Datamining en privacy:
Anything you do can be used against you

Een onderzoek naar in hoeverre Privacy Preserving datamining als een oplossing kan fungeren voor de problematiek die voortvloeit uit datamining.

Naam:	Bart van der Laan
Studentnummer:	3145093
Instelling:	Universiteit Utrecht
Studie:	MA Nieuwe Media en Digitale Cultuur
Onderdeel:	Masterscriptie
Datum:	11-05-2011
Docent:	Erna Kotkamp

Inhoudsopgave

Samenvatting

Om bruikbare informatie te verkrijgen uit de grote hoeveelheden data die commerciële- en overheidsinstellingen opslaan passen deze instellingen datamining toe. Datamining is proces van het analyseren van data vanuit verschillende perspectieven om zo het samen te vatten in bruikbare informatie (Cocx, 2009). Op datamining is veel kritiek, zowel op websites als in literatuur. De angst bestaat dat privacy geschonden wordt wanneer datamining wordt toegepast of wanneer geregistreerde persoonsgegevens worden doorgespeeld aan derden. Binnen het ontwikkelingsgebied van datamining en databases wordt deze kritiek opgemerkt en worden oplossingen bedacht om privacy te beschermen en toch datamining als analysemiddel te kunnen toepassen. De onderzoeksrichting wordt *Privacy Preserving datamining (PPDM)* genoemd (Han en Kamber, 2009, Brookshear, 2010 en Agrawal en Srikant, 2000). Hierbij wordt er vanuit gegaan dat wanneer individuele data niet toegankelijk is, privacy gewaarborgd is (Agrawal en Srikant, 2000). De opgeslagen data worden afgeschermd voor diegenen die analyses willen doen. Analisten kunnen wel datamining toepassen op datasets maar kunnen niet de individuele data benaderen. Kernvraag is *in hoeverre Privacy Preserving datamining als een oplossing kan fungeren voor de problematiek die voortvloeit uit datamining.*

In dit onderzoek zijn drie cases geanalyseerd om een antwoord te formuleren op deze vraag, de Bonuskaart van de Albert Heijn, het Herkenningsdienstsysteem en de database van Facebook. Reacties op nieuwsberichten over deze systemen en datamining zijn geanalyseerd om de problematiek die voorvloeit uit datamining inzichtelijk te maken. Iedere keer is gekeken naar in hoeverre Privacy Preserving datamining een oplossing is voor deze problematiek.

De belangrijkste kritiek op de Bonuskaart van de Albert Heijn is de schending van de privacy van klanten omdat analyse op koopgedrag plaatsvindt of informatie aan derden wordt doorgespeeld zonder dat de betreffende persoon hier van weet (Leenheer, 2009). Privacy Preserving datamining kan uitkomst bieden bij bescherming van de ingevoerde persoonsgegevens bij de gewone Bonuskaart en bescherming van combinaties tussen gegevens bij de anonieme variant van de Bonuskaart. Privacy Preserving datamining beschermt de consument echter niet tegen de analyse van persoonsgegevens en inzet van deze resultaten. Wanneer de Albert Heijn individuen gaat benaderen met persoonlijke aanbiedingen beschermt PPDM het individu niet. Tevens beschermt PPDM niet tegen het verlies (en eventuele verkoop) van de persoonsgegevens die zijn opgeslagen door Albert Heijn.

Het Herkenningdienstsysteem is een strafbladen systeem waarop Cocx (2009) datamining probeert toe te passen. Zoals Cocx zelf stelt heeft deze methode verregaande gevolgen voor privacy van diegenen waarvan gegevens in de database zijn opgeslagen (Cocx, 2009; p. 153). Men kan hierop immers beleid gaan bepalen (Olsthoorn, 2009). Hier kan worden gesteld dat Privacy Preserving datamining in mindere mate een antwoord vormt op de problematiek die voortvloeit uit datamining op het HKS. Men geeft aan de mogelijkheden van datamining voor politie en wetshandhaving te zien. Echter zien gebruikers van websites ook problemen. Vooral de problematiek die gaat over de overheid als een controle staat en het aanmerken van verdachten voordat zij een delict hebben gepleegd, heeft Privacy Preserving datamining geen antwoord op. Wanneer uit een analyse blijkt dat personen die delict X vaak plegen ook vaak delict Y plegen, kan dit gegeven worden ingezet om personen die delict X al hebben gepleegd te surveilleren en eventueel te arresteren voor het nog niet gepleegde delict Y. De oorspronkelijke data kan worden geanonimiseerd door de toepassing van PPDM, analyses kunnen worden gedaan, maar wanneer het individu kan wordt benaderd biedt PPDM geen bescherming.

Sociale Netwerk Site Facebook registreert veel gegevens van gebruikers. Profielen van deze gebruikers worden samengesteld door de toepassing van datamining. Advertenties die vervolgens worden aangeboden zijn gebaseerd op deze profilering (Facebook, 2011b), Facebook werpt hiermee vragen op ten aanzien van waarborgen van privacy. Ook is veel van de informatie die men achterlaat op zijn profiel publiekelijk toegankelijk en kan daarom in het bezit komen van onwelwillende derde partijen. Privacy Preserving datamining kan in het geval van Facebook niet een eenduidige oplossing bieden voor de gestelde problematiek van datamining. Dit komt onder andere doordat de discussies over privacy omvangrijk, complex, subjectief en divers zijn. Privacy Preserving datamining richt zich op het beschermen van de oorspronkelijke data in een database. Een online sociaal netwerk, waarbij

veel van die informatie ook publiekelijk toegankelijk is, leent zich zeer goed voor datamining zo blijkt uit het advertentieprogramma van Facebook. Het doorverkopen van gegevens is iets waar PPDM geen antwoord op vormt. Tevens is het beschermen van de oorspronkelijke data complex aangezien veel van deze informatie toch al (deels-) openbaar is.

Op de vraag in hoeverre Privacy Preserving datamining een oplossing is voor de problematiek die voortvloeit uit datamining kan worden gesteld dat Privacy Preserving datamining in mindere mate een oplossing is voor deze problematiek. PPDM kan helpen opgeslagen persoonsgegevens af te schermen. Maar;

1. PPDM kan niet als oplossing fungeren wanneer algemene analyses individueel worden ingezet;

De persoonsgegevens kunnen worden afgeschermd. Echter wanneer de resultaten van de analyses ingezet worden tegen het individu, zoals bij het doen van persoonlijke aanbiedingen (Bonuskaart & Facebook) of bij het voorspellen van criminele carrièrepaden (HKS), biedt PPDM geen uitkomst. Datamining beschermt individuele data maar kan een individu niet beschermen tegen de inzet van de resultaten uit datamining.

2. PPDM kan niet helpen tegen de keuze voor het opslaan van gegevens en het spanningsveld tussen doelstellingen van een organisatie, het juridische kader en een technologische oplossing;

Wanneer een organisatie als doel heeft om gegevens door te verkopen of wanneer politie criminele carrièrepaden wil gaan voorspellen staan deze doelstellingen altijd boven de keuze voor een oplossing als PPDM. Het kiezen voor een bepaalde technologische oplossing staat in verhouding tot de doelstelling van een bepaalde organisatie en de juridische randvoorwaarden.

Een deel van de problematiek kan worden ondervangen door transparantie te creëren in welke gegevens worden opgeslagen, wat met deze gegevens wordt gedaan en welke gevolgen dit kan hebben voor burger en consument. Tevens dient de burger of de consument beter op de hoogte zijn van rechten en plichten ten aanzien van gegevensopslag. Er dient hierbij inzicht te worden gecreëerd in hoeverre datamining te plaatsen valt onder de Wet Bescherming Persoonsgegevens door een politiek en maatschappelijk debat.

Inleiding

"We are data rich, but information poor." (Han en Kamber, 2009)

De opslag van data is de afgelopen decennia exponentieel toegenomen (Tienkamp, 2010). Hierdoor is het voor bedrijven en overheden steeds lastiger geworden om uit data bruikbare informatie te halen. Een softwarematige methode om informatie te verkrijgen uit databasegegevens is datamining. Datamining, of *Knowledge Discovery from Data* (KDD), is het proces van het analyseren van data vanuit verschillende perspectieven om zo het samen te vatten in bruikbare informatie (Cocx, 2009). Zo kunnen bedrijven uit data voorspellingen doen over het koopgedrag van consumenten en kunnen overheden burgers aanmerken die buiten gestelde normen of grenzen handelen. Datamining is dus het analyseren en vormen van data tot bruikbare informatie.

Datamining is controversieel, dit blijkt uit verschillende kritieken (zoals Solove, 2008 en Castells, 2001). Een voorbeeld hiervan is de toepassing van datamining om veiligheid van burgers te verhogen door vertrouwelijke gegevens te raadplegen of te doorzoeken. Potentieel toepassingsgebied is de toepassing van de algoritmes van datamining ten behoeve van wetshandhaving. Politieorganisaties hebben de afgelopen jaren veelvuldig data opgeslagen in hun informatiesystemen. Tim Cocx (2009) heeft in zijn proefschrift 'Algorithmic Tools for Data-Oriented Law Enforcement' geprobeerd de algoritmen en methodes van datamining toe te passen op een grote database van strafbladen, het Herkenningsdienstsysteem (HKS). Dit deed hij in nauwe samenwerking met de politie. Zijn resultaten bevatten een aantal bekende causale verbanden maar ook een aantal onverwachte;

> Vrouwen zijn significant vaker verslaafd aan drugs dan mannen;
> Mensen verdacht van doodslag zijn relatief vaak al veroordeeld wegens racisme;
> Joyriders nemen het ook niet zo nauw met de arbeidswetten en alcohol;
> Diefstal met geweld hangt vaak samen met wapenbezit;
> Mensen met Afrikaanse afkomst zijn vaak veroordeeld voor overtredingen inzake 'openbare veiligheid';
> Criminelen op het platteland begaan ook vaak verkeersmisdrijven.(Olsthoorn, 2009 en Cocx, 2009; p. 22).

Zoals Cocx zelf stelt heeft deze methode verregaande gevolgen voor privacy van diegenen waarvan gegevens in de database zijn opgeslagen (Cocx, 2009; p. 153). In een interview met Webwereld licht hij toe dat momenteel niets met deze data gedaan mag worden, maar dat hij zich kan voorstellen dat de politie hier wel wat mee zou willen doen. Men kan hierop immers beleid gaan bepalen (Olsthoorn, 2009). Data die door politie worden opgeslagen mag in Nederland alleen worden gebruikt voor het doel waarvoor ze zijn opgeslagen (Olsthoorn, 2009). Volgens Cocx (2009) mag datamining momenteel door de politie niet worden toegepast.

Datamining wordt ook toegepast voor commerciële doeleinden. Zo vergaren supermarkten data door klantenkaartsystemen (Leenheer, 2009). Als bekend voorbeeld geldt de Bonuskaart van de Albert Heijn (Ah.nl, 2011). Klanten dienen hierbij persoonsgegevens op te geven om een klantenkaart te verkrijgen. Met deze klantenkaart kunnen zij kortingen krijgen op allerlei producten. De Albert Heijn kan deze persoonsgegevens vervolgens koppelen aan het koopgedrag van klanten en hierop inspelen. Er kunnen bijvoorbeeld gerichte aanbiedingen gedaan worden en wellicht in de toekomst individuele prijzen gemaakt worden. Datamining kan inzicht geven in de producten die bepaalde typen klanten kopen. Zo kunnen supermarkten bepalen dat wanneer product X wordt gekocht, product Y ook vaak door dezelfde consument wordt gekocht. Daarom zal de supermarkt deze producten bij elkaar in de schappen plaatsen.

De belangrijkste kritiek op de Bonuskaart van de Albert Heijn is de schending van de privacy van klanten omdat analyse op koopgedrag plaatsvindt of informatie aan derden wordt doorgespeeld zonder dat de betreffende persoon hier van weet (Leenheer, 2009). Het College Bescherming Persoonsgegevens heeft in 1999 reeds aangegeven dat anonimiteit moet worden gewaarborgd, hiertoe heeft Albert Heijn een anonieme variant van de Bonuskaart op de markt gebracht (CBP, 1999). Deze geeft wel korting, maar de consument behoeft hiervoor geen gegevens achter te laten.

Een ander actueel onderwerp in het kader van datamining zijn Sociale Netwerk Sites als Facebook. Men laat op sociale media veelal een spoor van persoonlijke informatie achter. Dit stelt bijvoorbeeld Facebook in staat om inzicht te krijgen in online gedrag en Facebook kan deze informatie commercieel inzetten. Facebook maakt gebruik van datamining voor de analyse van de grote hoeveelheden opgeslagen gebruikersdata (Alphenaar, 2010). Door toepassing van datamining kunnen zij gebruikersgedrag filteren op interesses (Facebook, 2011a). Advertenties die vervolgens worden aangeboden zijn gebaseerd op deze profilering (Facebook, 2011b), Facebook werpt hiermee vragen op ten aanzien van waarborgen van privacy. Ook is veel van deze informatie publiekelijk toegankelijk en kan derhalve in het bezit komen van onwelwillende derde partijen.

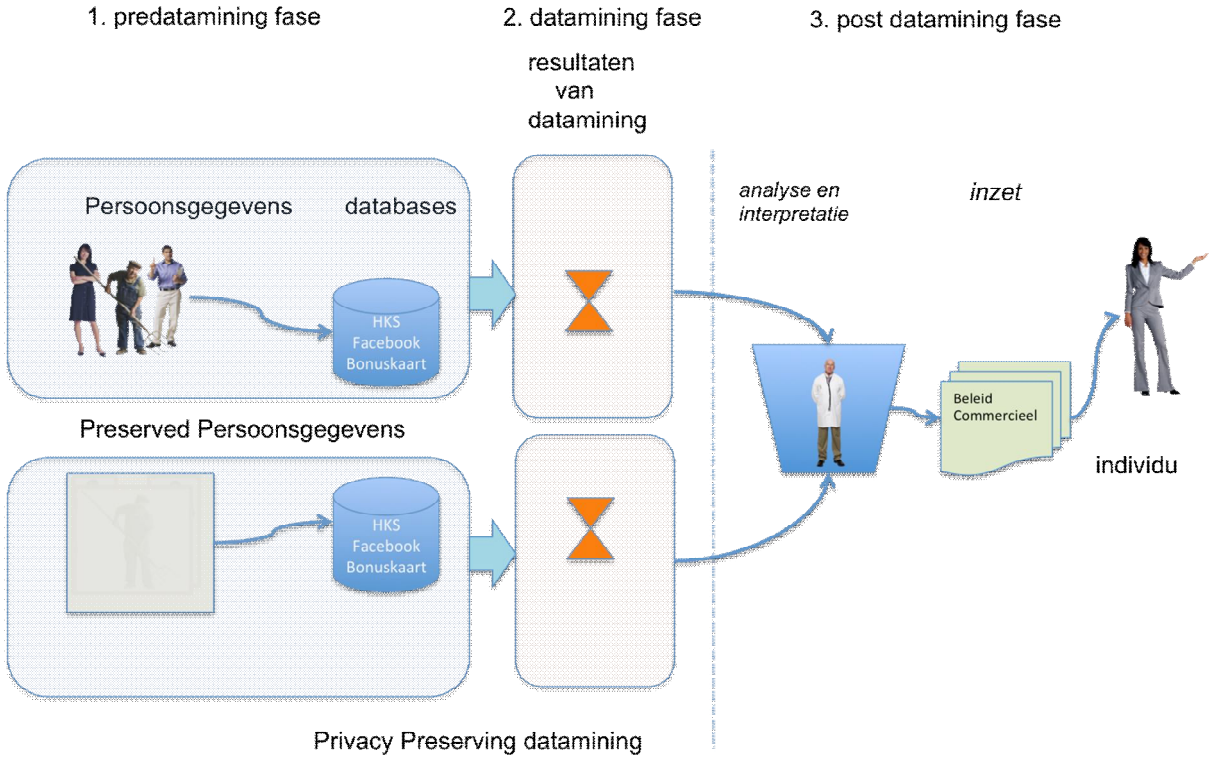
In het kader van de spanning tussen datamining en het waarborgen van privacy zijn meer voorbeelden te bedenken. Binnen het ontwikkelingsgebied van datamining en databases wordt deze kritiek opgemerkt en worden oplossingen bedacht om privacy te beschermen en toch datamining als analysemiddel te kunnen toepassen. Deze onderzoeksrichting wordt *Privacy Preserving datamining (PPDM)* genoemd (Han en Kamber, 2009, Brookshear, 2010 en Agrawal en Srikant, 2000). Hierbij wordt er vanuit gegaan dat wanneer individuele data niet toegankelijk is, privacy gewaarborgd is (Agrawal en Srikant, 2000). De opgeslagen data, zoals de gegevens die verzameld zijn met de Bonuskaart, worden afgeschermd voor diegenen die analyses willen doen. Analisten kunnen wel datamining toepassen op datasets maar kunnen niet de individuele data benaderen. In het voorbeeld van de Bonuskaart van AH betekent dit dat het koopgedrag van de individuele consument niet toegankelijk is maar dat alleen de data van alle consumenten die iets hebben gekocht met de Bonuskaart toegankelijk is.

Het maatschappelijk debat wordt niet uitsluitend over privacy schending gevoerd. Auteurs benoemen meer negatieve gevolgen. Zo kan datamining de overheid in staat stellen om de controle over burgers te vergroten (Castells, 2001), datamining zou burgers kunnen disciplineren (Elmer, 2003) en datamining zou discriminerende gevolgen kunnen hebben (Lyon, 2007). Kernvraag is *in hoeverre Privacy Preserving datamining als een oplossing kan fungeren voor de problematiek die voortvloeit uit datamining.*

Het onderzoek zal eerst methodologisch verantwoord worden. Om de kernvraag te beantwoorden wordt datamining eerst gedefinieerd in hoofdstuk één. Vervolgens worden in hoofdstuk twee de negatieve gevolgen van datamining nader omschreven vanuit de literatuur. Daarna worden er, in hoofdstuk drie, drie casestudies besproken; de Bonuskaart, het Herkenningsdienstsysteem (HKS) en de database van Facebook. Hierbij wordt gekeken naar wat de case inhoudt, welke negatieve gevolgen worden genoemd in reacties op nieuwsberichten die gaan over de cases en in hoeverre Privacy Preserving datamining als een oplossing kan fungeren voor de problematiek die voortvloeit uit datamining. Vervolgens worden, in hoofdstuk vier, de resultaten van de verschillende casestudies in verband gebracht met de onderzoeksvraag. Ten slotte zal in de conclusie en discussie teruggekomen worden op de hoofdvraag en zal deze beantwoord worden.

1. Methoden

Dit onderzoek bestaat uit twee delen. Het onderstaande figuur geldt als leidraad voor dit onderzoek. Om goed te kunnen begrijpen in hoeverre Privacy Preserving datamining als een oplossing kan fungeren voor de problematiek die voortvloeit uit datamining, dient datamining duidelijk gedefinieerd te worden. Ieder genoemd begrip in het onderstaande schema zal in dit onderzoek daarom nader worden toegelicht. Hierbij wordt gekeken naar hoe het proces van datamining eruit ziet en hoe bedrijven en organisaties analyses toepassen of uitvoeren. Deze uitleg dient ter onderbouwing van de geschetste casestudies. Vervolgens wordt uitgelegd hoe Privacy Preserving datamining in de praktijk werkt. Dit is van belang bij het bespreken en het interpreteren van de casestudies. In hoofdstuk 3 worden de effecten van datamining geschetst. Hierbij worden deze discussies gecategoriseerd en beoordeeld aan de hand van het onderstaand proces. Dat begint met de vastlegging van gegevens en eindigt met het definiëren van beleid en commerciële acties. Deze negatieve gevolgen zijn gecategoriseerd aan hand van onderscheidbare kenmerken. Deze categorisering is noodzakelijk om de resultaten uit casestudies onderling vergelijkbaar te maken. De reacties op nieuwsberichten die plaatsvinden op verschillende websites en die ter illustratie dienen zijn eveneens ingedeeld in deze categorieën. De vergelijking en analyse zijn beschreven in hoofdstuk 5.



Figuur 1: Fasering datamining en het bereik van Privacy Preserving datamining

1.1 Toepassingen

Er is gekozen voor drie verschillende toepassingsgebieden van de casestudies. De reden daarvoor is dat hiermee de problematiek vanuit drie verschillende actuele ingangshoeken belicht wordt. Deze zijn datamining in commercie, -wetshandhaving en -online sociale netwerksites. Met als respectievelijke cases de Bonuskaart van de Albert Heijn, het Herkenningsdienstsysteem van het KLPD en de database van Facebook. Discussies over datamining in deze toepassingsgebieden zijn zeer verschillend van aard ten aanzien het gebruik van gegevens. Naar verwachting zal de discussie in commercie zich richten op verkoopbaarheid van gegevens, binnen wetshandhaving op koppeling met andere bestanden en online sociale netwerksites op zowel de verkoopbaarheid van gegevens als de

koppeling met andere bestanden. Daarmee zal een eventuele toepassing van PPDM een andere rol gaan spelen. Bij de iedere case wordt het bovenstaande proces doorlopen. Er wordt gekeken naar discussies die gaan over de bescherming van de persoonsgegevens. Deze persoonsgegevens kunnen worden geanalyseerd door de toepassing van datamining. Vervolgens wordt gekeken naar discussies die gaan over de analyse en de inzet van de resultaten van de datamining analyse. Ook wordt gekeken naar de mogelijkheid van andere gegevens uit andere databases te koppelen aan de database uit de case. Daarnaast zijn er nog relativerende discussies die in dit schema niet worden besproken. In hoofdstuk 5 worden conclusies getrokken in relatie tot de centrale kernvraag en de toepasbaarheid dan wel effecten van PPDM.

Casus 1:De AH Bonuskaart

Als eerste wordt de Bonuskaart van Albert Heijn als een bekend voorbeeld voor datamining binnen commercie behandeld. De Bonuskaart is een klanten-loyaliteitssysteem. De klant kan korting krijgen op bepaalde producten van Albert Heijn in ruil voor gegevens die hij of zij achterlaat. Albert Heijn kan deze gegevens gebruiken om inzicht te krijgen in het koopgedrag haar consumenten (Albert Heijn, 2011a). Naar verwachting gaat hierover in de nabije toekomst een intensieve discussie ontstaan. In deze case wordt gekeken naar de reacties op twee nieuwsberichten. Het eerste bericht gaat over het delen van gegevens van Bonuskaart met justitie en het tweede bericht gaat over het implementeren van een RFID-chip in de Bonuskaart en de mogelijkheden die dat kan bieden. Deze nieuwsberichten, welke te vinden zijn in de bijlage, stonden op de websites; Webwereld.nl (2006) (alle reacties tot en met 7 juli 2006, N=17 reacties), Security.nl (2008) (alle reacties tot en met 14 oktober 2008, N=35 reacties) en Tweakers.net (2006) (alle reacties tot en met 7 juli 2006, N=101 reacties).

Casus 2: Het Herkenningsdienstsysteem

Als tweede casus is gekozen voor het Dienst Herkenningssysteem (HKS) vanwege de mogelijkheden met- en de omvang van HKS. Het HKS is een database met alle strafbladen, met daarbij allerlei persoonsgegevens waarvan de Korps Landelijke Politie Diensten en het Centraal Bureau voor de Statistiek eigenaar zijn. Dit systeem wordt gebruikt om jaarlijks inzicht te verkrijgen in algemene trends van criminaliteit in Nederland. Onlangs heeft de promovendus Tim Cocx (2009) getracht datamining toe te passen op dit systeem. Naar verwachting zal veel discussie ontstaan over het gebruik van datamining en concludering daaruit op een dergelijke soort database die gevoelige en persoonlijke informatie bevat. Hierover kwamen een aantal nieuwsberichten naar buiten (Webwereld, 2009a, 2009b) en er kwamen veel reacties hierop. De reden hiervoor is omdat het gaat over het HKS, primair over datamining, de nieuwsberichten stonden Nederlandse sites en meningen over deze situatie nauwelijks aanwezig zijn binnen de nieuwsberichten. Er is gekozen voor de analyse van reacties op dit nieuwsbericht op de sites Webwereld.nl (2009a) (alle reacties tot en met 9 december 2009, N=73 reacties) en Webwereld (2009b) (alle reacties tot en met 10 december 2009, N=52 reacties).

Casus 3: De database van Facebook

Als derde casus is gekozen voor Facebook, de database geeft immers een goed beeld over gebruik en voorkeuren bij gebruik van Internet . Ook hier geldt dat de omvang en de potentie van deze database voor datamining van belang is. Op Facebook laten deelnemers/vrienden opvallend veel informatie achter over zichzelf. Gebruikers kunnen wel via allerlei privacyinstellingen in zijn of haar profiel beveiliging creëren, om ervoor te zorgen dat niet iedereen zomaar kan meekijken met vertouwelijke gegevens. Er is gekozen voor twee nieuwsberichten over Facebook en hun privacybeleid. Facebook wordt momenteel veel bekritiseerd vanwege het nogal veranderlijke privacybeleid. Het eerste nieuwsbericht gaat over het advertentie systeem van Facebook waarmee Facebook de geaardheid van personen had getoond. In dit door Tweakers.net (2011a en b) geplaatste bericht zit een aspect van gebruik van datamining, het is geplaatst op Nederlandse websites. De website Tweakers.net (2011) (alle reacties tot en met 17 maart 2011, N=49 reacties), Tweakers.net (2010) (alle reacties tot en met 25 oktober 2010, N=92 reacties) plaatste recent dit bericht. Het tweede nieuwsbericht gaat over het privacy beleid van Facebook en hoe er omgegaan wordt met de opgeslagen data. Tevens het een en ander betrekking over de wijze van het analyseren van de data die zij tot haar beschikking heeft of de bronnen die tot haar beschikking staan.

1.2 Websites

Er is gezocht naar websites waarop nieuwsberichten zijn geplaatst die gaan over de genoemde casestudies. Er is gekozen voor websites met een technologisch karakter. Hiermee wordt bedoeld dat het onderwerp van de nieuwssite zich richt op nieuws en ontwikkelingen die gaan over technologie. Het gaat om de sites Webwereld.nl, Security.nl en Tweakers.net. Dit zijn de websites waarop de nieuwsberichten geplaatst zijn. Het onderwerp datamining is technologisch georiënteerd en complex. Verwacht wordt dat het publiek van deze websites enigszins technologische kennis heeft, wat diepgang kan geven aan de discussies over datamining. Het publiek heeft affiniteit met technologie en het zijn ervaren internet gebruikers. Daarbij moet worden opgemerkt dat Security.nl een site is waarop men over het algemeen kritisch is over privacy gerelateerde onderwerpen.

1.3 Nieuwsberichten en reacties

In hoofdstuk 4 worden de casestudies behandeld. In dit onderzoek is gekozen voor de analyse van reacties op nieuwsberichten die gaan over de cases. Analyse van reacties op nieuwsberichten die gaan over de cases maken discussies en de negatieve gevolgen van datamining goed inzichtelijk.
Natuurlijk is de keuze van deze nieuwsberichten van invloed op de discussies die daaruit volgen. De selectie van deze nieuwsberichten is daarom gebaseerd op een aantal objectieve of relevante kenmerken. Er is gekozen voor nieuwsberichten waar meningen zo min mogelijk naar voren komen, hieruit zijn enkele voorbeeld reacties gehaald. De criteria zijn gebruikt ter selectie van de nieuwsberichten:

1. Het nieuwsbericht moet relevant zijn voor de genoemde case door te kijken naar de inhoud;
2. Het nieuwsbericht heeft als kernonderwerp datamining;
3. Het nieuwsbericht dient geplaatst te zijn op een Nederlandse site vanwege culturele aspecten van privacy;
4. Het nieuwsbericht dient in enige mate een waardeoordeel te geven over privacy.

Het is hierbij van belang om te melden dat het begrip datamining vaak niet in nieuwsberichten of in de reacties op deze nieuwsberichten wordt genoemd. Er wordt veelal gekozen voor algemene beschrijvingen als ‘het ontdekken van patronen’ en ‘het doen van analyses’ waarmee dan toch datamining volgens bovenstaande definitie wordt bedoeld. De reacties op deze nieuwsberichten zijn gecategoriseerd aan de hand van de gekozen categorieën uit hoofdstuk 3. Er is gekeken naar op welke manier de reacties op de nieuwsberichten te plaatsen zijn in de categorieën discussies; 1) bescherming persoonsgegevens, 2) analyse en inzet, 3) inbreuk op privacy en 4) relativerende discussies. De keuze voor deze categorieën wordt verantwoord in hoofdstuk 3.

