

# Met algoritmes de nieuwe kabinetsformatie voorspellen?

In aanloop naar de Tweede Kamerverkiezingen presenteren politieke partijen hun verkiezingsprogramma. Deze zijn veelal online te vinden in pdf-formaat. VKA vroeg zich af of slimme data analyses wellicht inzicht kunnen bieden in hoe partijen zich tot elkaar verhouden.

Wat onderscheidt hen of verbindt hen juist? En zijn de verschillen überhaupt wel zo groot als we soms denken? Wat zijn de verbindende factoren? Welke partijen staan op basis van de verkiezingsprogramma's dan dicht bij elkaar, terwijl je dat niet zou verwachten? En natuurlijk: tot welk kabinet komt zo'n analyse?

## Voorbereiding

Om tot zo'n analyse te komen moeten we een aantal stappen doorlopen, te beginnen met de data preparatie. In meer eenvoudige taal; we moeten de tekst uit de verkiezingsprogramma's van de politieke partijen omzetten naar getallen, zodat de computer een berekening kan doen. Dit proces wordt ook wel *tekst mining* genoemd.

Als we dan eenmaal code hebben geschreven om dit te doen, dan volgen nog wat andere stappen. Zo wil je bijvoorbeeld dat bepaalde stopwoorden, zoals "de", "het" of "een", verwijderd worden. Dat zijn woorden die voor de analyse niet veel gaan toevoegen. Ook wordt de tekst 'gelijkgetrokken', door bijvoorbeeld hoofdletters klein te maken, en wordt er gecorrigeerd voor het feit dat sommige partijen een verkiezingsprogramma hebben van één A4, terwijl andere partijen meer dan tien pagina's geschreven hebben. Dat zijn allemaal belangrijke stappen, alvorens je begint met de analyse.

## Eerste analyse

Dan komen we aan bij onze eerste plot, namelijk een *commonality plot*. Een plot die laat zien welke woorden gedeeld worden door alle documenten die geanalyseerd zijn. In de context van de nieuwe verkiezingen kan het leuk zijn om de gemeenschappelijkheid in de verschillende programma's te visualiseren.



Een mooie zin die deze plot zou samenvatten is: "Nederland, waar mensen goede zorg en onderwijs krijgen."

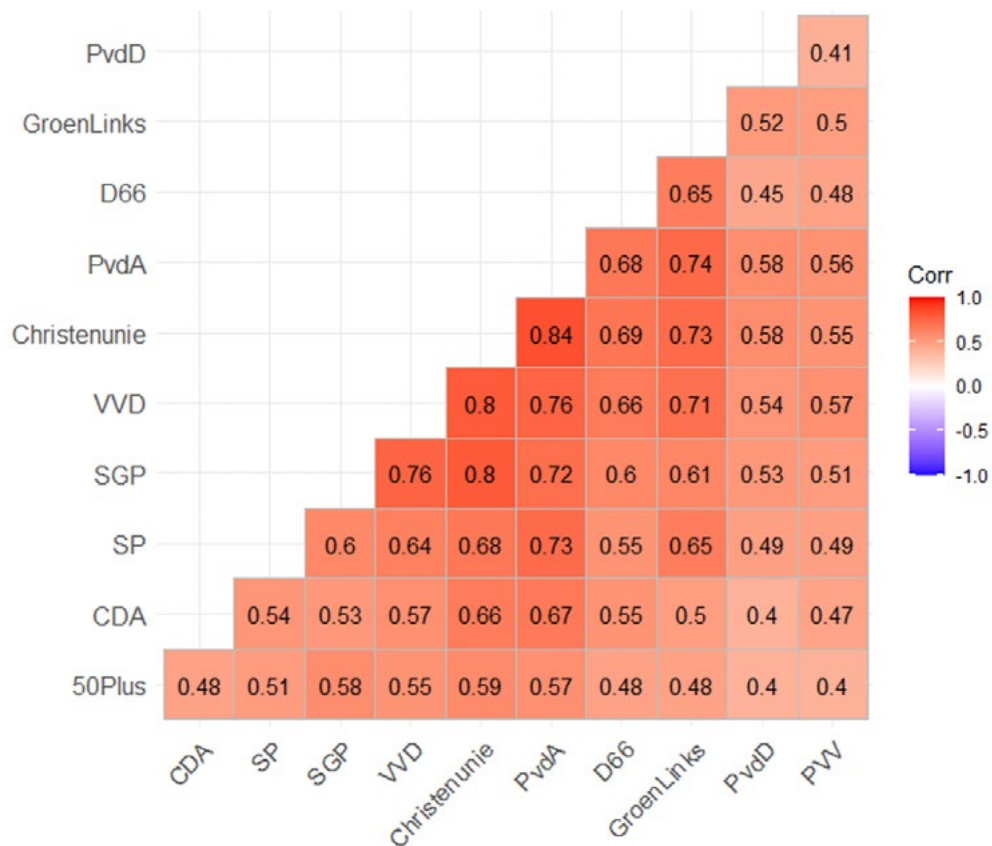
## Met data analyse de formatie starten?

