



Organisatieonderdeel Amsterdam  
Dienst Regional Informatie  
Bureau CIO

Contactpersoon 5.1.2.e  
Functie Dataminer  
Bezoekadres James Wattstraat 84  
1097 DM Amsterdam  
Telefoon 020 5.1.2.i  
E-mail 5.1.2.e@amsterdam.politie.nl  
Internet [www.politie.nl](http://www.politie.nl)

Datum 05 feburari 2013  
Status Concept  
Blad 1/16

## Criminaliteit Anticipatie Systeem (CAS)

Prospective Hotspots V2.0

Auteurs

5.1.2.e

**Datum** 05 feburari 2013  
**Status** Concept  
**Titel** Criminaliteit Anticipatie Systeem  
**Blad** 2/16



**Wijzigingsregister**

---

*Versie*    *Datum*            *Omschrijving*

Gezien

## Inhoud

	<b>Inleiding</b>	<b>4</b>
1.1	Achtergrond en aanleiding	4
1.2	Probleemschets en probleemstelling	5
1.2.1	<i>Prospective hotspots, huidige stand van zaken</i> <b>Fout! Bladwijzer niet gedefinieerd.</b>	
<b>2</b>	<b>Wensen</b>	<b>11</b>
2.1	Centrale wens	11
2.2	Overige wensen	11
<b>3</b>	<b>Aanpak</b>	<b>12</b>
3.1	Borging en automatisering	12
3.2	Verbeteringen aan analyse	12
3.3	Verbetering aan hanteerbaarheid	12
3.4	Verbreiding domein	13

## Inleiding

### 1.1 Achtergrond en aanleiding

In 2012 is door de afdeling Analyse en Onderzoek van bureau CIO in het kader van het onderzoek “Waar en wanneer het ertoe doet” een inzet-onafhankelijke pilot gedaan met de zogenaamde prospective hotspots-methode. Prospective hotspots zijn gebieden waarbij, gegeven bepaalde kenmerken van deze gebieden, de berekende kans op een crimineel incident hoger is dan normaal.

Deze kansen worden berekend met behulp van datamining-technieken. Datamining is een proces door middel waarvan praktisch bruikbare en interessante patronen kunnen worden ontdekt in grote hoeveelheden data. Voordeel van deze manier van het detecteren van hotspots is dat vrijwel alle informatie die beschikbaar is over een locatie betrokken kan worden bij de berekening. De datamining algoritmen bepalen zelf op basis van toegevoegde waarde of welke informatie wordt gebruikt in het uiteindelijke model.

De resulterende “vlakkenkaarten” kunnen vervolgens worden gebruikt om inzet te plannen. De inzetonafhankelijke pilot had goede resultaten, in een aantal districten (D1, D2, D4 en D5) is al capaciteit ingezet op deze manier. Desalniettemin zijn, niet in de laatste plaats dankzij de inzet in de praktijk, een aantal verbeterpunten genoemd.

De prospective hotspot-methode gemaakt ter illustratie van een onderzoek, maar krijgt steeds meer belangstelling. Om kwaliteit en efficiëntie te verbeteren is het belangrijk om een formeel project te benoemen om tot een generiek proces te komen. In het huidige document zal de prospective hotspot methode kort worden omschreven, en zullen de voorgenomen verbeterpunten worden genoemd. Aangezien is gebleken dat de naam ‘Prospective Hotspots’ door sommigen als een tongbreker wordt ervaren, krijgt het verbeterde proces een andere naam: het Criminaliteit Anticipatie Systeem (CAS).

