

Van algoritme naar opinie:
De impact van X-bots op politieke dynamiek en
peilingen in 2025

Een kwantitatieve analyse

Niffyn Trouw
14625970

Bachelor thesis
Credits: 18 EC

Bachelor *Informatiekunde*



University of Amsterdam
Faculty of Science
Science Park 900
1098 XH Amsterdam

Supervisor
Dr. T.R. Langerak

Informatics Institute
Faculty of Science
University of Amsterdam
Science Park 900
1098 XH Amsterdam

Semester 1, 2025-2026

Samenvatting

Dit onderzoek analyseert de invloed van geautomatiseerde accounts (bots) op het politiek sentiment op het platform X (voorheen Twitter) en de relatie daarvan met electorale verschuivingen in de aanloop naar de Nederlandse Tweede Kamerverkiezingen van 2025. In een gedigitaliseerde publieke sfeer fungeert X als een centrale arena voor politiek debat, maar de opkomst van bots roept vragen op over de authenticiteit van dit debat en de mate waarin online sentiment een oprechte afspiegeling vormt van de publieke opinie. De focus van dit onderzoek ligt op de Partij voor de Vrijheid (PVV) en Democraten66 (D66), die als ideologische tegenpolen dienen binnen het Nederlandse politieke landschap. Aan de hand van een dataset van 31.032 berichten, verzameld tussen juni en oktober 2025, is de samenhang tussen het online sentiment en de dagelijkse trends in de Peilingwijzer onderzocht. Hierbij is gebruikgemaakt van geavanceerde computationele methoden, waaronder Large Language Models (LLM's) zoals Qwen3:4b en Mistral voor sentimentanalyse en annotatie, en een XGBoost-model voor de detectie van botactiviteit. De methodologie omvat een Distance Correlation (dCor) test om niet-lineaire afhankelijkheden tussen sentiment en peilingdata te identificeren, evenals een analyse van de versterking van berichten door bots via een gewogen bereiksscore. Hiermee wordt beoogd vast te stellen of bots een significant effect hebben op de publieke opinievorming en of zij specifieke narratieven rondom politieke gebeurtenissen disproportioneel versterken. De resultaten van dit onderzoek dragen bij aan de wetenschappelijke kennis over digitale beïnvloeding in de Nederlandse context en bieden inzichten voor de bescherming van democratische meningsvorming tegen kunstmatige manipulatie. Uit het onderzoek blijkt dat er een niet-significante statistische samenhang is tussen het online sentiment op X en de politieke peilingen, een direct causaal verband kan niet onomstotelijk bewezen worden. Hierdoor hebben X-bots weinig tot geen invloed op de verkiezingen en zorgen hun activiteit en sentiment voor weinig verandering in de peilingen.

Trefwoorden: Verkiezingen, X.com, Bot, Twitter, Peilingen, Sentiment

Inhoudsopgave

1	Introductie	2
1.1	Probleemstelling en onderzoeks vragen	3
1.2	Maatschappelijke relevantie	4
1.3	Wetenschappelijke relevantie	4
1.4	Hypotheses	5
2	Theoretisch Kader	6
2.1	De rol van X in de politiek	6
2.2	Het online politieke landschap	7
2.3	Sociale media versus traditionele media	9
2.4	Keuze voor PVV en D66	9
3	Methode	10
3.1	Voorbereiding	10
3.1.1	Technische implementatie van analyse	10
3.1.2	Dataverzameling	10
3.1.3	API instellingen	11
3.1.4	Datastructuur	11
3.1.5	Data behandeling en schoonmaak	12
3.2	Sentimentanalyse en annotatie	12
3.3	Correlatie tussen X-sentiment en peilingen	13
3.3.1	Correlatiemethodes en assumpties	14
3.3.2	Pearson correlatietoets	15
3.3.3	Spearman correlatietoets	15
3.3.4	dCor correlatietoets	15
3.4	Bot bericht detectie	16
3.4.1	Tekst Embedding en tokenisatie	16
3.4.2	XGBoost modelkeuze en optimalisatie	16
3.5	Bereik van berichten	17
3.5.1	Bereik score	18
3.5.2	Verkiezingsvoorspelling en validatie	19
3.6	Zoeken naar trefwoorden	20
4	Resultaten	21
4.1	Beschrijvende statistieken	21

4.2	dCor: X-sentiment & peilingen	22
4.3	Bot detectie en hun sentiment	26
4.4	Trefwoorden analyse	27
5	Discussie	29
5.1	Bevindingen en beantwoording onderzoeksvergadering	29
5.2	Limitaties	30
5.3	Toekomstig Onderzoek	32
6	Conclusie	33
6.1	Correlatie	33
6.2	Botversterking	33
6.3	Trefwoordfrequentie na gebeurtenissen	34
A	Appendix	36
A.1	Github pagina voor data, notebook en software-werkomgeving . . .	36
A.2	Taalfouten corrigeren	36
A.3	Annotatie voor berichten	36
A.4	De belangrijkste gebeurtenissen	37

1 Introductie

Sociale media is in korte tijd uitgegroeid tot een dominante infrastructuur voor publieke communicatie waarin informatieproductie en consumptie grotendeels plaatsvinden buiten traditionele institutionele platformen (Poell e.a., 2019). Deze platformen faciliteren niet alleen sociale interactie, maar beïnvloeden ook in toenemende mate de manier waarop maatschappelijke en politieke betekenis wordt gevormd (Bijleveld e.a., 2024). Vooral de digitale publieke sfeer heeft zich in het afgelopen decennium ontwikkeld tot een centraal strijdtonel voor politieke macht en invloed. Waar politieke meningsvorming traditioneel werd gestuurd via geregelde media zoals kranten, radio en televisie, speelt het politieke debat zich tegenwoordig in toenemende mate af op sociale mediapлатформen. Met name X (voorheen Twitter) heeft zich ontwikkeld tot een arena waarin politici, journalisten en burgers in real time met elkaar interacteren (Buccoliero e.a., 2020). Deze verschuiving heeft het democratisch debat toegankelijker en dynamischer gemaakt, maar roept tegelijkertijd fundamentele vragen op over de authenticiteit en betrouwbaarheid van de online politieke discussie.

In de aanloop naar de Tweede Kamerverkiezingen van 2025 is de invloed van X groter dan ooit, aangezien het platform niet alleen fungeert als communicatiemiddel, maar ook als informele graadmeter voor politiek sentiment (Buccoliero e.a., 2020). Tegelijkertijd staat deze functie onder druk door de toenemende aanwezigheid van geautomatiseerde accounts, zogenoemde 'bots', die in staat zijn om op grote schaal politieke boodschappen te verspreiden en te versterken. Bots zijn geprogrammeerde accounts die autonoom opereren en menselijke interactie op sociale mediapлатформen imiteren door berichten te plaatsen, te herplaatsen, te liken of te reageren, vaak met een hoge frequentie en volgens vooraf ingestelde patronen (Ferrara e.a., 2016). In tegenstelling tot menselijke gebruikers zijn bots niet gebonden aan cognitieve of temporele beperkingen, waardoor zij continu actief kunnen zijn en disproportioneel bijdragen aan de zichtbaarheid van bepaalde politieke discussies. Hierdoor ontstaat onzekerheid over de vraag of het waarneembare online sentiment een oerchte afspiegeling vormt van maatschappelijke voorkeuren, of het resultaat is van doelbewuste, algoritmisch aangestuurde manipulatie. Deze spanning raakt aan de kern van democratische legitimiteit en maakt inzicht in de rol van bots in het politieke ecosysteem urgent.

Binnen dit onderzoek wordt sentiment gedefinieerd als de gemiddelde waardering die gebruikers op X uitdrukken ten opzichte van specifieke politieke partijen. Dit begrip staat in nauwe relatie tot publieke opinie, die hier wordt opgevat als de electorale bereidheid van stemgerechtigden om op een bepaald moment op een partij te stemmen. In de Nederlandse context wordt deze publieke opinie empirisch benaderd via peilingdata, in het bijzonder de Peilingwijzer (Louwerse, 2025). Deze combineert de uitkomsten van drie onafhankelijke panelen: IPSOS I&O, EenVandaag en Verian tot een gewogen gemiddelde als schatting van de electorale steun als percentage van de totale stemgerechtigde bevolking (Louwerse, 2025). Door sentiment op X te relateren aan deze peilingen ontstaat de mogelijkheid om online dynamieken te koppelen aan offline electorale verschuivingen.

1.1 Probleemstelling en onderzoeks vragen

De focus van dit onderzoek ligt op de Partij voor de Vrijheid (PVV) en Democraten 66 (D66). Deze afbakening is zowel empirisch als inhoudelijk gemotiveerd. De nipte electorale verkiezingsuitslag van 2025 positioneert hen als relevante tegenpolen binnen het hedendaagse Nederlandse politieke landschap (Schmidt, 2025). Daarnaast vertegenwoordigen zij duidelijk verschillende ideologieën, respectievelijk conservatief-populistisch en progressief-liberaal, wat de interpretatie van online sentiment en interactiepatronen vergemakkelijkt (Middendorp, 1992). Door deze scherpe tegenstelling zijn positieve en negatieve uitingen en partijgebonden mobilisatie door bots methodologisch beter te onderscheiden.

Ondanks de groeiende aandacht voor de rol van bots in politieke communicatie is er binnen de Nederlandse context nog beperkt empirisch bewijs voor de relatie tussen geautomatiseerd online sentiment en daadwerkelijke electorale verschuivingen. Met name ontbreekt systematisch onderzoek dat deze relatie meetbaar en tijdsgebonden in kaart brengt. Dit onderzoek wenst deze kenniskloof te vullen door in de periode van juni 2025 tot en met oktober 2025 de invloed van X-bots op het sentiment rondom PVV en D66 te analyseren en deze te relateren aan veranderingen in peilingen.