‘Off-topic’-reacties zijn uitgesloten van dit onderzoek. Deze reacties gaan over andere onderwerpen dan het onderwerp van het nieuwsbericht. Belangrijk hierbij is dat bij Tweakers.net een niveau-filter aanwezig is. Tweakers.net laat gebruikers het niveau van reacties op nieuwsberichten bepalen. Veel reacties zijn aangemerkt door andere gebruikers als off-topic. Echter wordt er in dit onderzoek soms juist wel gebruik gemaakt van deze berichten. Of iets off-topic is wordt gekwalificeerd door Tweakers als wat weinig informationele waarde bevat. Deze berichten kunnen wel gaan over het onderwerp van de nieuwsberichten maar hebben weinig inhoudelijke toevoeging op de discussies. In mijn onderzoek is er echter gekozen om berichten die gaan over de case mee te nemen, ondanks dat deze berichten soms weinig informatieve waarden hebben. Er wordt hierbij uitgegaan van voorgaande berichtgeving. Dit is gedaan omdat bepaalde berichten soms karakteriserend zijn voor de groep discussies. Berichten die niet gaan over de case zijn wel uitgesloten en worden in mijn onderzoek gedefinieerd als ‘Off-topic’. De berichten zijn niet kwantitatief geanalyseerd omdat een kwantiteit slechts een beperkte toevoeging zal hebben in het schetsten van de problematiek die voortvloeit uit datamining.

1.4 Structurering cases

Bij iedere case wordt er eerst uitvoeriger besproken wat de case is. Er wordt gekeken naar de doelstellingen en het privacybeleid van de organisaties die beheerder zijn van de databases. Zo wordt er bij de Bonuskaart gekeken naar de doelstellingen van de Albert Heijn met de gegevens die zij opslaan. Vervolgens worden de reacties op de nieuwsberichten uiteengezet. De reacties zijn onderverdeeld in de vier categorieën die worden toegelicht in hoofdstuk 3. Daarna wordt er gekeken

naar wat er zal veranderen wanneer Privacy Preserving datamining in deze situaties wordt toegepast. Dient deze ontwikkeling als een oplossing voor de problematiek die voorvloeit uit datamining?

2. Datamining

In dit hoofdstuk wordt besproken wat datamining is, waarom het wordt toegepast en uit welke fasering, zoals die getoond is in Figuur 1 uit hoofdstuk één, datamining bestaat. Daarna wordt er besproken wat Privacy Preserving datamining is. PPDM wordt gedetailleerd toegelicht als instrument om datamining als fenomeen beter hanteerbaar te maken.

2.1 Wat is datamining?

Volgens Cocx (2009) is datamining het proces van het analyseren van data vanuit verschillende perspectieven om het zo samen te vatten in bruikbare informatie. Brookshear (2009) definieert datamining als “technieken [die] worden gebruikt om patronen in gegevens te ontdekken” (p. 401). Data opgeslagen in databases kan worden omgezet in informatie die voor de organisatie van belang is. De methoden en algoritmen die worden toegepast binnen datamining zijn bruikbaar voor verschillende toepassingen. Zo is datamining belangrijk voor marketingdoeleinden maar wordt het bijvoorbeeld ook binnen het genetische onderzoek gebruikt om patronen in DNA te ontdekken door datasets te combineren (Brookshear, 2009; p. 401). Datamining wordt beschreven als een proces (Han en Kamber, 2006; DeRosa, 2004; Siebes, 1996), een methode (Brookshear, 2009), een algoritme (Wu et. al., 2008) en als een systeem (Han en Kamber, 2006). In dit hoofdstuk wordt datamining als een proces beschreven. Dit is van belang voor het bespreken van de casestudies omdat (delen van) de discussies zich toespitsen op bepaalde fasen uit het datamining proces.

De discussie rondom datamining toont zich bij de introductie van de term. De metafoor datamining verwijst naar mijnwerken (Hearst, 1999). Volgens verschillende auteurs is de term datamining in de Engelse opvatting eigenlijk verkeerd gekozen (Hearst, 1999 ; Han en Kamber, 2006). Bij mijnwerken gaat het bijvoorbeeld om ‘Goud winnen’ (Gold mining) in plaats van ‘Gesteente winnen” (Rock mining). Goud wordt gewonnen door gesteente of zand te ‘minen’, wat betekent dat vanuit dit gesteente of zand erts wordt gehaald om goud te genereren. Zonder de handeling ‘mining’ is het gesteente waardeloos. ‘Gold mining’ impliceert dus het winnen van erts vanuit gesteente. Datamining zou een verkeerd gekozen term zijn omdat dit impliceert dat data uit databases wordt gehaald. Datamining zou betekenen dat de ‘miner’ data wint uit databases, echter in deze poogt de ‘miner’ juist informatie en kennis te halen uit deze databases anders gezegd “meaningful data”. Bijvoorbeeld voor de analisten van. Albert Heijn is het niet relevant om kassabonnen te halen uit de Bonuskaart database, maar juist het koopgedrag van individuele of grote groepen consumenten te meten. De data wordt gebruikt voor een nieuw doel zoals het creëren van bijvoorbeeld klantprofielen bij Albert Heijn Bonuskaart, het meten van navigatiegedrag van gebruikers van sociale netwerksite Facebook en het doen van correlatie analyses op strafblad systemen.

Een betere omschrijving zou zijn ‘knowledge mining from data’. Volgens Han en Kamber (2006) zou dit te lang zijn wat tot een afkorting zou leiden; ‘knowledge mining’. Echter zou dit niet tot verbeelding spreken, aangezien een belangrijk element in datamining de grote hoeveelheden data zijn die gebruikt worden. Om deze reden werd datamining als term populair (Hearst, 1999). Een synoniem voor datamining is ‘Knowledge Discovery from Data’ (KDD), wat een betere omschrijving is van wat de methoden en algoritmen in essentie eigenlijk doen. Dat kennisvergaring uit (publieke) databases discussies oplevert blijkt uit de gekozen casestudies.

De ontwikkeling van datamining hangt nauw samen met de ontwikkeling van de database zelf (Brookshear, 2009). De methoden en algoritmen van datamining maken actief gebruik van databases. Het verschil tussen datamining en een database zit volgens Brookshear (2009) in de manier van het vergaren van informatie. Bij databases kan data worden opgevraagd door opdrachten te geven. Door alleen data op te vragen uit datasets wordt zeker nog geen datamining toegepast. Dergelijke opvragen hebben een bekende vraagstelling. De gebruiker zoekt naar specifieke data volgens bekende patronen in tegenstelling tot datamining, waarin de gebruiker probeert onbekende patronen op te sporen.

2.2 Data rich, information poor

Waarom wordt datamining überhaupt toegepast? Dit heeft te maken met “we are data rich, but information poor” (Han en Kamber, 2006; p. 4). Data opslag heeft zijn oorsprong in de jaren ‘60. Voorheen werd data ook opgeslagen, alleen in andere vormen. Han en Kamber (2006) laten zien dat vanaf de jaren ’70 de ontwikkeling van de databases een groeispurt kreeg. De database modellen hebben veel invloed gehad op de ontwikkeling van de databases en daarmee op de ontwikkeling van datamining. Een databasemodel is een manier van structureren van de data in een dataset (Brookshear, 2009).

Doordat databases meer en meer werden ingezet door commerciële- en overheidsorganisaties nam ook de hoeveelheid opgeslagen data enorm toe. In de late jaren ’80 kwamen steeds meer vragen ten aanzien van het gebruik van opgeslagen data. Volgens Han en Kamber (2006) ontstonden zogenaamde ‘datatombes’ die niet of nauwelijks werden bezocht. Slechts delen van de databases werden gebruikt, ondanks dat organisaties veel meer zouden kunnen weten over bijvoorbeeld consument en burger. Data werd alleen gebruikt voor het voorafgestelde doel waarvoor ze werden opgeslagen. Maar denkbaar is dat in het voorbeeld van de database van Facebook juist de combinaties van data tot interessante commerciële inzichten kunnen leiden. De hoeveelheid te analyseren data ligt ver boven het menselijk bevattingsvermogen.

Bedrijven en organisaties kwamen in een “data rich, information poor” situatie (Han en Kamber, 2006; p. 5). Ze hadden veel data, maar wisten nog niets. Krachtige analysetools en software werd nodig om nieuwe betekenisvolle informatie uit databases te halen. Datamining werd daarmee een belangrijk middel om dergelijke informatie uit de grote hoeveelheden data te halen. De kracht van datamining is dat juist de, tot dusverre “dode” datatombes konden worden aangesproken, waarbij gepoogd wordt patronen te ontdekken die nog niet bekend waren. Het gat tussen data en echte informatie wordt dus ingevuld met datamining. De toepassing om op geautomatiseerde wijze nieuwe patronen te ontwikkelen en daarmee informatie controversieel zoals duidelijk wordt bij het bespreken van de casestudies.

2.3 Hoe werkt Datamining?

Diverse auteurs beschrijven de inhoud van het proces van datamining op verschillende wijze (Han & Kamber, 2006 & Brookshear, 2009). Als overeenkomst is aan gegeven dat zij allen drie fasen benoemen, de ‘pre datamining-fase’, de ‘datamining-fase’ en de ‘post datamining-fase’. Figuur 1 (uit hoofdstuk één) geeft een schematische weergave van deze processen. Deze fasering van datamining is van belang bij het bespreken van de casestudies deze worden hieronder nader uitgewerkt.

2.3.1. Pre datamining-fase

De ‘pre datamining-fase’ beslaat het proces van het selecteren van geschikte data en het vaststellen van een onderzoeksvraag. Deze fase wordt ook wel de ‘data preprocessing’ genoemd (Han en Kamber, 2006). De fase wordt omschreven als “gathering and processing the data” (DeRosa, 2004; p. 9). Om datamining te kunnen toepassen is een ‘mining-vraag’ en een ‘mining-bestand’ nodig (Siebes, 1996). De mining-vraag beschrijft welke soort informatie nodig is en het mining-bestand is de database waaruit de mining-vraag beantwoord gaat worden (Siebes, 1996). In dit geval wordt het begrip database gebruikt als een verwijzing naar de daadwerkelijke gegevens set. Eerst dient vast te worden gesteld welke databases de bruikbare data hebben. Wanneer iemand bijvoorbeeld wil kijken of klanten met bepaalde demografische kenmerken van een bepaalde supermarkt vaak product X kopen in combinatie met Product Y heeft hij/zij verschillende bronnen ter beschikking. Namelijk de database met de klantgegevens en de database met transacties, gekochte producten en klantnummers. Doordat deze databases onderling niet gekoppeld zijn is het moeilijk om hierop analyses te doen en daarmee de mining-vraagstelling te beantwoorden. De informatie dient te worden opgeslagen in een nieuwe (statische-) database waarin deze gegevens wel gekoppeld kunnen worden (Brookshear, 2009).

Dit proces van aggregatie en reallocatie wordt ook wel ‘datawarehousing’ genoemd (Han en Kamber, 2006; DeRosa, 2004; Siebes, 1996 en Brookshear, 2009). Het verschil tussen een datawarehouse en een database is het dynamisch karakter van de database. In de database kan continu nieuwe data

worden toegevoegd waardoor het voor een analist moeilijk is om een analyse hierop te doen. In een datawarehouse is de data statisch en is het resultaat van opslag uit verschillende databases (Brookshear, 2009). Voordelen van datawarehousing zijn efficiëntie en nauwkeurigheid van de analyse en daarom wordt vaak gekozen voor 'datawarehousing' (DeRosa, 2004).

In datawarehousing zijn drie detailfasen te benoemen; 'Extract', 'Transform' en 'Load' (ETL) (Vassiliadis et. al. 2002). Om de juiste data te selecteren moet irrelevante data eerst verwijderd worden (Extract). Han en Kamber (2006) noemen dit proces ook wel het verwijderen van storende data. Wanneer de data is geselecteerd en is 'schoongemaakt', dient vervolgens de data vertaald te worden naar één formaat (Transform). Zo kan het voorkomen dat in de ene database de regel met het adres van de klant 'adres' wordt genoemd en in de andere database deze regel 'adresgegevens'. Deze regels worden dan veranderd in een formaat. In deze fase worden de bestanden ook klaar gemaakt voor datamining. Daarna kan de data in de statische 'datawarehouse' worden opgeslagen (Load) (Vassiliadis et. al. 2002). Op deze database kan datamining worden toegepast. Figuur 1 laat zien dat persoonsgegevens in databases worden gestopt.

2.3.2. Datamining-fase

Datamining is het proces van het analyseren van data vanuit verschillend perspectief om deze zo samen te vatten in bruikbare informatie (Cocx, 2009). Afhankelijk van de mining-vraag wordt een bijbehorende methode gekozen. Brookshear (2009) onderscheidt verschillende methoden voor analyse; beschrijven van categorieën, onderscheiden van categorieën, clusteranalyse, associatie analyse, outlier analyse en de sequentiële analyse. Figuur 1 laat zien dat op de geselecteerde database datamining kan worden toegepast. De zandlopers laten het proces van datamining zien. Het is een algoritme wat kan worden toegepast in een database systeem. In bijlage 2 staat een overzicht van de meest gebruikte analyses en de meest gebruikte algoritme bij het toepassen van datamining. In bijlage 3 staat een voorbeeld van een Apriori analyse.

2.3.3. Post datamining-fase

In de datamining-fase worden de resultaten gecreëerd. Deze dienen echter nog te worden geïnterpreteerd. Dit gebeurt in de 'post datamining-fase', "[it] involves conducting the searches, interpreting the results, and making decisions about how to use these results" (DeRosa, 2004; p. 12). Figuur 1 geeft een overzicht van de belangrijkste handelingen in deze fase. Een analist dient de resultaten van datamining te analyseren en te interpreteren. Vervolgens kunnen de resultaten door organisaties worden ingezet. Commerciële organisaties kunnen de resultaten commercieel inzetten door bijvoorbeeld bezoekers van websites te voorzien met persoonlijke advertenties en overheidsinstellingen kunnen de resultaten beleidsmatig inzetten.

Op basis van de analyses van datamining kan een voorspelling worden gemaakt over het gedrag van mensen, en kunnen modellen over het gedrag van mensen worden gemaakt (Siebes, 1996). Niet een noodzakelijke stap maar wel een gewenste stap, is het reflecteren op de resultaten. Levert de analyse de juiste informatie op en zijn hierin drogredenen te herkennen (DeRosa, 2004). De informatie dient voornamelijk geïnterpreteerd te worden door de analist. Echter wordt in veel commerciële toepassingen van datamining deze interpretatie volledig automatisch gedaan (DeRosa, 2004). Over beleidstoepassingen zoals bij het HKS kan dit proces van interpretatie en inzet een lange termijn bestrijken. De vraag is dan of deze resultaten mogen worden toegepast als bijvoorbeeld strafrechtelijk midden of dat van uit privacy perspectief wel wenselijk is (Cocx, 2009).

Han en Kamber (2006) laten in hun geschiedenis beschrijving zien de discussie over datamining en gevolgen voor maatschappij recentelijk meer belangrijk is geworden. Het oneigenlijke gebruik van data en problematiek die daarmee verband houdt, leidde ertoe dat een aantal informatici Privacy Preserving datamining ontwikkelden (zoals Agrawal en Srikant, 2000).

2.4 Privacy Preserving datamining

De discussies omtrent datamining richten zich met name op privacy. Privacy wordt geschonden wanneer datamining wordt toegepast. Uit de casestudies zal duidelijker worden welke aspecten van

privacy worden geschonden. Eerst wordt de technologie PPDM nader besproken. Om de problematiek tegemoet te komen is een ontwikkelingsrichting *Privacy Preserving datamining* gaande (Han en Kamber, 2006).

Privacy Preserving datamining wordt ook wel 'Privacy enhanced datamining' genoemd of 'Privacy sensitive datamining' (Han en Kamber, 2006). Het is data exploratie methodiek waarbij gepoogd wordt de privacy van de oorspronkelijke data te behouden. Privacy Preserving datamining houdt zich bezig met de vraag of we accurate analyses kunnen doen zonder dat de precieze informatie van individuen wordt geraadpleegd (Agrawal en Srikant, 2000).

Deze informatie kunnen zowel persoonsgegevens als andere data bevatten. Voor de Bonuskaart van Albert Heijn zou dat betekenen dat Albert Heijn wel het koopgedrag van consumenten kan analyseren, maar dat de opgeslagen persoonsgegevens, die consument dient op te geven bij het aanvragen van de kaart niet in relatie met elkaar gebracht kan worden. Privacy Preserving datamining kijkt daarmee dus naar de totale dataset en niet naar individuele persoonsgegevens.

Bij PPDM wordt de nadruk gelegd op de toegang tot de individuele persoonsgegevens. De gegevens die mensen achter laten op bijvoorbeeld Facebook of bij de aanvraag van een Bonuskaart Zoals Figuur 1 laat zien worden persoonsgegevens, die men moet achterlaten voor bijvoorbeeld het aanvragen van een Bonuskaart, afgeschermd. Privacy Preserving datamining wordt vervolgens toegepast op de afgeschermde persoonsgegevens. Het overige proces van datamining, zoals deze uiteen is gezet in Figuur 1, verloopt hetzelfde. De stippellijn illustreert het bereik van PPDM. De persoonsgegevens in datasets worden beschermd. Hierbij wordt vanuit gegaan van het principe dat wanneer de toegang niet mogelijk is, privacy van deze data en daarmee van deze personen wordt behouden. Deze probleemstelling wordt nader besproken tijdens de casestudies. Het zal blijken dat de discussies over privacy en datamining niet alleen gaan over persoonsgegevens die men dient achter te laten maar ook over de analyse van deze gegevens en de inzet van de resultaten van analyse.

Er zijn globaal twee soorten methoden om PPDM toe te passen; de datamodificatie-methoden en 'Secure Multi Party Computation' (Aggarwal en Yu 2008; Vaiyda et. al., 2006). De data modificatiemethoden richten zich op het vervormen van de oorspronkelijke data. Aggarwal en Yo (2008) onderscheiden een aantal methoden; randomisatie, transformatie, aggregatie, generalisatie en suppressie (Aggarwal en Yu, 2008). In het geval van toepassing op Albert Heijn casus zou dit betekenen dat de Bonuskaart data wordt gemodificeerd. De analist kan alleen gemodificeerde data inzien. Een voorbeeld van randomisatie kan zijn dat kassabonnen van consument X worden vervangen door de kassabonnen van consument Y worden en andersom. Op deze manier kan de analist de individuele gegevens wel inzien maar is het voor hem onmogelijk om de consument te koppelen aan de juiste kassabon. Een ander voorbeeld van randomisatie kan zijn dat de woonlocaties van consument X en Y worden gewisseld. In de ontwikkelingsrichting van Privacy Preserving datamining wordt vanuit gegaan dat privacy in dit geval behouden blijft. De analist kan nog steeds analyseren en patronen in data ontdekken.

2.5.1 Datamodificatie

Binnen de data modificatie-methoden zijn een aantal subfases te onderscheiden (Aggarwal en Yu, 2008). 1) Data wordt in een database gezet door het proces van datawarehousing en 2) vervolgens wordt een 'laag' aangebracht over deze data (Aggarwal en Yo, 2008). Deze laag betreft dezelfde data die in de database staat maar dan op een gemodificeerde wijze. De analist kan contact maken met deze laag maar niet met de oorspronkelijke database. Op die manier kan de analist alleen de gemodificeerde data raadplegen. De analist kan daarom nooit de persoonsgegevens koppelen aan een consument. Tijdens de toepassing van datamining wordt de data teruggezet naar de oorspronkelijke structurering van de database zodat de analyse toch wordt gedaan op de oorspronkelijke niet gemodificeerde data. Dit leidt ertoe dat de oorspronkelijke data beschermd blijft en dat toch analyses kunnen worden gaan op grote hoeveelheden data zonder dat privacy geschonden wordt.

2.5.2 Secure Multi Party Computation

Secure Multi Party Computation is een manier waarbij data afkomstig van verschillende partijen wordt gecombineerd om daarop datamining toe te passen (Aggarwal en Yu, 2008). Privacy bescherming wordt geboden door de integratie van encryptie en Multi Layered Security. Data blijft bij de oorspronkelijk partij in bezit- en is alleen toegankelijk voor diegene waarbij de data is opgeslagen. De toegang wordt gerealiseerd door lagen van toegang tot de data. Partijen kunnen daardoor alleen het eigen deel van de data raadplegen. Alleen dat deel wat door hen zelf is opgeslagen en verzameld is voor hen toegankelijk. Hiertoe zullen partijen naar verwachting een overeenkomst hebben gesloten. Hier is immers, als het goed is, een overeenkomst mee gesloten. Data wordt niet doorgespeeld aan derden maar kunnen toch datamining analyses worden gemaakt. Data kan worden gecombineerd zonder dat data aan derden inzichtelijk worden gemaakt. De kracht van datamining is datasets te combineren en algoritmen hierop los te laten zodat nieuwe en onverwachte patronen in de ruwe data kunnen worden ontdekt (Vaiyda et. al., 2006). Een andere oplossing is dat binnen een organisatie databases op verschillende niveaus toegang hebben door middel van autorisatie structuren. De organisatie zou ervoor kunnen kiezen om de analisten alleen toegang te geven voor bepaalde delen van de data. Zodat de persoonsgegevens door deze analisten niet kunnen worden geraadpleegd. Dit kunnen organisaties bereiken door Multi Layered Security toe te passen. De toegangniveaus dienen bepaald te worden door de eigenaren van databases.

Nu duidelijker is geworden wat datamining is, waarom het wordt toegepast, hoe het werkt en wat Privacy Preserving datamining is kan worden gekeken naar de problemen die voorvloeien uit de toepassing van datamining. Dit is kort aangehaald in hoofdstuk 2 maar wordt nu uitgebreid beschreven in hoofdstuk 3 en bij het bespreken van de casestudies in hoofdstuk 4. Het zal blijken dat Privacy Preserving datamining voor slechts- gedeeltelijk negatieve effecten die samenhangen met datamining een uitkomst kan bieden.

3. Problematiek die voortvloeit uit datamining

Om een antwoord te kunnen formuleren op de vraag in hoeverre Privacy Preserving datamining als een oplossing kan fungeren voor de problematiek die voortvloeit uit datamining dient uitvoerig besproken te worden wat die negatieve gevolgen zijn van datamining. Vervolgens kunnen deze negatieve gevolgen worden vergeleken met de hiervoor genoemde oplossing, PPDM, voor deze negatieve gevolgen. De negatieve gevolgen zijn ingedeeld in vier categorieën met discussies. Hieronder wordt eerst besproken waarom de negatieve gevolgen zijn ingedeeld in deze vier categorieën, vervolgens worden de categorieën besproken en worden meningen en negatieve gevolgen die worden genoemd in literatuur over datamining ingedeeld in volgens deze categorieën. In het navolgende hoofdstuk wordt een verdiepingslag op deze negatieve gevolgen gemaakt aan de hand van drie casestudies.

3.1 Categorisering en verantwoording

De ongewenste effecten die voortkomen uit datamining richten zich vooral op problematiek omtrent bepaalde aspecten van privacy (Solove, 2008). Privacy is een complex begrip (Lyon, 2007). Hieronder worden de discussies gecategoriseerd en wordt er laten zien waarom privacy een complex begrip is. Vervolgens wordt hier verantwoord waarom de discussies over datamining en privacy zijn ingedeeld in de categorieën.

3.1.1 Definitie van privacy

Tim Cocx (2009) stelt in zijn proefschrift dat de methoden en algoritmen van datamining controversieel zijn en toenemende bezorgdheid over privacy in de hand werkt. Overheden kunnen immers beleid gaan bepalen op de analyses die door Datamining worden gedaan. Zo vraagt Cocx (2009) zich af wat kan gebeuren als een overheid uitgaat van het patroon dat vrouwen vaker verslaafd zijn aan alcohol dan mannen. En wat er zou kunnen gebeuren wanneer bedrijven de gegevens commercieel gaan inzetten en de resultaten van datamining doorverkopen aan derden. Hij neemt aan dat er privacy wordt ingeleverd bij de toepassing van datamining, de vraagstelling is verbreed tot of en hoe eventuele privacy schending zich dan voltrekt.

> pri·va·cy [prajvesie] de; v(m) de mogelijkheid om in eigen omgeving helemaal zichzelf te zijn (VanDale, 2010)

VanDale (2010) definieert privacy als “de mogelijkheid om in eigen omgeving helemaal zichzelf te zijn” VanDale (2010) bespreekt een bepaalde gedraging van een persoon en is een subjectieve benadering van het begrip privacy. Privacy is, zoals Lyon (2007) laat zien, een veel complexer begrip waarvan de betekenis zowel historisch als cultureel kan verschillen (p. 203). Auteurs definiëren het begrip daarom ook verschillend. De discussie over privacy is een zoektocht naar wat privaat is en wat publiek. Een ruime interpretatie en een eerste definitie van wat privacy komt van Warren en Brandeis (1890); ze definiëren privacy als “The Right to be left alone”. Onder deze opvatting kunnen vele aspecten van het dagelijkse leven vallen. Volgens Solove (2008) is dit een onwerkbare opvatting van privacy, aangezien deze niet kan worden getoetst.

3.1.2 Dimensies van privacy

Privacy heeft een vijftal dimensies; 1) een lichamelijke dimensie, 2) een informationele dimensie, 3) een communicationele dimensie, 4) een sociale dimensie en 5) een ruimtelijke dimensie (Lyon, 2007). Lyon (2007) licht deze dimensies toe. De dimensies vertonen raakvlakken en onderlinge overlap. Een lichamelijke benadering van privacy gaat over de vrijheid van het lichaam en de bescherming hiervan. In Nederland is dit onder andere geregeld in de grondwet, hierover later meer. Deze dimensie gaat ook bijvoorbeeld over het lichaam als privé omgeving, naakt zijn thuis is een individuele keuze en het lichaam wordt dan als privaat gezien. De informationele dimensie gaat over privacy gevoelige gegevens en de controle hierover. De communicationele dimensie van privacy gaat over de controle over communicatie van een persoon. Het gaat hierbij bijvoorbeeld om privé gesprekken, waarbij het de bedoeling is dat niemand meeluistert. De communicatie vormt dan een privaat gebied. De sociale

dimensie gaat over gevoelens van privacy en vooral over de verhoudingen tot anderen. Privacy wordt vaak gedefinieerd in relatie tot een ander. Ook vragen over de inmenging van overheden in het privé leven van burgers valt onder de sociale dimensie. De ruimtelijke dimensie heeft betrekking tot een fysieke locatie. De discussies gaan dan over welke ruimtes gezien worden als privaat en welke ruimtes worden gezien als publiek. Surveillance camera's zijn een belangrijk onderdeel van deze discussie. Een camera op een marktplein wordt doorgaans gezien door burgers als gewoon maar een camera in iemands badkamer wordt niet geaccepteerd.

3.1.3 Sociale- en informationele dimensie van privacy

Bij de discussies over datamining worden met name vragen gesteld over de sociale dimensie en de informationele dimensie van privacy (Solove, 2008 ; Brookshear, 2009). De informationele dimensie die Lyon (2007) noemt heeft vaak betrekking op de discussie over wat privacy gevoelige informatie is. Net als over privacy zelf verschillen daarover de meningen en is de betekenis hiervan afhankelijk van de historische en sociale context (zoals Inness, 1992 ; Solove, 2008). Julie Inness (1992) beschrijft privacy gevoelige informatie als 'intieme' informatie. Het recht op privacy uit zich volgens deze auteur in tot de toegang tot deze informatie en beslissingen nemen over deze informatie (Inness, 1992 in: Solove, 2008). Het begrip 'intiem' is in deze context subjectief. Dit levert volgens Solove (2008) problemen op aangezien we veel informatie niet zien als intiem. Denk hierbij aan gegevens van banken en ID nummers. Dit zijn geen intieme gegevens maar zijn wel privaat en kunnen daarmee nog steeds privacy gevoelig zijn. Wanneer we het begrip 'intiem' breder zouden hanteren zou het een synoniem kunnen worden van wat privacy is. De sociale dimensie van privacy heeft betrekking op het gevoel dat mensen hebben wanneer ze het hebben over privacy. Een belangrijk kenmerk van privacy is dat het altijd gedefinieerd wordt in relatie tot een ander of tot de uitsluiting van een ander zoals dat bij de definitie van VanDale (2010) toonbaar wordt.

