aan de relevante accounting en auditing standaarden, de audit van slechte kwaliteit zal zijn (Krishnan & Schauer, 2001). De kritiek op deze stroming richt zich op het feit dat niet het voldoen aan de relevante standaarden het doel van de audit is. Het doel van de audit is namelijk zorgen voor kwalitatief goede financiële verslaggeving (Tritschler, 2014). Ondanks deze kritiek gebruikt een grote groep wetenschappers deze of een soortgelijke definitie voor audit kwaliteit (Copley & Doucet, 1993; Aldhizer et al., 1995; Krishnan & Schauer, 2001; Niemann, 2004).

De definities van de twee stromingen sluiten elkaar niet uit. De Government Accountability Office (GAO) definieert audit kwaliteit dan ook als een combinatie van de twee eerder genoemde stromingen (Tritschler, 2014). In deze thesis zal niet over audit kwaliteit gesproken worden als een combinatie van de definities zoals de GAO, maar doormiddel van een hiërarchische combinatie. De definitie van DeAngelo (1981) zal leidend zijn, het naleven van de standaarden is daarentegen een belangrijke factor bij het behalen van assurance en valt daarom ook onder audit kwaliteit.

4.4 Verwachtingskloof

Er bestaat een kloof tussen hoe de samenleving de kwaliteit van audit beoordeeld en wat de samenleving voor kwaliteit van audit verwacht, deze kloof is de 'audit expectation-performance gap' (Porter et al., 2012).



Figuur 1: 'Audit expectation-performance gap' van Porter et al., 2012

De 'audit expectation-performance gap' bestaat uit 3 componenten. Allereerst is er een kloof tussen wat de samenleving verwacht van auditors en wat er redelijkerwijs kan worden verwacht van de auditors. Deze verwachtingskloof bestaat uit de onredelijke verwachtingen van de samenleving. Daarnaast is er een verwachtingskloof tussen hoe de samenleving de kwaliteit van audit beoordeeld en wat er redelijkerwijs kan worden verwacht van de auditors. Deze verwachtingskloof bestaat enerzijds uit gebrekkige standaarden en anderzijds uit het gebrekkig uitvoeren van de audits (Porter et al., 2012).

4.5 Kwaliteit volgens toezichthouders

De AFM en MCA geven aan dat de kwaliteit van de jaarrekeningcontroles niet hoog genoeg is (AFM, 2017; MCA, 2018). Zoals eerder in dit hoofdstuk al bleek is er geen duidelijke definitie van audit kwaliteit. De NBA heeft daarom ook een 'green paper' opgesteld met als doel audit kwaliteit te definiëren. In dit paper wordt audit kwaliteit onderverdeeld in drie verschillende dimensies (NBA, 2017).

De eerste dimensie die de NBA benoemt is de externe dimensie van audit kwaliteit. Binnen deze dimensie is allereerst het handelen in publiekelijk belang belangrijk. Dat houdt in dat

bepaalde onjuistheden worden geïdentificeerd en dat de accountant daar opvolging aan geeft. Ook compliance met standaarden is van belang binnen de externe dimensie. Compliance met standaarden betekend specifiek compliance aan de Wet toezicht accountantsorganisaties en het Besluit toezicht accountantsorganisaties. Tot slot is het bij de externe dimensie van audit kwaliteit van belang dat er toegevoegde waarde ontstaat bij het bedrijf waar de jaarrekeningcontrole plaatsvindt.

De tweede dimensie die de NBA benoemt is de interne dimensie van audit kwaliteit. Hierbij zijn de structuur, gezag en toezicht van belang voor audit kwaliteit. De structuur bestaat uit de wijze waarop de activiteiten zijn verdeeld binnen het accountantskantoor. Het gezag heeft betrekking op de wijze van besluitvorming binnen accountantskantoren en de afweging die moet gemaakt worden tussen centralisatie en decentralisatie. Op basis van de Wet toezicht accountantsorganisaties is een bepaalde mate van centralisatie dan ook vereist. Toezicht heeft betrekking op de mate waarin managers toezicht houden op de medewerkers en deze indien er iets mis gaat ook bijsturen.

De derde dimensie die de NBA benoemt is de individuele dimensie van audit kwaliteit. De individuele dimensie heeft betrekking op de relatie tussen de prestatie van een accountant en audit kwaliteit. Indien de prestatie beter is, zal ook de audit kwaliteit beter zijn. Uit onderzoek blijkt dat voor prestatie drie aspecten van belang zijn: motivatie, mogelijkheden en vaardigheden.