De centrale onderzoeks vrag luidt: *Welke invloed hebben X-bots uitgeoefend op de publieke opinie in de aanloop naar de Tweede Kamerverkiezingen in 2025, en zorgen hun activiteit en sentiment voor veranderingen in de peilingen voor de PVV en D66?*

Ter beantwoording van deze hoofdvraag zijn de volgende deelvragen geformuleerd:

1. Is er een significante correlatie tussen het sentiment op X en de trends in de politieke peilingen voor de PVV en D66?
2. In welke mate versterken bots specifieke politieke berichten, en verschilt dit per partij?
3. Hoe verandert de frequentie van berichten met specifieke trefwoorden rondom belangrijke politieke gebeurtenissen, en welke rol spelen bots hierin?

1.2 Maatschappelijke relevantie

De maatschappelijke relevantie van dit onderzoek ligt in de bescherming van de democratische meningsvorming. Indien botgestuurde netwerken erin slagen om het politieke sentiment op X systematisch te beïnvloeden, ontstaat het risico op een vertekend beeld van maatschappelijke steun. Dit kan leiden tot een zogenoemd 'bandwagon-effect', waarbij kiezers hun voorkeur afstemmen op waargenomen populariteit die in werkelijkheid kunstmatig is geconstrueerd (Farjam, 2021). In een tijdperk waarin zowel buitenlandse immenging als binnenlandse desinformatie een structureel onderdeel zijn geworden van het politieke landschap, is het van groots belang om empirisch vast te stellen welk effect bots daadwerkelijk hebben bij verschuivingen in publieke opinie en peilingen (Irwin e.a., 2021). Inzicht hierin is essentieel voor beleidsmakers, media en burgers die streven naar een weerbaar en transparant democratisch proces.

1.3 Wetenschappelijke relevantie

Wetenschappelijk gezien sluit dit onderzoek aan bij bestaande literatuur over botdetectie en sentimentanalyse in politieke contexten, maar adresseert het tevens duidelijke gaten in de literatuur. Onderzoek naar de Nederlandse politiek is namelijk schaars. Eerder onderzoek van Sanders en van den Bosch, 2020 maakte gebruik van een dataset van circa 17.000 handmatig geannoteerde berichten om botgedrag en sentiment te classificeren. Dit onderzoek bouwt hierop voort door een grotere dataset van 31.032 berichten te analyseren en gebruik te maken van recente Large Language Models (LLM's), waaronder Qwen3:4b en Mistral, voor automatische annotatie en correctie (Vaswani e.a., 2017). Deze aanpak vergroot zowel de schaalbaarheid als de reproduceerbaarheid van de analyse en vermindert subjectiviteit dat bij handmatige codering hoort. Daarnaast onderscheidt dit onderzoek zich door sentiment en botactiviteit expliciet te koppelen aan peilingdata

rond bepaalde cruciale momenten in de politiek, zoals de val van het kabinet en de Gaza-protesten, waardoor de dynamiek tussen online discussie en electorale ontwikkeling beter inzichtelijk wordt.

1.4 Hypotheses

Op basis van het theoretisch kader en eerder onderzoek naar politieke communicatie op sociale media zijn de volgende hypotheses geformuleerd:

- **H0:** X-bots hebben geen invloed op de publieke opinie, want (a) er is geen statistisch significante samenhang aantoonbaar tussen het sentiment op X en electorale steun en/of (b) versterking van politieke berichten door botactiviteit is statistisch onaantoonbaar en/of (c) veranderingen in trefwoordenfrequentie voor PVV en D66, waarbij met name rond belangrijke politieke gebeurtenissen geen duidelijke patronen worden weergeven.
- **H1:** X-bots hebben wel invloed op de publieke opinie, want (a) er is een statistisch significante samenhang waarneembaar tussen het sentiment op X electorale steun **en** (b) versterking van politieke berichten door botactiviteit is statistisch waarneembaar **en** (c) veranderingen in trefwoordenfrequentie voor PVV en D66, waarbij met name rond belangrijke politieke gebeurtenissen, weergeven duidelijke patronen.

Om de causaliteit tussen het X-sentiment en de stembereidheid voor partijen (gemeten aan de hand van peilingen) te onderzoeken, wordt gekeken naar drie elementen: statistische correlatie, theoretische onderbouwing uit de literatuur en temporele volgorde. Temporele volgordeanalyse via trefwoordanalyse en statistische correlatietoetsen voor onderzoek naar de causaliteit van X-sentiment en publieke oordelen werd al eerder gebruikt (Khan e.a., 2021). In combinatie met een solide theoretische basis zijn deze drie elementen noodzakelijk om een mogelijk causaal verband tussen sentiment en peilingen te kunnen vaststellen (D'Agostino McGowan e.a., 2024). De hypotheses zijn daarom zodanig geformuleerd dat pas van causaliteit kan worden gesproken wanneer aan alle drie de voorwaarden wordt voldaan.

2 Theoretisch Kader

Binnen dit onderzoek staan politieke communicatie, sociale media en digitale beïnvloeding centraal. Het theoretisch kader behandelt daarom de aannames, kennis en methodiek waarop dit onderzoek voortbouwt.

2.1 De rol van X in de politiek

Na de overname van Twitter door Elon Musk in oktober 2022 is het platform zowel inhoudelijk als institutioneel ingrijpend veranderd. Verschillende studies en rapporten signaleren een versoepeling van moderatieregels en een toegenomen zichtbaarheid van conservatieve en populistische stemmen (Hickey e.a., 2025). Deze bevindingen sluiten aan bij eerder werk van Mudde, 2019, die stelt dat populistische en radicaal rechtse bewegingen zich strategisch verplaatsen naar platformen waar conflict, emotie en polarisatie worden beloond.

Hoewel deze studies overeenkomen in hun observatie van een ideologische verschuiving, verschillen zij in hun verklaringen. Waar Hickey e.a., 2025 vooral wijst op institutionele veranderingen in moderatie en platformbeleid, benadrukt Mudde, 2019 bredere communicatieve strategieën van populistische actoren. Deze verschillende invalshoeken maken duidelijk dat de verschuiving op X niet eenduidig kan worden verklaard en dat structurele en actor gedreven factoren samenkommen. Dit onderzoek sluit hierbij aan door niet alleen content, maar ook versterkingsmechanismen zoals bereik en engagement te analyseren. X functioneert hierdoor steeds meer als een ecosysteem waarin populistische narratieven worden versterkt. Dit vormt een relevante context voor dit onderzoek, aangezien de PVV sterk aansluit bij deze communicatiestijl. De resultaten van dit onderzoek zullen hierdoor naar verwachting sterker naar voren komen voor de PVV, aangezien de analyse uitsluitend gericht is op X en andere sociale mediaplatforms, die ook invloed op de publieke opinie kunnen uitoefenen, buiten beschouwing blijven. Hoewel meerdere sociale mediaplatforms een rol spelen in politieke communicatie, wordt X in de literatuur consequent aangewezen als het meest invloedrijke platform voor politieke discussie doordat het ook andere mediaplatformen beïnvloedt (Naranjo-Vinueza e.a., 2025). In tegenstelling tot Instagram of TikTok, waar visuele en persoonlijke content domineert, fungeert X als een hybride platform waarin politici, journalisten en burgers direct met elkaar interacteren (Buccoliero e.a., 2020).

Volgens Theocharis e.a., 2016 vervult X bovendien een duidelijke agenda functie: onderwerpen die op X domineren worden sneller opgepikt door journalisten en traditionele media. Andere onderzoekers beweren dat deze 'agenda functie' contextafhankelijk is en verschilt per land en mediasysteem (Seethaler & Melischek, 2019). Deze discussie laat het belang van onderzoek zien binnen een specifieke nationale context. Dit onderzoek gaat in op de nationale context van deze 'agenda functie' voor Nederland door officiële en geverifieerde X accounts van populaire nieuwsoutlets niet uit te sluiten van de dataset, zodat artikelen van journalisten en traditionele media ook een bijdrage leveren aan de meting van het sentiment op X. Door deze 'agenda functie' werkt X ook als een versterker tussen online en offline publieke opinie, wat het platform bijzonder geschikt maakt voor het bestuderen van electorale dynamiek.

2.2 Het online politieke landschap

Waar authentieke berichten van echte gebruikers doorgaans voortkomen uit individuele overtuigingen, ervaringen of emoties, zijn botberichten instrumenteel van aard en gericht op het strategisch vergroten van bereik, engagement en waargenomen populariteit (Varol e.a., 2017). Bots en door mensen geplaatste berichten zijn te onderscheiden aan de hand van metadata en gedragskenmerken. Botaccounts vertonen namelijk vaak een uitzonderlijk hoge postfrequentie, een constante activiteit over meerdere tijdzones en hebben een onevenwichtige verhouding tussen volgers en gevolgde accounts (Varol e.a., 2017). Daarnaast maken zij relatief vaak gebruik van herhaalde of sterk gelijkende tekstpatronen, geautomatiseerde herplaatsingen en beperkte inhoudelijke variatie (Varol e.a., 2017). Menselijke accounts daarentegen vertonen doorgaans meer temporele variatie, contextspecifieke interactie en een grotere diversiteit in taalgebruik en onderwerpkeuze. Door metadata-analyse en gedragsanalyse te combineren, kan met redelijke betrouwbaarheid onderscheid worden gemaakt tussen botgedreven en authentieke berichten volgens Ferrara e.a., 2016, maar dit kan nooit met volledige zekerheid worden vastgesteld, doordat sommige authentieke accounts worden gehackt en ingezet binnen een botnetwerk (Cresci e.a., 2023). Bots kunnen een vertekend beeld creëren van maatschappelijke steun en consensus, zonder dat dit correleert met daadwerkelijk gedeelde publieke opinies. Eerder onderzoek toont aan dat dergelijke botgedreven versterking niet alleen de informatiedynamiek beïnvloedt, maar ook de perceptie van authenticiteit en urgentie van politieke standpunten kan veranderen (Shao e.a., 2018). Deze ontwikkeling ondervindt daarmee de betrouwbaarheid van online sentiment als indicator voor publieke opinie en vormt een fundamentele uitdaging voor zowel democratische meningsvorming als onderzoek naar politieke communicatie.