3.1.4 Categorisering van privacy

Discussies over privacy die gaan over deze informationele en sociale dimensies zijn in dit onderzoek onder verdeeld in vier categorieën. De keuze voor deze categorieën is gemaakt aan hand van de reacties op de nieuwsberichten over de cases en op basis van bepaalde kenmerken;

1) Discussies die gaan over de bescherming van de persoonsgegevens;
2) Discussies die gaan over de analyse en inzet van deze persoonsgegevens;
3) Discussies die gaan over een inbreuk op privacy;
4) Relativerende discussies.

Bij iedere reactie op een nieuwsbericht is er gekeken naar welke fase van datamining zij aanspreken. Deze reacties zijn gegroepeerd in de bovenstaande categorieën. Discussies die gaan over de bescherming van de persoonsgegevens richten zich vooral op de gegevens die men ter beschikking stelt aan bedrijven en overheden. Het gaat bijvoorbeeld over het rechtmatige gebruik van de persoonsgegevens voor opslag en datamining. De discussies gaan over of men zomaar mag alles registreren en analyseren.

Discussies die gaan over de analyse en inzet van deze persoonsgegevens gaan over het analyseren van deze persoonsgegevens en het interpreteren van de resultaten. Deze discussies gaan ook over het inzetten van gegevens voor bijvoorbeeld commerciële of wetshandhaving doeleinden. De discussies gaan over wat mag worden gedaan met deze analyses en waarvoor deze analyses mogen ingezet.

Discussies die gaan over de inbreuk op privacy gaan over het doorverkopen van de persoonsgegevens aan derden en het koppelen van gegevens aan andere bronnen van data. Door middel van datawarehousing kunnen databases worden gekoppeld aan andere databases. Hierop kan weer datamining worden toegepast. De discussies gaan over of de gegevens doorverkocht worden en wat allemaal aan de persoonsgegevens gekoppeld kan worden.

Relativerende discussies zijn reacties op nieuwsberichten en auteurs die de problematiek van datamining niet onderkennen of er geen problemen mee hebben dat datamining wordt toegepast. Logischerwijs is Privacy Preserving datamining hier niet toepasbaar aangezien de probleemstelling

niet onderkend wordt of voor deze laatste categorie betrokkenen geen issue vormt. Andere gebruikers en auteurs bekritiseren dit standpunt.

Er bestaat een zekere mate van overlap tussen de eerste drie genoemde categorieën. Soms gaan discussies over de inzet van de persoonsgegevens juist met name over het doorverkopen aan derden en gaan discussies over de bescherming van de persoonsgegevens ook over de het inzetten van deze gegevens voor commerciële doeleinden. De indeling die gegeven is helpt om beter te begrijpen waarvoor Privacy Preserving datamining als een oplossing fungeert. Het kan helpen te begrijpen welke categorieën discussies wel worden aangesproken en welke niet. Deze categorisering worden hierna gehanteerd als kader bij het bespreken van de casestudies. Uit de casestudies zal blijken dat de kern van de discussies vooral betrekking heeft op datamining en privacy.

3.2 Bescherming persoonsgegevens

Discussies over bescherming van de persoonsgegevens richten zich met name op vragen over wat men wel of niet mag opslaan en gebruik en wat men wel of niet mag gebruiken (Custers, 2004, Vedder, 1998). De discussies richten zich op wat de bedrijven en organisaties willen of kunnen doen met de gegevens en wat ze mogen doen met de gegevens.

3.2.1 Privacy in juridisch kader geplaatst

Zo laat jurist Custers (2004) zien dat ons recht op privacy vaak te kort schiet bij het beschermen van groepsgegevens. Ons recht is erop gericht, volgens Custers (2004), om individuele persoonsgegevens te beschermen. Datamining maakt gebruik van grote datasets, waar op basis van kenmerken analyses kunnen worden gedaan. Groepsgegevens, die gedeeltelijke anoniem kunnen zijn, worden niet beschermd onder de Wet Bescherming Persoonsgegevens. Ook allerlei andere specifieke wetten helpen hierbij niet. Custers (2004) pleit voor een bescherming van groepsgegevens, door een apart artikel op te nemen in deze WBP. Vedder (1998) laat zien dat het juridische debat over datamining zich inderdaad lijkt te focussen op deze individuele gegevens. Vedder (1998) geeft aan dat de persoonsgegevens niet zozeer meer aandacht verdienen maar dat door datamining juist de discussie aandacht dient te hebben voor groepsgegevens. Door het proces van datamining kunnen persoonsgegevens, die oorspronkelijk vielen onder de bescherming van het WBP, opeens deze bescherming missen. De technologie zorgt ervoor dat algemene profielen worden gegenereerd op basis van persoonsgegevens. Deze profielen zijn niet zonder meer direct te herleiden naar individuen. Ook Vedder (1998) geeft aan dat ons recht op privacy te kort schiet in bescherming tegen technologieën als datamining. Cocx (2009), geeft aan dat wie een exploratief onderzoek doet naar de mogelijkheden van datamining voor doeleinden voor wetshandhaving datamining niet mag toepassen volgende de Nederlandse wet. De data die is opgeslagen binnen het Herkenningsdienstsysteem mag alleen worden gebruikt in het kader van de doelstellingen waarvoor de data is opgeslagen. Echter kan het natuurlijk zo zijn dat termijn een nieuw systeem ontwikkeld wordt waarbij doelstellingen als datamining wel zal worden opgenomen. Dan ontstaat een nieuwe situatie zoals Vedder (1998) en Custers (2004) aangeven en waarmee het recht op privacy in Nederland verder geproblematiseerd wordt.

3.2.2 Privacy in wetgeving

In Nederland wordt het recht op privacy op verschillende plekken juridisch gedefinieerd maar het is bovenal ook een grondrecht. Privacy is in juridische zin in Nederland; “het recht op eerbieding van [iemands] persoonlijke levensfeer” (art. 10 lid 1 Gw). Daaronder valt de bescherming van persoonsgegevens (art. 10 lid 2 Gw) en bescherming op de aanvraag van deze persoonsgegevens. Ook het recht op onaantastbaarheid van het lichaam (art. 11 Gw) en het recht op een eigen woning (art. 12 Gw) valt onder het recht op privacy. In internationale verdragen is privacy eveneens een bijzonder goed. Zo is privacy ook gedefinieerd in art. 8 van het Europese Verdrag van de Rechten van de Mens als;

> “Everyone has the right to respect for his private and family life, his home and his correspondence”. (art. 8 lid 1 EVRM)

In het EVRM worden echter ook beperkingen aangegeven ten aanzien van dit recht. Ook in Nederland kleven hieraan beperkingen zoals bij een strafrechtelijk onderzoek. In Nederland wordt de voorkeur gegeven aan toepassing van het Nederlands recht in het geval van privacy gerelateerde zaken. Soms kan bij bijzondere omstandigheden een beroep gedaan worden op dit Europese recht (art. 96 en 97 Gw).

3.2.3 Privacy en de Grondwet

De invulling van het grondrecht privacy in de zin van art. 10 Gw, zoals hierboven uiteen is gezet, en de naleving daarvan is onder meer geregeld in de Wet Bescherming Persoonsgegevens (Wbp). De overige artikelen in de Grondwet (art. 11 en 12 Gw) met betrekking tot privacy zijn onder meer geregeld in het wetboek van Strafvordering en het wetboek van Strafrecht. Het recht op privacy zoals deze is geformuleerd in art. 10 van de grondwet valt uiteen in een aantal subgrondrechten. Dat zijn;

> "De wet stelt regels ter bescherming van de persoonlijke levenssfeer in verband met het vastleggen en verstrekken van persoonsgegevens." (art. 10 lid 2 Gw)

> "De wet stelt regels inzake de aanspraken van personen op kennisneming van over hen vastgelegde gegevens en van het gebruik dat daarvan wordt gemaakt, alsmede op verbetering van zodanige gegevens." (art. 10 lid 3 Gw)

Deze twee rechten worden verder uitgewerkt in de WBP. Voor de toepassing van datamining voor commerciële doeleinden zijn veelal persoonlijke gegevens of persoonsgegevens nodig. Dit lijkt in strijd met de bovenstaande grondrechten. Echter zijn bepalingen gemaakt binnen de WBP om persoonsgegevens toch te verkrijgen en te gebruiken voor deze doeleinden. In deze wet wordt bepaald welke plichten organisaties hebben voor het omgaan persoonsgegevens.

Hier wordt gekeken of de data die gebruikt wordt voor datamining kunnen vallen onder de bescherming van de Wet Bescherming Persoonsgegevens en als dit het geval is waar wordt dan een inbreuk gemaakt op het recht van bescherming van de persoonsgegevens. De belangrijkste artikelen uit dit wetboek worden besproken. wordt hier uitgegaan van een commerciële inzet van Datamining want voor binnen de wetshandhaving zijn veel uitzonderingen.

De reikwijdte van de Wet Bescherming Persoonsgegevens wordt aangetoond in art. 2 WBP;

> "Deze wet is van toepassing op de geheel of gedeeltelijk geautomatiseerde verwerking van persoonsgegevens, alsmede de niet geautomatiseerde verwerking van persoonsgegevens die in een bestand zijn opgenomen of die bestemd zijn om daarin te worden opgenomen." (art. 2 lid 1 Wbp).

Hieruit vloeien een aantal rechtsvoorwaarden voort. Het dienen 1) persoonsgegevens te zijn, 2) die opgenomen zijn in een bestand of daarvoor bestemd zijn, 3) die (niet-)geautomatiseerd (geheel/gedeeltelijk) verwerkt worden. Het gevolg is dat de gebruik van persoonsgegevens wordt beschermd onder deze wet.

3.2.4 Datamining in juridisch kader

Omdat, zoals uit hoofdstuk één bleek, dat datamining vooral wordt gebruikt om patronen in gedragingen van mensen te voorspellen is het aannemelijk dat de gegevens persoonsgegevens zijn in de zin van art. 1 sub a WBP. Dit wetsartikel stelt namelijk dat dit gegevens zijn die verwijzen naar een identificeerbaar persoon. Er zijn twee soorten gegevens; indirecte gegevens en bijzondere gegevens. Indirecte gegevens zijn gegevens die iets vertellen over een persoon, dat kan een bepaalde aankoop zijn maar ook zijn of haar geslacht. Bijzondere gegevens worden gedefinieerd in art. 16 WBP en gaan vooral over persoonlijke gronden als bijvoorbeeld ras en seksuele voorkeur. De opslag van de bijzondere gegevens is verboden tenzij daarvoor uitzonderingen zijn. Uitgaande dat binnen commerciële datamining deze bijzondere persoonsgegevens niet worden geregistreerd en voldoet datamining in deze toepassing aan de eis van de bescherming persoonsgegevens volgens de wet.

Datamining en art.1 WBP

Een bestand is een gestructureerd geheel van persoonsgegevens (art. 1 sub c WBP). Het opslaan van gegevens in een bepaalde database is automatisch een gestructureerd geheel. De gegevens worden opgeslagen en kunnen worden gekoppeld aan elkaar. In het proces van datawarehousing gebeurt deze structurering opnieuw. Met verwerken worden iedere handelingen of ieder geheel van handelingen bedoeld ten aanzien van de persoonsgegevens (art. 1 sub b WBP). Dit kan een proces zijn van analyseren maar ook van opvragen en raadplegen. Opvallend is dat de WBP vooral kijkt naar verwerken en niet zozeer naar de opslag van persoonsgegevens. Verwerken is een breder begrip dan alleen opslag. Onder verwerken valt wel de opslag. Datamining analyseert en structureert de gegevens en daarom is datamining aan dit verwerken toe te kennen.

Datamining en art.2 WBP

Het geautomatiseerde proces van datamining is eveneens aanwezig in de zin van art. 2 WBP. Het proces van datawarehousing is immers een geautomatiseerd proces. Ook de filtering hierin en de herstructurering is te plaatsen onder deze rechtsvoorwaarde. Datamining zelf is bovenal een geautomatiseerde verwerking van gegevens. Juist dit automatische karakter levert vragen op over de verdere inzet hiervan. De data die wordt gebruikt in datamining valt binnen het bereik van dit wetboek art. 1 en 2 Wbp en is daarmee beschermd volgens de Wet Bescherming Persoonsgegevens. Alszodanig is geen sprake van privacy schending in de zin der wet indien voldaan wordt aan de gestelde voorwaarden.

3.2.5 Voorwaarden voor gebruik (opslag en verwerking) van beschermde gegevens

De wet formuleert een aantal voorwaarden waaraan organisaties moeten voldoen bij het gebruik van deze beschermde gegevens. Ten eerste dient de organisatie toestemming te vragen aan de betrokkene en mogen de gegevens alleen worden opslagen en verwerken als daar een goede reden voor is. Ten tweede dient een organisatie uitvoerig te formuleren met welk doel de persoonsgegevens worden opgeslagen. Ten derde is hierbij belangrijk, dat de gegevens alleen worden gebruikt ten behoeve van dit vooraf geformuleerde doel. Wanneer de organisatie de persoonsgegevens in wil zetten voor een ander commercieel doel dient de organisatie dit opnieuw vooraf kenbaar te maken aan de betrokkenen. Ook wanneer ze deze gegevens willen doorspelen aan derden. Ook mag hierbij niet meer en liet langer gegevens worden opgeslagen dan voor dit doelnoodzakelijk is. Ten vierde mogen de gegevens mogen niet verder worden verwerkt op een wijze die niet overeenkomt met het voorafgestelde doel. Ten vijfde heeft de organisatie een geheimhoudingsplicht en moet de gegevens beschermen. De wet definieert bovendien een aantal rechten die betrokkene hebben ten aanzien van deze gegevens. Ze mogen de gegevens ten alle tijden en binnen vier weken opvragen, corrigeren en verbieden.

3.2.6 Privacy problematiek en datamining

De problematiek van datamining ontstaat vooral bij het formuleren van de doelstellingen van de dataopslag (art. 9, 10 en 11 WBP). Datamining vindt patronen in data waarvan men eerder nog niet had gezien dat deze patronen bestonden. Art. 9 lid 1 en lid 2 WBP is het meest problematisch bij datamining. Het artikel stelt namelijk dat persoonsgegevens worden niet verder mogen worden verwerkt op een wijze die onverenigbaar is met de doeleinden waarvoor ze zijn verkregen" (Art. 9 lid 1 WBP).

De verwerking van de gegevens lijken soms verder te gaan dan het oorspronkelijk doel waarvoor de gegevens zijn opgeslagen. Dit is afhankelijk van de aard van de betreffende gegevens en het geformuleerde doel. Wanneer bedrijven bijvoorbeeld bepaalde individuele prijzen gaan maken is dit wellicht niet hetgeen wat de betrokkenen willen. Wanneer de opslag in database A alleen voor opslaan bedoeld is en deze wordt gecombineerd met database B dat gaat dit voorbij aan het eigenlijke doel van de data opslag.

Zoals laten zien is richten de discussies uit de categorie Bescherming Persoonsgegevens zich vooral op wat wel mag en wat niet mag worden gedaan met gegevens, die worden opgeslagen in tal van

databases. Vedder (1998) en Custers (2004) geven aan dat ons recht zich te veel richt op de bescherming van individuele gegevens die worden opgeslagen. Toepassingen zoals datamining zijn niet eenduidig te plaatsen onder de bescherming van het WBP. Een juridische en politieke discussie dient verdere invulling te geven aan rol van datamining in het recht op privacy. Tevens zijn tal van uitzonderingen op deze wet denkbaar waarbij, vooral in het geval van Wetshandhaving, deze tot problemen kunnen leiden. In het bespreken van het HKS case wordt dieper ingegaan op deze strafrechtelijke uitzonderingen in het WBP.

3.3 Analyse en Inzet

De volgende vraag is dan waar de problematiek ten aanzien van privacy bij de toepassing van datamining ontstaat. Discussies over de analyse en inzet van de gegevens die worden opgeslagen in databases gaan over het dataminen zelf en de interpretatie van de analyses. Hierbij wordt vooral gericht op de 'datamining-fase' en op de 'post datamining-fase'. Wat zouden bedrijven en overheden mogelijker wijze kunnen die met de resultaten die voorkomen uit datamining

Bij discussies over datamining in relatie tot privacy gaat het vaak over het idee van 'Big Brother is watching you', zoals beschreven is in het boek *1984* van George Orwell (1949). Bij het bespreken van de casestudies zal duidelijk worden dat deze Big Brother-reactie men name bij de verhouding tussen burger en overheid wordt aangehaald. Maar ook bij de commerciële toepassing van datamining blijkt dit het geval. Dit toekomstbeeld van een totalitaire staat die zijn burgers te allen tijde in de gaten houdt wordt vaak aangehaald bij discussie over privacy, maar ook bij discussie over datamining (Solove, 2008). Door middel van data opslag en door fysieke surveillance zou de overheid alles over de burger weten en in de gaten houden.

Een andere vergelijking die wordt aangehaald in discussie die gaan over de analyse en inzet van resultaten van datamining is die van 'Minority Report'. Datamining zou bedrijven en overheden namelijk in staat stellen voorspellingen te doen over het gedrag van burgers en consumenten. Minority Report is een roman die gaat over een situatie in 2054 waarin politie in staat is mensen op te pakken al voordat zij een bepaald delict hebben gepleegd (Mena, 2003). Datamining zou politie en justitie in staat stellen om voorspellingen te doen over criminele carrièrepaden (Cocx, 2009). Aan hand van het analyseren van gegevens over strafbladen kunnen voorspellingen worden gedaan over het patroon waarin men delicten pleegt. Wanneer een persoon delict x en y heeft gepleegd zou datamining kunnen uitwijzen dat de kans groot is dat deze persoon ook delict z zou plegen. Mena (2003) geeft aan dat datamining kan leiden tot 'false positives', de modellen die worden gegeneerd met de data zijn niet perfect. Enerzijds vanwege fouten in de data en anderzijds omdat iemand die ondanks een perfecte overeenkomst op alle kenmerken nog steeds geen dader hoeft te zijn. Mena (2003) waarschuwt voor dergelijke interpretatie van de analyses die voorkomen uit datamining.

In het kader van datamining is het concept van dataveillance van belang. Privacy wordt vaak in reacties op nieuwsberichten in verband gebracht met veiligheid (Lyon, 2007). Bij deze hypothese wordt vanuit gegaan dat overheden inbreuk zouden mogen maken op privacy om veiligheid te bevorderen. Surveillance is hiervoor een belangrijk middel. Overheden houden burgers in de gaten. Elmer (2003) laat zien dat dataveillance (een term van Clarke, 1988 in; Elmer, 2003) gaat over het toepassen van de handeling van surveillance in een database. Het 'kijken naar' wordt het 'kijken in'. Lyon (2007) laat zien dat dataveillance een hybride term is wat betekent dat men niet direct wordt bekeken maar data over personen wordt geanalyseerd. Hieraan kunnen conclusies worden gekoppeld. Een risico van deze vorm van surveillance is dat het mogelijk discriminerende tendensen kan krijgen. Op basis van categorieën in persoonlijke data worden immers patronen ontdekt. De vraag is volgens Lyon (2007) wie deze categorieën definieert. Gevolgen kunnen zijn dat men sociale stereotypen hanteert waardoor uitsluiting en privileges kunnen worden gegeven op basis van sociale kenmerken. De potentiële consument evenals de potentiële overtreder van de wet worden geclassificeerd aan hand van bepaalde kenmerken. Daarnaast kunnen de gemiddelde waarden gaan gelden als de norm, men dient zich te gedragen volgens deze norm (Pieters et. al. 2010). Door het proces van selectie en normalisering kunnen bepaalde groepen met bepaalde kenmerken anders behandeld worden dan andere groepen zonder deze kenmerken.

De voorafgaande punten richtten zich met op de verhouding tussen burger en overheid. Ook is de verhouding tussen consument en bedrijf van belang in andere discussies die gaan over de analyse van de persoonsgegevens en de inzet van de resultaten van datamining. Men kan immers op basis

van data voorspellingen doen over bijvoorbeeld het koopgedrag van consumenten. Consumenten kunnen vervolgens worden benaderd met allerlei persoonlijke aanbiedingen. Dit wordt uitgebreider besproken wanneer de cases van de Bonuskaart en die van Facebook wordt besproken. De gebruikers van de websites ervaren dit als een inbreuk op zijn of haar privacy. De discussies gaan over merkbare gevolgen van datamining en men speculeert over mogelijke negatieve gevolgen van dataopslag en datamining.

3.4 Inbreuk op privacy

De serie discussies die gaan over datamining en de inbreuk op privacy zijn discussies die vooral gaan over het doorverkopen gegevens aan derden, beveiligingen van databases en de koppeling van gegevens aan andere bronnen. Datawarehousing wordt hier in meerdere mate van belang. Gegevens die zijn doorverkocht kunnen worden gekoppeld aan nieuwe gegevens. Hierop kan opnieuw datamining worden toegepast en leiden tot nieuwe inzichten.

Blogger Jan-Willem Alphenaar (2011) laat zien dat profielen van gebruikers van sociale media tot grondstoffen van de datamining industrie gemaakt. Waarbij partijen als Facebook de leverancier worden van grote hoeveelheden informatie en adverteerders afnemers zijn om commercieel belang te winnen. Hij laat zien dat de eigenaren van sociale media op een potentiele goudmijn zitten. Doordat partijen als Facebook, Linkedin en Hyves veel weten van consumenten, ze weten wat hen beweegt en bezighoudt, kunnen zij veel omzet genereren door het doorverkopen van deze gegevens aan derden. Alpenaar (2011) laat zien dat de resultaten van datamining kunnen dienen als een verkoopmiddel, maar ook de gegevens zelf zijn een belangrijke bron van inkomsten. Gebruikers van deze Sociale media accepteren de voorwaarden bij het aanmaken van een profiel. Hier geldt dat de doelstellingen van de data uitgebreid geformuleerd dient te zijn in de algemene voorwaarden. In deze doelstellingen dienen de Sociale media op te nemen dat ook datamining wordt toegepast.

Een ander belangrijk element in de discussies over de inbreuk op privacy betreft het beveiligen van de grote hoeveelheden data. Jabocs (2005) laat zien dat grote databases moeilijk te beveiligen zijn en dat de kans op misbruik groot is waarbij gegevens in verkeerde handen komen. Het element van identiteitsdiefstal “identity theft” speelt hierbij een rol. Er bestaan grote hoeveelheden gedetailleerde informatie in de databases over consument en burger (de gemiddelde Nederlander is in ca. 265 databases geregistreerd NOS2011) . Deze persoonsgegevens verwijzen vaak direct naar een individu. Jacobs (2005) geeft aan dat men moet oppassen voor de grote hoeveelheid informatie opslag omdat deze makkelijk in verkeerde handen kan komen. Tevens geeft hij aan dat identiteitsdiefstal momenteel de snelst stijgende vorm is van criminaliteit. De auteur geeft aan dat privacy vaak in commerciële instellingen een technologische benadering krijgt. Het is een formaliteit wat ondervangen kan worden met de juiste technologische bescherming en protocollen (Jacobs, 2005).

3.5 Relativerende discussies

De relativerende discussies die gaan over datamining en privacy zijn te karakteriseren met reacties als dat men niets te verbergen heeft, dat privacy toch al niet meer bestaat en dat men het waardeert dat aanbiedingen worden gedaan en dat overheden datamining toepast om criminaliteit te verkleinen. Het ‘nothing to hide’ argument is veel gebruikte reactie voor de verhoudingen tussen privacy en overheden (Solove, 2008). Een andere reactie op de discussies over privacy en datamining is de reactie dat privacy toch al niet meer bestaat en dat we dat maar moeten accepteren (McNealy in: Castells, 2001). Ook zijn auteurs die aangeven de waarde in te zien van datamining om criminaliteit te verkleinen (Cocx, 2009). Andere auteurs, veel uit economische hoek, richten zich op de verbeteringen van loyaliteitsprogramma’s om zo meer commercieel gewin te behalen.

Een onderdeel van de discussie over de sociale dimensie van privacy is het argument ‘nothing to hide’, die Solove (2008) bespreekt. Vaak wordt de problematiek omtrent privacy afgedaan met het argument; “ik heb niets te verbergen”. Solove (2008) schrijft over Datamining dat men ook hier de discussie over de inbreuk op privacy vaak afdoet met de reactie dat men niets te verbergen heeft. Wanneer iemand niets te verbergen heeft, zou een inbreuk op zijn privacy dus ook geen probleem

zijn. Dit argument staat in sterke verhouding gesteld met veiligheid en met moraliteit. Overheden zouden inbreuk mogen maken op privacy in het kader van het bereiken van veiligheid.

Het 'nothing to hide'-argument zou gelden omdat men zich juridisch en moreel 'goed' zou gedragen en dus dan het ook niet erg zou zijn wanneer hij gesurveilleerd wordt. Solove (2008) relateert dit argument aan Datamining toegepast door de overheid. Wanneer overheden gegevens opslaan van hun burgers en deze analyseren zouden mensen deze inbreuk op hun privacy vaak afdoen met de reactie dat ze niets te verbergen hebben. Men vertrouwt de overheid in de afhandeling van privacy gevoelige gegevens en ze gaan vanuit dat wanneer zij niets fout hebben gedaan zij ook geen problemen krijgen. Ze gaan vanuit dat deze informatie besloten ligt in de databases van de overheid; oftewel niet toegankelijk voor derden. Dit vertrouwen uit zich ook in de verwachting van het soort informatie dat wordt opgeslagen. Solove (2008) laat zien dat mensen verwachten dat overheden niet naaktfoto's van hun burgers opslaan. Wanneer mensen in deze database staan en het proces van datamining in gang wordt gezet achten mensen volgens Solove (2008) de kans dat data van een individu wordt geraadpleegd niet groot. Ze gaan vanuit dat door de omvang van de data men nooit individueel patronen kan ontdekken. dienen dus computermatige filters te worden toegepast om opvallende patronen te ontdekken. Wanneer iemand niets te verbergen heeft zal hij dan ook niet bij de selectie van filtering komen, dus zal zijn data niet worden geraadpleegd. wordt bovendien vanuit gegaan dat de informatie niet per se gevoelig is dus dat dan de inbreuk op privacy laag is. De waarde veiligheid ligt dan hoger dan de waarde privacy. Het probleem met dit 'nothing-to-hide' argument is dat het aandacht geeft aan slechts delen van het debat, het ondervangt bepaalde delen maar vergeet andere problemen te benoemen (Solove, 2008; p. 772). Het is een te smal begrip van een concept als privacy. De problemen van datamining projecten zijn te omvangrijk om te ondervangen met "ik heb niets te verbergen".

In dit hoofdstuk zijn de voornaamste discussies over datamining en privacy aangehaald. In het kader van de vraagstelling *in hoeverre Privacy Preserving datamining een oplossing kan bieden voor de negatieve effecten van datamining* kan worden getoetst of dit het geval is. Echter is de verwachting dat de inhoud van concrete situaties bepalend is voor de discussies over privacy. Hierna worden drie situaties geschetst waarbij wordt gekeken naar de voornaamste discussies over deze situaties. Vervolgens wordt gekeken of Privacy Preserving datamining kan dienen als een oplossing voor de discussies in die situaties. De analyse die is gedaan in dit hoofdstuk zal worden vergeleken met de casestudies in hoofdstuk vier. Op die manier kan gekeken worden op een algemener niveau of Privacy Preserving datamining alle problematiek ondervangt. Veel van de elementen uit de bovengenoemde discussies zullen terug te vinden zijn in de reacties op nieuwsberichten die geanalyseerd worden.