Buiten de aspecten van audit kwaliteit die worden besproken in het 'green paper' van de NBA, is er nog een ander aspect waar de toezichthouder AFM veel waarde aan hecht. De AFM gaf in het rapport *'Uitkomsten onderzoek kwaliteit wettelijke controles Big 4-accountantsorganisaties'* (2014) aan dat zij constateren dat er een onvoldoende professioneel kritische grondhouding onder accountants is. Volgens de AFM wordt er onvoldoende kritisch gekeken of de controle-informatie niet in tegenspraak is met andere controle-informatie en wordt er onvoldoende kritisch gekeken of er nog aanvullende controlewerkzaamheden dienen te worden uitgevoerd. Daarnaast is de accountant onvoldoende kritisch richting de informatie die het management heeft gedeeld. Deze onvoldoende kritische houding heeft bijvoorbeeld betrekking op schattingen van het management voor ontwikkelingen in personeel, kosten, opbrengsten en de markt. Daarnaast

wordt er niet voldoende kritisch gekeken naar de berekening van voorzieningen en dient er kritischer te worden gekeken naar de verklaring van het management, bijvoorbeeld waarom een bepaalde betalingstermijn is overschreden.

4.6 Samenvatting

In dit hoofdstuk is de derde deelvraag beantwoord, deze vraag luidt: Wat is de definitie van kwaliteit met betrekking tot de jaarrekeningcontrole? Er is eerst gekeken naar de verschillende stromingen over audit kwaliteit, nadat deze stromingen zijn besproken is er gekozen om als definitie van audit kwaliteit het behaalde niveau van assurance vast te stellen. Daarnaast is onderzocht waaruit de verwachtingskloof van audit kwaliteit bestaat en op welke wijze audit kwaliteit wordt gezien door de toezichthouders.

Hoofdstuk 5: Incorporatie van big data en big data analytics in de jaarrekeningcontrole

5.1 Inleiding

In hoofdstuk twee is bekeken waar big data uit bestaat en welke big data analyse technieken er zijn. In dit hoofdstuk zal de vierde deelvraag worden beantwoord en zal worden gekeken hoe big data en big data analytics geïncorporeerd kunnen worden in de jaarrekeningcontrole.

5.2 Big data als audit bewijs

Eén van de toepassingen van big data in auditing is als audit bewijs. De audit moet voldoende en passend bewijs gebruiken (AICPA, 2014), waar passend kan worden onderverdeeld in betrouwbaar en relevant (Hoogduin et al., 2015). Er dient dus te worden gekeken of het gebruik van big data: meer, betrouwbaarder en relevanter audit bewijs oplevert.

Eén van de kenmerken van big data is dat het ongestructureerde data zoals afbeeldingen en video's bevat (Gandomi & Haider, 2015). Door de technologieën die big data biedt kunnen verschillende gebeurtenissen nu ook worden gecontroleerd op basis van ongestructureerde data (Moffitt & Vasarhelyi, 2013). Dit zorgt ervoor dat er meer audit bewijs beschikbaar is. Daarnaast kan big data ook nieuw audit bewijs opleveren. Zo zijn er big datatechnieken die automatisch e-mails kunnen controleren om te kijken of er wordt gefraudeerd (Holton, 2009). Zonder big data analytics zouden al deze e-mails stuk voor stuk dienen te worden gecontroleerd voor een auditor wat niet realistisch is.

Big data verhoogt niet enkel de hoeveelheid audit bewijs, maar kan ook de betrouwbaarheid van het bewijs verhogen. Een voorbeeld is het transporteren van goederen, om te verifiëren of goederen zijn getransporteerd worden transportdocumenten bekeken. Door gebruik te maken van big data en big data analytics kunnen nieuwe technieken worden gebruikt die de betrouwbaarheid van de controle verhogen. Voorbeelden hiervan zijn GPS-trackers (Hoogduin et al., 2015) en beveiligingsbeelden (Kogan et al., 2015). Een andere wijze waarop de big data en big data analytics te betrouwbaarheid van audit bewijs verhogen is door middel van tekstanalyse. De verschillende tekstanalysetechnieken, zoals clustering, die door big data

analytics mogelijk word gemaakt kunnen teksten automatisch samenvatten zodat de auditor enkele belangrijke zaken direct kan zien (Dhillon & Modha, 2001). Door middel van deze 'automated report generation' technieken wordt meer informatie aan het licht gebracht dan bij een handmatige samenvatting (Hoogduin et al., 2015). Ook kan de betrouwbaarheid van audit bewijs worden versterkt door het gebruik van externe informatie. Het aantal hits op internet kan een goede graadmeter zijn voor de populariteit en het aantal verkopen van een product (Kogan et al., 2015). Indien een product negatief in het licht staat in de media, maar de verkoopcijfers stijgen volgens het bedrijf zou de auditor vraagtekens kunnen stellen of deze cijfers wel juist zijn (Hoogduin et al., 2015). Big data kan dus in veel opzichten de betrouwbaarheid van audit bewijs verhogen. Wel dient de kanttekening geplaatst te worden dat indien er veel 'ruis' in de data aanwezig is, het analyseren ervan tot onbetrouwbaarder audit bewijs kan lijden omdat 'ruis' leidt tot data van een lage kwaliteit (Hoogduin et al., 2015).