Een centrale aanname binnen dit onderzoek is dat online sentiment samenhangt met stemintenties. Volgens het onderzoek van Valentino e.a., 2011 leidt de theorie van priming ertoe dat herhaalde emotionele prikkels de politieke evaluatie van burgers beïnvloeden. Positief sentiment kan leiden tot verhoogde sympathie, terwijl negatief sentiment afkeer en demobilisatie kan veroorzaken. Eerder onderzoek van Tumasjan e.a., 2010 toont aan dat geaggregeerd sentiment op sociale media statistisch samenhangt met verkiezingsuitslagen en peilingen, maar dit onderzoek is gedaan in een tijd waar menselijk activiteit nog relatief weinig gedigitaliseerd is (Ruess e.a., 2023). Veel invloedrijke studies naar de relatie tussen sociale media, sentiment en publieke opinie zijn uitgevoerd in een periode waarin geautomatiseerde accounts, algoritmische versterking en grootschalige manipulatie een aanzienlijk kleinere rol speelden dan in het huidige digitale ecosysteem (Ruess e.a., 2023). De toegenomen aanwezigheid van bots en andere vormen van geautomatiseerde beïnvloeding verslechtert daarmee de representativiteit van online data en stelt de betrouwbaarheid van conclusies die hierop zijn gebaseerd ter discussie. Dit betekent niet dat eerdere bevindingen ongeldig zijn, maar wel dat zij, gezien de veranderde technologische en institutionele context, met de nodige voorzichtigheid geïnterpreteerd moeten worden. Het huidige onderzoek gaat dan ook ervan uit dat bestaande inzichten opnieuw beoordeeld moeten worden gezien hedendaagse vormen van digitale beïnvloeding. Toch kan worden gesteld dat, ondanks de veranderde technologische context, niet alleen de richting van het sentiment een rol speelt, maar ook de waargenomen populariteit ervan. Eerder onderzoek naar politieke communicatie laat zien dat perceptie van steun en zichtbaarheid een effect kan hebben op politieke oordeelsvorming, los van de inhoud van een bepaald bericht (Farjam, 2021). Dit kan worden verklaard aan de hand van klassieke theorieën zoals het 'bandwagon effect' en de 'spiral of silence', wat inhoudt dat individuen hun voorkeuren mede afstemmen op wat zij als dominant of maatschappelijk gedragen beschouwen (Farjam, 2021; Noelle-Neumann, 1974). Hoewel sociale media deze mechanismen hebben versterkt en versneld, blijven de onderliggende psychologische processen grotendeels intact. Dit suggereert dat bestaande theoretische inzichten over populariteit en sociale beïnvloeding nog steeds relevant zijn, maar in een digitale context opnieuw en kritisch moeten worden toegepast. Dit onderzoek zoekt uit of dit in het huidige digitale politieke klimaat nog steeds is.

Niet elk bericht heeft dezelfde impact. Literatuur over digitale politieke communicatie benadrukt dat zichtbaarheid, bereik en engagement cruciale oorzaken zijn van invloed (Fujiwara e.a., 2024). Berichten met hoge view counts, likes en herplaatsingen bereiken niet alleen meer gebruikers, maar worden door algoritmen verder versterkt (Narayanan, 2023). Luceri e.a., 2019 toont aan dat bots worden ingezet voor metadata manipulatie door frequent te posten, strategisch te herplaatsen en engagement te simuleren om moderatieteams te misleiden en de zichtbaarheid en

geloofwaardigheid van misleidende content kunstmatig te vergroten. Hierdoor ontstaat een vertekend beeld van maatschappelijke steun. Dit verantwoordt de keuze in dit onderzoek om bereik expliciet mee te nemen in de analyse via een gewogen bereiksmaat.

2.3 Sociale media versus traditionele media

In de hedendaagse digitale samenleving is de invloed van sociale media op politieke meningsvorming aanzienlijk toegenomen ten opzichte van traditionele media (Congge e.a., 2023). Waar traditionele media opereert via selectie van de redactie, worden sociale mediaplatformen gedreven door algoritmische logica en directe participatie (Bennett & Pfetsch, 2018). Chadwick, 2017 beschrijft dit als een hybride mediasysteem waarin sociale media niet slechts aanvullend zijn, maar richtinggevend voor het publieke debat.

Niet alle onderzoekers zijn het eens over de mate waarin traditionele media hun dominante positie hebben verloren. Sommige studies wijzen erop dat journalistieke normen nog steeds een cruciale rol spelen in meningsvorming. Deze spanning in de literatuur benadrukt juist de noodzaak om platformspecifiek onderzoek te doen om te bestuderen of sociale mediaplatformen daadwerkelijk veel invloed hebben op de publieke opinie. X neemt binnen dit hybride systeem een unieke positie in door de directe interactie tussen politici en publiek, waardoor een groter deel van de populatie wordt geanalyseerd en ook hun mening wordt meegenomen, wat het platform relevant maakt voor dit onderzoek.

2.4 Keuze voor PVV en D66

De focus op PVV en D66 is zowel theoretisch als empirisch gemotiveerd. Ideologisch vertegenwoordigen zij tegengestelde posities binnen het Nederlandse politieke spectrum, respectievelijk conservatief populistisch en progressief liberaal (Middendorp, 1992). Deze tegenstelling maakt vergelijking analytisch relevant, omdat verschillen in online sentiment eenvoudig aan de desbetreffende ideologie en politieke partij kunnen worden toegeschreven.

Daarnaast waren de electorale verhoudingen tussen deze partijen in aanloop naar de verkiezingen van 2025 uitzonderlijk nipt (Schmidt, 2025). Onderzoek naar electorale volatiliteit laat zien dat kleine verschuivingen in publieke perceptie juist in dergelijke contexten politiek betekenisvol zijn (Dalton, 2018). Door deze partijen te vergelijken, draagt dit onderzoek bij aan het begrijpen van hoe digitale beïnvloeding kan doorwerken in een competitief en gepolariseerd politiek landschap.

3 Methode

Om de onderzoeksvragen systematisch en empirisch te beantwoorden, is een kwantitatieve onderzoeksbenadering gehanteerd. Deze benadering combineert statistische analysetechnieken met computationele methoden uit de machine learning om grootschalige sociale mediadata op een reproduceerbare en controleerbare wijze te analyseren. Door gebruik te maken van geautomatiseerde analysemethoden wordt beoogd subjectiviteit te minimaliseren en patronen in de data op een consistente manier bloot te leggen.

3.1 Voorbereiding

3.1.1 Technische implementatie van analyse

De volledige data-analyse is uitgevoerd binnen Jupyter Notebooks, die draaiden in een gecontroleerde Anaconda omgeving. Deze opzet maakt het mogelijk om dependencies, library versies en configuraties vast te leggen, waardoor de reproduceerbaarheid en transparantie van het onderzoek gewaarborgt wordt. Voor de classificatie van de data is een geoptimaliseerd machine learning model ingezet dat specifiek is getraind voor de detectie van door bots gegenereerde X berichten. Hierbij is aandacht besteed aan zowel de tekstuele kenmerken van berichten als aan relevante metadata.

De volledige dataset, inclusief de configuratiebestanden van de Anaconda-omgeving en de volledige broncode van het gebruikte machine learning-model, zijn uitgebreid gedocumenteerd en publiekelijk toegankelijk gemaakt op Github. De referentie hiernaartoe staat in de appendix. Hierdoor wordt het mogelijk voor andere onderzoekers om de analyse te repliceren, te verifiëren of verder uit te breiden.

3.1.2 Dataverzameling

De dataset bestaat uit in totaal 31.032 berichten, die zijn verzameld via een betaalde API omgeving (Apify, z.d.). Tijdens de extractie is expliciet gefilterd op berichten met een relatief hoog bereik, gemeten aan de hand van de "view count". Deze selectie is gebaseerd op de theoretische veronderstelling dat berichten met een grotere zichtbaarheid een grotere potentiële invloed uitoefenen op de publieke opinievorming (Apify, z.d.; Fujiwara e.a., 2024). Door deze focus wordt beoogd

de analyse te richten op content die daadwerkelijk door een substantieel publiek is waargenomen.

3.1.3 API instellingen

De dataverzameling is gefaciliteerd via een API-configuratie met specifiek gedefinieerde parameters om de relevantie en kwaliteit van de dataset te maximaliseren. Hierbij is gebruikgemaakt van een ingebouwd filtermechanisme dat berichten van zelfgeïdentificeerde bots automatisch uitsluit. De steekproef is samengesteld door op willekeurige tijdstippen gedurende de dag data te selecteren, om temporele bias zoveel mogelijk te reduceren.

De zoekstrategie was gebaseerd op een trefwoordenfilter dat specifiek gericht was op het Nederlandse politieke landschap. Termen zoals “Wilders” (voor PVV-gerelateerde content) en “Jetten” (voor D66-gerelateerde content) zijn gebruikt, aangevuld met meer algemene politieke terminologie. Hoewel deze methode effectief is gebleken voor het aggregeren van relevante data, is het belangrijk dat er wordt opgemerkt dat de dataset een beperkte mate van ruis bevat in de vorm van irrelevante berichten die wel voldeden aan de trefwoordcriteria, maar inhoudelijk buiten de scope van het onderzoek vallen. De data is verzameld in de periode van 1 juni tot en met 27 oktober, met een gemiddelde frequentie van circa 200 tot 300 berichten per dag. Zowel de omvang van de dataset als de afbakening van de tijdsperiode zijn deels ingegeven door budgettaire restricties bij het gebruik van de API (Apify, z.d.).

3.1.4 Datastructuur

Per verzameld bericht zijn meerdere metadata attributen geselecteerd die relevant zijn voor zowel descriptieve analyse als modeltraining. Deze attributen vormen gezamenlijk de basis voor de latere statistische en machine learninganalyses:

- Tekstuele inhoud: `fullText`.
- Gebruikersstatistieken: aantal volgers (`author/followers`) en het aantal gevolgde accounts (`author/following`).
- Engagement (meta)data: aantal favorieten (`bookmarkCount`), weergaven (`viewCount`), retweets (`retweetCount`), reacties (`replyCount`), citaten (`quoteCount`) en likes (`likeCount`).
- Temporele data: dag van publicatie (`createdAt`).