4. Casestudies

De casestudies die hierna zullen worden besproken gaan over de Bonuskaart van Albert Heijn, het Herkenningsdienstsysteem van het KLPD en het CBS en de database van Facebook. Bij iedere casestudie wordt eerst gekeken naar de problematiek. De cases worden geïntroduceerd met de nieuwsberichten. Daarna wordt ingegaan op wat de case is. Vervolgens wordt gekeken naar de doelstellingen van de organisatie voor de database en gegevensopslag. Dit is noodzakelijk om inzicht te genereren in waarom dergelijke gegevensopslag plaatsvindt en wat het bedrijf of de overheidsinstelling doet met deze gegevens. Vervolgens wordt gekeken naar het privacybeleid van de opgeslagen gegevens. Dit geeft nog meer inzicht in wat het bedrijf of de overheidsinstelling voor ogen heeft met deze gegevens. Daarna worden de discussies die ontstaan over datamining, de case en privacy aan de hand van de vier genoemde categorieën besproken. Dit is noodzakelijk om een goede vergelijking tussen enerzijds de analyse uit hoofdstuk twee te maken en anderzijds om de cases onderling te vergelijken. Tot slot wordt iedere keer gekeken in het kader van de hoofdvraag in hoeverre Privacy Preserving datamining kan fungeren als een oplossing voor de discussies die worden geschetst. Iedere keer wordt de vraag gesteld over wat wordt gedaan met de achter te laten persoonsgegevens, waarom dit wordt gedaan, welke discussies dit oplevert en of Privacy Preserving datamining als een oplossing kan fungeren voor deze problematiek.

4.1 Casestudie 1: de Bonuskaart

In juli 2006 circuleerde er een gerucht dat een RFID-chip gecombineerd werd met de Bonuskaart. De Bonuskaart met de barcode zou worden vervangen door een Bonuskaart met de RFID-chip. Dit zou Albert Heijn mogelijk in staat stellen om consumenten op afbetaling te laten betalen (Webwereld, 2006) en om gericht aanbiedingen te tonen op beeldschermen wanneer de consument met de kaart zou langslopen (Tweakers, 2006). Albert Heijn gaf desgevraagd aan dat zij nog geen concrete plannen had om deze technologie in te voeren.

Een andere situatie ontstond in 2008 toen bekend werd dat Albert Heijn verplicht werd om gegevens van de Bonuskaart te delen met justitie (Security, 2008), bekendgemaakt door de nieuwszender BNR. Jaarlijks zou Albert Heijn meerdere malen zijn data delen met justitie. Het zou gaan om individuele feiten waarbij deze informatie zou dienen om in kaart te krijgen wanneer een bepaalde persoon op welke plaats en op welk tijdstip was.

4.1.1 Doelstellingen en privacybeleid Bonuskaart

De Bonuskaart is een klantenkaart om kortingen te verkrijgen bij de supermarkten van Albert Heijn en werd in 1997 geïntroduceerd (Ah.nl, 2011). Klanten konden bij de filialen van Albert Heijn deze kaart aanvragen en moesten hiervoor allerlei persoonsgegevens achterlaten. Het doel van de kaart voor de supermarkt is om het koopgedrag van consumenten te meten, te voorspellen en klantloyaliteit te bewerkstelligen (Leenheer, 2006). Albert Heijn probeert het gebruik van hun producten en diensten door consumenten te inventariseren en te analyseren (CBP, 1999). Om inzicht te krijgen in het koopgedrag van hun klanten probeert Albert Heijn op de verzamelde gegevens door middel van de Bonuskaart statistische analyses te doen (Ah.nl, 2011). Hiervoor wordt datamining toegepast. Ook kan de Bonuskaart worden gekoppeld aan een Airmileskaart. Dit is een spaarsysteem waarmee met punten bepaalde producten worden gekocht. De Airmileskaart is gekoppeld aan meerdere winkels. Deze casestudie beperkt zich tot de inzet van de Bonuskaart.

Vervolgens probeert Albert Heijn de klantenkaarthouder te informeren over producten en diensten, voordelen, kortingen en om aanbiedingen te verstrekken. De klant dient hiervoor zijn adres, e-mailadres, naam, geslacht en leeftijd achter te laten. Middels het akkoord gaan met de algemene voorwaarden voor aanvragen van een Bonuskaart geeft de klant toestemming voor de opslag van zijn gegevens.

Er wordt in beperkte mate aan de klant bekend gemaakt wat wordt gedaan met geregistreerde gegevens. Op de website van Albert Heijn, bij aanvraag van de kaart en in de algemene voorwaarden wordt nauwelijks uitgeweid over de inzet van de gegevens en de statistische analyses die worden

uitgevoerd op de opgeslagen gegevens (Ah.nl, 2011). Dit blijkt ook uit discussies onder consumenten die ontstond vlak na de introductie van de kaart (CBP, 1999). Toen werden er vragen gesteld door consumenten over de noodzaak van de registratie van de gegevens. De registratiekamer (de voorloper van het College Bescherming Persoonsgegevens), die de zaak over de Bonuskaart en het achterlaten van consumentengegevens behandelde, achtte het een feit dat Albert Heijn de klant niet voldoende informeerde over de inzet van de persoonsgegevens (CBP, 1999). Ook ontstond discussie over het doel van registratie van de gegevens. Kort na de introductie van de kaart werd Albert Heijn onder meer vanwege verplicht deze reden een anonieme variant op de markt te brengen (CBP, 1999). Hiervoor hoeft de klant zijn gegevens niet achter te laten maar kan hij wel gebruik maken van de kortingen. Voor Albert Heijn betekent dit dat ze toch inzicht kunnen krijgen op het gebruik van hun producten en diensten maar dat ze persoonsgegevens niet kunnen koppelen aan dit koopgedrag.

Een andere aanleiding die zorgde voor de introductie van de anonieme variant van de Bonuskaart gaat over vragen als de noodzakelijkheid van gegevensregistraties voor het beoogde doel. Het doel van de registratie van de gegevens is 'het gebruik van de producten en diensten door consumenten te inventariseren en te analyseren' en 'de klantenkaarthouder te informeren over producten en diensten, voordelen, kortingen en aanbiedingen te verstrekken' (Ah.nl, 2011). Voor het eerste doel van de Bonuskaart achtte de Registratiekamer registratie van persoonsgegevens niet noodzakelijk (CBP, 1999). Zonder persoonsgegevens kan Albert Heijn ook het gebruik van hun producten en diensten inventariseren en analyseren. Met de anonieme variant kan de supermarkt dus nog steeds het koopgedrag analyseren. Voor het benaderen van klanten door middel van post zijn echter wel persoonsgegevens nodig. Het daardoor mogelijk niet kunnen verkrijgen van persoonlijke kortingen komt dan voor rekening van de klant, zo oordeelt de registratiekamer (CBP, 1999). De klant mag kiezen om zijn gegevens achter te laten en daarmee te profiteren van (persoonlijke) kortingen via de post of e-mail, of mag kiezen om geen gegevens achter te laten en dus ook niet te profiteren van de kortingen per post of e-mail. Sinds kort heeft Albert Heijn op zijn website een mogelijkheid om een persoonlijk profiel aan te maken. Hieraan is de Bonuskaart gekoppeld. Hiermee kan de klant ook benaderd worden voor persoonlijke aanbiedingen en kan de klant zijn eerder gekochte producten waarvoor hij een Bonuskaart heeft gebruikt inzien. Op basis van deze lijst probeert Albert Heijn nieuwe persoonlijke aanbiedingen aan te dragen.

De gegevens die voortkomen uit het gebruik van de Bonuskaart zijn onder gebracht in een databank bij het College Bescherming Persoonsgegevens en opvraagbaar via de site van deze instantie (CBP, 1999). Op die manier voorziet Albert Heijn de klant met het recht tot inzicht in zijn gegevens. De gegevens die worden verzameld door de anonieme variant van de Bonuskaart zijn niet beschermd onder de Wet Bescherming Persoonsgegevens.

4.1.2 Discussies

De berichten zijn gecategoriseerd aan de hand van inhoudelijkheid van het bericht. Daarbij is gekeken naar welk aspect van de discussie omtrent privacy en de Bonuskaart de berichten zijn toe te schrijven. zijn vier soorten berichten te categoriseren; 1) berichten die gaan over de bescherming van persoonsgegevens, 2) berichten die gaan over de analyse en inzet van deze persoonsgegevens, 3) berichten die gaan over de inbreuk op privacy en 4) berichten die vooral relativerend of ontkennend zijn. Belangrijk hierbij is te vermelden dat deze berichten vaak te plaatsen zijn onder meerdere categorieën en onderling overlap vertonen.

4.1.3 Bescherming persoonsgegevens

De eerste categorie betreft berichten die gaan over de opgeslagen gegevens en het achterlaten van persoonsgegevens en gaan met name om de oorspronkelijke data. Men dient zijn eerder genoemde persoonsgegevens achter te laten in ruil voor het gebruik van de Bonuskaart en het verkrijgen van korting. Wat deze berichten kenmerkt is dat de berichten zich richten op de beschikbaarheid van, de toegankelijkheid tot en de noodzaak van het achterlaten van deze persoonsgegevens die worden aangenomen als privacy gevoelige gegevens. Deze berichten gaan bijvoorbeeld over de opmerkingen dat een anonieme variant beschikbaar is van de Bonuskaart. Ook worden vragen gesteld bij de noodzaak van de registraties en waarom Albert Heijn deze gegevens nodig heeft. Er wordt ook gekeken naar inzicht in de gegevens die worden gekoppeld aan de Bonuskaart. Het valt de gebruikers

op dat deze gegevens moeilijk inzichtelijk zijn. Via de website kan de gebruiker van de Bonuskaart wel zijn of haar eigen koopgeschiedenis inzien maar men gelooft niet dat dit alles is wat Albert Heijn kan inzien en niet meer. Voorbeelden van berichten uit deze categorie zijn;

> "Je kan invullen wat je wilt. Niemand die controleert of het je echte gegevens zijn. Zo heb ik ingevuld dat ik een 99-jarige vrouw ben met 7 kinderen onder de 12 jaar :-) En mijn adres was geloof ik ergens in Baskenland." (Goeievraag, 2010)

> "[…] je kunt een anonieme bonuskaart nemen. Dan vraag je zo'n kaart aan, maar je vult niet je gegevens in. Dan doet hij het gewoon en kun je van alle bonussen meegenieten. (Goeievraag, 2010)

4.1.4 Analyse en Inzet

De tweede categorie betreft berichten die gaan over de inzet en de potentie van deze persoonsgegevens; wat zou Albert Heijn mogelijkerwijs kunnen doen met deze gegevens? Dit zijn vaak speculerende berichten over wat allemaal gebeurt met de gegevens die personen dienen in te vullen bij aanvraag van de Bonuskaart. In het geval van de RFID-chip discussie wordt gedacht aan mogelijke scenario's die kunnen voortkomen uit een dergelijke technologie. Juist in combinatie met de analyses van het koopgedrag zou Albert Heijn in staat gesteld kunnen worden "Minority Report-achtige praktijken" uit te zetten (Tweakers, 2006). Zo wordt het scenario geschetst dat als consumenten een bepaald koopgedrag hebben, zij door middel van lcd-schermen in de winkel worden benaderd met persoonlijke aanbiedingen. Als zij dan tijdens het winkelen langs zo'n scherm lopen zouden ze alleen aanbiedingen zien die voor hen aantrekkelijk zijn;

> "Ow hell, ik neem die kaart echt nooit meer mee als strax allemaal van die beeldschermen me achtervolgen met reclame op! De mens word[t] nog eens paranoia..." (Tweakers, 2006)

> "Zouden ze zoiets ook kunnen gebruiken om het patroon waarin mensen de winkel doorgaan te kunnen vastleggen en analyseren? Ik kan me voorstellen dat, dat erg interessante informatie op kan leveren." (Tweakers, 2006)

Ook wordt vaak een vergelijking gemaakt met Big Brother-achtige praktijken. Dit verwijst naar de in Hoofdstuk twee genoemde Orwelliaanse discussie betreffende angst voor controle van een totalitaire organisatie die alles van de burger en consument weet. Deze berichten gaan ook over de inzet van de achtergelaten gegevens en de analyses die hierop worden gedaan. Tevens worden hierbij concrete voorbeelden aangehaald van merkbare gevolgen voor gebruikers van de Bonuskaart.

Men ziet de potentie in van de ontwikkelingen van de Bonuskaart en het huidige systeem van de Bonuskaart. Een belangrijk element van de berichten die gaan over de inzet van de persoonsgegevens is de speculatie over de combinatie van andere gegevens met overige instanties, zoals justitie en politie. De gebruikers zien de mogelijkheden van een geïntegreerd systeem waarbij potentiële terroristen worden aangemerkt door een bepaald koopgedrag;

> "Misschien moet je eens naar de big picture kijken.. "Ze" kunnen je van voor tot achter doorlichten .. denk aan:
> Internet log gegevens.
> Mobiele telefoon logs, en plaatsbepaling.
> Bank, pin transacties.
> Creditcard transacties.
> Bonuskaarten
>
> Al met al is als je die dingen samenvoegd een aardig plaatje over jouw persoon mogelijk.... […]" (Security, 2008)

> "De mensen die met een 'anonieme' bonuskaart denken veilig te zijn hebben het fout. Stel jij wordt verdacht van iets, jij wordt gearresteerd, jouw portemonnee wordt in beslag genomen en men vind jouw bonuskaart." (Security, 2008)

Er worden hardop vragen gesteld bij de combinaties van de systemen. Ook is er discussie over de mogelijkheden van dit systeem en de doorontwikkeling hiervan. Momenteel gaat het om het meten van koopgedrag en inzicht krijgen in het gebruik van producten en diensten van Albert Heijn, maar in de toekomst is wellicht meer mogelijk. De gebruikers zien zeker potentie voor Albert Heijn in de mogelijkheid van het combineren van systemen en databanken.

4.1.5 Inbreuk op privacy

De derde categorie omvat berichten die worden gekenmerkt door discussies die gaan over de inbreuk op privacy. Het zijn discussies die gaan over het ongebreideld doorverkopen aan van deze gegevens aan derden en vragen over de beveiliging van het systeem achter de Bonuskaart. Gebruikers op de websites speculeren over de mogelijkheden van deze gegevens voor andere partijen. Zo wordt het combineren van de analyses van het koopgedrag met andere gegevens genoemd als een voorbeeld. Ook wordt gekeken naar de combinatie van gegevens van justitie met die van de Bonuskaart. Middels datawarehousing kan deze informatie aan elkaar worden gekoppeld. Op die manier kan de overheid alles van een consument en burger meten. Dit ervaren gebruikers van de websites als een probleem.

> "Ze zien dat meteen of je de waarheid spreekt, als jij als crimineel zegt; Ik was thuis ten tijde van de moord op X, checken ze ff je Bonus kaart, en daar zien ze aan dat je nog ff flink wat boodschappen tussendoor hebt gedaan en je dus liegt. Ofzo!" (Security, 2009)

Herhaaldelijk wordt als probleem de doorverkoop van informatie aan derden opgebracht in de discussie. .Deze partijen zouden opnieuw combinaties kunnen maken met hun eigen gegevens en de gegevens van de Bonuskaart. Dit is een element wat ondervangen zou moeten zijn door de algemene voorwaarden van de Bonuskaart. In deze algemene voorwaarden van de Bonuskaart staat dat Albert Heijn in principe de gegevens niet doorzet aan derden en dit alleen wordt gedaan wanneer Albert Heijn de noodzaak heeft om dit te doen of het belang hoog acht (in het kader van noodzakelijkheid) bij bijvoorbeeld een online bestelling. Toch is hier discussie over en bestaat een angst dat de gegevens in 'verkeerde handen' komen. Een gebrek aan transparantie en ervaringen uit het verleden kunnen hiervoor de oorzaak zijn. Ook hier gaat het om de mogelijkheid om analyses te doen in combinatie met andere systemen.

4.1.6 Relativerende discussies

De laatste categorie zijn relativerende of ontkennende berichten. De gebruikers die deze berichten plaatsen geven aan het niet erg te vinden dat Albert Heijn zijn of haar gegevens en koopgedrag kent. Tevens geven ze aan dat privacy niet zozeer geschonden wordt wanneer koopgedrag wordt bekeken. De gebruikers geven de voorkeur aan kortingen en kunnen persoonlijke aanbiedingen (soms) waarderen. De gebruikers proberen de Orwelliaanse visie van andere gebruikers te relativeren met reacties als;

> "[…] als ik teveel shoarma koop, ben ik dan potentieel terrorist?" (Security, 2008)

> "Oh mijn God, oh mijn God! De supermarkt weet wat voor boodschappen ik doe  Dat weet de eigenaar van mijn buurtsuper ook van mij, who cares.. alleen maar handig." (Tweakers, 2006)

De berichten zijn ingedeeld op basis van bepaalde aspecten van het geheel van de mogelijkheden van de Bonuskaart. Het is opvallend dat in deze berichten nauwelijks wordt verwezen naar het begrip privacy zelf. Impliciet wordt vanuit gegaan dat wanneer bepaalde scenario's zich voordoen, wanneer gegevens worden achtergelaten of koopgedrag wordt geanalyseerd een zekere mate van inbreuk wordt gedaan op privacy. Juist de relativerende of ontkennende berichten stellen vragen bij dit begrip; wordt wel inbreuk gemaakt op privacy wanneer koopgedrag wordt geanalyseerd of gekoppeld aan andere databanken? Wanneer privacy expliciet wordt genoemd wordt nauwelijks ingegaan op wat privacy is en waarom de gegevens privacy gevoelig zijn.

4.1.7 Privacy Preserving datamining

Wat zou gebeuren wanneer Privacy Preserving datamining zou worden toegepast in plaats van datamining? Wat zou veranderen wanneer het kooppatroon niet door datamining wordt geanalyseerd maar juist door Privacy Preserving datamining? Zou de problematiek en de discussies over privacy en de Bonuskaart worden opgelost? bestaan een zekere mate van discrepantie bestaat tussen de aangedragen technologische oplossing en de discussie omtrent de Bonuskaart. Privacy Preserving Datamining lijkt slechts op delen van de problematiek toepasbaar.

4.1.8 Bescherming Persoonsgegevens

Zoals eerder is uitgelegd richt Privacy Preserving datamining zich met name op het beschermen van de oorspronkelijke data. De oorspronkelijke gegevensset, in dit geval de database met daarin het Bonuskaart nummer, de persoonsgegevens en de daaraan gekoppelde transacties (en alle kenmerken die daarbij worden geregistreerd), wordt niet meer toegankelijk gemaakt. In het geval van transformerende technieken binnen Privacy Preserving datamining wordt de oorspronkelijke data vervormd, zodat degene die analyseert een vertekening van de oorspronkelijke dataset ziet. De data in deze set staat zou daarom niet te herleiden zijn tot het individu. Interessant hierbij in het kader van de Bonuskaart is dat de anonieme variant van de Bonuskaart in principe dit probleem oplost echter op een geheel andere wijze. Het koopgedrag kan worden geanalyseerd maar daaraan kan vervolgens niet een individu worden gekoppeld. Echter, uit de berichten uit de tweede categorie (berichten over de inzet van de persoonsgegevens), blijkt dat mensen aangeven dat door andere gegevens te koppelen aan de anonieme variant de Bonuskaart toch te herleiden is naar een individu. Wanneer iemand zijn anonieme Bonuskaart geeft bij het afrekenen en vervolgens met zijn bankpas betaalt kan de anonieme kaart toch worden gekoppeld aan een persoon. Een gebrek aan transparantie over de inzet van de gegevens, verzameld door de anonieme variant van de Bonuskaart, kan een oorzaak van dit denken zijn. Wanneer de dataset van een anonieme variant van de Bonuskaart wordt gemodificeerd kunnen de gegevens dus daarom niet worden herleid naar een individu. Privacy Preserving datamining kan derhalve als een oplossing fungeren voor het probleem van het koppelen van gegevens van de anonieme kaart aan bijvoorbeeld andere persoonlijk gegevens als een bankpas.

4.1.9 Analyse en Inzet

Het voornaamste probleem met Privacy Preserving datamining is de speculative berichtgeving over de inzet van de persoonsgegevens. De methode is gericht op de bescherming van de oorspronkelijke data tijdens het proces van datamining. Het is gericht op de bescherming van de persoonsgegevens van individuen die wordt gebruikt bij het analyseren van het kooppatroon en het ontdekken van associaties. Zoals uit de tweede categorie berichten bleek, zien gebruikers juist de inzet van deze patronen en de combinaties met andere data, als het probleem. Dit is het resultaat van analyses door middel van ongeconditioneerde datamining. Ondanks dat de oorspronkelijke data wordt beschermd door Privacy Preserving datamining, wordt de persoon niet ‘beschermd’ tegen de analyses die worden gedaan door datamining. Juist de uitkomsten van deze analyses vormen voor de gebruikers een probleem. Zo gaf een gebruiker aan dat hij een keer condooms had gekocht en hij vervolgens allerlei seksgerelateerde aanbiedingen kreeg. Dit werd door deze gebruiker en andere gebruikers gezien als ‘Big Brother’-praktijken. Ook kan in het geval van de overige discussies deze modificatie technieken van Privacy Preserving datamining in mindere mate tot een oplossing leiden.

4.1.10 Inbreuk op privacy

Privacy Preserving datamining kan de persoonsgegevens niet beschermen tegen eventueel verlies en de doorverkoop aan derden. Gebruikers op de websites zien het probleem van de koppeling van de gegevens die worden opgeslagen in de database van de Bonuskaart met een database die door Justitie gebruikt wordt. Juist deze koppeling kan leiden tot ‘Big Brother’- of ‘Minority Report’-achtige praktijken. Zoals is gesteld kan Privacy Preserving datamining alleen bescherming bieden aan de persoonsgegevens die worden opgeslagen. Als een bepaalde doelstelling van organisatie het delen van deze gegevens aan derden is, kan Privacy Preserving datamining geen uitkomst bieden. Privacy Preserving datamining is een middel die kan worden ingezet wanneer behoefte is om de persoonsgegevens te beschermen maar als de organisatie tot doel gegevens aan derden ter beschikking te stellen of te verkopen schiet Privacy Preserving datamining als middel te kort. Privacy Preserving datamining is een technologische oplossing voor een cultureel complex probleem.

Wanneer een oplossing niet aansluit met de doelstellingen van een organisatie zal deze oplossing ook niet kunnen worden gehanteerd. Privacy Preserving datamining kan alleen helpen wanneer organisaties bewust ervoor kiezen persoonsgegevens te beschermen.

4.1.11 Relativerende discussies

Problematisch in relativerende discussies omtrent de Bonuskaart, privacy en datamining is, dat deze gebruikers de ernst van de situatie niet erkennen of het niet erg vinden wanneer analyses worden gedaan op hun gegevens of hun gegevens worden gedeeld met derden. Hier blijkt het 'nothing-to-hide' argument de dominante factor. In dit geval kan Privacy Preserving datamining geen oplossing zijn voor deze gebruikers. Ze geven immers zelf aan dat ze het niet erg vinden wanneer privacy geschonden wordt of beoordelen het niet als een schending van hun privacy.

4.1.12 Conclusie

Wanneer we Privacy Preserving datamining als mogelijke oplossing voor negatieve gevolgen van datamining proberen te koppelen aan de discussies over de Bonuskaart, privacy en datamining ontstaat er een discrepantie tussen de technologische oplossing en de discussies omtrent privacy. Privacy Preserving datamining kan uitkomst bieden bij bescherming van de ingevoerde persoonsgegevens bij de gewone Bonuskaart en bescherming van combinaties tussen gegevens bij de anonieme variant van de Bonuskaart. Privacy Preserving datamining beschermt de consument echter niet tegen de analyse van persoonsgegevens en inzet van deze resultaten. Tevens beschermt het niet tegen het verlies van gegevens. De doelstellingen van een organisatie staan boven technologische oplossingen als Privacy Preserving datamining. Het delen van persoonsgegevens met derden kan een organisatie financieel gewin geven. Het is de verantwoordelijkheid van degene die data in beheer heeft, in dit geval is dat Albert Heijn, om te bepalen wat de juiste inzet is van deze persoonsgegevens en om te bepalen of deze gegevens worden gedeeld met derden dit alles uiteraard binnen het wettelijk kader dat nog steeds in ontwikkeling is.

4.2 Casestudie 2: Het Herkenningsdienstsysteem

Recent verkende Tim Cocx (2009) de mogelijkheden van datamining voor strafrechtelijk onderzoek. Hij analyseerde daarvoor gegevens uit een strafbladensysteem en keek naar welke delicten “vaak” door dezelfde personen werden gepleegd. Vervolgens keek hij naar de mogelijkheden van datamining om probeerde te voorspellen welke volgende stappen veroordeelden zouden maken in hun criminele carrière. Tim Cocx (2009) gaf aan dat datamining niet mag worden toegepast op het Herkenningsdienstsysteem (HKS) omdat de gegevens die zijn opgeslagen hierin alleen zijn bedoeld voor strafrechtelijk onderzoek en het inzicht creëren in algemene trends in criminaliteit in Nederland. Echter is zeer de vraag of datamining niet mag worden toegepast door de politie een complexe juridische vraag. Zoals jurist Engelfriet (2009) aangeeft toetst Cocx (2009) dit aan hand van de Wet Bescherming Persoonsgegevens. Engelfriet (2009) laat echter zien een toetsing nodig is aan de Wet Politiegegevens en het Wetboek van Strafvordering. Een belangrijk punt wat Engelfriet (2009) aangeeft is dat de focus niet zozeer moet liggen wel dan niet toepassen van datamining door politie datamining maar op de vraag wat de politie kan doen met de gevonden resultaten. Wat kan de politie bijvoorbeeld gaan doen met het resultaat dat joyrijders over het algemeen vaker in aanraking komen met alcohol?

De reacties op deze nieuwsberichten gaan met name over de vraag wat gedaan kan worden met de gegevens die verkregen zijn uit het analyseren van het HKS door middel van datamining. Eerst worden de doelstellingen van en het privacybeleid omtrent het Herkenningsdienstsysteem geschetst. Er wordt hierbij ingegaan op de uitzonderingen die worden genoemd in de WBP. Vervolgens wordt een overzicht gegeven van discussies over de nieuwsberichten met als onderwerp datamining en het HKS. Vervolgens wordt gekeken of Privacy Preserving datamining een oplossing kan zijn voor de beschreven problematiek.

4.2.1. Doelstellingen en privacybeleid van het HKS

Het HKS is een strafbladen- en verdachtengegevenssysteem waarbij een overzicht kan worden gegenereerd van deze regionaal gebonden gegevens. Het HKS bevat informatie over personen tegen wie een proces-verbaal is opgemaakt maar gegevens over misdrijven. Het kan zijn dat diverse personen meerdere misdaden plegen en het kan zijn dat misdaden door meerdere personen gepleegd worden. Deze combinaties worden gemaakt in het HKS. Het systeem geeft informatie over de persoon zoals, geboortedatum, geboorteland, nationaliteit en geslacht maar ook informatie over het delict. Dit kan informatie zijn zoals die genoteerd is bij de aangifte. Problematisch in het kader van de doelstellingen van het HKS, het genereren van een algemeen beeld van criminaliteit in Nederland, is de volledigheid van dit systeem. Zo worden HALT registraties en bijzondere opsporingsdiensten als de Douane niet opgenomen in dit systeem (CBS, 2011a). Bepaalde typen geregistreerde misdrijven, zoals economische en financiële delicten, zijn daarom ondervertegenwoordigd in dit systeem (CBS, 2011b).

Het doel van het HKS is het beantwoorden van algemene vragen over-, de oorzaak van en trends in criminaliteit. Tevens moet het HKS een Landelijke Criminaliteits-Kaart genereren (Cocxs, 2009). Deze kaart is een jaarlijks rapport wat gegenereerd wordt door dienst IPOL (onderdeel van het KLPD) waarbij wordt gericht op criminaliteitsgebieden, sociodemografische kenmerken en wetshandhaving beleid (CBS, 2011a). Er wordt hierbij gericht op de aard, omvang en trend in criminaliteit (Kruize en Kool, 2002). Tevens heeft het HKS als doel dat het ondersteuning biedt aan de uitvoering recherchetaken (Kruize en Kool, 2002). Met behulp van het HKS kunnen kenmerken van verdachte personen worden opgezocht. Het HKS systeem genereert een uitkomst van verdachten die overeenkomsten hebben met deze kenmerken (Kruize en Kool, 2002). Het HKS is het resultaat van 27 regionale Herkenningdienstsystemen, die worden gevuld door 27 politiekorpsen. Het landelijke HKS combineert deze gegevens. Het HKS is gestart in 1986 en wordt momenteel gewerkt aan vervanging van dit systeem (Nationale Ombudsman, 2010).