Het laatste kenmerk van audit bewijs dat nog getoetst dient te worden is relevantie. Het gebruik van big data en big data analytics is relevant voor het verzamelen van audit bewijs omdat het unieke informatie kan verschaffen ten opzichte van een traditionele audit (Hoogduin et al., 2015). De relevantie blijkt ook uit de veranderende bedrijfsprocessen die dienen te worden gecontroleerd. De processen waar bedrijven gebruik van maken zijn sterk gedigitaliseerd, hierdoor dienen auditor ook op een ander manier de jaarrekeningcontrole te doen (Hoogduin et al., 2015).

5.3 Voorspellingen op basis van big data

Big data analytics geven de mogelijkheid om bepaalde voorspellingen te doen, die zonder deze technieken niet mogelijk waren (Madden, 2006). Dit komt doordat traditionele voorspelmechanismen niet in staat zijn de hoeveelheid, snelheid en complexiteit van big data te analyseren (Arribas-Bel, 2014). Data mining technieken kunnen worden gebruikt om voorspellingen te doen op basis van big data (Rey & Wells, 2013; Varian, 2014). Het voorspellen op basis van big data gebeurt momenteel al bij het maken van weersverwachtingen (Hamm, 2013), maar ook bij het voorspellen van financiële rendementen (Alessi et al., 2009). Het gebruik van zulke voorspellende technieken kan ook gebruikt worden binnen de auditing om voorspellingen te doen over bepaalde cijfers of controleprocedures. Zo

kunnen er methoden worden ontwikkeld die voorspellen waar de risico's liggen bij de audit van een bedrijf. Doordat de risicocomponenten kunnen worden voorspeld, kan de auditor extra aandacht aan dat onderdeel van de audit besteden (Cao et al., 2015).

5.4 Outliers en fraudedetectie

Big data analytics bieden de mogelijkheid om outliers in een dataset te vinden (Kantardzic, 2011). Outliers zijn data die een afwijking hebben ten opzichte van de rest van de data of afwijken van de normale data (Fayyad et al., 1996). Het vinden van outliers kan dan ook heel nuttig zijn bij fraudedetectie (Bolton & Hand, 2001). Het gebruik van big data analytics om outliers te detecteren kan dan ook heel nuttig zijn binnen het auditproces (Thiprungsri & Vasarhelyi, 2011). Het detecteren van outliers binnen auditing is mogelijk op basis van classificatietechnieken waarin er in de 'training set' twee groepen, frauduleuze bedrijven en niet frauduleuze bedrijven, worden aangeduid (Fanning et al., 1995; Deshmukh et al., 1997; Bakar et al., 2006). Deze methode is echter niet bruikbaar in de echte wereld, omdat er geen 'training set' kan worden gecreëerd (Thiprungsri & Vasarhelyi, 2011). Clustering is daarentegen wel een methode die gebruikt kan worden voor fraudedetectie (Thiprungsri & Vasarhelyi, 2011). Clustering is een speciale vorm van classification (Jain & Dubbes, 1988), clustering maakt namelijk gebruik van artificial intelligence in de vorm dat het algoritme zelf leert welke groepen/clusters er bestaan (Jain, 2010). Door middel van clustering is het dan ook mogelijk om data die outliers zijn van frauduleuze bedrijven of activiteiten te detecteren (Thiprungsri & Vaserhelyi, 2011).

5.5 Big data en de kwaliteit van een audit

In hoofdstuk vier is audit kwaliteit gedefinieerd als het niveau van assurance dat geboden kan worden bij een audit. Er dient nu te worden gekeken of het gebruik van big data en big data analytics de assurance van een audit verhoogt.

Big data en big data analytics kunnen helpen bij het creëren van audit bewijs (Hoogduin et al., 2015). Big data zorgt er namelijk voor dat er meer, betrouwbaarder en relevanter audit bewijs ontstaat. Een auditor bereikt assurance indien hij voldoende audit bewijs heeft. Indien er dus

meer, betrouwbaarder en relevanter audit bewijs ontstaat zal ook het niveau van assurance stijgen.

Het voorspellende aspect van big data kan de auditor de mogelijkheid geven om te kijken wat risicovolle procedures en transacties zijn binnen een bedrijf (Cao et al., 2015). Indien de auditor meer aandacht besteed aan risicovolle procedures zal er een hoger niveau van assurance worden bereikt, omdat er een grotere kans is dat een onjuistheid wordt ontdekt. Daarnaast blijkt volgens de AFM dat de accountant onvoldoende kritisch is tegenover verschillende schattingen van het management over bijvoorbeeld: de ontwikkeling van kosten, opbrengsten en de hoogte van voorzieningen. Het gebruik van big data analytics kan helpen bij het beter schatten van deze zaken, zodat de accountant een hulpmiddel krijgt voor professioneel kritische grondhouding.