3.1.5 Data behandeling en schoonmaak

De ruwe tekstdata bevat diverse elementen die ongeschikt zijn voor directe analyse, zoals emoji's, mediaverwijzingen en andere vormen van ruis. Met behulp van de Python-library `re` (Regex) van Python Software Foundation, z.d. is een systematische schoonmaakpijplijn geïmplementeerd die bestaat uit de volgende stappen:

- Verwijderen van gebruikerstags: @ (@gebruikersnaam).
- Strippen van de # - symbolen bij hashtags (met behoud van de tekstuele inhoud)
- Verwijderen van URL's, hyperlinks, interpunctie, overtollige witregels en meidiabestanden (zoals foto's).
- Normalisatie van de tekst door conversie naar lowercase en het verwijderen van veelvoorkomende stopwoorden.

Na deze initiële schoonmaak is de tekst verder geoptimaliseerd voor annotatie middels het open source Large Language Model (LLM) Mistral, aangestuurd via Ollama (Ollama, z.d.-a). Mistral is geselecteerd vanwege de relatief hoge verwerkingssnelheid in vergelijking met andere modellen van vergelijkbare omvang. Het model is ingezet om taalfouten te corrigeren en de algemene leesbaarheid te verbeteren, wat de consistentie en betrouwbaarheid van de uiteindelijke annotatie ten goede komt (Vadlapati, 2023). De exacte prompt die voor dit proces is gebruikt, is opgenomen in de appendix.

3.2 Sentimentanalyse en annotatie

Voor het beantwoorden van deelvraag 1 en 2 is gebruikgemaakt van het Qwen3:4b model voor de geautomatiseerde annotatie van de dataset. Dit model is gekozen vanwege de sterke redeneervaardigheden in combinatie met de mogelijkheid om binnen de beschikbare rekenkracht en geheugenlimieten te opereren (Ollama, z.d.-b; White e.a., 2025). De inzet van een Large Language Model (LLM) voor annotatie draagt bij aan een hoge mate van consistentie en objectiviteit en reduceert de potentiële bias die inherent is aan handmatige annotatie door menselijke codeurs (Gilardi e.a., 2023). Met betrekking tot de hoofdvraag geldt dat, om de invloed van berichten te kunnen meten, vastgesteld moet worden op wie het bericht is gericht en of het een positieve of negatieve lading heeft, zodat de effecten accuraat gemeten kunnen worden.

De volledige prompt die is gebruikt voor de analyse van de tekstuele berichten uit de API voor het meten van het sentiment op X is opgenomen in de appendix. Op basis van deze prompt zijn de volgende variabelen gegenereerd:

- **sentiment**: de sentimentwaarde van het bericht, gekwantificeerd op een Likert-achtige ordinale schaal met de numerieke categorieën -2, -1, 0, 1 en 2. Negatieve waarden duiden op een negatief sentiment, positieve waarden op een positief sentiment (Likert, 1932). De waarde 0 is toegekend aan neutrale, beschrijvende berichten, berichten over andere politieke partijen of inhoud die als irrelevant zijn geklassificeerd. Deze numerieke schaal maakt het mogelijk om gemiddelde sentimentwaarden over de gehele dataset te berekenen.
- **party_target**: een categorische variabele die aangeeft op welke partij het bericht betrekking heeft. De mogelijke waarden zijn 'PVV', 'D66', 'both' (indien beide partijen worden genoemd) en 'None' (indien geen van beide partijen explicet wordt benoemd).

Bij de annotatie is expliciet onderscheid gemaakt tussen impliciete en expliciete berichtgeving. Dit onderscheid is relevant aangezien bestaand onderzoek suggerert dat impliciete boodschappen doorgaans een minder directe invloed hebben op publieke opinievorming dan expliciete boodschappen (Zerback e.a., 2015). Ter illustratie van het annotatieproces door het Qwen3 model zijn representatieve voorbeeldberichten weergegeven in Tabel 3.1.

Tabel 3.1: Annotatiewijze LLM

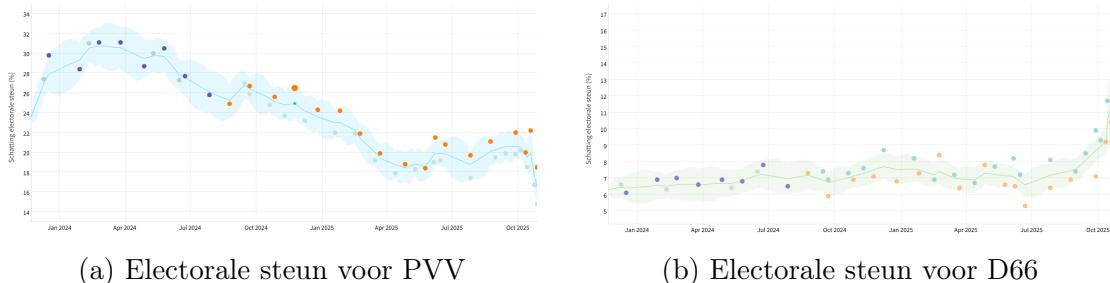
Voorbeeldbericht	Output LLM
Rob Jetten is een held!	2
Rob Jetten komt op televisie.	1
Jimmy Dijk is de weg kwijt!	0
Wat een mooi weer vandaag.	0
Wilders stottert tijdens het debat.	-1
Wilders' retoriek is gevvaarlijk.	-2

3.3 Correlatie tussen X-sentiment en peilingen

Ter beantwoording van deelvraag 1 is de relatie tussen de sentiment geannoteerde berichten op X en de publieke opinie empirisch onderzocht. Publieke opinie is in dit onderzoek gekwantificeerd aan de hand van opiniepeilingen, die fungeren als een meetstaf van de electorale steun per politieke partij op dagbasis. Deze peilingen bieden geen exacte meting van stemintenties, maar worden in de politieke

literatuur algemeen aanvaard als een valide meetstaf voor veranderingen in publieke voorkeuren over de tijd (Jennings & Wlezien, 2018). Het kwantificeren van publieke opinie vormt een essentiële stap in de beantwoording van de hoofdvraag, aangezien het vergeleken wordt met het online sentiment op X.

De gehanteerde peilingdata zijn afkomstig uit drie onafhankelijke panelonderzoeken, die elk gebruik maken van eigen steekproeven en methodologische aannames. Om de invloed van paneel specifieke ruis en meetfouten te reduceren, zijn de dagelijkse peilingen geaggregeerd door het gemiddelde te berekenen. Deze aggregatie zorgt voor een stabielere en robuustere schatting van de electorale steun en sluit aan bij gangbare methoden in vergelijkend peilingsonderzoek. Dit samenvoegingsproces en de resulterende trends zijn visueel weergegeven in Figuur 3.1 (Louwerse, 2025).



Figuur 3.1: Visualisatie van de electorale steun per partij op basis van geaggregeerde peilingdata.

De datapunten voor de periode van 1 juni 2025 tot en met 27 oktober zijn handmatig overgenomen en gestructureerd per partij. Deze handmatige extractie was noodzakelijk vanwege de afwezigheid van een direct downloadbaar dataformaat dat bruikbaar was met de gebruikte analysetools. De verkregen tijdreeksen zijn vervolgens gekoppeld aan het dagelijkse gemiddelde sentiment van de geannoteerde berichten, zodat een correlatieanalyse mogelijk wordt gemaakt tussen online sentiment en electorale steun.

3.3.1 Correlatiemethodes en assumpties

Voor de analyse van de samenhang tussen het sentiment van berichten en de electorale steun is gekozen voor de Distance Correlation (dCor) test (Székely e.a., 2007). De keuze voor deze methode is vooral door de empirische eigenschappen van de verzamelde data, waaronder de niet normale verdeling en het ontbreken van een eenduidige monotone relatie. Voorafgaand aan de toepassing van de dCor test zijn

meerdere statistische assumpties onderzocht om vast te stellen welke correlatiemethode het meest geschikt is.

3.3.2 Pearson correlatietoets

De Pearson correlatie veronderstelt een lineaire samenhang tussen variabelen en vereist dat de data normaal verdeeld is. Om deze assumptie te toetsen, zijn de Shapiro-Wilk test en aanvullende Q-Q plots toegepast (Shapiro & Wilk, 1965). Uit deze analyses bleek dat zowel de sentimentdata als de peilingdata significant afwijken van een normale verdeling. Op basis hiervan is geconcludeerd dat het gebruik van een Pearson correlatietoets methodologisch niet verantwoord was.

3.3.3 Spearman correlatietoets

- Monotoniciteit: De Spearman-correlatie vereist dat de relatie tussen de variabelen monotone eigenschappen vertoont. Door middel van visuele inspectie van scatterplots is onderzocht of sprake was van een consistente stijging of daling in de datapunten (Hauke & Kossowski, 2011). Deze inspectie liet geen eenduidige monotone relatie zien, dus dit beperkt de toepasbaarheid van de Spearman correlatietoets.
- Autocorrelatie: Daarnaast is voor alle relevante variabelen een correlogram opgesteld met behulp van de Autocorrelatie Functie (ACF) om te onderzoeken of de tijdreeksen een systematisch, monotoon verloop vertonen. Hoewel alleen de sentimentdata van D66 enige afwijkende patronen lieten zien, was dit onvoldoende om te voldoen aan de statistische voorwaarden voor een valide Spearman correlatietoets (Box e.a., 2015).

3.3.4 dCor correlatietoets

Gezien de afwezigheid van normaliteit en monotoniciteit is de dCor test het meest geschikte en valide instrument gebleken voor deze analyse (Schober e.a., 2018). De dCor test maakt het mogelijk om zowel lineaire als niet-lineaire afhankelijkheden tussen variabelen te detecteren. De methode vereist dat de data een duidelijk gedefinieerd bereik (minimum en maximum) heeft en dat de steekproefomvang voldoende groot is. Aan beide voorwaarden is in dit onderzoek voldaan.