Interessant in het kader van datamining is het combineren van de gegevens uit het HKS en het Sociaal Statistisch Bestand (SSB). Aan de hand van bovengenoemde gegevens koppelt het CBS gegevens uit het SSB. Het doel van deze koppeling is het genereren van inzicht in de bovengenoemde misdrijfgegevens en verdachten. Volgens het CBS (2011b) lukt dat momenteel bij 90 procent van de verdachten die geregistreerd staan in het HKS. Op basis van de gegevens die zijn

opgeslagen in het HKS wordt gekeken in het SSB welke informatie hieraan gekoppeld wordt. Het SSB is niet één bestand, het zijn meerdere bestanden die het resultaat zijn van vraag naar informatie uit verschillende bronnen. Bronnen kunnen zijn; Gemeentelijke Basis Administratie (GBA), Aangiftegegevens Loonbelasting en Zorgtoeslaggevens (CBS, 2011b). Dit zijn bronnen die van verschillende overheidsinstanties afkomstig zijn. Afhankelijk van de onderzoeksvraag en het doel van gebruik van de informatie wordt één of meerdere bestanden gegenereerd. De betreffende overheidsinstanties kunnen deze combineren met hun eigen data. In het geval van het HKS worden gegevens uit het GBA, zoals woonlocatie, burgerlijke staat en familiegegevens, ingebracht in het HKS. Dit kan aansluiten bij het doel om inzicht te krijgen in de landelijke criminaliteit.

Tim Cocx (2009) laat zien dat een aantal wetten voor de toepassing van datamining binnen wetshandhaving van belang is; namelijk de Wet Bescherming persoonsgegevens en de Wet politieregisters. In aanvulling hierop is ook het Wetboek van Strafvordering en het Wetboek van Strafrecht van belang. In principe wordt de data over personen in politieregisters beschermd onder de Wet Bescherming persoonsgegevens. Het zijn veelal gegevens over personen, de gegevens worden verwerkt en er is sprake van een bestand. Er zijn echter uitzonderingen op deze bescherming en in bijzondere gevallen mag op het recht op privacy in relatie tot persoonsgegevens inbreuk op worden gemaakt. Namelijk in het kader van strafrechtelijk onderzoek vervalt deze bescherming.

In het kader van strafrechtelijk onderzoek kan worden afgeweken van de bescherming van de persoonsgegevens onder de Wet Bescherming persoonsgegevens (art. 2 lid 2 Wbp juncto art. 4 Wpg). Dit betekent dat de politie meer vrijheid heeft in de omgang van data. Toch zijn ook hier enige beperkingen opgelegd aan de autoriteiten. Alleen in het belang van strafrechtelijk onderzoek mag inbreuk worden gepleegd op het recht van privacy. De aard en het doel van een strafrechtelijk onderzoek is hierbij leidend. Uitsluitend indien aard en het doel van dit strafrechtelijk onderzoek daartoe rechtvaardigt kan data worden verwerkt zonder bijvoorbeeld toestemming van personen en behoeven daarmee de geformuleerde plichten uit het WPB door autoriteiten niet te worden nageleefd. De Wet politieregisters stelt dat het aanleggen van een politieregister slechts plaats mag vinden voor een bepaald doel en voor zover noodzakelijk is voor de goede uitvoering van de politietaak.

De geregistreerde gegevens in het HKS vallen onder een uitzondering op de Wet Politieregisters (Engelfriet, 2009). De gegevens zijn verzameld in het kader van strafrechtelijke doelstellingen, namelijk het opsporen van verdachten. Het gebruik van het HKS in het kader van het recherchewerk sluit aan bij de uitvoering van de politietaak. Door kenmerken uit een bepaalde zaak in te voeren in het HKS kunnen verdachten worden gevonden die overeenkomsten hebben met deze kenmerken..

4.2.2. Problematiek

Wat is het probleem van een database met een dergelijke omvang en de mogelijkheden van datamining? Onlangs waarschuwde de Algemene Inlichtingen en Veiligheidsdienst (AIVD) voor de toenemende mate van de verzameldrift van overheden ten aanzien van burgers (AIVD, 2009; AIVD, 2010 en Tienkamp, 2010). Het gevaar volgens de AIVD zit in het relationele karakter van de databases. In navolging van deze problematiek stelt de AIVD dat ook gevaren schuilen in Social Engineering en datamining (Tienkamp, 2010). Aan de hand van algoritmen en methodes om informatie te analyseren kunnen patronen worden ontdekt in deze verzameling van informatie. Vervolgens kunnen wellicht voorspellingen worden gedaan over het gedrag van burgers en kan beleid door politie, justitie en politiek hierop worden gebaseerd.

Datamining op een database met een grote omvang kan interessant zijn voor politie.. Daarbij kan inzicht worden gegenereerd in 1) welke soort delicten vaak worden gepleegd in combinatie met andere soorten delicten, 2) criminele carrièrepaden 3) voorspelling van vervolgstappen in individuele criminele carrièrepaden, 4) aanwijzen van burgers die buiten ‘de norm’ handelen (Cocx, 2009). Het HKS is een systeem wat momenteel actief wordt gebruikt voor recherchewerk en wordt tevens geanalyseerd voor de Landelijke Criminaliteitskaart. Wanneer datamining actief wordt toegepast kan het HKS gaan dienen als een middel voor beleid, strategie en preventie. De potentie van datamining voor een systeem met dergelijke omvang is groot. Tevens is het combineren van gegevens uit andere bronnen met het HKS interessant voor politie. Inzicht in het handelen van verdachten kan daarmee worden vergroot.

Maar mogen justitie en politie dergelijke analyses wel uitvoeren? Het is de vraag of datamining valt onder de uitzonderingen die worden gesteld in het Wbp. Uitzonderingen op de Wbp mogen worden gemaakt in het kader van de politietaak. Hiervoor dient gekeken te worden naar wat de politietaak inhoudt (art 1, 4 en 6 Wpg). De politietaak is geformuleerd in art. 2 Politiewet 1993; “de politie heeft als taak in ondergeschiktheid aan het bevoegde gezag en in overeenstemming met de geldende rechtsregels te zorgen voor de daadwerkelijke handhaving van de rechtsorde en het verlenen van hulp aan hen die deze behoeven”. Zoals Fijnaut et. al. (2007) laten zien is de formulering ‘handhaving van de rechtsorde’ breed te interpreteren; opsporing en preventie zijn eveneens onderdelen van deze handhaving. Naeyé (2009) laat zien dat door de discussie over de politietaken de aandacht deels uitgaat naar daadwerkelijke handhaving en deels naar preventie. De politietaken die Naeyé noemt zijn; “het uitoefenen van preventief onderzoek, het corrigerend toespreken […] het nemen van fysieke maatregelen en het uitoefenen van dwang of geweld” (Naeyé, 2009). Bij preventie zijn vooral de verhoudingen tussen andere overheidsorganen van belang. Hier kan samenwerking plaatsvinden en politie en de politietaak is dan onderdeel van een veiligheidsketen, waarbij organisaties als een keten informatie winnen over criminaliteit. Volgens Cocx (2009) sluit datamining niet aan bij deze geformuleerde politietaak. Of datamining binnen de politie taak valt is niet eenduidig vast te stellen vanwege de brede definitie van en de onenigheid over wat deze politietaak inhoudt. Uit de onderstaande analyse zal blijken gebruikers in reacties op de nieuwsberichten eveneens vragen stellen bij de taak van de politie en de rol van datamining binnen deze politietaak. Privacy blijkt ook hier weer het voornaamste onderdeel van discussie. Vervolgens wordt gekeken of Privacy Preserving datamining een antwoord kan zijn voor de discussies over het HKS inzake datamining en privacy.

4.2.3. Discussies

Nu worden discussies geschetst over het HKS, dataopslag, datamining en privacy. Dit gebeurt aan hand van dezelfde categorisering van discussies die hiervoor is genoemd; bescherming persoonsgegevens, Analyse en Inzet, inbreuk op privacy en relativerende reacties. Het betreft de nieuwsberichten op de websites Webwereld.nl (2009a) en Webwereld.nl (2009b). Een verantwoording van de keuze van deze platformen is opgenomen in de methodologie in hoofdstuk 1. Het blijkt dat de focus bij de discussies omtrent datamining, het HKS en privacy met name ligt op de categorie ‘Analyse en Inzet’.

4.2.4 Bescherming persoonsgegevens

Discussies over de bescherming van de persoonsgegevens gaan met name over waarom zoveel data opgeslagen dient opgeslagen te worden, welke data wordt opgeslagen en over de toestemming voor het gebruik hiervan. Relatief weinig discussies gaan over de categorie bescherming persoonsgegevens. Onder gebruikers bestaat een zekere mate van angst, dat hoe meer data politie en justitie genereren hoe meer informatie zij tot haar beschikking krijgt. De informatie kan vervolgens weer leiden tot een inzet in een strafrechtelijk onderzoek.

Ook discussieert men over de toestemming van opslag van data in het HKS. Men gaat er vanuit dat politie toestemming heeft om al deze gegevens te verzamelen en op te slaan. Tevens gaan de gebruikers van de websites vanuit dat de politie gegevens uit verschillende bronnen haalt om haar eigen gegevens aan te vullen. Zo gaan de gebruikers er vanuit dat gegevens die beschikbaar zijn bij andere overheidsinstellingen ook beschikbaar zijn bij de politie. Een aantal gebruikers stelt hierbij de vraag in hoeverre al deze data zomaar mag worden gebruikt voor datamining. Andere gebruikers reageren hierop met de stelling dat het inderdaad de vraag is of het mag, maar dat als het niet mag, de overheid alleen nog maar een juridisch kader hoeft te genereren om het alsnog mogelijk te maken. Een voorbeeld van een bericht dat gaat over de toestemming van de opslag van data is de volgende;

> “Ongeacht de toezeggingen die gedaan zijn en/of gedaan worden in de toekomst, denk ik te kunnen stellen dat we vanuit kunnen gaan dat het toch wel gebruikt gaat worden.Er hoeft slechts 1 iemand met bevoegdheden te zijn die het als een ultiem middel ziet misdaad tegen te gaan en huppa, de wet wordt aangepast.
> ik hou mijn hart vast. Echt. Al dan niet met toestemming, dit gaat misbruikt worden. Niet door de overheid, dan wel door derden die toegang hebben/krijgen. Ik baal hiervan, en er is weinig wat ik hier aan kan doen.” (Webwereld, 2009b)

> "Waarom wordt het dan toch gedaan. Gewoon met jullie handjes er van af blijven, nog niet eens dit onderzoek laten doen of de techniek laten ontwikkelen om dit onderzoek te kunnen doen." (Webwereld, 2009b)

In navolging hiervan volgt een andere discussie die gaat over wát wordt opgeslagen. De gebruikers gaan vanuit dat politie en justitie veel gegevens tot hun beschikking hebben. Gebruikers stellen hier vragen over de noodzaak van de opslag van gegevens en hebben het, zoals in het bovenstaande voorbeeld, over de onmacht tegenover dergelijke autoriteiten. De gebruikers geven aan dat opslag door autoriteiten niet voorkomen kan worden en dat individuen weinig kunnen uitrichten tegen dergelijke praktijken. Recentelijk kwam een discussie over het opnemen van foto's van hangjongeren in het HKS. Deze discussie laat zien dat het niet onomstreden is wat de reikwijdte van gegevensopslag is voor de politie. Tevens geven sommige gebruikers aan dat een fout in de database wellicht kan leiden tot een verkeerde veroordeling;

> "[…]kortom datamining zorgt voor dat de verkeerde mensen achter de tralies verdwijnen! Dat maakt de zaak alleen maar onveiliger […] (Webwereld, 2009a)

4.2.5 Analyse en Inzet

De groep discussie waarover het meest wordt gediscussieerd en met name de aandacht ligt in de berichten van gebruikers is die van 'Analyse en Inzet'. Deze berichten gaan met name over de rol van politie en datamining, over dat ommekeer plaatsvindt in bewijslast en verdenkingen, dat verdenkingen minder noodzakelijk zijn, over een controlestaat en over vragen over de noodzaak van dergelijke analyses.

Er worden door deze groep gebruikers vragen gesteld over wat de politie allemaal mag doen en behoort te doen. Ze discussiëren over of datamining plaats mag vinden binnen de politietaak. Er wordt gekeken of de analyses die worden gedaan door Cocx (2009) waarden gaan hebben die relevant zijn voor de rol van de politie. Tevens volgt een semi-juridische analyse over wat de politie wel en niet mag gebruiken als bewijs. De aandacht binnen deze discussies gaan met name over de datamining als bewijs voor strafrechtelijk onderzoek en of de politie dergelijke analyses mag uitvoeren;

> "Zie je wel, het mag...
> Dus we zijn al in de politiestaat beland. Alleen verkeren we nog steeds in de ontkenningsfase." (Webwereld.nl, 2009b)

> "Dat het dataminen gebeurt, daar doe je gewoon weinig of niets tegen. Maar juist daarom is een spanningsveld tussen justitie en burger." (Webwereld, 2009b)

In navolging hiervan volgen uitgebreide discussies over de rechtelijke macht en over de rol van verdenkingen. Een aantal gebruikers geeft aan dat bij de komst van dergelijke analyses de rollen van verdachte en bewijs worden omgedraaid;

> "Als ze alle data van een verdachte door spitten is dat goed. Als ze alle data doorspitten opzoek naar mogelijke verdachten, kom je op zeer glad ijs." (Webwereld, 2009b)

Men dient momenteel eerste verdachte te zijn om bewijzen tegen hem te zoeken. Er dient een redelijke mate van een verdenking te zijn (art. 24 Sv). Gebruikers geven aan dat dit bij datamining wordt omgedraaid. Hierbij wordt volgens hen, gekeken naar data en vervolgens wordt gekeken naar verdenkingen die kunnen worden herleid uit deze data, waarna verdachten kunnen worden aangewezen. Men kan derhalve op basis van theoretische statistiek worden aangemerkt als verdachte. Met andere woorden wanneer iemand zich buiten de normen gedraagt, kan hij/zij in de problemen komen. Deze gebruikers zien hier absoluut de ernst van in. Ze zijn bang dat redelijke verdenkingen minder nodig zijn om bijvoorbeeld iemands huis te gaan doorzoeken, Met andere woorden statistiek gaat gelden als bewijs.

> "Men gaat niet in deze database op zoek naar een bepaald persoon. Maar men doet juist het omgekeerde: alle op de persoon herleidbare gegevens worden eruit gesloopt en worden

> algemene partronen afgeleid. Puur en alleen om te kijken of er relaties zijn tussen bepaalde soorten gedrag of kenmerken. Het is geen techniek die naar één dader zoekt." (Webwereld, 2009a)

Autoriteiten zoals Politie, politiek en justitie worden hier in meerdere mate vergeleken met een totalitaire controle staat. De vergelijking met Minority Report gaat over de situatie waarin op basis van statistisch onderzoek wordt bepaald of iemand een bepaald misdrijf heeft gepleegd. Ook bijvoorbeeld voordat hij überhaupt iets heeft gedaan. Wanneer iemand zou voldoen aan bepaalde kenmerken dan kan hij al verdacht zijn omdat datamining dat in het verleden heeft uitgewezen;

> "[…] jij zit niet in de database, maar je onschuldige gedrag komt overeen met dat van een groep die wel in zit. En opeens heb jij ook een sticker "verdacht" opgeplakt gekregen." (Webwereld, 2009a)

Deze reacties gaan vooral over het zoeken naar overeenkomsten tussen delicten, verdachten en statistiek. Door middel van data kan de overheid nu alles zien en meten volgens deze gebruikers. Een voorbeeld van een dergelijke reactie is;

> "Hoe meer data hoe meer big brother is watching you. […] Als voorbeeld,er is iets gebeurd, daarvoor is het bewezen dat de persoon in kwestie mannelijk is, boven de 2 meter, donker haar heeft, en Nederlands spreekt. Dat zou inhouden dat bijna alle basketballers in Nederland, maar ook ikzelf, tot de verdachten behoren." (Webwereld, 2009b)

4.2.6 Inbreuk op privacy

De derde categorie betreft berichten die gaan over de inbreuk op privacy. Discussies uit deze categorie gaan met name over angst dat de gegevens in verkeerde handen komen en het koppelen van allerlei overige bronnen aan het HKS.

Onder gebruikers bestaat een zekere mate van angst dat de gegevens die worden geregistreerd in het HKS en worden gekoppeld aan andere bronnen deze in handen komen van derden. De gebruikers geven hierbij aan dat ze verwachten dat het HKS niet genoeg beveiligd is en dat het voor hackers eenvoudig zou zijn om de persoonsgegevens en allerlei andere bronnen tot hun beschikking kunnen krijgen. Gebruikers uiten hierover hun zorgen;

> "Onze overheid wil meer en meer weten, je kan het niet/moeilijk ontwijken, ze maken misbruik van en kunnen het voor geen meter beveiligen." (Webwereld, 2009a)

De veelvoudige koppeling van allerlei bronnen door justitie en politie kan ervoor zorgen dat de overheid vrijwel alles van burgers zou kunnen weten. Een veelvoorkomend voorbeeld wat gebruikers aanhalen is de combinatie met de OV-chipkaart. Deze elektronische vervoerspas zou gedetailleerde informatie kunnen geven over de beweging en locatie van gebruikers. Wanneer die aan delicten gekoppeld zou worden kan de politie eenvoudig deze verdachten kunnen vergelijken met de plaats van delict. Ook al zou een persoon zich daar niet of nauwelijks begeven;

> "Dit is dus de reden waarom wij met zijn allen de straat op moeten gaan tegen EPD, rekeningrijden, OV chipkaart, dataretentiewetten, RFID chips, en meer van dit soort ongein.
>
> Ik kan met 100% zekerheid zeggen dat er met de tijd een systeem komt dat alle gegevens van iedereen gaat dataminen." (Webwereld, 2009a)

4.2.7 Relativerende reacties

De relativerende reacties gaan met name over het in hoofdstuk twee genoemde 'nothing-to-hide'-argument. Deze gebruikers vinden dergelijke analyses niet erg omdat ze niets te verbergen hebben voor de overheid en politie.

Een andere groep gebruikers ziet de mogelijkheid van datamining binnen de politie als een voordeel. Deze gebruikers kijken naar de mogelijkheden en eventuele kansen op een lager criminaliteitsniveau in Nederland. Ze zien de waarde in van de analyses en zien mogelijkheden in het combineren en verbeteren van de data door toevoeging van andere bronnen.

> "Hoe meer data hoe beter. Er is met dit onderzoek immers bewezen dat uit data wel degelijk berekend kan worden of iemand crimineel gedrag gaat vertonen in de toekomst.
> Dat klinkt als een goed argument om nog meer data op te slaan. Van iedereen, niet van reeds veroordeelden." (Webwereld, 2009a)

De overige relativerende berichten houden in dat men verstandig moet gaan stemmen om te voorkomen dat datamining door politie toegepast gaat worden. Deze gebruikers richten zich op een oplossing voor de eventuele toekomstige problematiek. Bepaalde politieke partijen en organisaties proberen zich te verzetten tegen de toepassing van datamining op databases als die van het HKS.

4.2.8 Privacy Preserving datamining

Zoals eerder is gesteld ligt de nadruk bij de discussie over het HKS, datamining en privacy met name op de Analyse en Inzet van de geregistreerde gegevens. Cocx (2009) heeft een verkennend onderzoek gedaan naar de mogelijkheden van een technologie als datamining. De discussies die hieruit volgen richten zich met name over de inzet van de resultaten om voorspellingen te doen over het gedrag van burgers en de mogelijkheden om hier andere bronnen aan te koppelen. Privacy Preserving datamining is een technologische oplossing voor een complex cultureel probleem. Privacy Preserving datamining kan ook hier niet fungeren als een oplossing. Ook hier blijkt een discrepantie te bestaan tussen de oplossing van Privacy Preserving datamining en de genoemde problematiek van datamining.

4.2.9 Bescherming persoonsgegevens

De individuele gegevens van burgers worden vastgelegd in tal van systemen. Het HKS maakt volgens de doelstellingen gebruik van demografische gegevens uit het SSB die worden gehaald uit de bron GBA. Discussies over datamining, het HKS en privacy gaan met name over vragen waarom opslag nodig is, waarom deze hoeveelheden opslag noodzakelijk zijn, wat wordt opgeslagen en of toestemming is voor deze opslag. Privacy Preserving datamining zorgt hier opnieuw in mindere mate voor een oplossing. De discussies die gaan over wat wordt opgeslagen gaan met name over welke combinaties van bronnen gemaakt worden. Dat men geregistreerd staat in enige database die ter beschikking staat bij de overheid wordt niet bediscussieerd. Men gaat hier vanuit dat men opgeslagen staat. Het gaat bij het HKS om welke combinaties met andere systemen gemaakt kunnen worden. Wat eveneens een probleem is, is de vraag in hoeverre datamining te plaatsen is onder de politietaak. bestaat onder gebruikers van websites discussie over in hoeverre datamining inderdaad daaronder valt. Een politieke discussie is nodig om hierover uitsluitsel te geven. Ook hierbij moet worden gemeld dat Privacy Preserving datamining geen oplossing is.

4.2.10 Analyse en Inzet

Doelstellingen die het HKS heeft gaan over het genereren van de Landelijke Criminaliteitskaart en het ondersteunen van recherchewerk. Bij de Landelijke Criminaliteitskaart wordt door de analisten gebruik gemaakt van een geanonimiseerde versie van het HKS (CBS, 2011b). De analisten krijgen geen individuele records te zien, dus hoeven zij alleen maar te zoeken naar patronen in deze data. In die zin is Privacy Preserving datamining een overbodige stap want de data is al geanonimiseerd. Ook wanneer data uit andere bronnen wordt gekoppeld is PPDM overbodig. Dit aangezien ook deze bronnen, zoals het GBA, geanonimiseerd worden (CBS, 2011b). Hier geldt bovendien dat PPDM de burger niet kan beschermen tegen interpretatie en inzet van de analyses.

Analyses zoals Cocx (2009) die heeft gedaan en de mogelijkheden die PPDM biedt is ook hier geen oplossing voor de problematiek. De discussies richten zich namelijk met name op het aanmerken van personen als verdachten aan de hand van cijfers die zijn gegenereerd door datamining. Deze analyses zijn gedaan op een geanonimiseerde database. Wanneer blijkt dat een persoon met een bepaalde culturele achtergrond vaker een bepaald delict pleegt kan dit aanleiding zijn voor politie tot

verdenken, of een bevestiging van een bepaald vermoeden. Privacy Preserving datamining richt zich op de bescherming van de persoonsgegevens en andere gegevens die zijn opgeslagen in databases zodat algemene analyse alsnog gemaakt worden. Privacy Preserving datamining kan burgers niet beschermen tegen interpretatie van resultaten en wat met voorspellingen van gedrag van burgers wordt gedaan. Niet op basis van beleid maar ook niet op basis van individuele gevallen. wordt in het recherchewerk wel gekeken naar individuele records in het HKS. Immers wordt aan de hand van bepaalde kenmerken gekeken of een delict of een verdachte gekoppeld kan worden aan deze kenmerken. Dit gaat om het identificeren van een individu. De oorspronkelijke data is in het kader van de politietaak van belang. Wanneer politie de data gaat inzetten voor wetshandhaving dient zij zeer kritisch te kijken naar de analyses die zijn gedaan.

4.2.11 Inbreuk op privacy

In het geval van inbreuk op privacy geldt dat gebruikers over het algemeen aangeven angsten te hebben dat de geregistreerde gegevens in handen komen van criminelen. Ook de AIVD waarschuwde voor de toenemende mate van dataopslag. Deze data kan immers worden gecombineerd met nieuwe data en vervolgens kunnen nieuwe analyses op worden gedaan. De fase datawarehousing is een belangrijke stap in het combineren van dergelijke datasets. Privacy Preserving datamining kan niet helpen dat bepaalde systemen worden gehackt of in verkeerde handen komen. Overheden dienen de beveiliging van hun systemen te optimaliseren. Gebruikers geven hierbij aan dat ze ervaren dat de overheid hier niet vaak in slaagt. Eveneens kan Privacy Preserving datamining burgers niet beschermen tegen het combineren van systemen van andere overheidsinstellingen dan de politie. Politie dient hier zelf kritisch naar te kijken en de politiek dient regels te formuleren voor de omgang met grote hoeveelheden datasets en de uitwisseling van deze datasets.

4.2.12 Relativerende reacties

Opnieuw geldt dat Privacy Preserving datamining geen antwoord heeft op discussies die gaan over dat men het niet erg vindt dat datamining wordt toegepast. Men gaat vanuit dat ze niets te verbergen hebben. Andere gebruikers geven aan dat deze argumenten niet gelden in het geval van datamining. Men kan immers plots wel wat te verbergen hebben als hij voldoet aan een bepaalde statische analyse. Op deze problematiek heeft Privacy Preserving datamining geen antwoord omdat dit zich richt op de bescherming van de oorspronkelijke dataset die is opgeslagen.

4.2.13 Conclusie

In het geval van het Herkenningsdienstsysteem kan worden gesteld dat Privacy Preserving datamining in mindere mate een antwoord vormt door de discussies die voortkomen uit het HKS, datamining en privacy. Men geeft aan de mogelijkheden en problemen te zien voor datamining voor wetshandhaving. Met name de problematiek die gaat over de overheid als een controle staat en het aanmerken van verdachten voordat zij een delict hebben gepleegd, heeft Privacy Preserving datamining geen antwoord op. De oorspronkelijke data kan worden geanonimiseerd maar analyses worden toch gedaan. Het gaat juist om de inzet van deze analyses ter aanvulling van het politiewerk. Men kan bijvoorbeeld wanneer hij bepaalde sociodemografische kenmerken heeft opeens in aanmerking komen als verdachte wanneer deze kenmerken worden geassocieerd met een bepaald delict. Er blijkt ook hier een zekere mate van discrepantie te ontstaan tussen de aangereikte oplossing Privacy Preserving datamining en het complexe debat over datamining en privacy.

4.3 Casestudie 3: de Facebook-database

In 2009 ontwikkelden MIT studenten de software ‘Gaydar’ (Schroeder, 2009). Dit is een programma om te kijken of gebruikers op de sociale netwerksite Facebook homoseksueel zijn. Dit wordt berekend aan hand van de vrienden van een bepaald profiel. Wanneer een persoon zijn seksuele voorkeur niet heeft getoond op zijn profiel kan dit programma schatten of deze persoon homoseksueel is door te kijken naar zijn vrienden en hun seksuele voorkeur. De studenten berekenden dat gebruikers met een

homoseksuele geaardheid significant meer homoseksuele vrienden hebben dan heteroseksuele gebruikers. Aan de hand van deze berekening werd een voorspellend model gemaakt waarbij seksuele geaardheid kan worden geschat. Om dit te bereiken hebben ze 70.000 Facebook accounts van verschillende scholen gedownload. De studenten ontwikkelden dit om een discussie op gang te brengen over privacy en wat men wel en niet plaatst op openbare profielen. Iedereen kan immers gebruik maken van deze data. Om te voorkomen dat anderen opnieuw duizenden of miljoenen gebruikersprofielen zomaar downloaden, zoals in het geval van de 'Gaydar', heeft Facebook extra privacy- en veiligheidsmaatregelen genomen.

Ook hier blijkt datamining interessant om informatie te halen uit de grote hoeveelheden data die wordt opgeslagen op Facebook. Mensen veranderen bijvoorbeeld voortdurend hun profielstatus of vertellen waar ze mee bezig zijn. Zo bleek dat Facebook goed inzicht heeft in wanneer mensen relaties verbreken (Alphenaar, 2011). Deze informatie kan vanuit commercieel en justitieel oogpunt aantrekkelijk zijn.

4.3.1. Doelstellingen en privacybeleid van de Facebook-database

Facebook is een sociale netwerksite die is gestart in 2004 en is inmiddels wereldwijd de grootste. Het heeft meer dan een half miljard gebruikers (Facebook, 2011a). De data die Facebook in zijn bezit heeft is daarom zeer omvangrijk. Datamining kan een belangrijk middel zijn om inzicht te krijgen in deze grote hoeveelheden data en kan helpen om nieuwe informatie te ontdekken. Om inzicht te krijgen in hoe datamining wordt toegepast door Facebook, dient gekeken te worden naar het privacybeleid van Facebook.

Facebook heeft een uitgebreid privacybeleid over hoe zij omgaan met de gegevens van gebruikers. Ook geven zij adviezen en reiken methoden aan waarmee gebruikers hun privacy beter kunnen beschermen. Facebook heeft een aantal doelstellingen met de informatie die gebruikers achterlaten op hun profiel. Facebook wil een veilige, efficiënte en sociale ervaring creëren voor gebruikers (Facebook, 2011b). Facebook gebruikt de data om bijvoorbeeld contact op te nemen met gebruikers, om service te verlenen, om vriend suggesties te doen en om te zorgen dat anderen de desbetreffende gebruikers vinden.