Wij hebben een analyse hoe de politieke partijen zich tot elkaar verhouden en welke woorden een partij het meest kenmerkend is ten opzichte van andere partijen. Deze analyse is gebaseerd op de woordkeuze en de frequentie van deze woorden. Wij vonden het leuk te zien dat op basis van de tekst van verkiezingsprogramma's sommige partijen dicht bij elkaar worden geplott, terwijl je dat niet zou verwachten zoals SP – VVD en CU – D66. Bekijk de onderstaande figuur in hoeverre je je herkent in het beeld dat je zelf van politieke partijen hebt.



Dan komen we bij de laatste plot aan die we gemaakt hebben op basis van de beschikbare verkiezingsprogramma's. Kunnen we op basis van woordkeuze en frequentie van woorden, de correlaties berekenen tussen verschillende partijen?

Dan kan zeker. Hoewel het niet inhoudelijk hoeft te betekenen dat partijen op elkaar lijken, kunnen we wel concluderen dat partijen een bepaalde voorkeur hebben voor bepaalde woorden en die ook met een bepaalde frequentie gebruiken in hun programma's. Als we dan een correlatieplot maken, dus een plot waarmee we laten zien in hoeverre politieke partijen op elkaar lijken kijkend naar hun woordgebruik, dan komen we tot de volgende visualisatie:



Er zijn hier wat interessante verbanden te zien. Allereerst dat de grootste correlatie die we zien bestaat uit de ChristenUnie en de PvdA. Twee partijen die doorgaans in het politieke spectrum niet heel ver van elkaar verwijderd zijn en dat wordt ook bevestigd door de correlaties tussen hun verkiezingsprogramma's. Verder zien we, kijkend naar de laagste correlaties, dat vooral 50Plus een erg lage correlatie laat zien met de PvdD en de PVV.

De geluiden dat de VVD meer naar het midden toeschuift lijkt bevestigd te worden door de correlaties van deze partij met de andere verkiezingsprogramma's. De VVD laat de grootste correlatie zien met de ChristenUnie, met als tweede de politieke partijen de PvdA en de SGP. Gezien de voorspellingen door de Ipsos rondom het aantal verwachte zetels van elke politieke partij - met inachtneming van de bovenstaande correlaties - zouden we tot een voorspelling kunnen komen rondom de nieuwe kabinetsformatie.

Met de aanname dat de VVD de grootste partij gaat worden in de nieuwe kabinetsformatie, komen wij tot de volgende voorspelling voor een nieuw kabinet:

1. **VVD met 38 zetels**
2. **ChristenUnie met 5 zetels**
3. **PvdA met 12 zetels**
4. **SGP met 3 zetels**
5. **GroenLinks met 13 zetels**
6. **D66 met 14 zetels**

Daarmee komen we op een totaal van 85 zetels (als de voorspellingen van de Ipsos uitkomen) en de correlaties op basis van de verschillende verkiezingsprogramma's wellicht iets gaan verklappen over de uitkomst van de kabinetsformatie. In hoeverre komt dit overeen met je eigen verwachtingen?

### Meer weten?

Heb je een complex probleem dat vraagt om een slimme oplossing of is er iets anders dat je in een volgend artikel graag besproken zou zien? Stuur een mailtje naar: [bas.bloemberg@vka.nl](mailto:bas.bloemberg@vka.nl) of [stefan.deblj@vka.nl](mailto:stefan.deblj@vka.nl).