Een belangrijk aandachtspunt bij de interpretatie van de dCor resultaten is de aannname van onafhankelijkheid (Székely e.a., 2007). Aangezien de onderzochte variabelen tijdreeksen betreffen, is er volgens de literatuur vaak sprake van temporale afhankelijkheid (Granger e.a., 2001). Deze tijdsafhankelijkheid kan leiden

tot het risico op schijncorrelaties, waarbij samenhang wordt waargenomen zonder dat sprake is van een causale relatie (Granger e.a., 2001).

3.4 Bot bericht detectie

Ter beantwoording van deelvraag 2 is een machine learning model ontwikkeld en getraind voor de identificatie van door bots gegenereerde berichten. Eerder onderzoek toont aan dat een combinatie van tekstuele kenmerken en metadata essentieel is voor een accurate detectie van botactiviteit op sociale media (Luceri e.a., 2019). In dit onderzoek is daarom explicet gekozen voor een hybride benadering waarin beide informatietypen worden geïntegreerd. Met betrekking tot de hoofdvraag geldt dat, om de invloed van X-bots te kunnen meten, eerst objectief moet worden vastgesteld welke berichten afkomstig zijn van bots en welke niet.

3.4.1 Tekst Embedding en tokenisatie

Voor de verwerking van tekstuele data is gebruikgemaakt van RobBERT, een gespecialiseerd Nederlandstalig taalmodel dat is gebaseerd op de RoBERTa-architectuur (Delobelle e.a., 2020; Liu e.a., 2019). RobBERT is geselecteerd vanwege zijn aantoonbaar sterke prestaties op het gebied van tokenisatie en het genereren van semantisch rijke embeddings voor Nederlandstalige Natural Language Processing (NLP) taken (Delobelle e.a., 2020).

Aangezien machine learning-modellen een vaste en consistente inputgrootte vereisen, is de lengte van de berichten gestandaardiseerd (Garreta & Moncecchi, 2013). Op basis van een analyse van de distributie van tokenlengtes is de mediaan gekozen als referentiepunt. Berichten die korter waren dan deze mediaanlengte zijn aangevuld met nulvectoren (padding), terwijl langere berichten zijn ingekort tot de mediaanlengte door middel van truncation. Deze aanpak minimaliseert informatieverlies en voorkomt dat extreem lange of korte berichten onevenredige invloed uitoefenen op het model.

3.4.2 XGBoost modelkeuze en optimalisatie

Voor de uiteindelijke classificatietaak is gekozen voor het XGBoost algoritme. Ter validatie van de labels en ter beperking van onderzoeksbias zijn 100 willekeurige berichten handmatig geannoteerd door twee onafhankelijke beoordelaars. Hierbij zijn de labels 0 (authentiek bericht) en 1 (bot bericht) toegekend.

Het XGBoost model is geoptimaliseerd met behulp van een GridSearch procedure, waarbij zowel accuraatheid als de ROC-AUC curve zijn gebruikt om de optimale balans tussen de True Positive Rate (TPR) en False Positive Rate (FPR) te bepalen (Hoo e.a., 2017; Huang e.a., 2012). De volgende hyperparameters zijn systematisch geëvalueerd:

- `max_depth`: [4, 6, 8]
- `learning_rate`: [0.01, 0.05, 0.1]
- `subsample`: [0.8, 1.0]
- `colsample_bytree`: [0.8, 1.0]
- `n_estimators`: [100, 300, 500]

Er is bewust gekozen voor de parameter `tree_method='hist'`. Dit histogram gebaseerde algoritme verhoogt de reken en geheugenefficiëntie aanzienlijk bij grotere datasets, zonder aantoonbaar verlies aan modelprestaties (Chen, 2016). De output van het model is vastgelegd met `objective='binary:logistic'`, wat resulteert in een binaire classificatie met een probabilistische output tussen 0 en 1 (Chen, 2016). Als evaluatiemetric is `eval_metric='logloss'` gehanteerd, aangezien deze maat gevoeliger is voor foutieve maar overzekere voorspellingen dan enkel accuraatheid en direct aansluit bij de logistische optimalisatie van het model.

3.5 Bereik van berichten

Voor deelvraag 2 is de mate van versterking van bot berichten in vergelijking met authentieke berichten systematisch onderzocht. Hierbij is niet alleen gekeken naar het absolute aantal berichten, maar ook naar de wijze waarop deze berichten zich verdelen over de onderzochte politieke partijen en het daaraan gekoppelde sentiment. Deze benadering maakt het mogelijk om inzicht te verkrijgen over botactiviteit en als deze activiteit zich disproportioneel richt op specifieke partijen en of deze activiteit gepaard gaat met een afwijkend sentimentprofiel ten opzichte van authentieke gebruikers. Deze analyse draagt bij aan de hoofdvraag door te laten zien welke politieke partijen door X-bots worden versterkt en in welke mate.

Door het sentiment expliciet te koppelen aan de bedoelde partij (`party_target`) is geanalyseerd welke politieke partij het frequentst doelwit is van bot activiteit en of er significante verschillen bestaan in de aard en toon van de berichtgeving tussen partijen. Deze koppeling stelt in staat om patronen te identificeren in zowel positieve als negatieve beïnvloeding. Daarnaast is het procentuele verschil bere-

kend tussen het gemiddelde sentiment van authentieke berichten en dat van bot berichten, waardoor de mate waarin bots bijdragen aan een vertekend sentimentlandschap kwantitatief kan worden vastgesteld.

3.5.1 Bereik score

Om de potentiële impact van de berichten op een objectieve en vergelijkbare wijze te meten, is een bereik score berekend op basis van Formule 3.1. Deze formule is geïnspireerd op het werk uit Sanders en van den Bosch, 2020, waarin het aandeel van berichten per partij ten opzichte van de totale hoeveelheid berichten wordt gebruikt als meetstaf voor electorale invloed. In dit onderzoek is deze benadering verder verfijnd door de introductie van een vermenigvuldigingsfactor die is gebaseerd op het gemiddelde sentiment van de gehele periode per partij.

Deze modificatie is gebaseerd op de theoretische assumptie dat de effectiviteit van politieke communicatie niet uitsluitend afhankelijk is van het kwantitatieve bereik, maar ook van de kwalitatieve toon van de boodschap. Positief of lovend sentiment kan bijdragen aan een gunstigere perceptie van een partij, terwijl bekritiserende of vijandige berichten juist een negatieve invloed kunnen uitoefenen op politieke voorkeuren (Manser, 2016). Door het gemiddelde sentiment over de gehele onderzoeksperiode te integreren in de berekening, wordt het bereik gewogen naar de aard van de berichtgeving.

Door deze koppeling van sentiment en berichtvolume per partij ontstaat een gewogen bereik score die een indicatie geeft van de potentiële electorale invloed van zowel authentieke als door bots gegenereerde berichten. Deze score reflecteert daarmee niet alleen de zichtbaarheid, maar ook de richting van de invloed op het politieke discussie.

$$\text{Bereik score}_{\text{partij}} = \frac{N_{\text{partij}} \times (1 + \bar{S}_{\text{partij}})}{N_{\text{totaal}}} \times 100 \quad (3.1)$$

met:

N_{partij} = hoeveelheid bericht gericht op de partij

\bar{S}_{partij} = gemiddelde sentiment van berichten gericht op de partij

N_{totaal} = totale hoeveelheid van berichten in de dataset

3.5.2 Verkiezingsvoorspelling en validatie

De oorspronkelijke formule voor het berekenen van de electorale invloed, zoals beschreven in de studie van Sanders en van den Bosch, 2020, is weergegeven in Formule 3.2. Deze originele benadering houdt uitsluitend rekening met het relatieve aandeel van berichten per partij, zonder differentiatie naar sentiment of toon van de berichtgeving.

$$\text{Originele bereik score}_{\text{partij}} = \frac{N_{\text{partij}}}{N_{\text{totaal}}} \times 100 \quad (3.2)$$

met:

N_{partij} = hoeveelheid bericht gericht op de partij

N_{totaal} = totale hoeveelheid van berichten in de dataset

In dit onderzoek is deze originele formule vergeleken met de aangepaste variant waarin sentiment expliciet wordt meegenomen. Om de precisie en voorspellende waarde van beide benaderingen te kwantificeren, is de *Mean Absolute Error* (MAE) gehanteerd als primaire evaluatiemetriek, zoals aangegeven in Formule 3.3 (Wun & Pearn, 1991) en ook gebruikt in Sanders en van den Bosch, 2020. De MAE biedt een directe en intuïtieve maat voor de gemiddelde absolute afwijking tussen de voorspelde electorale steun en de feitelijke peilingwaarden.

Een lagere MAE duidt op een hogere accuraatheid van het model en daarmee op een betere aansluiting tussen de berekende bereik score en de empirisch waargenomen electorale uitkomsten. Deze evaluatie maakt het mogelijk om de robuustheid en meerwaarde van de aangepaste formule te vergelijken met de resultaten uit de bestaande literatuur en om vast te stellen in welke mate sentimentdata afkomstig van X een betrouwbare indicator vormen voor electorale steun.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (3.3)$$

met:

n = aantal observaties

y_i = werkelijke percentage van de stemmen voor observatie i

\hat{y}_i = voorspelde waarde vanuit het model voor observatie i

3.6 Zoeken naar trefwoorden

Ter beantwoording van deelvraag 3 is een evenementanalyse uitgevoerd om de impact van significante politieke en maatschappelijke gebeurtenissen op gesprekken op X te kwantificeren. In de periode van 1 juni tot en met 27 oktober is voor vijf geselecteerde gebeurtenissen onderzocht of er sprake was van een statistisch relevante toename in het gebruik van aan het evenement gerelateerde trefwoorden 5 dagen voor en na een gebeurtenis. Deze aanpak maakt het mogelijk om tijdelijke pieken in aandacht en discussie te relateren aan concrete externe gebeurtenissen. Deze analyse draagt bij aan de hoofdvraag door te onderzoeken in welke mate specifieke politieke gebeurtenissen leiden tot een toename van activiteit op X, en of er een causaal verband zichtbaar is in de temporele volgorde vóór en na een gebeurtenis, inclusief eventuele stijgingen of dalingen in het gebruik van relevante trefwoorden. Met betrekking tot de hoofdvraag kan de causaliteit van bepaalde invloeden die X-bots mogelijk hebben gehad worden aangetoond.