Facebook stelt gebruikers in staat om privacy instellingen voor een profiel zelf te laten bepalen. Gebruikers kunnen ervoor kiezen om de gegevens te tonen aan iedereen, alleen aan vrienden, alleen aan vrienden en vrienden van deze vrienden of alleen zichtbaar voor de gebruiker zelf. Dit zijn instellingen die gebruikers per element in het profiel kunnen personaliseren. Deze elementen zijn; dingen die men deelt, dingen die anderen delen, en contactgegevens. Voorbeelden van dingen die een gebruiker deelt zijn statussen van relaties, seksuele voorkeur en berichten van gebruikers. Men kan bijvoorbeeld voor kiezen om zijn seksuele voorkeur niet aan iedereen te tonen en om zijn berichten wel aan iedereen te tonen. Voorbeelden van 'dingen die anderen delen' zijn foto's waarin gebruikers zijn 'getagd' en instellingen van de mogelijkheid om anderen te laten reageren op een persoon. 'Tagging' is de mogelijkheid voor gebruikers om vrienden die ze herkennen op foto's kenbaar te maken. Wanneer een persoon op een foto wordt 'getagd', wordt de naam en het profiel van de 'getagde' gekoppeld aan de foto. Deze 'getagde' gebruiker krijgt hiervan een bericht en kan deze weigeren. Ook hierbij kan de persoon ervoor kiezen om 'getagde' foto's alleen zichtbaar te maken voor zichzelf. Voorbeelden van contactgegevens zijn het e-mailadres van de gebruiker en het woonadres van de gebruiker. Ook hiervoor geldt dat de gebruiker per onderdeel deze privacy instellingen kan aanpassen. Daarnaast biedt Facebook de mogelijkheid om de gebruikers te laten kiezen of hun profiel wordt geïndexeerd door andere partijen of niet. Hiermee kan de gebruiker bijvoorbeeld aangeven of hij zijn of haar profiel getoond wordt in de Google zoekresultaten, wanneer een persoon of bijvoorbeeld een werkgever zoekt naar zijn of haar naam. Tevens kunnen gebruikers door hen te kiezen andere gebruikers blokkeren.

Toepassing van datamining op de facebook database vormt een interessant fenomeen om hier trends te ontdekken in het gedrag van de consumenten. Data kan worden geanalyseerd om patronen te vinden in wat beweegt onder gebruikers van Facebook. Vervolgens kan dit commercieel worden ingezet om gebruikers te voorzien van persoonlijke aanbiedingen. Zo liet Facebook in een jaarschema zien waarin werd getoond op welke dag in het jaar men het meest zijn relatie verbreekt (Alphenaar, 2011). Dit kunnen voor adverteerder mogelijke interessante momenten zijn om in te zetten met

campagnes afhankelijk van hun eigen producten. Facebook krijgt door hun grote hoeveelheden data inzicht van de levenscyclus van mensen. Iemand zijn status kan bijvoorbeeld wijzigen in getrouwd, gescheiden of overleden.

Ook gebruikt Facebook informatie van gebruikers om persoonlijke advertenties aan te bieden (Facebook, 2011c). Vooral in dit geval is datamining een uitkomst. Facebook kan bijvoorbeeld interesses clusteren en categoriseren en bepaalde gebruikers hieraan koppelen. Facebook kan verdienen aan deze geclusterde informatie door deze geclusterde gebruikers te voorzien van advertenties van geïnteresseerde adverteerders. Een voorbeeld wat Facebook zelf geeft is dat wanneer een gebruiker aangeeft dat hij interesse heeft in voetbal dat aan deze persoonlijke advertenties over voetbalkleding wordt getoond in posities bij 'Gesponsord' bij zijn of haar profiel. Een case die Facebook geeft is dat gebruiker die hun relatiestatus op verloofd zetten benaderd kunnen worden met persoonlijke advertenties van bijvoorbeeld een trouwfotobedrijf. Hiervoor gebruikt Facebook opgeslagen profiel data. Dit is zowel de door de gebruiker afgeschermde data als de niet afgeschermde data. Facebook gebruikt deze data, ondanks dat men kiest voor toegankelijkheid alleen voor de gebruiker zelf, om adverteerders gebruikers persoonlijke aanbiedingen te doen. Wanneer iemand bijvoorbeeld aangeeft dat het niet getoond mag worden dat hij verloofd is kan deze gebruiker toch worden voorzien van advertenties over trouwfoto's. Adverteerders die geïnteresseerd zijn in bepaalde groepen gebruikers dienen bepaalde karakteristieken aan te merken waarop hun advertentie wordt getoond.

In navolging van dit advertentiesysteem kan Facebook ervoor zorgen dat bepaalde advertenties worden voorzien van gebruikersprofielen. Wanneer iemand connectie heeft met zijn of haar favoriete band en deze band maakt gebruik van het advertentie systeem kunnen vrienden die deze advertentie zien het profiel van de gebruiker naast deze advertentie zien. Gebruikers kunnen kiezen voor een 'opt-out' uit deze functionaliteit. Dit betekent dat gebruikers in principe deelnemer zijn hieraan maar dat ze achteraf kunnen kiezen om hiermee te stoppen.

Ook kan de informatie die wordt opgeslagen in de database van Facebook worden gedeeld met derden. Dit gebeurt alleen in het geval van goedkeuring van de desbetreffende gebruiker. Gebruikers kunnen persoonlijk worden benaderd door adverteerders wanneer zij goedkeuring hiervoor hebben gegeven. Dit gebeurt op basis van 'Opt-in'. Gebruikers nemen hieraan in principe niet deel maar kunnen dit wel doen door de deelname goed te keuren.

4.3.2. Problematiek

Er zijn gevallen bekend waar informatie door Facebook is doorgegeven aan derden zonder toestemming van de desbetreffende gebruikers. Wanneer iemand bijvoorbeeld iets bestelt door middel van Facebook dient zijn gegevens te worden gedeeld met derden. Volgens het privacybeleid van Facebook gaat dit om partijen die de privacy voorwaarden van Facebook hebben geaccepteerd. wordt niet uitgelegd welke voorwaarden dit zijn en in welke gevallen wat precies wordt geregistreerd en wordt doorgegeven aan derden. Ook wil Facebook ervoor zorgen dat zij meer naamsbekendheid genereert. Daarom wordt af en toe gebundelde informatie gedeeld aan derden of worden advertenties getoond op andere sites. Facebook streeft er hierbij naar dat individuele profielen niet herleid kunnen worden, maar adverteerders dienen een cookie te plaatsen op de computer van de gebruikers. Zo kan het zijn dat een bepaalde site wordt 'geliked' door gebruikers en dat hierbij profielfoto's worden getoond van gebruikers. Eigenaren van andere sites kunnen een 'like-button' of een 'like-box' implementeren. Hiermee kunnen op Facebook ingelogde gebruikers aangeven dat ze de site of elementen op de site leuk vinden. Dit wordt vervolgens aangegeven op hun profiel waardoor vrienden kunnen zien dat ze deze leuk vinden en kunnen interesse hierin opwekken. Op die manier kan deze website en Facebook bekendheid genereren. Opnieuw wordt deze informatie gedeeld met derden en kan het interessant zijn om aan de like-gebruikers data van derden te koppelen.

In het verlengde hiervan kwam recentelijk kwam een discussie op gang omtrent contactinformatie op Facebook. Door middel van een applicatie konden derden eenvoudig aan de telefoonnummers en adressen van gebruikers komen. Facebook staat software toe, zoals bijvoorbeeld Farmville, die kan worden gekoppeld aan Facebook. De ontwikkelaars van deze software zouden met een druk op de knop aan de informatie van deze gebruikers kunnen komen. Gebruikers zouden toestemming hiervoor moeten geven en op een knop in hun profiel drukken, er werd verwacht dat men deze voorwaarden vaak over het hoofd zou zien. Na enkele dagen van kritiek werd deze functionaliteit weer verwijderd.

Echter werd de functionaliteit opnieuw toegevoegd aan Facebook. Dit keer met extra beveiligingsmogelijkheden. In de vernieuwde functionaliteit werd de gebruiker meer mogelijkheden geboden om de voorwaarden van dit delen van contactinformatie in te zien. Datamining kan voor deze ontwikkelaars interessant zijn omdat nu hun eigen data gekoppeld kan worden aan de contactinformatie van Facebook. Hieruit kunnen nieuwe patronen worden gehaald en kan leiden tot verdere inzet. Ook hierover blijkt weer veel discussie gaande.

4.3.3 Discussies

Nu worden discussies geschetst over Facebook, dataopslag, datamining en privacy. Dit gebeurt aan hand van dezelfde categorieën van discussies die hiervoor zijn genoemd; bescherming persoonsgegevens, Analyse en Inzet, inbreuk op privacy en relativerende reacties. De websites die geanalyseerd, zijn Tweakers.net (2010) en Tweakers.nl (2011).

4.3.4 Bescherming persoonsgegevens

Berichten omtrent Facebook, datamining en privacy uit de categorie 'bescherming persoonsgegevens' gaan over de gegevens die men achterlaat op de profielen op Facebook. Deze groep discussies gaat over de bescherming van de oorspronkelijke gegevens en hoe gebruikers hiermee om gaan en dienen te gaan. Er zijn verschillende discussies hierover gaande; discussies over wat men wel en wat men niet moet achterlaten op Facebook, discussies over het veranderende privacybeleid, discussies over het rechtsgebied, discussies over de privacy instellingen in het account en discussies over de complexheid van deze instellingen.

Discussies die gaan over wat wel en niet dient te worden opgeslagen op het profiel gaan over wat men online allemaal achterlaat. Men laat volgens deze gebruikers een voetspoor achter van al zijn gegevens en het is voor anderen goed inzichtelijk wat men allemaal beweegt. Zo geven gebruikers het voorbeeld van werkgevers of potentiële werkgevers die kunnen inzien wat men allemaal online doet. De gebruikers geven aan dat gebruikers zeer voorzichtig moeten zijn met wat men allemaal online zet, want het kan zomaar in verkeerde handen vallen. Tevens geven gebruikers aan dat wanneer een persoon niet wil dat informatie wordt gedeeld met anderen, dat ze deze dan ook niet zouden moeten invullen op Facebook.

Andere gebruikers geven aan dat ze hun profiel en hun privacy instellingen goed hebben ingesteld om zo te voorkomen dat ongewenste andere gebruikers hier misbruik van maken of elementen kunnen zien waarvan gebruikers niet willen dat ze gezien worden. Dit zijn de eerder genoemde instellingen om te kiezen wat aan wie allemaal wordt getoond. Echter geven een aantal gebruikers van de websites aan dat deze regels voortdurend aan verandering onderhevig zijn waardoor het soms lastig is om de privacy instellingen naar wens te houden;

> "Ik heb m'n privacy instellingen heel goed getweaked bij Facebook, maar alsnog veranderen ze de regels best vaak waardoor mensen/bedrijven/apps die ik niet expliciet heb geautoriseerd bij m'n gegevens kunnen." (Tweakers.net, 2011)

Een andere groep berichten gaat over de ingewikkeldheid van de privacy instellingen van de profielen. Sommige gebruikers gaan vanuit dat de instellingen die gebruikers tot hun beheer krijgen met opzet complex zijn gemaakt. Dit zou zo gedaan zijn door Facebook zodat ze legaal deze informatie zouden kunnen delen met derden.

Facebook zorgt tevens voor, volgens deze gebruikers, dat veel opties van te voren op 'opt-out' staan. Dit betekent dat men zich voor iedere functionaliteit moet afmelden als zij daar geen gebruik van willen te maken. Tevens maakt Facebook niet overal even duidelijk waar de gebruikers zich kunnen afmelden van allerlei functionaliteiten en bepaalde gebruikers weigeren. Ook stelt men vragen bij de eigendom van de gegevens. Wanneer iemand iets achterlaat is hij dan nog beheerder van die gegevens of wordt Facebook eigenaar? Tevens is het hierbij niet altijd duidelijk welk recht van toepassing is op de gegevens die online staan.

4.3.4 Analyse en Inzet

De tweede groep discussies gaat over de Analyse en Inzet van de gegevens die gebruikers achterlaten op Facebook. Deze groep houden speculerende discussies in over wat met de gegevens wordt gedaan en wat er allemaal wordt geanalyseerd. Dit zijn discussies over concrete en merkbare gevolgen op Facebook en de advertenties, discussies over merkbare gevolgen buiten Facebook en discussies over onduidelijkheden wat Facebook allemaal kan analyseren. Men speculeert over wat Facebook allemaal analyseert van profielen en hoe het advertentie systeem werkt;

> "Naja, een tijd terug had een vriend van me dat ook!
> Hij vroeg zich af waarom er allemaal reclame specifiek voor homo's op zijn Facebook pagina stonden.
> Toen bleek dat er per ongeluk was ingesteld dat hij geïnteresseerd was in mannen!
> (Tweakers.net, 2010)

Het bovenstaande bericht is een voorbeeld van een dergelijk bericht. De gebruikers van de websites kijken naar wat men merkt en hoe Facebook aan deze informatie komt. Datamining is hier gebruikt om gebruikers te categoriseren en te clusteren en vervolgens te voorzien van advertenties die aantrekkelijk zijn voor deze categorieën. Ook worden hier vragen gesteld over de wenselijkheid van advertenties wanneer iemand bijvoorbeeld zijn seksuele voorkeur heeft afgeschermd. De gebruikers geven aan het niet prettig te vinden wanneer Facebook deze informatie alsnog gebruikt om deze gebruikers te voorzien met advertenties op basis van afschermde informatie. Andere gebruikers geven hierop aan dat wanneer men wil dat deze informatie niet wordt gedeeld met anderen, de gebruikers deze dan ook beter niet online kunnen zetten. Tevens wordt gekeken naar andere merkbare gevolgen van deze analyses. Dit zijn vooral merkbare gevolgen op andere sites. Bepaalde gebruikers geven aan dat hun vrienden bepaalde interesses hebben en dat deze gebruikers worden voorzien van advertenties op basis van deze interesses;

> "[…][V]oorbeeld: ik ga wel een sushi eten met mijn vrienden. mijn email adres wordt dus geassocieerd met mensen die dat wel eens doen ook als ik daar zelf niets over vertel. [V]ervolgens log ik ergens in waar mijn email geregistreerd staat en krijg ik allemaal sushi-ads. Gelukkig laat ik mij niet in met viespeuken anders had ik hele andere adds te zien gekregen" (Tweakers.net, 2011)

De gegevens worden gekoppeld aan advertenties op externe sites volgens de gebruikers van de websites. Facebook combineert waarschijnlijk zijn informatie met een extern advertentie systeem. Het email adres wordt gekoppeld aan iemand persoonlijke voorkeuren en berichten. Tevens wordt dit gekoppeld aan een cookie op een computer van een gebruiker. De externe site maakt contact met dit cookie en laat vervolgens advertenties zien op basis van deze interesses en voorkeuren.

> "[…] als jij de cookie van facebook nog hebt staan en je komt op een pagina met "like" button registreerd facebook dat jij die pagina bezocht OOK ALS JE ER NIET OP KLIKT! Dus jij leest je in over Libie op cnn.com? Facebook weet het. Jij eet op woensdag wel een pizzaatje? Facebook weet het. Jij bezocht een pagina met informatie over hoe je met aids moet leren leven? Facebook weet het. Jij leest je in over een bepaalde psychologische aandoening? Facebook weet het. Jij leest hier nu dit artikel en facebook weet het." (Tweakers.net, 2010)

Tevens zijn er veel speculerende berichten over wat Facebook allemaal kan weten en welke informatie buiten Facebook gekoppeld kan worden aan de Facebook data. Omdat Facebook weinig loslaat over de koppeling van gegevens aan derden, is het een speculerend karakter wat deze berichten domineert.

4.3.5 Inbreuk op privacy

Hetgeen waarover de meeste discussie gaande is, is over de categorie 'inbreuk op privacy'. Dit zijn berichten die zich richten op een inbreuk op privacy in een formele zin. De gegevens die men achterlaat op Facebook worden doorgegeven aan derden en dit wordt niet gewaardeerd door gebruikers. Tevens staan hier ook speculerende berichten in over wat Facebook doorgeeft aan derden.

Een belangrijke mogelijkheid in dit verband is de koppeling van de software aan Facebook. Zo kunnen ontwikkelaars allerlei Facebook-games ontwikkelen of allerlei sociale media-software, als bijvoorbeeld Hoot-suit, hieraan toevoegen. Een voorbeeld van een Facebook-game is Farmville, waarbij gebruikers een virtuele boerderij moeten beheren en zijn vrienden kan inschakelen om hem te helpen.

Gebruikers van websites geven aan dat ze het niet waarderen wanneer deze gegevens worden gegeven aan deze software ontwikkelaars. Veel belangrijke en persoonlijke data kan worden gegeven en gecombineerd met andere data die deze software ontwikkelaars in hun bezit hebben. Tevens zijn gebruikers bang dat deze ontwikkelaars op hun beurt deze gegevens weer doorgeven aan derden. Datawarehousing kan ervoor zorgen dat bestaande data wordt gekoppeld aan gegevens van derden. Hierop kunnen opnieuw analyses worden gedaan en dit kan leiden tot nieuwe inzichten. Net als bij de Bonuskaart zijn er gebruikers die mogelijkheden zien voor opsporing- en preventie-instanties om te zien waar gebruikers waren in het geval van een delict. Facebook geeft overigens aan dat ze aan dergelijke onderzoeken medewerking verlenen.

> "Punt hier is natuurlijk ook dat niet iedereen beseft in hoeverre pers. gegevens worden 'gedeeld' met derden. Hoe duidelijk is 'een' facebook hierover? Er is zoiets als consumentenbescherming tegen al te geld-geile bedrijven, iig in Nederland. En da's goed in mijn ogen. Er ligt een verantwoordelijkheid bij bedrijven die gratis diensten aanbieden. De relatie met de klant is minder helder dan bij transacties waar geld wordt betaald. De klant moet goed opletten, het bedrijf echter ook." (Tweakers.net, 2010)

Ook zijn er gevallen waarbij het niet gaat over wat Facebook doet met de opgeslagen gegevens maar de beschikbaarheid of gaten in de beveiliging van deze gegevens voor illegale derden. Hoewel Facebook probeert de gegevens te beschermen van dit soort partijen geven gebruikers aan dat de bescherming nooit volledig is. Tevens geven ze hierbij aan dat deel van de gegevens openbaar is en beschikbaar is te vinden door eenvoudig vriend van een van de vrienden van een bepaalde gebruiker te worden.

4.3.6 Relativerende reacties

Tot slot zijn er nog de relativerende reacties op de discussies die ontstaan over Facebook, privacy en datamining. Dit zijn gebruikers die het probleem van de discussies niet inzien, die aangeven dat deelname aan Facebook geen verplichting is en dat er voldoende oplossingen zijn binnen Facebook om privacy te garanderen. Deze mensen geven aan dat ze het over het algemeen niet erg vinden dat dergelijke analyses worden gedaan. Ook geven ze aan het niet erg te vinden dat hun gegevens mogelijk gekoppeld worden aan de gegevens van derden.

> "Zelf al zou dit het geval zijn, mij boeit het niet egt. Ik krijg overal waar ik kijk reclame advertenties gericht op iets van mij, dat kan werk, hobby of wat dan ook zijn."(Tweakers.net, 2011)

Een andere vorm van relativeren is dat men vaak aangeeft dat deelname aan Facebook geen verplichting is.

> "[…] Facebook verplicht jou niet om alle gegevens juist in te vullen. Je mag alle namen invullen, gegevens afschermen van partijen en je profiel verwijderen." (Tweakers.net, 2010)

Andere bespreken dat het geen gemis is wanneer een account niet bestaat op Facebook. Daarnaast poneert men dat niet alles wat hij/zij doet hoeft gemeld behoeft te worden op het Internet. Door geen account aan te maken of door select om te gaan met het plaatsen van gegevens en berichten zou privacy behouden kunnen blijven. Op die manier wordt de digitale voetafdruk kleiner volgens deze personen.

Er wordt tevens gekeken naar oplossingen voor de privacy problematiek omtrent Facebook. Facebook biedt tenslotte allerlei mogelijkheden om accounts af te schermen. Tegelijkertijd komt er veel kritiek los

op een dergelijke reactie. Gebruikers tonen aan dat de mogelijke afscherming van de gegevens niet altijd werkt. Zo had een gebruiker zijn gegevens afgeschermd voor andere gebruikers maar ging dit testen door een fictief profiel aan te maken. Wat bleek was dat niet alle privacy instellingen even goed werkten en hij daardoor deze vertrouwelijke gegevens toch kon benaderen. Andere gebruikers reageren vooral kritisch omdat er toch analyses gedaan worden op basis van profielen ondanks dat men zijn of haar gegevens afschermt.

4.3.7 Privacy Preserving Datamining

Omdat bepaalde informatie eenvoudig via openbare media te vinden is , is het discutabel of Privacy Preserving datamining kan helpen ter afscherming alle persoonlijke gegevens in het geval van de databank van Facebook. De focus van de discussies over Facebook, datamining en privacy ligt met name op de derde categorie inbreuk, er is met name aandacht voor het doorverkopen van gegevens aan derden waartegen Privacy Preserving datamining in mindere mate kan beschermen. Ook hier blijkt een discrepantie te bestaan tussen deze datamining oplossing en de discussies die gaan over Facebook, datamining en privacy.

4.3.8 Bescherming persoonsgegevens

Een mogelijkheid om de gegevens te beschermen en toch analyses te doen op de grote hoeveelheden data is het transformeren van de oorspronkelijke data door middel van Privacy Preserving datamining. De resultaten kunnen dan toch worden ingezet voor commerciële doeleinden. Dit is een mogelijkheid voor Facebook om de gegevens te beschermen. De discussies focusseren zich op wat men wel en niet achter laat op Facebook. Datamining kan ervoor zorgen dat op grote schaal inzichtelijk wordt wat men uitvoert en wat de gebruiker beweegt. Aan de hand hiervan kunnen adverteerders gebruikers benaderen met persoonlijke gerichte advertenties. Privacy Preserving datamining kan ervoor zorgen dat toch informatie gewonnen kan worden zonder dat de profielen geraadpleegd worden. Facebook gebruikt de gegevens uitsluitend om datamining op toe te passen , In het geval van service verlening zullen de Facebook medewerkers toch toegang moeten hebben tot persoonlijke gegevens van de gebruiker. Deze data afschermen voor Facebook-medewerkers kan hier tot problemen leiden bij het verlenen van een goede service. Tevens is het zeer de vraag waarom het wenselijk zou zijn voor Facebook om profielinformatie af te schermen wanneer deze toch al (deels-) openbaar is.

4.3.9 Analyse en Inzet

De profielgegevens worden geanalyseerd en ingezet voor het advertentie programma van Facebook. Het voorbeeld van de gebruiker dat hij achtervolgd werd met advertenties over sushi omdat zijn vrienden dit waarderen is een gevolg van datamining. Wanneer deze gebruiker de site verlaat maar ingelogd blijft kan zijn profiel worden gekoppeld aan cookies van andere sites waardoor advertenties toch kunnen worden getoond. Maar ook het inzicht dat men zijn relatie vaker verbreekt rond Valentijn dan in de rest van het jaar is een gevolg van de toepassing van datamining. De discussies over dit aspect gaan met name over dit advertentie systeem. De inzichten die worden gegeneerd door datamining worden ingezet voor commerciële doeleinden. De profielen van gebruikers worden geanalyseerd en ingezet. Privacy Preserving Datamining richt zich op het beschermen van de data die wordt gebruikt bij het dataminen. PPDM zou hier kunnen worden toegepast maar het zal gebruikers niet beschermen tegen de inzet van dergelijke gegevens. De individuele profielen kunnen immers worden gekoppeld aan statistieken.

Facebook is, volgens het privacy beleid, een mediator in het advertentieprogramma. Alleen zij hebben inzicht in de profielen van de gebruikers en adverteerders dienen op basis van algemene karakteristieken aan te geven bij welke profielen zij hun advertentie willen tonen. Dit gebeurt op basis van een keuze lijst. De adverteerders komen daardoor niet in aanraking met de persoonlijke informatie van de gebruikers, deze is alleen zichtbaar voor Facebook zelf. Het transformeren van de data door middel van Privacy Preserving Datamining aan de bron is in dit opzicht daarom niet per se noodzakelijk, Facebook heeft deze data door het ‘mediator’-model naar hun eigen zeggen afdoende beschermd. Adverteerders kunnen alleen karakteristieken aflezen.

4.3.10 Inbreuk op privacy

De discussie over Facebook, datamining en privacy focust met name op de derde categorie ‘inbreuk op privacy’ en het doorverkopen van de data in Facebook aan derden. Privacy Preserving datamining richt zich op bescherming van de oorspronkelijk data bij het toepassen van datamining. Data wordt getransformeerd zodat de analist die datamining toepast individuele data niet kan inzien. Op zich beschermt dit daarmee niet tegen doorverkoop aan derden. Een andere mogelijkheid van PPDM is het slechts in delen toegankelijk maken van de data door middel van Multi Layered Security, bepaalde partijen zouden bijvoorbeeld alleen algemene karakteristieken van de data in de Facebook database kunnen raadplegen. Dit is hoogstwaarschijnlijk momenteel het geval bij de database van Facebook en het gebruik van de data door softwareontwikkelaars. Facebook geeft hierbij aan dat het gaat om gebundelde en geanalyseerde data. Echter blijkt dat men ook bezig is met het doorgeven van contactinformatie, zoals telefoonnummers en adressen, aan derden. De recentelijke ontwikkeling waar softwareontwikkelaars inzicht zouden mogen krijgen in contact informatie van gebruikersprofielen geeft aan dat er een dergelijk systeem bestaat. Voorheen mochten ontwikkelaars alleen algemene, niet persoonlijke informatie, die eindgebruikers op ‘toegankelijk voor iedereen’ hebben gezet, raadplegen. Facebook is beheerder van deze database en kan de rechten van toegang bepalen. De ontwikkelaars en overige derden die deze data zouden afnemen kunnen deze data weer koppelen aan hun eigen database. Datamining kan vervolgens voor zorgen dat hierop analyses gedaan kunnen worden en daarmee nieuwe informatie te vergaren is. Privacy Preserving datamining en Multi Layered Security kan alleen helpen wanneer de eigenaar van de database de juiste toegangskeuzes maakt. De eigenaar blijft in het beheer van de data. Wanneer een partij beslist om persoonlijke informatie te delen, al dan niet door middel van een opt-in, beschermt deze oplossing niets. De keuze van het delen van informatie en het verkopen van profielen is een element waar alleen Facebook een beslissing over kan nemen.

Tevens is er een discussie gaande over het feit dat de profielinformatie toch al ten dele openbaar is. De profiel informatie is vaak openbaar of is vaak via vrienden of via vrienden van vrienden in te zien. Toch blijft de data in het beheer van Facebook en worden bijvoorbeeld softwareleveranciers die de data die ze tot hun beschikking hebben gekregen en doorverkocht hebben bestraft. Facebook probeert zijn data te beschermen tegen allerlei hackeraanvallen en andere gevaren waardoor de beschermde data in handen komt van andere partijen. Situaties als bijvoorbeeld de ‘Gaydar’ laten zien dat Facebook hier niet altijd in slaagt. Ook hier kan PPDM niet voor een oplossing zorgen aangezien het de profieldata is die wordt gedownload. Juist in deze oorspronkelijke data zijn deze illegale partijen geïnteresseerd en in het koppelen van deze data aan andere data zoals dat gebeurt in het proces van datawarehousing. In deze eerste fase van datamining (pre-datamining) biedt Privacy Preserving datamining geen bescherming tegen inbreuk van privacy.