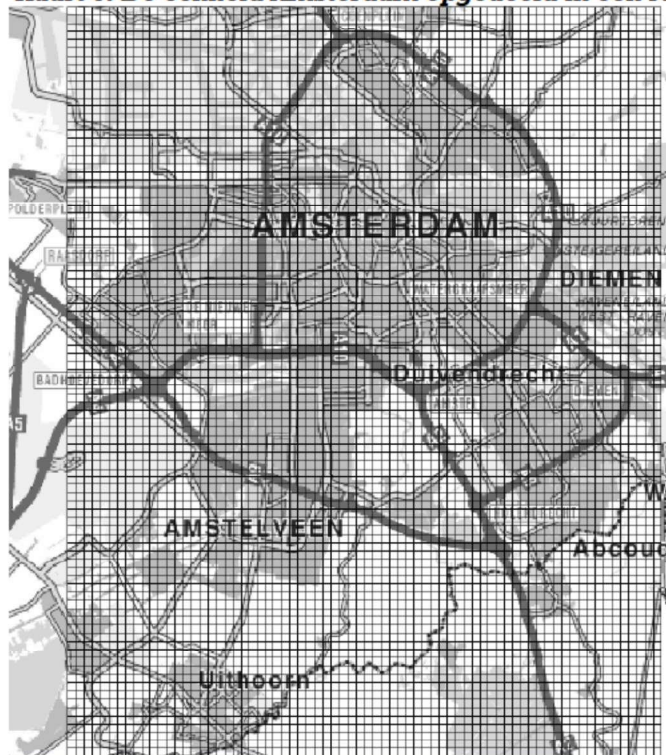
## 1.2 Probleemschets en probleemstelling

Bij de prospective-hotspotmethode is het uitdrukkelijk de bedoeling om een zo goed mogelijke voorspelling te maken van de locaties van incidenten, nog voordat deze plaatsvinden. Dit wordt gedaan door enerzijds meer informatie te gebruiken, en anderzijds door slimme rekenmethodes toe te passen. Individuele incidenten zijn niet te voorspellen, we kunnen immers niet in het hoofd van de dader kijken. Maar het is wel te voorspellen dat in bepaalde gebieden woninginbraken vaker voorkomen dan in andere.

De oorzaken hiervoor kunnen samenhangen met bouwtechnische aspecten van de wijk (bijvoorbeeld oudere huizen waar makkelijk binnen te komen is), of daderkenmerken (bijvoorbeeld vertrouwdheid met de omgeving omdat de dader in de buurt woont). Daarmee wordt het de opgave om kenmerken te vinden die gebieden met veel inbraken kunnen onderscheiden van gebieden met weinig inbraken.

Er zijn veel manieren om 'gebieden' te bepalen. In het politiesysteem wordt onderscheid gemaakt tussen districten, wijken en buurten. Daarnaast is er ook nog de gemeentelijke indeling en kunnen postcodegebieden worden gebruikt. Deze zijn echter minder geschikt voor de voorspellingstaak, omdat de gebieden vaak grillig van vorm zijn. Dit heeft als gevolg dat het moeilijker is om gebieds-kenmerken te koppelen; hoe simpeler de vorm, hoe eenvoudiger de koppelingstaak. Daarom werkt de prospective-hotspotmethode met vierkante gebieden. De kaart van de eenheid Amsterdam wordt dus opgedeeld in een raster.