Indien een significante toename in trefwoordgebruik werd geconstateerd, is de gemiddelde procentuele stijging berekend om de intensiteit van het effect vast te stellen. Deze maat biedt inzicht in de relatieve impact van verschillende gebeurtenissen en maakt onderlinge vergelijking mogelijk.

Om een zo objectief mogelijke selectie van de meest invloedrijke gebeurtenissen te waarborgen, is gebruikgemaakt van het online GPT-5.2 chatbotmodel. Dit model is geselecteerd vanwege zijn superieure prestaties in complexe redeneertaken en het vermogen om relevante gebeurtenissen te identificeren binnen grote hoeveelheden contextuele informatie (White e.a., 2025). De specifieke prompt die voor deze identificatie is gehanteerd, is opgenomen in de appendix. De door het model geïdentificeerde evenementen zijn:

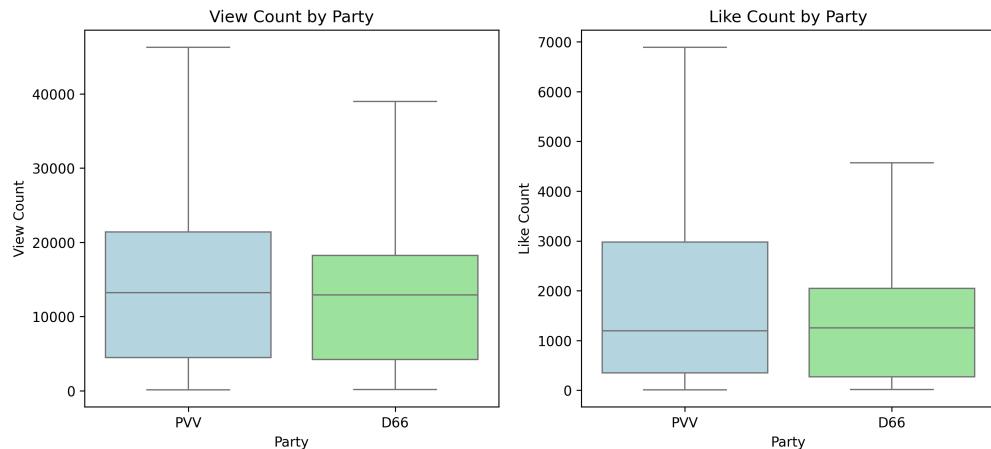
- **Val van de regering:** PVV verlaat de coalitie (3 juni).
- **Gaza-gerelateerde protesten:** De zogenaamde “Red Line demonstraties” (15 juni).
- **Kabinetscrisis:** NSC-ministers stappen op naar aanleiding van het debat over Israël sancties (22 augustus).
- **Campagnevoering:** Lijsttrekkersdebatten en de algemene campagnepiek, met een debat gehouden door EenVandaag als meest bekeken debat met lijsttrekkers van PVV en D66 (1 oktober) (AVROTROS, z.d.).
- **Electoraal slotstuk:** De aanloop naar de verkiezingsdag (21 oktober).

4 Resultaten

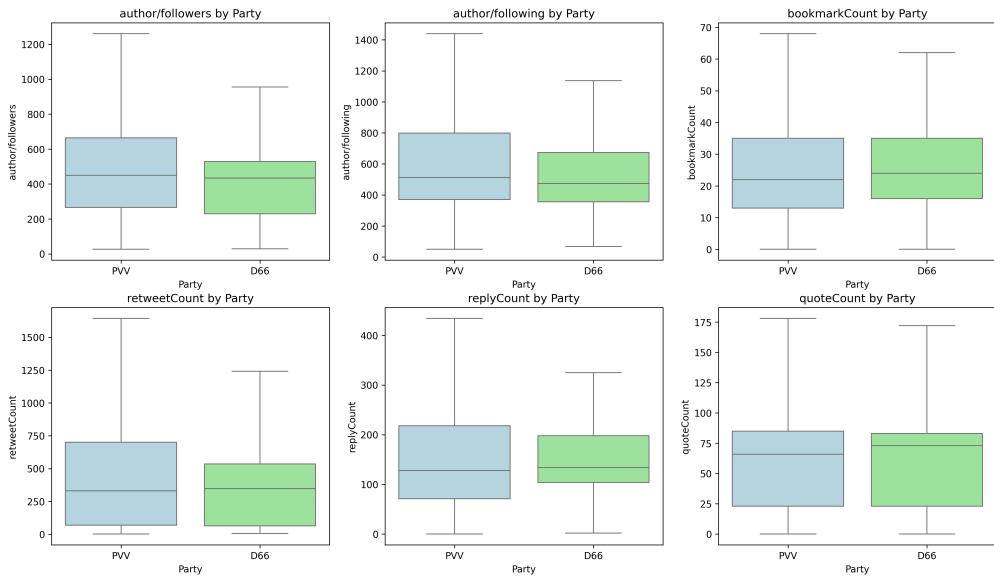
Dit hoofdstuk presenteert de empirische resultaten van het onderzoek. Eerst worden de descriptieve statistieken besproken, gevolgd door de analyse van de samenhang tussen X-sentiment en peilingen met behulp van distance correlation (dCor). Vervolgens worden de resultaten van de botdetectie en het bijbehorende sentiment gepresenteerd. Tot slot wordt ingegaan op de trefwoordenanalyse rondom belangrijke politieke gebeurtenissen.

4.1 Beschrijvende statistieken

Figuur 4.1 toont een visuele representatie van de data voor views en likes, terwijl Figuur 4.2 de overige metadata weergeeft. Voor beide figuren zijn uitschieters verwijderd, aangezien deze relatief infrequent voorkomen en extreem hoge waarden hebben ten opzichte van het merendeel van de observaties.



Figuur 4.1: Boxplot voor views en likes



Figuur 4.2: Boxplot voor de rest

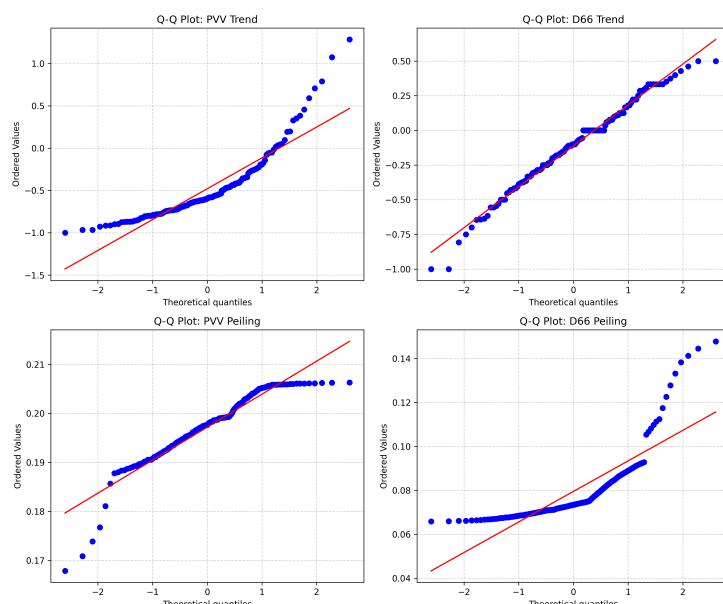
4.2 dCor: X-sentiment & peilingen

Voorafgaand aan de correlatieanalyse is getoetst of aan de assumenties van traditionele correlatiemethoden werd voldaan. De Shapiro-Wilktest toont aan of bepaalde data normaal verdeeld is, wat nodig is voor een Pearson toets. In Tabel 4.1 zijn de resultaten hiervan te zien voor sentiment en peilingdata. Figuur 4.3 ondersteunt deze bevindingen door middel van een visuele test met Q-Q-plots, waarmee wordt onderzocht of de data normaal verdeeld is. Daarnaast is onderzocht of een Spearman correlatietoets toepasbaar is, hiervoor moet gekeken worden of de data monotoon en onafhankelijk is. De analyse van de autocorrelatiefuncties in Figuur 4.4 is uitgevoerd om te onderzoeken of de data monotoniciteit vertoont. De sentimentdistributies per partij zijn weergegeven in Figuur 4.5 (a,b) om te kijken of er onregelmatigheden zijn qua sentiment en berichten naar een bepaalde partij. Ook wordt in dit figuur het verschil tussen menselijke accounts en bots zichtbaar (c). Ten slotte bevat dit figuur de relatie tussen peilingen en sentimentwaarden om een correlatietoets uit te voeren (d), de distancecorrelatie tussen peilingen en sentiment bedraagt respectievelijk 0,2144 voor de PVV en 0,2144 voor D66. Op basis van de gecombineerde dataset (bots en authentieke gebruikers) resulteerde de aangepaste bereikscore in een voorspelde electorale steun van 13.88% voor de PVV en 5.89% voor D66. Vergelijken met de werkelijke peilingen leidde dit tot een absolute fout van respectievelijk 2.9% voor de PVV en 11% voor D66. De gemiddelde MAE bedraagt hiermee 6.95 procentpunt.