4.3.11 Relativerende reacties

Zoals ook in het geval van de Bonuskaart en in het geval van de HKS database blijkt dat ook ten aanzien van Facebook er relativerende reacties zijn. Bij een dergelijke benadering waarbij uitgegaan wordt van goed vertrouwen. Deze personen zullen naar verwachting weinig tot geen inspanning verrichten om hun persoonlijke gegevens goed te beschermen. Een middel zoals PPDM biedt dan nauwelijks extra zekerheid omdat effectieve toepassing valt of staat met een zekere zelfdiscipline. Ga je laks om met je gegevens dan loop je een vergroot risico op potentieel misbruik. .Deze personen geven aan dat ze het niet erg vinden wanneer ze benaderd worden met persoonlijke aanbiedingen. Tevens wordt er hier vanuit gegaan dat men veel privacy instellingen in Facebook zelf kan beheren zodat niet iedereen alles kan zien. Dit wordt bekritiseerd met reacties dat deze instellingen soms niet werken. Daarnaast wordt vaak gezegd dat er op Facebook geen verplichte deelname is, wanneer iemand angstig zou zijn voor zijn of haar privacy dan hoeft hij of zij niet perse een account te nemen.

4.3.12 Conclusie

Privacy Preserving datamining kan in het geval van Facebook niet een eenduidige oplossing bieden voor de gestelde problematiek van potentieel misbruik. Dit komt onder andere doordat de discussies over privacy omvangrijk, complex, subjectief en divers zijn. Privacy Preserving datamining richt zich op het beschermen van de oorspronkelijke data in een database. Een online sociaal netwerk, waarbij veel van die informatie ook publiekelijk toegankelijk is, leent zich zeer goed voor datamining zo blijkt

uit het advertentieprogramma van Facebook. Ook kan deze informatie over levensfasen van gebruikers zeer interessant zijn voor derden. Het doorverkopen van gegevens is iets waar PPDM geen antwoord op geeft. Tevens is het beschermen van de oorspronkelijke data complex aangezien veel van deze informatie toch al (deels-) openbaar is. Daarmee is er in zekere mate een discrepantie tussen de softwarematige oplossing PPDM en de culturele discussies die gaan over Facebook, privacy en datamining.

5. Analyse en reflectie

In dit hoofdstuk zullen de resultaten uit de casestudies met elkaar in verband worden gebracht en gerelateerd aan het theoretisch kader zoals geschetst in hoofdstuk drie. PPDM biedt, zoals al is aangetoond, slechts ten dele de noodzakelijke extra bescherming van gegevens en spreekt ook zeker niet iedereen aan. Er is een discrepantie geconstateerd tussen de technische oplossing en de problematiek die door de gebruikers van de websites wordt genoemd. Daarom wordt ook gekeken naar een mogelijk andere oplossing voor de problematiek die met ongeconditioneerde datamining samenhangt. De eerder genoemde relativerende discussies worden buiten beschouwing gelaten. PPDM is geen oplossing voor deze groep discussies.

5.1 Bescherming Persoonsgegevens

Zoals eerder aangetoond, gaan de discussies in de categorie Bescherming Persoonsgegevens, over welke persoonsgegevens worden opgeslagen van gebruikers van de Bonuskaart, het HKS en Facebook. Bij de Bonuskaart gaat het met name over de vraag welke gegevens men achterlaat en waarom die achter gelaten moeten worden, terwijl gebruikers van de websites bij het HKS impliciet uitgaan van dat de overheid al veel data in beheer heeft. Bij Facebook richten de discussies zich vooral op de keuze wat men allemaal online dient te plaatsen en waar men rekening mee dient te houden.

Kenmerkend van de groep discussies over de bescherming van de persoonsgegevens is, dat vaak juridische vragen worden gesteld over wat wel en niet mag worden gedaan met de gegevens. Interessant hierbij is dat gebruikers in het geval van de Bonuskaart en Facebook wel akkoord gaan met de algemene voorwaarden, waarin een aantal van de doelstellingen voor de opslag van hun gegevens staan geformuleerd. Zoals in hoofdstuk twee aangetoond werd door Vedder (1998) en Custers (2004) zijn individuele gegevens wel beschermd worden onder de Wet Bescherming Persoonsgegevens maar dat groepsgegevens en resultaten die worden behaald met datamining niet worden beschermd. In het geval van het HKS kan worden gezegd dat specifieke uitzonderingen gelden op de WBP wanneer wordt gesproken over strafrechtelijk onderzoek naar verdachten. In navolging van Vedder (1998) en Custers (2004) wordt in dit onderzoek gesteld dat een politieke en juridische discussie moet plaatsvinden over de rol van datamining in de Wet Bescherming Persoonsgegevens.

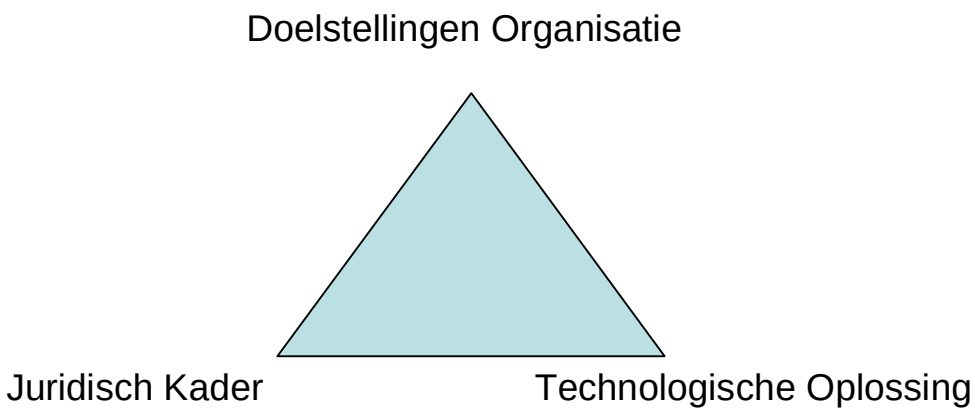
Privacy Preserving datamining zou bij uitstek voor de categorie Bescherming Persoonsgegevens een oplossing moeten bieden. Echter blijkt dat in discussies uit deze categorie veel meer problemen van andere aard genoemd worden waarvoor PPDM geen oplossing kan bieden. PPDM kan niet voorkomen dat gegevens überhaupt worden opgeslagen en PPDM kan discussies niet voorkomen die gaan over de vraag waarom die gegevens opgeslagen dienen te worden.

5.2 Analyse en Inzet

Een belangrijke categorie van discussies gaat over de analyse van de persoonsgegevens en de inzet van de data-analyse. Discussies uit deze categorie gaan vooral over de ‘datamining-fase’ en de ‘post datamining-fase’, zoals deze getoond zijn in Figuur 1 (uit hoofdstuk één). Privacy Preserving datamining beschermt de toegang tot persoonsgegevens bij het toepassen van datamining. Figuur 1 laat zien dat PPDM zich met name richt op de eerste en tweede fase van het gehele proces van datamining. Wanneer de datamining een resultaat oplevert wordt er overgaan naar de derde fase van datamining, ‘post datamining-fase’. In de discussies over de nieuwsberichten worden er met name vragen gesteld over wat de resultaten van de analyses zijn en wat gedaan kan worden met deze resultaten.

Vooral bij het HKS bleek deze groep van discussies dominant. Juist angsten voor ‘Big Brother’ en ‘Minority Report’-achtige praktijken waren onderwerp van discussie. Men zou al verdachte kunnen worden voordat men überhaupt iets heeft gedaan. De resultaten van datamining dienen te worden geanalyseerd en geïnterpreteerd. Privacy Preserving datamining kan burgers niet beschermen tegen de interpretatie van resultaten van datamining. Wanneer er fouten worden gemaakt in interpretatie van deze resultaten kan dat leiden tot individuele problematiek.

Gebruikers gaven aan dat door de grote hoeveelheden data die overheden tot hun beschikking hebben, veel van burgers meetbaar is. Ook wordt ditzelfde argument bij de Bonuskaart genoemd. Een ander element uit commercieel datamining zoals dat gebeurt in het geval van de Bonuskaart en de database van Facebook is de discussies over merkbare gevolgen van datamining. Men ontvangt persoonlijke advertenties of aanbiedingen via hun profiel of andere kanalen. Deze situatie wordt eveneens vergeleken met Big-Brother. Bedrijven weten alles van hun consument en zorgen ervoor dat persoonlijke aanbiedingen deze consument zal bereiken. De situatie die wordt genoemd in verband met Facebook is wanneer iemand zijn seksuele voorkeur afschermt voor andere gebruikers door middel van advertenties toch deze seksuele voorkeur wordt getoond. Individuen worden aangesproken op de resultaten van datamining op geanonimiseerde data. Zoals Figuur 1 laat zien worden door middel van PPDM persoonsgegevens beschermd, wordt datamining toegepast, maar wordt het individuen uiteindelijk aangesproken door de inzet van algemene analyses. Wanneer politie beleid gaat maken op de resultaten die voortkomen uit datamining op het HKS, biedt PPDM geen uitkomst. Als politie criminele carrièrepaden gaat voorspellen aan de hand van de resultaten van datamining worden algemene analyses toegepast op een individueel geval. Als de Albert Heijn klanten wil voorzien met persoonlijke aanbiedingen, worden algemene analyses worden toegepast op een individueel niveau. Juist het beschermen van het individu was het uitgangspunt voor het ontwikkelen van het PPDM. PPDM beschermt dit individu in dit geval niet tegen de inzet van de resultaten van datamining.



Figuur 2 Spanningsveld tussen het juridische kader, doelstellingen van een organisatie en technologische oplossingen

Belangrijk is het spanningsveld tussen het juridische kader, doelstellingen van een organisatie en de technologische oplossingen. Dit is geïllustreerd in Figuur 2. Privacy Preserving datamining blijkt ingezet te worden als een technologische oplossing voor een cultureel probleem. Problematisch bij de keuze voor een oplossing als PPDM, zijn de doelstellingen van organisaties. Wanneer organisaties bepaalde doelstellingen hebben voor het opslaan en gebruik van data, staan deze doelstellingen boven een oplossing als Privacy Preserving datamining. Organisaties zullen in sommige gevallen niet kiezen voor een oplossing als PPDM omdat het niet aansluit bij hun doelstellingen. Wanneer politie datamining wil inzetten om criminaliteit tegen te gaan zal een oplossing als PPDM niet wenselijk zijn. Het gaat de politie juist om individuen die potentieel crimineel handelen. Het beschermen van gegevens een potentieel crimineel is daarom niet wenselijk. Of het mag, bepaald het juridisch kader. Bij het bespreken van het HKS is duidelijk geworden dat datamining niet eenduidig te plaatsen is in de politietaak. Het toepassen van datamining in het kader van strafrechtelijk onderzoek vindt plaats binnen het eerder geschetste juridisch kader. De keuze voor een technologisch oplossing vindt altijd plaats in relatie tot de doelstellingen van een organisatie en de juridische randvoorwaarden. De mate waarin een technologische oplossing succesvol is wordt bepaald binnen het genoemde spanningsveld waarbij de andere factoren juridisch kader en doelstelling voor het gebruik en opslag al dan niet doorslaggevend zijn.

5.3 Inbreuk op privacy

Privacy Preserving datamining kan burgers en consumenten niet kan beschermen tegen een formele inbreuk op privacy. Als organisaties, als de Albert Heijn, gegevens gaan doorverkopen of koppelen aan derden, heeft PPDM geen antwoord. Opnieuw geldt hierbij dat Privacy Preserving datamining en technologische oplossing is waar overheden en bedrijven voor kunnen kiezen. PPDM kan helpen, maar is zeker niet een volledige oplossing om negatieve effecten van datamining te voorkomen. Doelstellingen over het gebruik van gegevens en het doorverkopen van gegevens zullen dominanter zijn dan de keuze voor een technologische oplossing om privacy te beschermen. Zo blijkt hierbij opnieuw een discrepantie te ontstaat tussen de technologische oplossing en de veelheid aan problemen die gebruikers in reacties op nieuwsberichten noemen.

Het delen van gegevens uit de database van de Bonuskaart met justitie is een inbreuk op privacy in formele zin. Het WBP wordt overtreden maar de strafrechtelijke uitzondering hierin maakt het toch mogelijk. Gebruikers vergelijken de inzage van justitie in het Bonuskaart systeem met Big Brother. Overheden kunnen door het combineren van grote hoeveelheden data in principe van alles van consument en burger te weten komen. Er dient daarom een juridische en ethische discussie te komen over of datamining/datawarehousing en hoe deze te plaatsen valt binnen de taak van politie. Privacy Preserving datamining kan gegevens uit deze database niet beschermen tegen het delen met justitie of met bedrijven.

Zoals in het geval van de database van Facebook worden de resultaten van datamining gedeeld met derden. Dit bleek uit een actuele situatie waarin softwareontwikkelaars door slechts één druk op de knop de contactgegevens van gebruikers konden opvragen. Discussies omtrent Facebook, privacy en datamining focussen met name op dit aspect van datamining. Door middel van datawarehousing kunnen deze softwareontwikkelaars hun eigen data koppelen aan de data van contactgegevens uit de database van Facebook. Gebruikers van de websites gaan vanuit dat adverteerders tegen betaling deze contactgegevens kunnen verkrijgen. Privacy Preserving datamining beschermt gebruikers in ieder geval niet tegen eventuele doorverkoop. Hierbij geldt opnieuw dat wanneer Facebook de doelstellingen hebben op persoonsgegevens door te verkopen aan derden een technologische oplossing als PPDM geen oplossing is. Ook hierbij geldt het juridische kader wat het delen van gegevens wel of niet mogelijk maakt.

5.4 Oplossingen

Uit de casestudies blijkt dat Privacy Preserving datamining slechts delen van de genoemde problematiek in reacties op de nieuwsberichten kan helpen oplossen. Alleen de achter gelaten persoonsgegevens kunnen worden beschermd maar wanneer algemene analyses individueel worden ingezet biedt PPDM geen oplossing. Daarom wordt nu gekeken naar een bredere oplossing voor de genoemde problematiek.

Veel discussies omtrent privacy, zoals bij de Bonuskaart en bij het HKS, komen voort uit onwetendheid over wat met de gegevens wordt gedaan. Er dient meer transparantie te komen in het gebruik van de gegevens. Voor velen is het een ‘black box’ wat er gebeurt met zijn of haar gegevens. Overheden en bedrijven dienen transparanter te zijn in de doelstellingen met betrekking tot de gegevens. Hoewel dit een verplichting is wanneer persoonsgegevens worden beschermd onder de Wet Bescherming Persoonsgegevens, blijkt dat dit bij Albert Heijn niet eenvoudig te achterhalen is. Tevens resteert de vraag welke rol datamining speelt binnen de WBP. Zoals Vedder (1998) en Custers (2004) aangaven dient er een politiek debat te komen over de bescherming van groepsgegevens.

In het kader van strafrechtelijke wetshandhaving blijkt men te speculeren over waar zijn/haar gegevens worden geregistreerd en wat over hem/haar wordt geregistreerd. Tevens dient er duidelijker geformuleerd te worden wat het doel is van het opslaan van de gegevens. Wanneer inzichtelijk wordt welke analyses bedrijven en overheden toepassen op databases dan wordt een deel van de speculerende discussie ondervangen. Momenteel blijven de doelstellingen van organisaties in het gebruik van gegevens en de samenhangende registraties vaag. Dit is het geval bij de Bonuskaart maar ook het geval bij het HKS. Wanneer deze doelstellingen concreter en toetsbaar worden geformuleerd kunnen consumenten en burgers zelf beter bepalen wat er met hun gegevens wordt

gedaan en of zij dat wenselijk achten. Ook moet beter worden verwoord wat de mogelijke gevolgen zijn van de analyses voor de consument of burger. Zowel bij de Bonuskaart, bij het HKS en bij de database van Facebook wordt druk gespeculeerd over welke combinaties worden gelegd met andere databases.

Tevens is het voor consumenten en burgers niet altijd duidelijk wat de rechten en plichten ten aanzien van zijn of haar persoonsgegevens zijn. Wanneer het duidelijker is dat consumenten inzicht mogen krijgen in bepaalde systemen om te kunnen zien wat er met hun gegevens gebeurt wordt een deel van de discussie over datamining en privacy ondervangen. Dit heeft ook betrekking op data opslag in het algemeen.

Conclusie

Eerst is in dit onderzoek een algemeen kader geschetst van wat datamining is, waarom het wordt toegepast en uit welke fase datamining bestaat. Dit is geïllustreerd in Figuur 1. Vervolgens is gekeken naar wat Privacy Preserving datamining is. Datamining blijkt uit discussies op nieuwsberichten en uit literatuur controversieel. Privacy zelf is een complex begrip waarvan de betekenis cultureel en historisch kan verschillen (Lyon, 2007). De discussies over privacy en datamining zijn in te delen in vier categorieën. De eerste groep discussies gaat over de Bescherming van de persoonsgegevens die zijn opgeslagen, de tweede groep discussies gaat over de analyse van deze gegevens en de inzet van de resultaten van de analyses, de derde groep discussies gaat over een inbreuk op privacy en de vierde groep discussies is te kenmerken als relativerend. Reacties uit de eerste categorie gaan bijvoorbeeld over de vraag welke gegevens opgeslagen worden en of dit überhaupt wel mag. Reacties uit de tweede categorie gaan over wat allemaal geanalyseerd kan worden en wat hiermee gedaan kan worden. Reactie uit de derde categorie gaan bijvoorbeeld over het doorverkopen van gegevens aan derden of het koppelen van databronnen. De laatste categorie gaat vaak over het in hoofdstuk twee toegelichte 'nothing-to-hide'-argument.

Terugkomend op de hoofdvraag *in hoeverre Privacy Preserving datamining als een oplossing kan fungeren voor de problematiek die voortvloeit uit datamining* kan worden geconcludeerd dat een discrepantie bestaat tussen de oplossing van Privacy Preserving datamining en de groepen discussies. Aan de hand van drie casestudies, de Bonuskaart, het HKS en de database van Facebook, is aangetoond dat Privacy Preserving datamining slechts in enkele gevallen een oplossing kan zijn. Privacy Preserving datamining richt zich op het beschermen van de oorspronkelijke persoonsgegevens.

Privacy Preserving datamining modificeert persoonsgegevens zodat organisaties gegevens van personen niet kunnen raadplegen. Op de totale dataset van deze gemodificeerde gegevens kunnen datamining analyses worden gedaan. Er is in dit verband nog geen privacyprobleem. Alleen wanneer deze gegevens nog niet geanonimiseerd zijn, kan Privacy Preserving datamining zorgen dat data niet meer te herleiden is tot een persoon. Zoals uit de casestudies blijkt is de problematiek aanzienlijk complexer. De doelstellingen die organisaties nastreven zijn het belangrijkste leidende principe voor hoe en wat er met gegevens zal worden gedaan.

Bij de analyse, interpretatie en toepassing van de resultaten echter, kan er wel een probleem ontstaan omdat algemene resultaten specifiek toegepast kunnen worden. Persoonsgegevens worden beschermd door PPDM. Analyses op totale, geanonimiseerde data worden gedaan door de toepassing van datamining. Alleen wordt de inzet van de resultaten van de analyses op het individu gedaan. Organisaties, als de Albert Heijn, Facebook en politie, kunnen namelijk aan hand van de resultaten van datamining persoonlijk adverteren en individuele criminele carrière paden voorspellen. Privacy Preserving datamining kan het niet voorkomen dat de analyses die worden gedaan met datamining worden ingezet voor strafrechtelijk onderzoek. Ook kan Privacy Preserving datamining men niet beschermen tegen eventuele doorverkoop van gegevens aan derden. Doelstellingen, zoals het doorverkopen van persoonsgegevens, van organisaties staan boven de keuze voor een technologische oplossing als PPDM.

Een deel van de problematiek kan worden ondervangen door transparantie te creëren in welke gegevens worden opgeslagen, wat met deze gegevens wordt gedaan en welke gevolgen dit kan hebben voor burger en consument. Tevens dient de burger of de consument beter op de hoogte zijn van rechten en plichten ten aanzien van gegevensopslag. Er dient hierbij inzicht te worden gecreëerd in hoeverre datamining te plaatsen valt onder de Wet Bescherming Persoonsgegevens door een politiek en maatschappelijk debat. Daarnaast dienen de doelstellingen van organisaties voor het opslaan van gegevens en systemen, concreet en toetsbaar te worden geformuleerd. Door transparantie te creëren in de opslag van de gegevens wordt voor een gedeelte ingespeeld op elementen waar Privacy Preserving datamining te kort schiet.

Literatuur

Agrawal, Rakesh en Srikant Ramakrishnan. “Privacy Percerving Data Mining”. [2000] *IBM Almaden Research Centre*. Geraadpleegd op 1 april 2011 via http://www.almaden.ibm.com/cs/projects/iis/hdb/Publications/papers/sigmod00_privacy.pdf

Aggarwal, Charu C. en Yo, Philip S. *Privacy Preserving Data Mining: Models and Algorithms*. Springer Science en Business Media. New York; 2008.

Ah.nl. “Algemene voorwaarden ah.nl, AH Wijndomein, AH Fotoservice en AH Telecom”. [2011a] *Albert Heijn*. Geraadpleegd op 8 april 2011 via http://www.ah.nl/artikel?trg=article.disclaimer

Ah.nl. “AH Bonuskaart”. [2011b] *Albert Heijn Bonuskaart*. Geraadpleegd op 8 april 2011 via http://www.ah.nl/bonuskaart

Algemene Inlichtingen en Veiligheidsdienst. “Jaarverslag AIVD 2009”. [2009] *AIVD; Ministerie van Binnenlandse Zaken en Koninkrijksrelaties*. Geraadpleegd op 26-10-2010 via http://www.jaarverslag.aivd.nl/downloads/Jaarverslag_2009_AIVD.pdf

Algemene Inlichtingen en Veiligheidsdienst. “Kwetsbaarheidanalyse Spionage”. [2010] *AIVD; Ministerie van Binnenlandse Zaken en Koninkrijksrelaties*. Geraadpleegd op 26-10-2010 via https://www.aivd.nl/@126104/item-126104

Alphenaar, Jan Willem. “Wij zijn verworden tot grondstoffen in de datamining industrie” [2010] *Marketingfacts.nl*. Geraadpleegd op 10 april 2011 via http://www.marketingfacts.nl/berichten/20110217_wij_zijn_verworden_tot_grondstoffen_in_de_datamining_industrie/

Brookshear, Glenn, J. *Inleiding Informatica. 8e editie*. Pearson Education Benelux, Amsterdam; 2009.

Castells, Manuel. *The Internet Galaxy; Reflections on the Internet, Business and Society*. Oxford University Press; 2001.

Cocx, Tim. “Algorithmic Tools for Data-Oriented Law Enforcement”. [2009] *Institute for Programming research and Algorithmics*. Geraadpleegd op 13-4-2011 via https://openaccess.leidenuniv.nl/bitstream/1887/14450/5/thesis.pdf

College Bescherming Persoonsgegevens. “Albert Heijn: Bonuskaart”. [1999] *College Bescherming Persoonsgegevens.* Geraadpleegd op 8-1-2011 via http://www.cbpweb.nl/Pages/uit_97V0034.aspx

Centraal Bureau voor de Statistiek. “SSB informatie, 2010”. [2010] *Centraal Bureau voor de Statistiek* SSB informatie. Geraadpleegd op 17 april 2011 via http://www.cbs.nl/nl-NL/menu/informatie/onderzoekers/ssb/ssb-info-medio-07.htm

Centraal Bureau voor de Statistiek. “Verdachten misdrijven methode”. [2011] *Centraal Bureau voor de Statistiek* Verdachten Misdrijven Methode. Geraadpleegd op 16 april 2011 via http://www.cbs.nl/nl-NL/menu/themas/veiligheid-recht/methoden/dataverzameling/korte-onderzoeksbeschrijvingen/verdachten-misdrijven-methode.htm

Custers, Bart. “Data mining bedreigt privacy”. [2004] *Rechtennieuws.nl*. Geraadpleegd op 10 april 2011 via http://rechtennieuws.nl/1332/data-mining-bedreigt-privacy.html

DeRosa, Mary. “Data Mining and Data Analysis for Counterterrorism”. *Centre for Strategic and International Studies*; 2004.

Engelfriet, Arnoud. “Politie mag niet doen met datamining van eigen bestand”. [2009] *Ius Mentis*. Geraadpleegd op 13 april 2011 via http://blog.iusmentis.com/2009/12/08/politie-mag-niets-doen-met-datamining-van-eigen-bestand/

Elmer, Greg. “A diagram of panoptic surveillance”. *New media & society*, SAGE Publications London; 2003.

Facebook. “Privacy Policy”. [2011a] *Facebook.com*. Geraadpleegd op 9 april 2011 via http://www.facebook.com/policy.php

Facebook. “Algemeen informatie Facebook”. [2011b] *Facebook.com*. Geraadpleegd op 12 april 2011 via http://www.facebook.com/facebook#!/facebook?sk=info

Facebook. “Facebook Advertising” [2011b] *Facebook.com*. Geraadpleegd op 13-4-2011 via http://www.facebook.com/advertising/

Fijnaut et al. Politie: *Studies over haar werking en organisatie*. Kluwer, Deventer: 2007.

Han, Jiawei en Kamber, Micheline. *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems, Morgan Kaufmann Publishers, 2006.

Hearst, Marti A. “Untangling Text Data Mining”. *School of Information Management & Systems University of California*, Berkeley; 1999.

Jacobs, Bart. “Select before you Collect”. *Institute for Computing and Information Sciences aan de Radboud Universiteit*. Mens en maatschappij 1006 AA 54 12, Nijmegen: 2005.

Kruize, Peter. en Kool, M. “Analyse van criminele loopbanen met behulp van politiegegevens”. [2002] *Website voor de politie*. Geraadpleegd 28 maart 2011 via http://www.websitevoordepolitie.nl/archief/analyse-van-criminele-loopbanen-met-behulp-van-politiegegevens-58.html

Michiels et. al. *Kluwer Collegebunde*l. Kluwer; Deventer; 2011.

Leenheer, Jorna. “Loyaliteitsprogramma’s: Zinvol CRM-instrument?”. Tijdschrift voor Economie en Management Vol. LI, 2; 2006.

Lyon, David. *Surveillance Studies: An Overview*. Polity Press, Camebridge; 2007.

Mena, Jesús. *Investigative data mining for security and criminal detection*. Elsevier Science, Burlington; 2003.

Naeyé, Jan. *Het Nederlandse politierecht; tekst en commentaar 2009-2010*. Kluwer, Deventer; 2009.

Nationale Ombudsman. “Ombudsman maakt zich zorgen over onjuiste gegevens”. [2010] *De Nationale Ombudsman*. Geraadpleegd op via http://www.nationaleombudsman-nieuws.nl/nieuws/2010/ombudsman-maakt-zich-zorgen-over-onjuiste-gegevens

Nederlands Vennootschap van Burgermeesters. “Ter Horst vindt registratie van hangjongeren een goede zaak”. [2011] *website voor het Nedelands Vennootschap van Burgermeesters*. Geraadpleegd op 17-4-2011 via http://www.burgemeesters.nl/node/654

Olsthoorn, Peter. “Politie test datamining criminelendatabank”. [2009] *Webwereld.nl*. Geraadpleegd op 12-1-2011 via http://webwereld.nl/nieuws/64515/politie-test-datamining-criminelendatabank.html

Pieters et. al. “Inzicht en toezicht: Controle in de kennissamenleving”. In: Jaarboek Kennissamenleving. Uitgeverij Aksant, Amsterdam; 2010.

Schroeder, Stan. “GAYDAR: Your Facebook Friends Can Reveal Your Sexual Orientation”. [2009] *Mashable.com*. Geraadpleegd op 10 april 2011 via http://mashable.com/2009/09/21/facebook-friends-sexual-orientation/

Security.nl. “AH deelt gegevens bonuskaart met Justitie”. [2008] *Security.nl*. Geraadpleegd op 17 april 2011 via http://www.security.nl/artikel/23260/AH_deelt_gegevens_bonuskaart_met_justitie.html

Siebes, Arno. “Schatgraven in Databases”. *Centrum voor Wiskunde en Informatica*, Amsterdam; 1996.