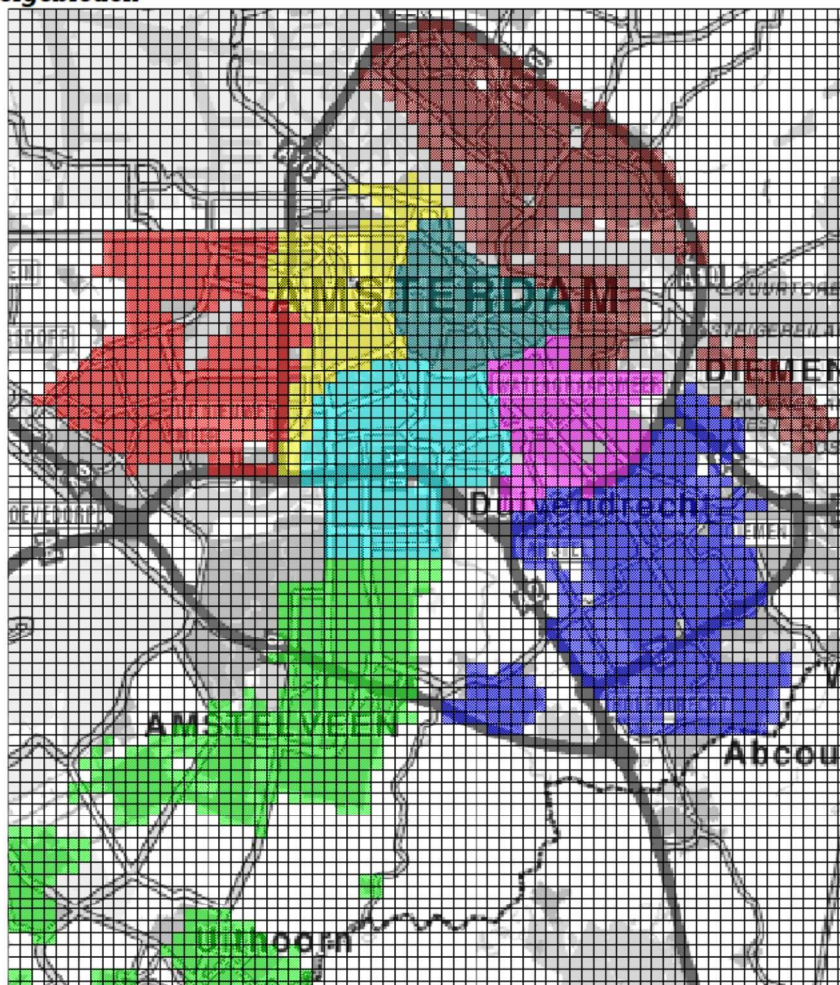
***Kaart 3: De eenheid Amsterdam opgedeeld in een raster***



Het gebruikte raster is opgebouwd uit vierkante gebieden van 250 bij 250 meter. Het is direct al duidelijk dat een groot gedeelte van de vierkante gebieden

eigenlijk niet interessant zijn voor deze vraag, omdat het a priori onwaarschijnlijk is dat er incidenten worden gepleegd en daarom niet interessant zijn voor de inzet van surveillances. Voorbeelden zijn open water, weilanden, parken enz. Om de classificatietask zo effectief mogelijk te maken zijn deze gebieden geëxcludeerd van de analyse.

***Kaart 4: Illustratie van de verdeling van de eenheid Amsterdam in voorspelgebieden<sup>1</sup>***



Bij ieder van de overblijvende vakjes is het mogelijk om een groot aantal kenmerken te bepalen. Enerzijds statische kenmerken van het vakje: kenmerken die niet sterk zullen variëren met het verstrijken van de tijd. Voorbeelden zijn socio-economische status en demografische kenmerken volgens het CBS, het aantal winkels, cafés, coffeeshops enz. binnen het vakje. Anderzijds is het ook mogelijk om dynamische kenmerken van het vakje te koppelen. Dynamische kenmerken zijn kenmerken die juist sterk zullen variëren met het verstrijken van de tijd. Voorbeelden zijn het aantal inbraken in de afgelopen maand, tijd sinds laatste inbraak, tijd van het jaar, tijdstip van de dag enz.

