

# Wat is CAS

In Nederland werkt de politie sinds 2013 met CAS (Criminaliteitsanticipatiesysteem), in eerste instantie in Amsterdam en later in heel Nederland. Het systeem deelt een gebied op in vakjes van 125 bij 125 meter. Vakjes waarvan de kans op een incident vooraf als laag kan worden ingeschat, zoals weilanden en open water, worden verwijderd. Van de overblijvende vakjes wordt een grote hoeveelheid gegevens verzameld: criminaliteitshistorie, afstand tot bekende verdachten, en daarnaast ook demografische en socio-economische gegevens via het CBS. Van ieder vakje wordt op verschillende peilmomenten geregistreerd welke gegevens op dat moment bekend zijn. Vervolgens wordt vastgelegd wat er:

- in de week na de peiling aan incidenten heeft plaatsgevonden (tbv de actuele voorspelling).
- in de week daarna heeft plaatsgevonden (tbv de 1-weekse vroegplanning)
- in de week zes weken na het peilmoment heeft plaatsgevonden (tbv de 6-weekse vroegplanning)

Per vakje wordt twee jaar historie gemeten, onderscheiden in wekelijkse peilmomenten. Dit resulteert in 104 peilingen per gebiedje. Om een uitspraak te doen over welke combinaties van kenmerken indicatief zijn voor criminaliteit in de nabije toekomst, wordt gebruikt gemaakt van logistische regressie. Dat is een machine learning algoritme, dat in staat is om patronen te leren herkennen.

## CAS-Dataset

Zie hiervoor het document 'Pseudocode CAS.pdf'. Kort gezegd wordt voor ieder gebied (basisteam, district of eenheid) een analysebestand gemaakt met features die grotendeels gebaseerd zijn op politiedata en voor een klein gedeelte op data van het CBS. Naast deze features bevat het bestand bovenstaande drie target variabelen. Zowel de features als de target-variabelen hebben betrekking op de speerpunten die door basisteam, district of eenheid zijn aangevraagd.

## CAS-Modeleren

Voor ieder speerpunt wordt vervolgens voor iedere targetvariant een classificatiemodel gemaakt (logistische regressie). De dataset wordt at random opgesplitst in een training en een testset. De trainingset wordt gemaakt om het model te maken, de testset om de performance van het model te meten (via de AUC van de gainchart).

## CAS HTML genereren

Met behulp van het gemaakte model wordt voor iedere combinatie van vakje, speerpunt en targetvariant een kans gegenereerd. Er wordt gekeken of de performance op de testset aan de minimale drempelwaarde voldoet (0.65). Wordt deze performance niet gehaald, dan wordt het speerpunt niet getoond op de kaart. Wordt deze performance wel gehaald, dan wordt de top 3 procent van de vakjes met de hoogste risico's getoond.