Ook het vinden van outliers die kunnen leiden tot fraudedetectie verbeterd de audit kwaliteit. Het vinden van de outliers zorgt er namelijk voor dat onjuistheden kunnen worden gedetecteerd. En het detecteren van onjuistheden leidt weer tot assurance en een hogere audit kwaliteit (DeAngelo, 1981).

5.6 Samenvatting

In dit hoofdstuk is gekeken op welke wijze big data en big data analytics kunnen worden geïncorporeerd in de jaarrekeningcontrole. Hierbij is onderzocht of big data als audit bewijs kan dienen, of big data kan helpen bij verschillende voorspellingen en wat de rol van big data kan zijn binnen fraudedetectie. Tot slot is in dit hoofdstuk gekeken naar wat de relatie is tussen big data en audit kwaliteit.

Hoofdstuk 6: Conclusie en discussie

6.1 Conclusie

In deze thesis is de vraag beantwoord of het gebruik van big data en big data analytics de kwaliteit van een jaarrekeningcontrole kan verbeteren. Deze hoofdvraag is beantwoord aan de hand van vier deelvragen.

De eerste deelvraag is: Wat zijn big data en big data analytics? Big data wordt gekenmerkt door de drie klassieke V's en de drie nieuwe V's (Gandomi & Haider, 2015). Big data kan worden gedefinieerd als: Een verzameling van enorme datasets met een grote diversiteit aan type data met bepaalde kenmerken waardoor het lastig is deze te verwerken doormiddel van traditionele dataverwerkingssoftware (Chen & Zhang, 2014). Big data analytics zijn analyse technieken om big data mee te analyseren. In deze thesis is gekeken naar data mining en de verschillende taken die data mining kan verrichten. Er kan onderscheid worden gemaakt in de zes kerntaken van data mining. (Kantardzic, 2011; Fayyad et al., 1996). Een andere vorm van big data analytics was artificial intelligence. Binnen artificial intelligence kan onderscheidt gemaakt worden tussen: natural language processing, knowledge representation, automated reasoning en machine learning.

De tweede deelvraag is: Hoe ziet een 'traditionele' jaarrekeningcontrole eruit zonder het gebruik van big data en big data analytics? De jaarrekeningcontrole bestaat uit vier fases (Norgaard & Kühn, 2013):

De derde deelvraag is: Wat is de definitie van kwaliteit met betrekking tot een jaarrekeningcontrole? Er zijn twee stromingen met betrekking tot audit kwaliteit (Tritschler, 2014). De ene stroming ziet kwaliteit als het niveau van assurance dat wordt bereikt, terwijl de andere stroming de kwaliteit van audit baseert op de mate waarin de geldende standaarden worden nageleefd. In deze thesis is gekozen om audit kwaliteit te zien als het niveau van assurance dat wordt bereikt.

De vierde deelvraag is: Op welke wijze kunnen big data en big data analytics worden geïncorporeerd in de jaarrekeningcontrole? Hier is geconcludeerd dat big data en big data analytics kunnen zorgen voor meer, betrouwbaarder en relevant audit bewijs (Hoogduin et al., 2015). Daarnaast kunnen verschillende voorspellingen worden gedaan op basis van big

data analytics (Cao et al., 2015). Ook kunnen big data en big data analytics helpen bij het detecteren van frauduleuze activiteiten (Thiprungsri & Vaserhelyi, 2011).

Op basis van de deelvragen dient geconcludeerd te worden dat het gebruik van big data en big data analytics de kwaliteit van een jaarrekeningcontrole kan verbeteren. Er bestaan namelijk verschillende mogelijkheden tot incorporatie van big data en big data analytics in de jaarrekeningcontrole die het niveau van assurance verhogen en dus de kwaliteit van een jaarrekeningcontrole.

6.2 Discussie en vervolgonderzoek

In deze thesis is onderzocht of big data en big data analytics de kwaliteit van een jaarrekeningcontrole kunnen verhogen. Hierbij is enkel gekeken naar de definitie van audit kwaliteit als mate waarin assurance wordt bereikt. De NBA legt audit kwaliteit echter veel breder uit dan de gehanteerde definitie in deze thesis. Het gebruik van big data zal in de individuele dimensie bijdragen aan de prestaties van de accountant. Het vergroot namelijk de mogelijkheden die de accountant heeft om de jaarrekeningcontrole te doen en audit bewijs te vinden. Binnen de interne dimensie is er mogelijk ook een rol weggelegd voor big data en big data analytics. Binnen de interne dimensie is één van de belangrijke aspecten toezicht binnen de accountantskantoren. De big datatechnieken die worden gebruikt bij het controleren van andere bedrijven tijdens de audit zouden ook gebruikt kunnen worden om het interne toezicht te verbeteren binnen de accountantskantoren.