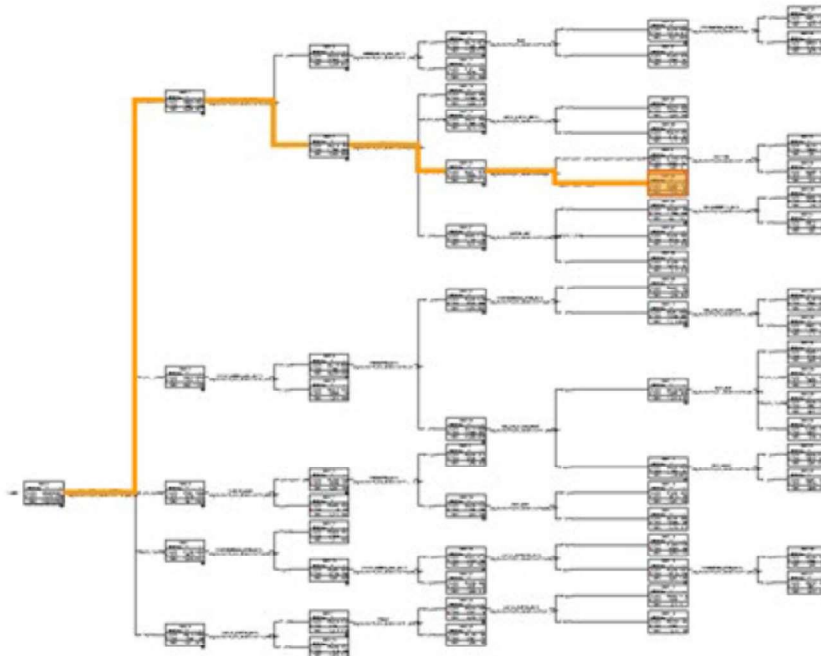
<sup>1</sup> De niet-gekleurde vakjes doen niet mee in de analyse.

Van ieder vakje is de inbraakenhistorie gedurende vier jaar bijgehouden. Iedere maand is geregistreerd of er binnen het vakje een inbraak heeft plaatsgevonden, en wat de statische en dynamische kenmerken aan dat vakje waren in de voorgaande maand. Op deze manier is er een set gegevens verkregen die mogelijkerewijs voorspellende waarde hebben voor het plaatsvinden van een inbraak.

Om de daadwerkelijke voorspelling te doen zijn datamining algoritmes toegepast die bedoeld zijn om classificaties te maken. Per district zijn, omdat verondersteld werd dat relevante factoren niet overal gelijk zijn, aparte modellen gemaakt. Voor District 1 Centrum en District 2 Noord is één model per district gemaakt. Voor District 3 Oost, District 4 Zuid en District 5 West zijn per district twee modellen gemaakt: het gedeelte dat het dichtst bij het centrum ligt is gescheiden van het perifere gedeelte. In kaart 4 is te zien wat de resulterende gebieden zijn. Voor de modelbouw zelf is een groot scala aan classificatiealgoritmes getest, maar uiteindelijk is gefocust op een tweetal: CHAID en een neurale netwerk.

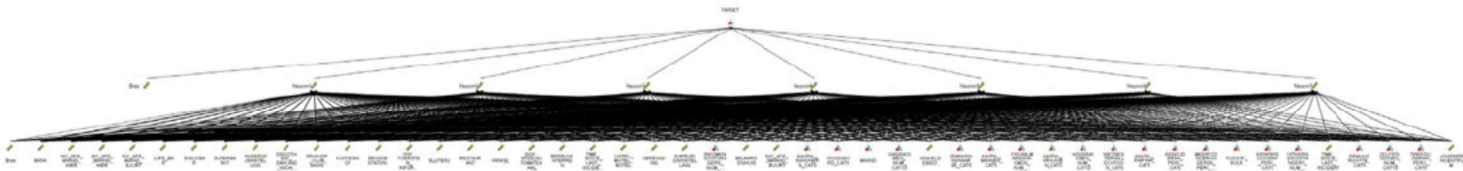
Kort gezegd is het voordeel van CHAID dat het een model oplevert dat een grove voorspelling maakt, maar wel vrij makkelijk uit te leggen is. Het heeft namelijk de structuur van een zogenaamde decision tree, waarbij het visueel inzichtelijk gemaakt kan worden welke kenmerken met een verhoogd risico gepaard gaan. Zo kan van ieder vakje worden beschreven *waarom* het wel of niet wordt geclassificeerd als een vakje waar inbraken zullen worden gepleegd (zie figuur 5 voor een voorbeeld). Uit ervaring blijkt dat het bij de introductie van dit soort methodes verstandig is om een dergelijke uitleg te kunnen geven. In onderstaand voorbeeld van een decision tree worden voorspellende kenmerken voor woninginbraken afgebeeld. De oranje aftakking geeft een kans op woninginbraken weer van 29,3%. Deze aftakking heeft de volgende kenmerken: 1) laatste inbraak minder dan een maand geleden; 2) in de omringende vakjes vonden meer dan vier inbraken plaats; 3) aantal inwoners in postcodegebied tussen 40 en 55 personen; 4) het is maart, oktober of november.