Tabel 4.1: Resultaten van de Shapiro-Wilk test voor normaliteit

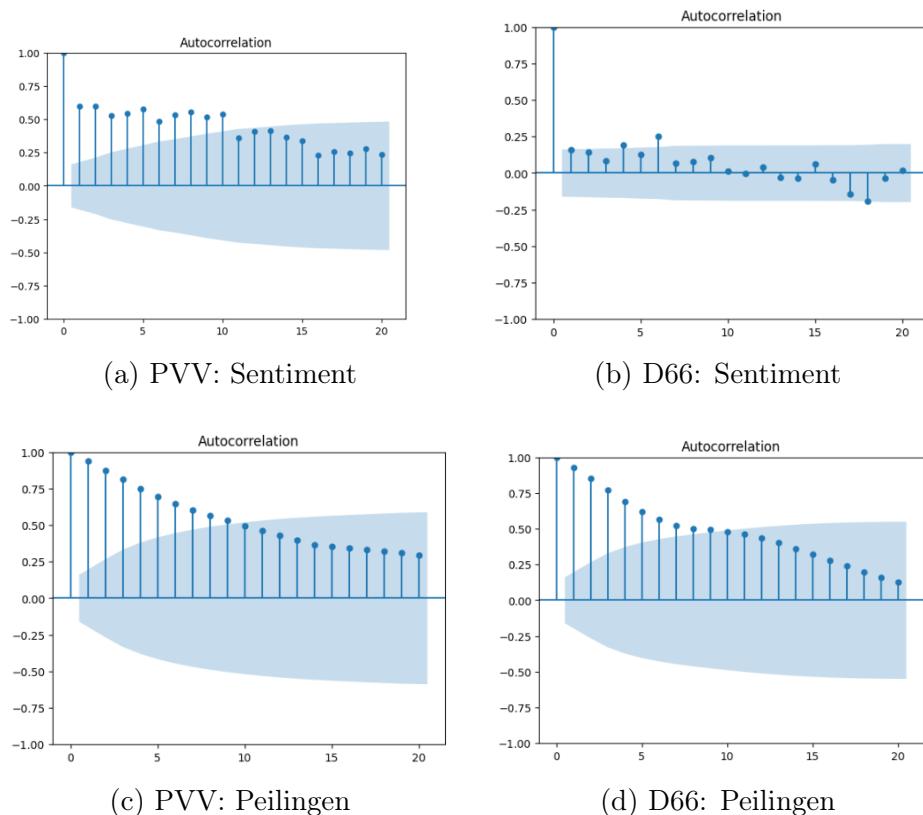
Variabele	W-Waarde	p-waarde
Sentiment PVV	0.8358	< 0.001
Sentiment D66	0.9857	0.1293
Peilingen PVV	0.9073	< 0.001
Peilingen D66	0.7032	< 0.001

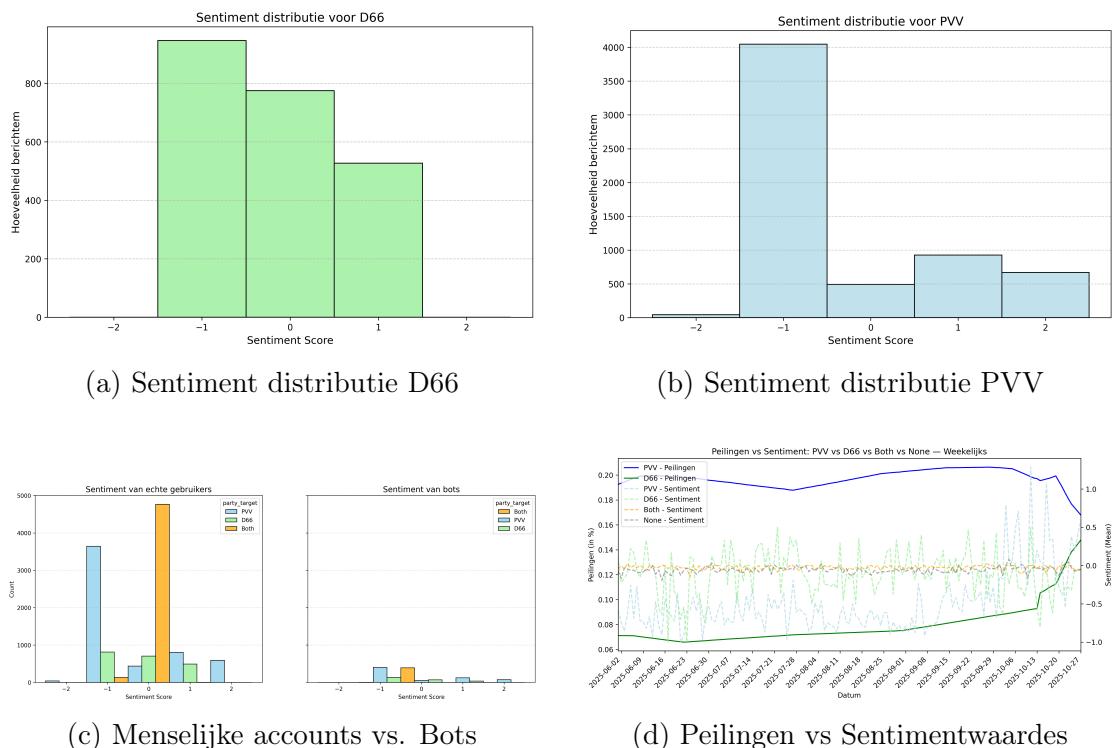
Opmerking: Een p-waarde < 0.05 duidt op een significante afwijking van een normale verdeling.



Figuur 4.3: Q-Q plot voor normaliteit

Figuur 4.4: Afhankelijkheidsanalyse van Peilingen en Trends

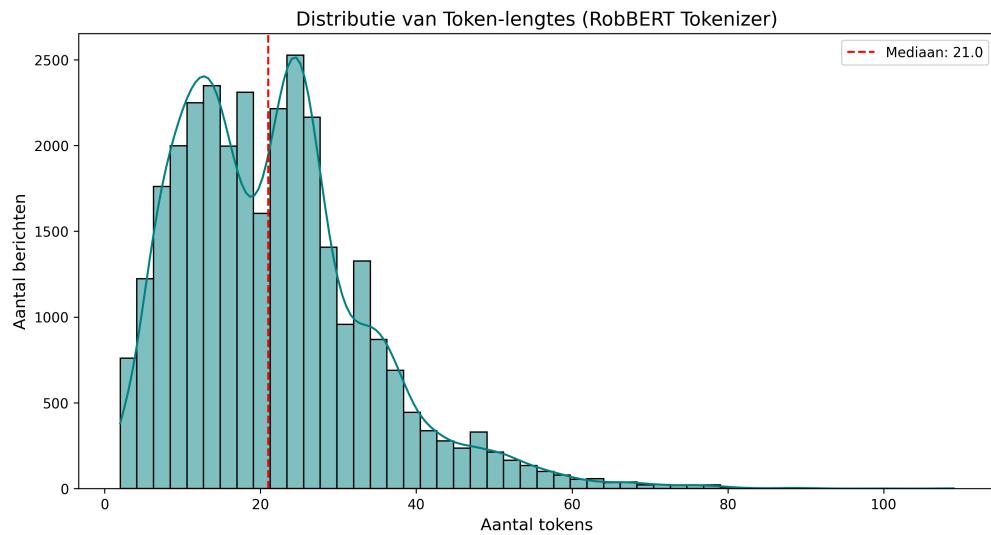




Figuur 4.5: Overzicht van sentimentdistributies per categorie

4.3 Bot detectie en hun sentiment

Voor de tekstuele representatie is gekozen voor padding en truncation op basis van de mediaan tokenlengte (21 tokens), om vertekening door extreem korte of lange berichten te voorkomen. De distributie van tokenlengtes is weergegeven in Figuur 4.6.



Figuur 4.6: Distributie van de hoeveelheid tokens

Het XGBoost-model is getraind met een gestratificeerde train-testverdeling (test size = 0.2). De optimale hyperparameters resulteerden in een maximale boomdiepte van 4, een learning rate van 0.01 en 300 estimators. Het model behaalde een logloss van 0.13837 en een accuracy van 0.95. De bijbehorende evaluatiemetrieken zijn weergegeven in Figuur 4.7.

	precision	recall	f1-score	support
0	0.95	1.00	0.97	18
1	1.00	0.50	0.67	2
accuracy			0.95	20
macro avg	0.97	0.75	0.82	20
weighted avg	0.95	0.95	0.94	20

Figuur 4.7: Evaluatiemetrieken

Tabel 4.2 toont de verdeling van berichten van menselijke accounts en bots, uitgesplitst naar partijgerichte berichtgeving. Hiermee wordt inzichtelijk welke partij gemiddeld meer haat of lovende berichten ontvangt en in welke mate deze berichten afkomstig zijn van mensen of bots.

Tabel 4.2: Gemiddeld sentiment per partijtarget: Echte gebruikers vs. Bots.

Partij Target	Echt (bot=0)	Bot (bot=1)
Both	-0.028	-0.013
D66	-0.160	-0.404
PVV	-0.313	-0.209

4.4 Trefwoorden analyse

In Tabel 4.3 is het gebruik van evenement specifieke trefwoorden weergegeven voor en na geselecteerde politieke gebeurtenissen. Dit is gedaan om te kijken of er een causaal verband is qua temporele volgorde.

Tabel 4.3: Overzicht van belangrijke politieke evenementen en bijbehorende trefwoorden.

Evenement	Trefwoorden	Voor / Bot%	Na / Bot%
Val van de regering, de PVV verlaat coalitie (3 juni)	"Val", "Vallen", "Gevallen"	4 ; 0%	16 ; 6.2%
Gaza gerelateerde "Rode lijn" protesten (15 juni 2025)	"Gaza", "Palestina", "Israël", "Rode lijn", "Protest"	0 ; 0%	6 ; 16.7%
NSC-ministers stappen op over Israël sanctiedebat (22 augustus)	"NSC", "Opstappen", "Stapt op", "Sanctie", "Isreal". "Minister"	18 ; 27.8%	16 ; 6.2%
Lijsttrekkersdebatten en campagnepiek (1 oktober)	"Lijsttrekker", "Campagne", "Debat", "Geert", "Wilders", "Rob", "Jetten"	70 ; 0%	280 ; 24.6%
Electoraal slotstuk (21 oktober)	"Verkiezingsdag", "Ik stem", "Stem op", "Kies", "Kies voor"	0 ; 0%	1 ; 0%

5 Discussie

Dit hoofdstuk bespreekt de bevindingen van het onderzoek in relatie tot de onderzoeks vragen. Daarbij wordt explicet aandacht besteed aan de beperkingen van de gebruikte data, methoden en technische infrastructuur. Tot slot worden concrete richtingen voor toekomstig onderzoek geformuleerd.