Een ander belangrijk onderwerp met betrekking tot de kwaliteit van de jaarrekeningcontrole is de professioneel kritische grondhouding van de accountant. Hoewel big data analytics kunnen helpen om bepaalde schattingen te maken, dient de accountant zelf kritisch te zijn over verschillende zaken. Of big data daarbij zou kunnen helpen of deze taak zou kunnen overnemen is niet in deze thesis onderzocht. Het is denkbaar dat artificial intelligence verder wordt ontwikkeld en zelfstandig in staat is om een professioneel kritische grondhouding aan te nemen waardoor de accountant overbodig zou worden. Het gebruik van artificial intelligence zal daarentegen mogelijk een negatief effect hebben op de 'perceived audit quality'. Hoewel de samenleving belang hecht aan de hogere mate van assurance die wellicht geboden kan worden door artificial intelligence, vinden mensen het fijn als een persoon een

verklaring afgeeft dat de jaarrekening kloppend is. Indien een persoon deze verklaring afgeeft is het namelijk mogelijk om iemand ter verantwoording te roepen indien zaken fout zijn gelopen. Bij artificial intelligence is het dan ook een discussie hoe de aansprakelijkheid geregeld moet worden. Big data en big data analytics kunnen de accountant helpen om een betere jaarrekeningcontrole te doen, maar het is geen vervanging van de professioneel kritische grondhouding van de accountant.

Daarnaast heeft het gebruik van big data binnen de jaarrekeningcontrole nog vele obstakels (Alles, 2015). Ook dient de kanttekening te worden geplaatst dat big data en big data analytics niet altijd geschikt, dan wel de beste opties, zijn voor sommige zaken binnen de jaarrekeningcontrole. Zo blijkt traditionele data mining minder geschikt voor auditing dan gespecificeerde computer geassisteerde auditing technieken (Wong, 2001).

Voor vervolgonderzoek zou het interessant zijn om daadwerkelijk big data analytics toe te passen in een auditproces om te kijken of de verschillende analyse technieken werkelijk bruikbaar audit bewijs kunnen leveren. Daarnaast is het ook interessant of het gebruik van big data een effect heeft op de 'perceived audit quality'. Daarnaast bestaat er een grote mate van onzekerheid tot welke mate artificial intelligence kan worden toegepast binnen het audit, het gebruik van artificial intelligence binnen de jaarrekeningcontrole is dan ook een interessant onderwerp voor vervolgonderzoek.

Literatuurlijst

- Abiteboul, S. (1997). Querying Semi-Structured Data. In Proceedings of 6th International Conference on Database Theory (pp. 1–18).
- Autoriteit Financiële Markten (AFM). (2014). *Uitkomsten onderzoek kwaliteit wettelijke controles Big 4-accountantsorganisaties*. Retrieved from <https://www.afm.nl/nl-nl/nieuws/2014/sep/rapport-controles-big4>
- Autoriteit Financiële Markten (AFM). (2017). *Kwaliteitsslag OOB-accountantsorganisaties gaat te langzaam*. Retrieved from <https://www.afm.nl/nl-nl/nieuws/2017/juni/kwaliteitslag-oob>
- Aldhizer III, G. R., Miller, J. R., & Moraglio, J. F. (1995). Common attributes of quality audits. *Journal of Accountancy*, 179(1), 61. Retrieved from <http://search.proquest.com/docview/206770113?accountid=26357>
- Alessi, L., Barigozzi, M. and Capasso, M. (2009). Forecasting large datasets with conditionally heteroskedastic dynamic common factor. European Central Bank Working Paper 1115.
- Alles, M. G. (2015). Drivers of the use and facilitators and obstacles of the evolution of big data by the audit profession. *Accounting Horizons*, 29(2), 439–449. <https://doi.org/10.2308/acch-51067>
- Arens, A. A., Elder, R. J. & Beasley, M. S. (2010). *Auditing and assurance services: An integrated approach: global edition (13th edition)*. Englewood Cliffs, NJ: Prentice-Hall.
- Arning, A., Agrawal, R. & Raghavan, P. (1996). A Linear Method for Deviation Detection in Large Databases. In *Complexity* (pp. 164–169).
- Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49, 45–53. <https://doi.org/10.1016/j.apgeog.2013.09.012>
- Association of Chartered Certified Accountants, (ACCA). (2011). *The value of audit: views from retail (private) investors*.
- Bakar, Z. A., Mohamad, R., Ahmad, A., & Deris, M. M. (2006). A comparative study for outlier detection techniques in data mining. *IEEE Conference on Cybernetics and Intelligent Systems*. <https://doi.org/10.1109/ICCIS.2006.252287>
- Bellman, R. (1978). *An introduction to artificial intelligence: Can computers think?* Boston, MA: Thomson Course Technology.
- Berry, M. J., & Linoff, G. (1997). *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc..
- Bolton, R. J. & Hand, D. J. (2001). Unsupervised Profiling Methods for Fraud Detection. *Proc. Credit Scoring and Credit Control VII*, 5–7. <https://doi.org/10.1.1.24.5743>