**Figuur 5: Voorbeeld van een decision tree**



Neurale netwerken leveren modellen die een precieze voorspelling kunnen maken, maar de werking van het model is moeilijk te verwoorden. Het is dus slecht mogelijk om uit te leggen welke factoren een rol spelen in de voorspelling. In een neuraal netwerk worden de verbanden tussen de voorspellende variabelen en de te voorspellen delicten weergegeven als verbindingsterkten tussen informatieknooppunten. Deze architectuur is gebaseerd op die van het menselijk brein, waar hersencellen informatie aan elkaar doorgeven via elektrische stroompjes. Vaak wordt de term 'black box' gebruikt om dit soort modellen te beschrijven: er worden kenmerken ingevoerd, en er komt een voorspelling uit, maar de gemaakte combinaties van informatie zijn dusdanig complex dat het volledig onduidelijk is wat er nu binnen het model zelf gebeurt. Zoals te zien in figuur 6, biedt dit weinig aanknoppen voor inhoudelijke interpretatie.

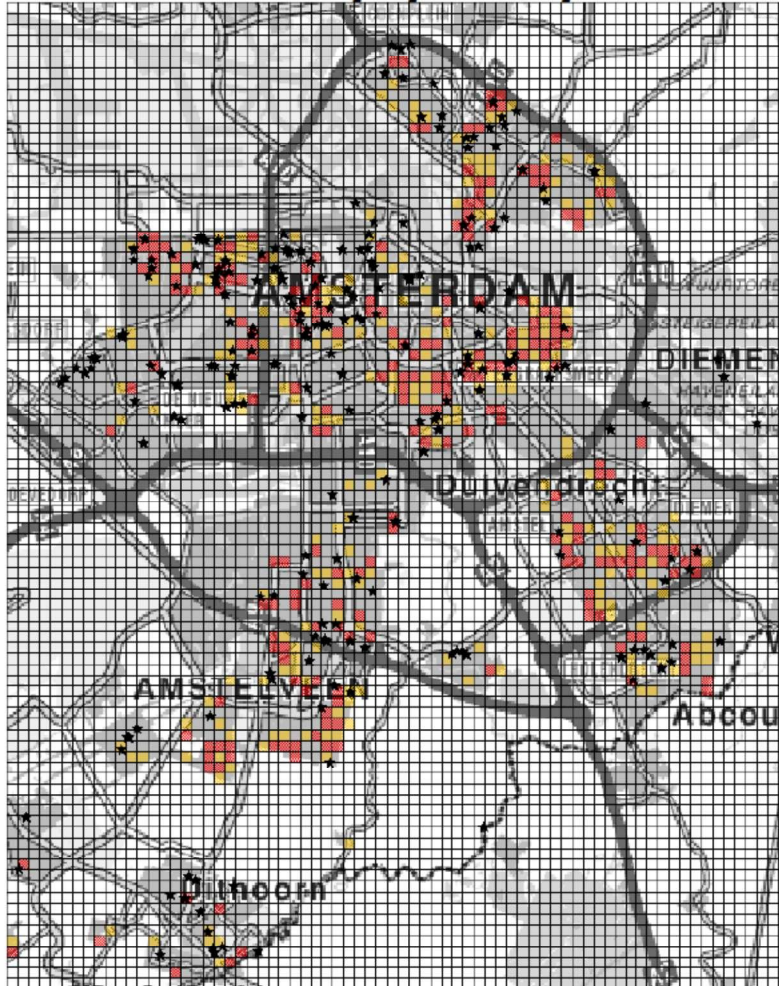
**Figuur 6: Voorbeeld van een neuraal netwerk**