Solove, Daniel. ““I’ve Got Nothing to Hide” and Other Misunderstandings of Privacy.” *George Washington University Law School*, 2008. Geraadpleegd op 8-1-2011 via http://papers.ssrn.com/sol3/papers.cfm?abstract_id=998565&rec=1&srcabs=930514

Tienkamp, Ewald. “Ook spionen waarschuwen voor verzamelsdrift”. [2010] *Bits of Freedom*. Geraadpleegd op 26-10-2010 via https://www.bof.nl/2010/09/17/ook-spionnen-waarschuwen-voor-verzameldrift/

Tweakers.net. “AH werkt aan RFID Bonuskaart”. [2006] *Tweakers.net*. Geraadpleegd op 17 april 2011 via http://tweakers.net/nieuws/43330/ah-werkt-aan-rfid-bonuskaart-update.html

Tweakers.net. “Facebook onthult geaardheid gebruikers via advertenties”. [2010] *Tweakers.net*. Geraadpleegd op 17 april 2011 via http://tweakers.net/nieuws/70393/facebook-onthult-geaardheid-gebruikers-via-advertenties.html

Tweakers.net. “EC privacybewakers moeten sterker staan tegen Facebook”. [2011] *Tweakers.net*. Geraadpleegd op 17 april 2011 via http://tweakers.net/nieuws/73276/ec-privacybewakers-eu-moeten-sterker-staan-tegen-facebook.html?mode=nested&max=10&niv=0&order=asc&page=1#reacties

Vaiyda et. al. “Privacy Preserving Data Mining”. *State University New Jersey*, Departement of Computer Science, New Jersey; 2006.

VanDale. “Privacy”. [2010] *VanDale.nl*. Geraadpleegd op 12 december 2010 via http://www.vandale.nl/vandale/zoekService.do?selectedDictionary=nn&selectedDictionaryName=Nederlands&searchQuery=privacy

Vassiliadis et. al. “Conceptual Modeling for ETL Processes”. *National Technical University of Athens*. DOLAP’02, November 8; 2002.

Vedder, Anton, H. “Het einde van de individualiteit? Groepsprofilering, datamining, brute pech en dom geluk”. *P&I Privacy en informatie;* 1998. Geraadpleegd op 17 april 2011 via http://arno.uvt.nl/show.cgi?fid=5073

Webwereld.nl. “Albert Heijn ontkent plannen RFID chip in Bonuskaart”. [2006] *Webwereld.nl*. Geraadpleegd op 17 april 2011 via http://webwereld.nl/nieuws/41876/albert-heijn-ontkent-plannen-rfid-chip-in-bonuskaart.html

Webwereld.nl. “Politie test datamining criminelendatabank”. [2009a] *Webwereld.nl*. Geraadpleegd op 17 april 2011 via http://webwereld.nl/nieuws/64515/politie-test-datamining-criminelendatabank.html

Webwereld.nl. “Politie gaat datamining zeker gebruiker”. [2009b] *Webwereld.nl*. Geraadpleegd op 17 april 2011 via http://webwereld.nl/nieuws/64531/-politie-gaat-datamining-zeker-gebruiken-.html

Wu et. al. “Top 10 algorithms in datamining”. *Knowledge Information Systems*, Springer verlag Londen; 2008.

Bijlage

1. Soorten methoden en algortimen voor datamining analyse
2. Voorbeeld Apriori analyse
3. Nieuwsberichten

Bijlage 1. Soorten methoden en algoritmen voor datamining analyse

Methoden voor analyse

Categorie beschrijvingstechniek is een methode om eigenschappen te koppelen aan gegevens. Brookshear (2009) noemt het voorbeeld van het vinden van kenmerken van mensen die kleine, zuinige autootjes willen kopen. Kenmerk X zegt dan iets over deze mensen, het is het zoeken naar overeenkomsten tussen data en het toekennen van kenmerken over deze data. Categorie onderscheidingstechniek is een manier om de informatie onderling te scheiden door het toekennen van kenmerken. Bijvoorbeeld mensen die op zoek zijn naar een nieuwe auto te scheiden van mensen die op zoek zijn naar een tweedehands auto. Het is het zoeken naar verschillen tussen de gegevens en het indelen van deze gegevens in categorieën.

Een clusteranalyse is het vinden van overeenkomsten tussen gegevens die tot op heden onbekend zijn. Gegevens blijken bepaalde overeenkomstige kenmerken te hebben die op voorhand nog niet waren vastgesteld. Het verschil met een categorieanalyse is dat in een clusteranalyse herkenning van toevallige overeenkomsten leidend is, terwijl bij een categorieanalyse deze overeenkomsten vooraf gedefinieerd zijn en de analist gericht op zoek gaat naar deze gegevens.

Een associatie analyse is een analyse waarbij gekeken wordt naar correlatie, m.a.w. er wordt gezocht naar koppelingen tussen gegevens. De outlier analyse dient gegevens te vinden die niet binnen gestelde normen vallen, het gegeven met kenmerk Z wijkt af van de grote groep die allemaal kenmerk X hebben. Of het gegeven dat continu kenmerk X heeft en plotseling kenmerk Z krijgt. Als voorbeeld geeft Brookshear (2009) het afwijken van iemands bestedingspatroon met de creditcard bij bijvoorbeeld fraude of diefstal. Brookshear (2009) geeft aan dat deze methode vooral binnen het onderzoek naar terrorisme toepasbaar is. De laatste vorm is de sequentiële patroon analyse. Het is het herkennen van gedragspatronen binnen databronnen gemeten over een langere tijd.

Er zijn verschillende algoritmen die kunnen voorzien in deze analyses. De top 10 meest gebruikte algoritmen worden genoemd. Op basis van kwantitatief onderzoek en interviews onder respondenten van bezoekers van een Datamining congres proberen Wu et. al. (2008) in kaart te brengen wat de top tien meest gebruikte algoritmen zijn; C4.5, *k*-Means, SVM, Apriori, EM, PageRank, AdaBoost, *k*NN, Naive Bayes, and CART. C4.5 is een algoritme wat gebruikt wordt voor classificatie en heeft veel verwantschap met CART. SVM staat voor ‘Support Vector Machines’, wat eveneens een manier is om data te classificeren. Wu et. al. (2008) noemen deze een van de meest accurate algoritmen. *k*-Means is een algoritme om een clusteranalyse mee uit te voeren, data wordt geclusterd en het wordt duidelijk welke data buiten de norm vallen. EM staat voor Expectation Maximisation en is eveneens een clusteranalyse. PageRank is een methode die wordt gebruikt door zoekmachines op het internet, kenmerken worden aan bepaalde websites gekoppeld om op basis hiervan belang te bepalen voor de zoekopdracht van de gebruiker. Het is daarom een classificatie analyse. AdaBoost is een ensemble analyse, en een van de weinige die beschikbaar is. *k*NN staat voor *k* Nearest Neighbor en zoekt naar een aantal objecten in de data set dat het meest dicht bij het test object (*k*) staat. Opnieuw een manier op data te classificeren alleen op basis van grote. Naive Bayes, ook wel stupid Bayes genoemd, is een makkelijke methode om data in te delen in categorieën welke door de gebruiker worden bepaald. Het algoritme categoriseert de data aan hand van de vooraf vastgelegde eigenschappen. Apriori is de meest bekende en veel gebruikte associatie analyse van Agrawal en Srikant (1994). Dit algoritme kan worden gebruikt om te bepalen in hoeverre consument die product x kopen ook product y kopen. In bijlage één staat een voorbeeld van een dergelijke analyse.

Bijlage 2. Voorbeeld Apriori analyse

Voorbeeld

Er zijn 5 verschillende artikelen te koop in een Supermarkt;
{1} Water
{2} Bruisend Water
{3} Cola
{4} Brood

Er zijn 4 transacties (transactie T, totaal transacties is volgens Apriori D) geweest, die de volgende producten kochten;

1. {1, 2}
2. {1, 3}
3. {1, 2, 4}
4. {1, 2, 4}

Er wordt geturft hoevaak ieder artikel afzonderlijk is gekocht;

{1} Water = 4 keer
{2} Bruisend Water = 3 keer
{3} Cola = 1 keer
{4} Brood = 2 keer

Degene die wil vaststellen of er een associatie te vinden is moet een frequentie bepalen, hij bepaalt hoe vaak een koop nodig is om als “vaak” te moeten worden geclassificeerd. In het voorbeeld gebruik ik dat wanneer een artikel 3 keer of meer is gekocht het vaak is gekocht. Zowel Water (1) als Bruisend Water (2) worden vaak gekocht. Dit is de eerste stap in filtering van alle data, dit zijn de geschikte kandidaten.

Vervolgens worden er paren gemaakt van de geschikte kandidaten, hoe vaak komen zij samen voor?

{1, 2} = 3

Wat en Bruisend Water werden 3 keer gekocht in combinatie met elkaar. De formule houdt hier op omdat er geen paren van 3 gemaakt kunnen worden omdat er maar twee kandidaten waren en nu slechts nog maar één. Dit eerste deel van de formule stelt de analist in staat om automatisch patronen te herkennen. Dat Water en Bruisend Water drie keer gekocht werden met elkaar was voor hem wellicht nog niet duidelijk.

De laatste stap is het formuleren van assiocatieregels. Dit is het vaststellen van een percentage wat gekenmerkt kan worden als een belangrijke overeenkomst. In mijn voorbeeld hanteer ik 50%. De formule ziet er alsvolgt uit;

{x} => {y}; Wanneer product X voorkomt hoe vaak komt daarvan product Y dan voor
{y} => {x}; Wanneer product Y voorkomt hoe vaak komt daarvan product X dan voor

{Water: 4 keer} => {Bruisend Water: 3 keer}; 3 delen door 4 = 75%

{Bruisend Water: 3 keer} => {Water: 3 keer}; 3 delen door 3 = 100%

De gevolgen van deze associatie analyse is dat de marketeer of de supermarkt kan zeggen dat wanneer iemand Water koopt deze dan ook vaak Bruisend Water koopt. En, dat wanneer iemand Bruisend Water koopt deze dan ook vaak Water koopt. De gevolgen zijn dan dat de supermarkten kunnen inspelen op de vraag van consumenten en speciale combinatie aanbiedingen kunnen toepassen.

Bijlage 3 Nieuwsberichten

Bonuskaart

Security.nl. “AH deelt gegevens bonuskaart met Justitie”. [2008] Security.nl. Geraadpleegd op 17 april 2011 via http://www.security.nl/artikel/23260/AH_deelt_gegevens_bonuskaart_met_justitie.html

> AH deelt gegevens bonuskaart met justitie
>
> Op klantenkaarten zoals de Albert Heijn bonuskaart staat dat de verzamelde gegevens niet ter beschikking aan derden worden gesteld, toch moeten grote winkelketens zeker tientallen keren per jaar klantgegevens delen met justitie, zo laat BNR weten. Hoe vaak er precies om de gegevens wordt gevraagd wil het OM niet vertellen. "Het kan belangrijke informatie bevatten over wat iemand precies heeft gedaan op welk tijdstip of wat ze precies hebben gekocht. Dat kan belangrijk zijn in een strafrechtelijk onderzoek," zegt Officier van Justitie Monique Korte.
>
> "Het is op zich geen nieuw feit. Het is wettelijk toegestaan om gegevens op te vragen," aldus Naima Azough van GroenLinks. Volgens de politica maken de supermarkten niet goed duidelijk aan hun klanten dat justitie de gegevens ook kan bekijken. Verder zou politie en justitie vertrouwelijk met deze gegevens om moeten gaan. "Daar heb ik eerlijk gezegd mijn twijfels over." Azough vindt dat als mensen niet langer verdacht zijn, de gegevens vernietigd moeten worden. Bij het OM zou men echter laks omgaan met gegevens. "Ik ben bang dat dit soort gegevens, boodschappenlijstjes en locaties waar je bent geweest, gewoon rondslingeren of in ieder geval niet vernietigd worden zoals zou moeten gebeuren."

Tweakers.net. “AH werkt aan RFID Bonuskaart”. [2006] Tweakers.net. Geraadpleegd op 17 april 2011 via http://tweakers.net/nieuws/43330/ah-werkt-aan-rfid-bonuskaart-update.html

> AH werkt aan RFID-Bonuskaart' - update
>
> Volgens vakblad Elsevier Retail werkt Albert Heijn aan een nieuwe versie van zijn Bonuskaart waarin een RFID-chip is verwerkt. Dergelijke chipjes kunnen door speciale apparatuur op een afstand van enkele meters worden 'gepinged' waarop ze een unieke identificatiecode uitzenden. Hierdoor is het niet langer nodig om de kaart langs een barcodelezer te halen. Tevens stelt de technologie de supermarktketen in staat om klanten onmiddelijk te herkennen als ze de winkel binnenlopen, waarmee 'Minority Report'-achtige scenario's mogelijk worden: de klant zou op basis van zijn koopgeschiedenis via beeldschermen op speciale aanbiedingen getrakteerd kunnen worden. Het is niet bekend of de supermarktketen dit soort mogelijkheden van RFID-technologie wil gaan benutten. Wel zou Albert Heijn van plan zijn om met behulp van de nieuwe kaarten klanten in staat te stellen om op afbetaling boodschappen te doen. Als de artikelen in de supermarkt ook gechipped worden, is het zelfs denkbaar dat de klant via een speciale kassa, die alle boodschappen in een keer uitleest, direct de tent uit kan lopen.
>
> Update 21:00 - Webwereld meldt dat Albert Heijn op de berichtgeving van Elsevier heeft gereageerd door te stellen dat er geen concrete plannen zijn om de volgende generatie Bonuskaarten van RFID-chips te voorzien. Pas als het bedrijf zelf duidelijkheid heeft welke kant het met de Bonuskaart op wil, zal het met meer informatie komen.

Webwereld.nl. “Albert Heijn ontkent plannen RFID chip in Bonuskaart”. [2006] Webwereld.nl. Geraadpleegd op 17 april 2011 via http://webwereld.nl/nieuws/41876/albert-heijn-ontkent-plannen-rfid-chip-in-bonuskaart.html

> Albert Heijn ontkent plannen RFID chip in Bonuskaart
>
> Een woordvoerder van Albert Heijn ontkent dat de supermarktketen bezig is met het ontwikkelen van een nieuwe bonuskaart met rfid-chip.

Elsevier Retail meldde woensdag dat Albert Heijn van plan is de huidige bonuskaart af te schaffen. Een nieuwe versie zou een intelligente chip bevatten waarmee klanten op afbetaling kunnen winkelen. Het vakblad haalde het voorbeeld aan van een Engelstalige supermarkt die sinds dit voorjaar een klantenkaart met rfid-chip gebruikt.

Diverse media namen het bericht over waarin gesuggereerd werd dat Albert Heijn de rfid-technologie ook gaat gebruiken. Albert Heijn-woordvoerster Els van Dijk ontkent tegenover Webwereld echter dat er plannen zijn voor het gebruik van rfid-chips. "Op dit moment zijn er geen concrete plannen om iets te doen met rfid", aldus Van Dijk. De zegsvrouw voegde toe dat het bedrijf plannen voor de bonuskaart pas naar buiten brengt, als het duidelijk weet wat het met de klantenpas wil.

Volgens Van Dijk is het gebruik van rfid echter nog toekomstmuziek en zal de draadloze chip eerder voor andere doeleinden worden ingezet. "Ahold en Albert Heijn zijn nog in een heel pril stadium aan het nadenken over hoe we ooit in de logistiek gebruik kunnen maken van rfid", aldus Van Dijk.

HKS

Webwereld.nl. “Politie test datamining criminelendatabank”. [2009a] Webwereld.nl. Geraadpleegd op 17 april 2011 via http://webwereld.nl/nieuws/64515/politie-test-datamining-criminelendatabank.html

Politie test datamining criminelendatabank

Een Leidse onderzoeker testte datamining op de database van 1 miljoen veroordeelde of verdachte Nederlanders. Met de onthullende resultaten mag de politie officieel niets doen.

De analyse van de gigantische database legt talloze significante en onbehaaglijke verbanden bloot tussen bepaalde bevolkingsgroepen en crimineel gedrag. Dit blijkt uit het proefschrift Algorithmic Tools for Data-Oriented Law Enforcement waarop Tim Cocx vorige week is gepromoveerd.

Het onderzoek is het project DALE (Data Assistance for Law Enforcement) en betaald door NWO. Informaticus Cocx liet algoritmes los op een database met alle Nederlanders die sinds 1998 zijn veroordeeld voor een misdrijf, of in de afgelopen zes maanden zijn verdacht van een misdrijf. Dat zijn gegevens van één miljoen landgenoten met hun demografische details en criminele verleden.

Misdaad en persoonskenmerken

"M'n promotor Joost Kok vroeg me dit onderzoek te gaan doen. Onze vakgroep werkte al nauw samen met de politie. De Dienst Nationale Recherche Informatie had behoefte aan verkennend onderzoek naar de mogelijkheden van datamining."

Dus toog Cocx aan de slag aan de Universiteit Leiden, waar ooit professor Wouter Buikhuisen werd weggestuurd na een openbare hetze wegens criminologisch onderzoek naar misdaad en persoonskenmerken. Buikhuisen is juist afgelopen maanden gerehabiliteerd en geeft een gastcollege in 2010.

Cocx vond onder meer de volgende sterke correlaties, sommige voorspelbaar, andere verrassend:

* De vrouwen in de database zijn significant vaker verslaafd aan drugs dan mannen.
* Mensen verdacht van doodslag zijn relatief vaak al veroordeeld wegens racisme.
* Joyriders nemen het ook niet zo nauw met de arbeidswetten en alcohol.
* Diefstal met geweld hangt vaak samen met wapenbezit.
* Afrikaanse afkomst en veroordeeld voor overtredingen inzake 'openbare veiligheid'. * Criminelen op het platteland begaan ook vaak verkeersmisdrijven.

Politie mag er niks mee

"Ik vond dat opmerkelijk", zegt Cocx. Wat kan de politie ermee? "Ik kan me voorstellen dat in sommige situaties er alarmbelletjes gaan rinkelen bij de politie. Je kunt er ook beleid op gaan baseren. Maar voorlopig wordt er niets mee gedaan omdat het niet toegestaan is. Je mag data wettelijk alleen gebruiken voor het doel waarvoor ze zijn verzameld. Dit heeft geen prioriteit."

Cocx zou er, op grond van zijn resultaten, ook voorzichtig mee omspringen: "In marketing is een succespercentage van 85 procent hoog en met succes toe te passen in campagnes. Maar voor opsporing is een foutenpercentage van 15 procent tegenover mensen die je onterecht verdenkt niet aanvaardbaar."

Minority Report

Privacy is en blijft een zorg. Het staat de toepassing van datamining dan ook in de weg. Cocx zelf heeft er niet zo veel mee: "Persoonlijk kan het me niet zo veel schelen. Maar op mijn vakgebied worden veel zorgen geuit over privacy met verwijzing naar Minority Report-achtige toestanden."

"Ik denk dat je van geval tot geval moet bekijken of nieuwe methoden in de opsporing met behulp van informatica toegepast kunnen worden zonder de privacy te schenden."

Accurate voorspellingen

Over Minority Report gesproken, het geautomatiseerd grasduinen in de databank levert behoorlijk accurate voorspellingen op, blijkt uit het hoofdstuk 'An Early Warning System for the Prediction of Criminal Career'.

"Een cluster van 10.000 criminele carrières uit de database kan als solide basis dienen voor extrapolatie in de tijd", concludeert Cocx. Zo kan een voorspellingsbetrouwbaarheid van maar liefst 88 procent worden bereikt. En dat percentage kan nog hoger, denkt de onderzoeker.

Webwereld.nl. “Politie gaat datamining zeker gebruiken”. [2009b] Webwereld.nl. Geraadpleegd op 17 april 2011 via http://webwereld.nl/nieuws/64531/-politie-gaat-datamining-zeker-gebruiken-.html

Politie gaat datamining zeker gebruiken

Het dataminen van de criminelendatabase door een Leidse onderzoeker is een voorbode van toekomstig gebruik. "Informatie kan en zal worden gebruikt."

Privacywaakhond Bits of Freedom (BOF) reageert op de promotie van een Leidse onderzoeker die datamining heeft gedaan op de database van 1 miljoen veroordeelde of verdachte Nederlanders. Officieel mag de politie niets doen met de onthullende resultaten, stelt onderzoeker Tim Cocx zelf.

Gebruik mag wel

Jurist Arnoud Engelfriet bestrijdt dat echter. Hij denkt dat de Leidse onderzoeker zich ten onrechte baseert op de Wet Bescherming Persoonsgegevens. Volgens Engelfriet is hier de Wet Politiegegevens van toepassing. Die staat wel degelijk gebruik en combinatie toe van persoonsgegevens waar de politie mee werkt. "Die verbanden mogen dan worden gebruikt voor de dagelijkse politietaak", blogt Engelfriet.

BOF-directeur Ot van Daalen ziet het somber in. "De zorg over gebruik van datamining als opsporingsinstrument is geheel terecht. Het risico op vergaande controle door de overheid en onjuiste profilering is groot. Datamining zou dus slechts in zeer uitzonderlijke omstandigheden toegestaan mogen zijn, en dan pas als het met voldoende waarborgen omkleed is."

Voorbode

"Dat dit nu in een wetenschappelijk kader gebeurd, is een voorbode van de toekomst", vertelt Van Daalen aan Webwereld. Hij vreest dat dergelijk gebruik van datamining door opsporingsinstanties gelegaliseerd, en dan gewoon gebruikt zal worden.

Engelfriet merkt op dat gebruik weliswaar mag, maar hij denkt niet dat het veel oplevert. "Wel vraag ik me af wat de politie kàn doen met deze gegevens. Want het is vast wel interessant als borreltafelfeitje dat 'joyriders het ook niet zo nauw nemen met de arbeidswetten en alcohol' maar wat moet je daarmee als agent?"

Ook SP-Kamerlid Arda Gerkens ziet het niet meteen somber in. "Ik ben niet bang dat de politie hiermee iets doet. Vergeet niet dat er nu überhaupt al criminologisch onderzoek wordt gedaan, wereldwijd. Op zich is dit niets nieuws."

Incomplete groep

Daarmee stelt zij niet dat er niets aan de hand is. "Ik vind het wel jammer dat dit kennelijk in opdracht van de politie is gedaan." Cocx vertelde Webwereld al dat zijn vakgroep nauw samenwerkt met de politie en dat zijn promotor Joost Kok hem gevraagd heeft dit onderzoek te doen.

Gerkens weet niet of de datamining bewust vanuit de politie is gepleegd of dat het slechts een verkennende vingeroefening is. Op zich maakt dat niet veel uit. "Beide zijn stappen in een richting waar de politie best wel heen wil. Maar die kant wil de politiek niet op."

"Vanuit de politiek vinden we dat dataming geen goed middel is. Het gaat hier immers om een incomplete groep, dat kun je niet extrapoleren naar de hele Nederlandse bevolking. Pas dus op voor conclusies", waarschuwt zij. "We zijn zeker niet tegen wetenschappelijk onderzoek, maar dit komt vanuit de politie." Dat is te dicht tegen de praktijk aan, erkent het SP-Kamerlid.

Verzamelwoede

Privacywaakhond BOF haakt aan: "Wij hebben altijd al gewezen op de informatie van burgers die wordt verzameld." Die verzamelwoede is er niet alleen bij bedrijven. Ook overheden vergaren veel data en maken daarbij eveneens gebruik van bedrijfsbestanden met data.

Van Daalen ziet het echter nog niet als een gepasseerd station. "Gebruik van dergelijke opsporingsmiddelen zal wel op grond van een wet zijn. De burger kan dus nog invloed hebben." Daarnaast houdt BOF ook een oogje in het zeil. "Wij zullen dit kritisch volgen", belooft Van Daalen.

Gradaties en waarborgen

Overigens is BOF niet 100 procent tégen gebruik van datamining. "Er zijn gradaties, ja." Van Daalen acht het voor ernstige misdrijven, zoals bijvoorbeeld pedofilie, wel geoorloofd. "Ik snap het sentiment daaromtrent wel. Maar zorg voor waarborgen. Kijk goed naar de groepen waar je datamining op pleegt."

De BOF-directeur wijst op de grote verschillen tussen veroordeelde, verdachte en onverdachte burgers. Een geuite verdenking of concrete beschuldiging kan zó beklijven. Zeker als 'het systeem', dus datamining van bepaalde informatie, dat lijkt te onderbouwen, waarschuwt hij.

Tweakers.net. “Facebook onthult geaardheid gebruikers via advertenties”. [2010] Tweakers.net. Geraadpleegd op 17 april 2011 via http://tweakers.net/nieuws/70393/facebook-onthult-geaardheid-gebruikers-via-advertenties.html

Facebook onthult geaardheid gebruikers via advertenties

Facebook onthult informatie over de geaardheid van gebruikers aan adverteerders door advertenties speciaal voor hen te tonen. Dat zou blijken uit een steekproef van onderzoekers. Facebook ontkent de aantijging.

Tijdens het experiment werd een aantal accounts aangemaakt met dezelfde profielinformatie, waarbij alleen de geaardheid verschilde. Het bleek dat homoseksuele Facebook-gebruikers andere advertenties te zien krijgen, staat in de paper, die is geschreven door een onderzoeker van Microsoft Research India en een Duitse onderzoeker van het Max Planck-instituut. Omdat de advertenties niet duidelijk gericht waren op homoseksuelen, zou klikken op de advertentie duidelijk maken aan de adverteerder dat de desbetreffende Facebook-gebruiker in zijn profiel heeft ingevuld dat hij homoseksueel is.

Facebook toont gerichte advertenties op basis van profielinformatie die niet publiekelijk zichtbaar is. De netwerksite ontkent tegenover ITWorld echter de aantijging door te stellen dat Facebook geen informatie aan adverteerders geeft, die hen identificeerbaar maakt. Bovendien zou de informatie niet gelinkt worden aan acties die uitgevoerd worden op de website van de adverteerder: wanneer een homoseksuele Facebook-gebruiker via een advertentie op een site komt, waar hij wat bestelt, zou de adverteerder niet registreren dat hij van een advertentie kwam die aan homoseksuele Facebook-gebruikers wordt getoond.

Volgens de onderzoekers is er sprake van een privacy-probleem, omdat de advertenties niet duidelijk gericht zijn op een bepaalde doelgroep. Het is onduidelijk of er ook met andere gevoelige profielinformatie wordt gewerkt om gerichte advertenties te tonen en wat er met die gegevens gebeurt. De grootte van de steekproef van de onderzoekers laat echter ruimte voor twijfel: er is gewerkt met slechts zes Facebook-profielen, waarvan een werd ingevuld als homoseksuele man.

Tweakers.net. “EC privacybewakers moeten sterker staan tegen Facebook”. [2011] Tweakers.net. Geraadpleegd op 17 april 2011 via http://tweakers.net/nieuws/73276/ec-privacybewakers-eu-moeten-sterker-staan-tegen-facebook.html?mode=nested&max=10&niv=0&order=asc&page=1#reacties

EC privacybewakers moeten sterker staan tegen Facebook

Privacywaakhonden in de EU moeten meer macht hebben om buitenlandse sites die zich op Europese burgers richten te onderzoeken en er tegen op te treden als ze EU-regels overtreden. Eurocommissaris Reding geeft Facebook als voorbeeld.

Eurocommissaris Viviane Reding van Justitie noemde in een toespraak voor het Privacy Platform vier 'pilaren' voor het vaststellen van privacyregels. De herziening van de Europese regels ter bescherming van data van burgers is een belangrijke doelstelling van Reding.

De vier pilaren zijn: 'het recht om vergeten te worden', transparantie, privacy by default en bescherming ongeacht de locatie van de data. Als voorbeeld noemt Reding 'een in de VS gevestigd social network-bedrijf met miljoenen actieve gebruikers in Europa'. Een dergelijk bedrijf, waarmee met name op Facebook wordt geduid, moet voldoen aan de EU-regels, volgens Reding.

Om die regels te handhaven krijgen de privacywaakhonden in de 27 lidstaten van de EU de macht om dergelijke databeheerders die zich op Europese burgers richten te onderzoeken en er juridisch tegen op te treden. Reding wil de onafhankelijkheid van de nationale autoriteiten die belast zijn met de bescherming van data versterken en hun bevoegdheden harmoniseren. In Nederland is het College Bescherming Persoonsgegevens hiervoor verantwoordelijk.

Reding vindt ook dat de verschillende waakhonden beter moeten samenwerken bij zaken die de grenzen van EU-landen overschrijden. Hierbij gaf ze als voorbeeld de privacyzorgen van 'online-plattegronddiensten met foto's van straten en de huizen van mensen'. Hiermee duidt ze op Google Streetview, dat in diverse Europese landen, waaronder Duitsland, op privacybezwaren stuitte. "Er is een meer gecoördineerde aanpak op EU-niveau nodig om dergelijke zaken op een consistente en effectieve manier aan te pakken", aldus Reding.