- Brachman, R. J., Levesque, H. J., & Reiter, R. (1992). Knowledge representation. Cambridge, MA: MIT press.
- Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density- based local outliers. *SIGMOD Record*, 29, 93-104.
- Burke, J. A., & Lee, H. (2015). Protecting the public interest through mandatory auditor firm rotation: A controversial issue. *Advances in Public Interest Accounting*, 18, 1–55. <https://doi.org/10.1108/S1041-706020150000018001>
- Cai, X., Nie, F., & Huang, H. (2013). Multi-View K -Means Clustering on Big Data. *IJCAI*, 2598–2604.
- Cao, M., Chychyla, R., & Stewart, T. (2015). Big data analytics in financial statement audits. *Accounting Horizons*, 29(2), 423–429. <https://doi.org/10.2308/acch-51068>
- Charniak, E. & McDermott, D. (1985). *Introduction to Artificial Intelligence*. MA, Boston: Addison Wesley Publ.
- Chen, P. C. L., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences*, 275, 314–347. <https://doi.org/10.1016/j.ins.2014.01.015>
- Choi, J., Kim, F., Kim, J., & Zang. Audit Office Size, Audit Quality and Audit Pricing. (2010). *Auditing: A Journal of Practice and Theory*. 29, (1), 73. Research Collection School Of Accountancy. Available at: https://ink.library.smu.edu.sg/soa_research/10
- Chowdhury, G. G. (2003). Natural language processing. *Annual review of information science and technology*, 37(1), 51-89.
- Collis, J., Jarvis, R., & Skerratt, L. (2004). The demand for the audit in small companies in the UK. *Accounting and Business Research*, 34(2), 87–100. <https://doi.org/10.1080/00014788.2004.9729955>
- Copley, P. A., & Doucet, M. S. (1993). The Impact of Competition on the Quality of Governmental Audits. *Auditing: A Journal of Practice and Theory*, 12(1), 88–98.
- Cukier, K. (2010). Data, data, everywhere. A special report on managing information. *The Economist*, 1–14. <https://doi.org/10.1038/nsmb0805-633>
- Dangare, S. C., & S. Apte, S. (2012). Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques. *International Journal of Computer Applications*, 47(10), 44–48. <https://doi.org/10.5120/7228-0076>
- DeAngelo, L. E. (1981). Auditor size and audit quality. *Journal of Accounting and Economics*, 3(3), 183–199. [https://doi.org/10.1016/0165-4101\(81\)90002-1](https://doi.org/10.1016/0165-4101(81)90002-1)
- DeAngelo, L. E. (1981). Auditor independence, “low balling”, and disclosure regulation. *Journal of Accounting and Economics*, 3(2), 113–127. [https://doi.org/10.1016/0165-4101\(81\)90009-4](https://doi.org/10.1016/0165-4101(81)90009-4)

- Deshmukh, A., & Talluru, L. (1998). A Rule-Based Fuzzy Reasoning System for Assessing the Risk of Management Fraud. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 7, 223–241. [https://doi.org/10.1002/\(SICI\)1099-1174\(199812\)7:4<223::AID-ISAF158>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1099-1174(199812)7:4<223::AID-ISAF158>3.0.CO;2-I)
- Dhillon, I. S., & Modha, D. S. (2001). Concept decompositions for large sparse text data using clustering. *Machine Learning*, 42(1–2), 143–175. <https://doi.org/10.1023/A:1007612920971>
- Djorgovski, S. G., Fayyad, U. M., & Weir, N. (1996). From Digitized Images to Online Catalogs. Data Mining a Sky Survey. *AI Magazine*, 17(2), 51–66.
- Elam, R. Rezaee, Z. Sharbatoghlie, & McMickle, P. (2002) Continuous Auditing: Building Automated Auditing Capability. *AUDITING: A Journal of Practice & Theory*, 21 (1), pp. 147-163. <https://doi.org/10.2308/aud.2002.21.1.147>
- Elliott, R. K. & Jacobson, P. D. (1998). Audit independence concepts: Certified Public Accountant. *The CPA Journal*, 68(12).
- Elliott, R. K., & Rogers, J. R. (1972). Relating Statistical Sampling To Audit Objectives. *Journal of Accountancy*, (July), 46–55. <https://doi.org/Article>
- Engel, E., Hayes, R. M., & Wang, X. (2007). The Sarbanes-Oxley Act and firms' going-private decisions. *Journal of Accounting and Economics*, 44(1–2), 116–145. <https://doi.org/10.1016/j.jacceco.2006.07.002>
- Fanning, K., Cogger, K. O., & Srivastava, R. (1995). *Detection of management fraud: a neural network approach*. In Proceedings the 11th Conference on Artificial Intelligence for Applications (pp. 220–223). IEEE Comput. Soc. Press. <https://doi.org/10.1109/CAIA.1995.378820>
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 37–54. <https://doi.org/10.1145/240455.240463>
- Ferrell, O. C., & Ferrell, L. (2011). The Responsibility and Accountability of CEOs: The Last Interview with Ken Lay. *Journal of Business Ethics*, 100(2), 209–219. <https://doi.org/10.1007/s10551-010-0675-y>
- Gandomi, A. & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, 35, (2), p.137-144.
- Hamm, S. (2013). *How Big Data Can Boost Weather Forecasting*. Retrieved from <https://www.wired.com/insights/2013/02/how-big-data-can-boost-weather-forecasting/>
- Han, J., Kamber, M., & Pei, J. (2011). Introduction In Data Mining (pp. 1–38). *Elsevier*. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Harrison, J. (2009). *Handbook of practical logic and automated reasoning*. Cambridge, MA: Cambridge University Press.

- Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Ullah Khan, S. (2015). The rise of “big data” on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115. <https://doi.org/10.1016/j.is.2014.07.006>
- Haugeland, J. (1985). *Artificial Intelligence: The Very Idea*. Cambridge: MIT Press
- Herrick, T. and Barrionuevo, A. (2002). “Were auditor and client too close-knit?” *Wall Street Journal*. 21 januari.
- Hilbert, M., & López, P. (2011). The world’s technological capacity to store, communicate, and compute information. *Science*, 332(6025), 60–65. <https://doi.org/10.1126/science.1200970>
- Holton, C. (2009). Identifying disgruntled employee systems fraud risk through text mining: A simple solution for a multi-billion dollar problem. *Decision Support Systems*, 46(4), 853–864. <https://doi.org/10.1016/j.dss.2008.11.013>
- Hoogduin, L., Yoon, K., & Zhang, L. (2015). Big data as complementary audit evidence. *Accounting Horizons*, 29(2), 431–438. <https://doi.org/10.2308/acch-51076>
- Jain, A. & Dubbes, R. (1988). *Algorithms for Clustering Data*. NJ, Englewood Cliffs: Prentice Hall
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Johnson, V. E., Khurana, I. K., & Reynolds, J. K. (2002). Audit-Firm Tenure and the Quality of Financial Reports. *Contemporary Accounting Research*, 19(4), 637–660. <https://doi.org/10.1506/LLTH-JXQV-8CEW-8MXD>
- Kantardzic, M., (2011). *Data Mining: Concepts, Models, Methods, and Algorithms, Second Edition*. NJ: Wiley E-book retrieved from <https://onlinelibrary.wiley.com/doi/book/10.1002/97811118029145>
- Knechel, R. W. (2009, September). *Audit Lessons from the Economic Crisis : Rethinking Audit Quality* [Inaugural Lecture]. Retrieved from <https://cris.maastrichtuniversity.nl/portal/files/980210/guid-d6188f1c-c82f-470d-9d58-6c4cd1226ac6-ASSET1.0>
- Kogan, A., Vasarhelyi, M. A., & Tuttle, B. M. (2015). Big data in accounting: An overview. *Accounting Horizons*, 29(2), 381–396. <https://doi.org/10.2308/acch-51071>
- Krishnan, J., & Schauer, P. C. (2001). Differences in quality among audit firms. *Journal of Accountancy*, 192, 85. Retrieved from <http://proquest.umi.com/pqdweb?did=75432274&Fmt=7&clientId=28929&RQT=309&VName=PQD>
- Kurzweil, R., Richter, R., & Schneider, M. L. (1990). *The age of intelligent machines* (Vol. 579). Cambridge, MA: MIT press.

- Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5(12), 2032–2033. <https://doi.org/10.14778/2367502.2367572>
- Luger, G. F & Stubblefield, W.A. (1993). *Artificial intelligence: structures and strategies for complex problem solving*. San Francisco, CA: Benjamin-Cummings Publishing Company.
- Lukoianova, T & Rubin, V. L. (2014). Veracity roadmap: Is big data objective, truthful and credible? In *Advances in Classification Research Online* (Vol. 24, pp. 4–15). *American Society for Information Science and Technology*. <https://doi.org/10.7152/acro.v24i1.14671>
- Madden, S. (2012). From Databases to Big Data. *Internet Computing, IEEE*, 16(3), 4–6. <https://doi.org/10.1109/MIC.2012.50>
- Mayhew, B. W., & Wilkins, M. S. (2003). Audit firm industry specialization as a differentiation strategy: Evidence from fees charged to firms going public. *Auditing. American Accounting Association*. <https://doi.org/10.2308/aud.2003.22.2.33>
- Mehta, R. (2017). *Big Data Analytics with Java: Data analysis, visualization & machine learning techniques*. VK, Birmingham: Packt Publishing
- Moffitt, K. C., & Vasarhelyi, M. A. (2013). AIS in an Age of Big Data. *Journal of Information Systems*, 27(2), 1–19. <https://doi.org/10.2308/isys-10372>
- Monitoring Commissie Accountancy (MCA). (2018). *Doorpakken!* Retrieved from http://www.monitoringaccountancy.nl/wp-content/uploads/2018/05/Tweede-verslag-Monitoring-Commissie-Accountancy-Doorpakken.pdf#_ga=2.61233708.1016494075.1532464542-451748922.1532464542
- Murdoch, T. B., & Detsky, A. S. (2013). The inevitable application of big data to health care. *JAMA - Journal of the American Medical Association*. <https://doi.org/10.1001/jama.2013.393>
- Nørgaard, H., & Kühn, T. (2013, September). *Risikobaseret tilgang til revision Hvordan får vi egentlig forholdt os praktisk til ISA 315?*
- Palmrose, Z. V. (1988). An Analysis of Auditor Litigation and Audit Service Quality. *The Accounting Review*, 63(1), 55–73. <https://doi.org/10.2307/247679>
- Porter, M. E., & Millar, V. E. (1985). *How information gives you competitive advantage*.
- PwC. (2013, January). *Understanding a financial statement audit*. Retrieved from <https://www.pwc.com/gx/en/audit-services/publications/assets/pwc-understanding-financial-statement-audit.pdf>
- Quinlan, J. (1986). Introduction of decision trees. *Machine Learning*, 1 (1), pp. 81-106.
- Rey, T., & Wells, C. (2012). Integrating data mining and forecasting. *OR-MS Today*, 39(6), 34. Retrieved from <https://ucd.idm.oclc.org/login?url=http://search.proquest.com/docview/1274383862?accountid=14507>

- Rich, E., & Knight, K. (1991). *Artificial intelligence*. New York, NY: McGraw-Hill Education.
- Robert, J. S., & Schalkoff, R. J. (1990). *Artificial Intelligence: An Engineering Approach*. New York, NY: McGraw-Hill Education.
- Romney, M., Steinbart, P. (2018). *Accounting information systems* (14^e ed.). Athenaeum
- Russell, S. & Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. NJ, Englewood Cliffs: Prentice Hall
- Russom, P. (2011). Big data analytics. *TDWI Best Practices Report*, Fourth Quarter, 19(4), 1–34.
- Sarle, W. S., Jain, A. K., & Dubes, R. C. (1990). Algorithms for Clustering Data. *Technometrics*, 32(2), 227. <https://doi.org/10.2307/1268876>
- Schalkoff, R. (1990). *Artificial Intelligence: An Engineering Approach*. NY, NY: McGraw-Hill
- Schroek, M., Shockley, R., Smart, J., Romero-Morales, D., Tufano, P. (2012). *Analytics: The real-world use of big data*. IBM Institute for Business Value.
- Sikos, L. F. (2015). *Big Data Applications*. In *Mastering Structured Data on the Semantic Web* (pp. 199–216). Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-1049-9_8
- Sumathi, S., & Sivanandam, S. N. (2006). *Introduction to Data Mining and its Applications* (Vol. 29). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-34351-6>
- Thiprungsri, S., & Vasarhelyi, M. A. (2011). Cluster analysis for anomaly detection in accounting data: An audit approach. *International Journal of Digital Accounting Research*, 11, 69–84. https://doi.org/10.4192/1577-8517-v11_4
- Titman, S., & Trueman, B. (1986). Information quality and the valuation of new issues. *Journal of Accounting and Economics*, 8(2), 159–172. [https://doi.org/10.1016/0165-4101\(86\)90016-9](https://doi.org/10.1016/0165-4101(86)90016-9)
- Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28(2), 3–28. <https://doi.org/10.1257/jep.28.2.3>
- Vesanto, J. & Hollmén, J. (2004). *Innovations in Intelligent Systems*. Duitsland, Berlijn: Springer-Verlag
- Villars, R.L., Olofson, C.W. and Eastwood, M., 2011. Big data: What it is and why you should care. White Paper, IDC, 14.
- Walker, R. (2015). *From big data to big profits*. VK, Oxford, Oxford University Press
- Williams, M., & Haugeland, J. (1987). *Artificial Intelligence: The Very Idea*. *Technology and Culture*, 28(3), 706. <https://doi.org/10.2307/3105016>
- Winston, P. H. (1992). *Artificial intelligence*, 3rd ed. Reading, MA: Addison-Wesley.

Wong, J. (2001). Data mining as a tool for internal auditors. *Internal Auditing*, 16 (1), pp. 21-25.

Zhou, K., Fu, C., & Yang, S. (2016). Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*. Elsevier Ltd.
<https://doi.org/10.1016/j.rser.2015.11.050>