In de praktijk kunnen beide modelsoorten worden gebruikt. De voorspelling van het neurale netwerk wordt gebruikt om het incidentenrisico te bepalen, en het CHAID-model wordt gebruikt om inzicht te krijgen in de factoren die samenhangen met het inbraakrisico. In de beschrijving van resultaten van deze pilot zullen echter alleen de resultaten van het neurale netwerk worden beschreven. Voor het beschreven doel zou het te ver reiken om een verklaring te geven van de gebieden met hogere risico's.

In het algemeen is het resultaat van een classificatiemodel een *risicoscore*. Dit is een getal dat meestal tussen 0 en 1 ligt. Als de risicoscore hoog genoeg is, dwz als deze boven een bepaalde drempelwaarde uitkomt, wordt een vakje geclassificeerd als een vakje met verhoogde kans op een incident en kan deze gekleurd worden weergegeven op de kaart. Er zijn veel manieren om deze drempelwaarden te bepalen. In kaart 5 is gekozen om per voorspelgebied de vakjes met de 10% hoogste risicoscores als rood weer te geven, en de 10% daarna als oranje.

**Kaart 5: Voorbeeld van een prospective-hotspotkaart**



Voorgaande kaart geeft de voorspelling weer voor woninginbraken in november 2012, in de middag (dit zijn de gekleurde vakjes). Deze voorspelling kan aan het begin van de maand worden gemaakt. Verder staan in de kaart ook de daadwerkelijke woninginbraken geplot (dit zijn de sterretjes). Te zien is dat ongeveer 35% van de inbraken plaatsvinden in de hoog-risico vakjes. Daarnaast vindt een zeer groot gedeelte van de overige inbraken plaats in naburige vakjes. Dit suggereert dat als ervoor gezorgd wordt dat in de gekleurde locaties politie aanwezig is, 1) de kans op een heterdaad bij een surveillance groter is en 2) dat, als een getuige melding doet van een inbraak, de politie sneller ter plaatse zal zijn.

De prospective-hotspotmethode kent de volgende voordelen ten opzichte van de klassieke hotspot methode:

- Duidelijk afgebakende risicogebieden.
- Mogelijkheid om meer informatie te gebruiken dan alleen delicten in het verleden. Ook bestaat de mogelijkheid om in de toekomst extra informatie toe te voegen als deze beschikbaar komt (woningkenmerken via Kadaster, buurtkenmerken via gemeente, enz.).
- Alleen relevante informatie wordt gebruikt voor de voorspelling. Het modelbouw algoritme verwijdert irrelevante informatie automatisch.

## **2 Wensen**

### **2.1 Centrale wens**

Het belangrijkste onderdeel van dit project is om CAS via DataDetective beschikbaar te stellen aan gebruikers, en om ervoor te zorgen dat de gebruikers regelmatig over ververste output beschikken.

### **2.2 Overige wensen**

Naast de centrale wens zijn er ook een aantal wensen uitgesproken ter verbetering van de huidige prospective hotspots:

- Verbetering van de analyse. Er zijn een aantal manieren geopperd waarop de voorspelling verbeterd zou kunnen worden.
- Verbetering van de hanteerbaarheid van de voorspellingen. Wat betreft het praktisch gebruik van het model zijn er ook een aantal verbeterpunten genoemd.
- Verbreding van het voorspellingsdomein. De wens is om meer soorten misdaad te kunnen voorspellen dan alleen woninginbraak.

