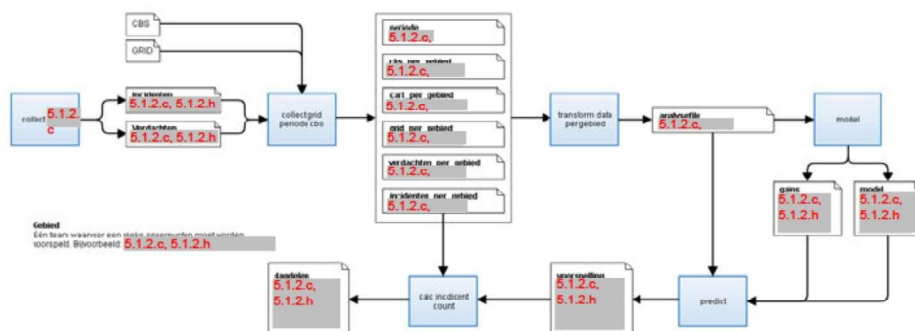


Verwerking



Features

Collect incidenten en verdachten

In deze job wordt een aantal tabellen uit 5.1.2.c, 5.1.2.h opgehaald en gejoined. Het eind resultaat is een tabel met incidenten met bijbehorende locaties en een tabel met verdachten met woonadressen. Deze tabellen worden beiden gefilterd op maatschappelijke klassen die voorkomen in de speerpunten.

Collect grid cbs periode

In deze job worden de grid, cbs en periode gegevens verzameld. 5.1.2.c, 5.1.2.h

Hier worden ook de vakjes net buiten het gebied meegenomen om ook aan de randen van een gebied de juiste waarden te kunnen bepalen. In deze job wordt ook het cartesisch product van de periode tabel en de grid tabel bepaald en voorzien van de gebieden.

Transform

In deze job wordt voor elk gebied voor elk speerpunt voor alle perioden in de afgelopen 2 jaar (het gaat hier om de wekelijkse perioden zoals in de periodetabel, zie Input) de historie opgebouwd, 5.1.2.c, 5.1.2.h

Hier worden ook de CBS gegevens aan gekoppeld en dit wordt per gebiedje opgeslagen. Dit is de analysetabel.

Model

Er is hier sprake van een classificatie probleem. Er moet voorspeld worden of er wel of niet iets gaat plaatsvinden in een hokje. Vanwege de performance is er voor gekozen gebruikt te maken van een logistic regression algoritme.

Voor elk gebiedje wordt het analysebestand zoals weggeschreven vanuit de transform stap opgehaald. 5.1.2.c, 5.1.2.h

De andere targets worden niet meegenomen, deze kennis is ook niet aanwezig op het moment dat er voorspeld gaat worden.

Als het model gemaakt is aan de hand van de trainset, kunnen met behulp van de testset de gains bepaald worden. Als eerste wordt de AUC bepaald, en vervolgens de 3% lift waarde. Deze 3% lift waarde geeft het percentage hits aan in de 3% hokjes met de hoogste score.

Zowel de gains als de score worden per gebiedje, per speerpunt, per target opgeslagen.

Voorspelling

Voorspelling

Voor elk van de 3 targets wordt voor elk gebiedje, voor elk speerpunt een voorspelling gedaan van de kans is dat er een incident gaat plaatsvinden.

5.1.2.c, 5.1.2.h

Trend

Voor elk van de 3% hokjes met de hoogste kans en alle hier aan grenzende hokjes wordt het aantal incidenten bepaald in de afgelopen 2 jaar per dagdeel per dag van de week. Vervolgens wordt aan de hand van deze tellingen opnieuw een hotspot verdeling bepaald, zodat nu de kleurintensiteit van de 3% hoogste hokjes kan worden getoond op basis van de trend op verschillende dagen en dagdelen. Met deze informatie kunnen de CAS tellingen en de dagdeelkaarten worden weergegeven.