### 3 Aanpak

#### 3.1 Borging en automatisering

De eerste activiteit van het project zal zich erop richten om CAS via DataDetective beschikbaar te maken aan de gebruikers. Daarnaast zullen een aantal Het zou zal te veel tijd kosten om eerst alle overige verbeteringen door te voeren en CAS daarna pas beschikbaar te maken. Het borgingsproces zal zodanig worden ingericht dat als versie 2.0 gereed is de overgang makkelijk moet kunnen plaatsvinden.

Op het moment worden de voorspelkaarten gemaakt door tabellen die met behulp van door IBM SPSS Modeler zijn gemaakt direct in te lezen in MapInfo. Aan het eind van het project zullen de Modeler-tabellen op een plek worden gezet waar DataDetective bij kan, zodat MapInfo via DataDetective de juiste informatie aangeboden krijgt.

Op deze manier worden de bewerkingen in MapInfo geautomatiseerd en kan de eindgebruiker direct aan de slag met de kaarten.

#### 3.2 Verbeteringen aan analyse

De volgende zaken zullen worden uitgevoerd om de analyse te verbeteren en zodoende de kwaliteit van de voorspellingen te verhogen:

- 

5.1.2.c, 5.1.2.d

- 

5.1.2.c, 5.1.2.d

#### 3.3 Verbetering aan hanteerbaarheid

Veel opmerkingen die naar aanleiding van de pilot in december zijn gemaakt hebben betrekking op het praktisch gebruik van de kaarten.

- Bij de huidige versie van prospective hotspots valt een gedeelte van de regio buiten het raster (te weten Kudelstaart en een gedeelte van Aalsmeer). Voor CAS zal het gebied dat door het raster wordt bestreken zo worden aangepast dat de hele regio gedekt wordt.
- Op het moment wordt een rastergrootte van 250 bij 250 meter genomen. Er is opgemerkt dat dit te groot is om in de praktijk op in te kunnen zetten. Het is echter zo dat als het raster te klein is genomen, geen betrouwbare voorspellingen meer gedaan kunnen worden. In dit project zal de huidige rastergrootte worden vergeleken met een grootte van 100 bij 100 en een grootte van 50 bij 50, om zo proefondervindelijk vast te kunnen stellen wat de beste rastergrootte is. Dit zal worden gedaan met behulp van lift- en gaincharts (dit zijn grafieken waarmee de kwaliteit van een datamining oplossing kan worden vastgesteld). Overigens kan dit pas plaatsvinden als de overige wijzigingen onder “Verbeteringen aan analyse” en “Verbetering aan hanteerbaarheid” al zijn uitgevoerd.
- Er worden bij de huidige versie van prospective hotspots 8 gebieden gemodelleerd, de wens is om één model per district gebied te maken ten behoeve van de reguliere planning, plus een model voor de hele regio ten behoeve van de planning voor de flexteams.
- In plaats van 4 dagdelen moeten 3 dagdelen worden gebruikt, die corresponderen met de diensttijden.
- De wens is uitgesproken om CAS wekelijks in plaats van maandelijks te draaien. In principe is dit mogelijk, maar eerst zal een studie moeten worden gedaan om te bepalen of er zo voldoende voorvallen beschikbaar zijn om een model mee te maken.
- Het is wenselijk om de kleuringsregels aan te passen. Op het moment wordt de kleur bepaald aan de hand van het deciel van de risicoscore van een vlakje, per dagdeel. Dit betekent dat er altijd evenveel vlakjes worden gekleurd (hoewel locatie en kleur wel kunnen verschillen). Het zou beter zijn om de kleuring plaats te laten vinden aan de hand van de absolute waarde van het vlakje, omdat zo het groter en kleiner worden van een hotspot kan worden gevisualiseerd.

### 3.4 Verbreiding domein

Een veelgestelde vraag is of het mogelijk is om andere soorten incidenten in kaart te brengen dan woninginbraak. Dit is in principe een herhalingsoefening, maar zal wel zijn weerslag hebben op het borgingsproces. In de eerste instantie zullen de HICs (straatroof, woninginbraak, overval en mishandeling) en de HVCs (inbraak auto, fietsendiefstal) worden voorspeld.

## 4 Planning

### 4.1 Deliverables

De volgende zaken zullen worden opgeleverd bij beëindiging van dit project:

- Een IBM-SPSS Modeler stream die de de benodigde MapInfo –files oplevert.
- Een proces waarmee DataDetective gebruikers direct over de kaarten kunnen beschikken.
- Een handleiding waarin beschreven wordt hoe de kaarten gebruikt moeten worden.

### 4.2 Planning

Oplevering zal plaatsvinden in twee fases. Omdat CAS al gedeeltelijk wordt ingezet is ervoor gekozen om eerst een versie op te leveren met de wijzigingen die de hoogste prioriteit hebben.

Versie 0.5 zal de volgende verbeteringen hebben:

1. Kaarten worden beschikbaar gemaakt via DataDetective;
2. Het raster zal worden aangepast en er zal worden bekeken of de rasterdichtheid naar 50 m<sup>2</sup> kan. De vergelijking met 100 m<sup>2</sup> zal in deze versie nog niet worden gemaakt.
3. Voor pleegtijden zal het midden tussen begintijd en eindtijd van een incident worden gebruikt.
4. De voorspelkaarten zullen in drie dagdelen worden aangeleverd die corresponderen met de dienstitijden. Dit in plaats van de huidige vier dagdelen.

Versie 1.0 zal alle verbeteringen hebben die in sectie 3 is omschreven.

De volgende capaciteit is ingeschat voor het project:

CAS versie	Onderdeel	Benodigd aantal dagen
CAS 0.5	Borging	2
	Van 8 naar 5 districtsmodellen + regiomodel	2
	Kwaliteitsverbetering pleegtijd	1
	Drie dagdelen ipv vier dagdelen	2
	Aanpassing raster	1
	Kwaliteitsvergelijking rastergrootte 250x250 versus 50x50 m	2
	Afronding en documentatie	2
	<i>Subtotaal CAS 0.5</i>	<i>12</i>
CAS 1.0	Borging	1
	Extra afgeleide variabelen	9
	Wekelijks laten draaien in plaats van maandelijks	2
	Kwaliteitsvergelijking wekelijks model vs maandelijks model	2
	Aanpassen kleuringsregels	2
	Verbreiding domein	5
	Afronding en documentatie	2
	<i>Subtotaal CAS 1.0</i>	<i>21</i>
<b>Totaal</b>		<b>35</b>

Als op 4 maart met het project gestart wordt, en het is mogelijk om twee dagen per week aan het project te werken, dan worden de volgende opleverdata geschat:

CAS 0.5: 9 april 2013

CAS 1.0: 1 juli 2013

#### 4.3 Belanghebbenden

Teams Informatie?  
TICC?

#### 4.4 Randvoorwaarden

Om op tijd te kunnen leveren is het nodig dat:

- Er gebruik gemaakt kan worden van datawarehouse BinK! ;
- SPSS Modeler beschikbaar is;
- Dat medewerkers van Sentient beschikbaar zijn om de borging via DataDetective te faciliteren;
- Dat er twee dagen per week aan CAS gewerkt kan worden.
- Dat benodigde informatie uit de operatie (hangplekken, vluchtroutes etc) tijdig beschikbaar is, dwz vóór 9 april 2013.

## 5 Projectorganisatie

### 5.1

Projectrol	Naam	Toelichting
Opdrachtgever	5.1.2.e	5.1.2.e
<b>Stuurgroepleden</b> Opdrachtgever Senior leverancier Senior gebruiker	5.1.2.e 5.1.2.e 5.1.2.e	5.1.2.e CIO Analyse & Onderzoek 5.1.2.e Team Informatie Zuid
Dataminer	5.1.2.e	CIO Analyse & Onderzoek, uitvoer project