5.1 Bevindingen en beantwoording onderzoeks vrag

Over het algemeen behaalt de PVV een groter gemiddeld bereik dan D66, met uitzondering van enkele variabelen zoals bookmarkCount en author/following. Daarentegen vertoont de PVV consistent de hoogste maximumwaarden.

De verdeling van sentimentlabels uit Figuur 4.5(c) laat zien dat het merendeel van de berichten als neutraal is geclasseerd, gevolgd door negatief sentiment en positief sentiment. Sterk positief en sterk negatief sentiment komen relatief weinig voor. Dit laat zien dat er over het algemeen meer negativiteit is, maar berichten met een extremer mening zijn vaker positiever. Wat betreft partijgerichte berichtgeving (Figuur 4.5(a,b)), richt het grootste deel van de berichten zich niet explicet op PVV of D66, terwijl PVV vaker wordt genoemd dan D66. Dit bevestigt de theorie van Hickey e.a., 2025 dat conservatieve en populistische content wordt bemoedigd (en daarmee dus ook de PVV) op X, aangezien de PVV als conservatief wordt geclasseerd (Middendorp, 1992). Het gemiddelde sentiment van authentieke gebruikers is positiever dan die van bots. Ook zijn bots aanzienlijk negatiever sentiment richting D66 (-0.404) dan authentieke gebruikers (-0.160) (zie Tabel 4.2). Voor PVV berichten is het patroon omgekeerd, bots zijn gemiddeld minder negatief dan menselijke accounts (zie Tabel 4.2). Dit bevestigt ook Hickey e.a., 2025, want de PVV wordt hierdoor opnieuw beoordeeld. Bij berichten die beide partijen noemen zijn de verschillen beperkt.

Een Spearman of Pearson correlatietoets was niet mogelijk door de uitslagen van de normaliteitstest en afhankelijkheidstest, dus uiteindelijk is er voor dCor gekozen. Voor de PVV werd een distance correlation van 0,2144 gevonden (zie Resultaten). Voor D66 bedroeg de distance correlation 0,2154 (zie Resultaten). Deze waarden duiden op een zwakke tot matige statistische samenhang tussen het gemiddelde dagelijkse sentiment op X en de peilingtrends voor beide partijen.

Rond de val van de regering op 3 juni is een duidelijke toename zichtbaar van vier naar zestien berichten (zie Tabel 4.3). De Gaza gerelateerde protesten op 15 juni laten eveneens een stijging zien, zij het op een lager absoluut niveau. Voor de NSC crisis in augustus is het aantal trefwoordvermeldingen voor en na het evenement vergelijkbaar, wat wijst op een reeds aanwezige discussie voorafgaand aan het moment zelf. De meest uitgesproken toename wordt waargenomen rond de lijsttrekkersdebatten en campagnepiek op 1 oktober, waarbij het aantal relevante berichten stijgt van 70 naar 280. Het electoraal slotstuk in oktober laat slechts een beperkte toename zien. Dit suggereert dat evenementen gepaard gaan met sterke pieken in politieke discussie op X, terwijl institutionele crises en protesten een meer gematigd effect hebben op trefwoordfrequentie. Vooral na evenementen neemt het gebruik van bots toe, wat zichtbaar is in het verhoogde botgehalte bij vrijwel alle evenementen na afloop.

Kortgezegd hebben X-bots nagenoeg geen effect op de publieke opinie. De statistische samenhang tussen het sentiment van de bots en de mening van het publiek is simpelweg te zwak. Ondanks dat de bots negatiever zijn over D66 en de PVV een groter platform geven, blijft de impact gering.

5.2 Limitaties

Een eerste belangrijke beperking van dit onderzoek betreft de omvang en samenstelling van de dataset die is gebruikt voor botdetectie en sentimentanalyse. Hoewel de dataset met ruim 31.032 berichten relatief groot is, was uitbreiding met meer data beperkt door budgettaire restricties bij het gebruik van de API. De kosten van grootschalige dataverzameling maakten het onhaalbaar om een langere tijdsperiode of een groter aantal berichten te analyseren. Hierdoor blijft de statistische kracht van de analyses beperkt en kunnen subtile effecten mogelijk onopgemerkt blijven.

Daarnaast bleek dat de API, ondanks het toepassen van trefwoordenfilters, niet volledig betrouwbaar was in het selecteren van uitsluitend relevante berichten. In de dataset zijn berichten opgenomen die geen expliciete filterwoorden bevatten, evenals berichten in andere talen dan het Nederlands. Hoewel een deel van deze ruis tijdens de data schoonmaak is verwijderd, kan niet worden uitgesloten dat irrelevante of verkeerd geïnterpreteerde berichten invloed hebben gehad op de sentiment en botclassificatie.

Een verdere beperking is de keuze om slechts twee politieke partijen, PVV en D66, binnen de scope van dit onderzoek op te nemen. Deze afbakening is methodologisch verdedigbaar, omdat het differentiëren tussen sentiment en partijdoelwit bij

meerdere partijen aanzienlijk complexer wordt. Wel is hierdoor de generaliseerbaarheid van de bevindingen naar het bredere politieke landschap beperkt. Het kan namelijk zijn dat dit ertoe leidt dat interacties en vergelijkingen tussen andere relevante partijen buiten beschouwing blijven. De classificatie van de variabele *party_target* bleek in de praktijk moeilijk, met name bij berichten die impliciet of indirect naar meerdere partijen verwijzen. Voor een nauwkeuriger onderscheid zou een geavanceerder model nodig zijn, specifiek getraind op partijherkenning in politieke context. De beschikbare hardware, bestaande uit een RTX 2060 Super GPU, een Intel i5-8400 CPU en 12 GB aan RAM, maakte het echter onmogelijk om grotere en complexere modellen te gebruiken dan Qwen3:4b (White e.a., 2025). Modellen met meer parameters en betere prestaties vereisen aanzienlijk meer rekenkracht en geheugen, wat de schaal en diepgang van de analyse heeft beperkt.

Ook bij de berekening van de bereikscoördinaten traden methodologische beperkingen op. Het zou wenselijk zijn om het gemiddelde sentiment te wegen per score van de Likert-achtige schaal (-2,-1, 0, 1, 2), zodat extreme sentimenten een grotere invloed hebben op de uiteindelijke score. In de praktijk bleek dit niet haalbaar, omdat een groot deel van de berichten een sentimentwaarde van nul had. Door deze grote hoeveelheid nulwaarden verloor de vermenigvuldigingsfactor zijn effect, wat de toegevoegde waarde van een dergelijke weging minimaliseerde. Dit probleem is nauw verbonden met de kwaliteit van de annotatie en had mogelijk voorkomen kunnen worden met een grotere dataset.

Verder moet rekening worden gehouden met mogelijke platformbias. Eerder onderzoek suggereert dat X sinds de overname door Elon Musk relatief meer rechtse en populistische content faciliteert en versterkt (Hickey e.a., 2025). Dit kan ertoe leiden dat rechtse partijen, zoals de PVV, structureel zichtbaarder zijn in de dataset. De waargenomen verschillen in bereik en sentiment kunnen daardoor gedeeltelijk het gevolg zijn van platformdynamiek in plaats van daadwerkelijke verschillen in publieke opinie. Deze bias kan verklaren waarom sommige resultaten extremer zijn voor de PVV.

Ten slotte kent de gebruikte correlatiemethode beperkingen. Hoewel distance correlation geschikt is voor het detecteren van zowel lineaire als non-lineaire samenhang, werden niet aan alle assumenties voldaan. Aangezien zowel sentiment als peilingen tijdreeksen zijn, bestaat het risico op schijncorrelaties als gevolg van gedeelde trends of externe invloeden.

5.3 Toekomstig Onderzoek

Toekomstig onderzoek kan op meerdere manieren voortbouwen op de bevindingen van deze studie. Allereerst zou uitbreiding van de dataset, zowel in omvang als in tijdsduur, de robuustheid van de resultaten aanzienlijk vergroten. Met meer data wordt het mogelijk om slimmere modellen te gebruiken en complexere bereikscore te formuleren en effectief toe te passen. Het gebruik van bepaalde LLM's en bepaalde machine learning technieken is voornamelijk toegepast om objectiviteit te waarborgen (Gilardi e.a., 2023). Handmatig geannoteerde datasets zijn vaak gevoelig voor de bias van de onderzoeker en vergen bovendien aanzienlijk meer tijd dan automatische annotatie door een computer (Gilardi e.a., 2023).

Daarnaast biedt het opnemen van meer politieke partijen interessante mogelijkheden voor vergelijkend onderzoek. Een multi-partijbenadering zou beter aansluiten bij het Nederlandse politieke landschap, maar vereist geavanceerdere modellen voor partijgerichte berichtgeving en meer rekenkracht dan in dit onderzoek beschikbaar is.

Ook methodologisch zijn er duidelijke uitbreidingsmogelijkheden. In plaats van klassieke correlatiemethoden zou toekomstig onderzoek gebruik kunnen maken van specifieke tijdreeksanalyses die wel beschikbaar zijn via Python. Deze technieken zijn wellicht beter voor het berekenen van temporele afhankelijkheid en kunnen helpen om schijncorrelaties te onderscheiden van echte sterke correlaties.

Ten slotte zou het combineren van verschillende databronnen, zoals andere sociale mediaplatforms of traditionele media, waardevolle context kunnen bieden. Door data van X te trianguleren met bijvoorbeeld nieuwsartikelen of zoekgedrag kan beter worden vastgesteld welk effect online sentiment kan hebben op maatschappelijke dynamiek.

6 Conclusie

Dit onderzoek heeft als doel om te onderzoeken welk effect het sentiment op X, in het bijzonder de activiteit van geautomatiseerde accounts (bots), samenhangt met veranderingen in de publieke opinie in de aanloop naar de Tweede Kamerverkiezingen van 2025. Door sentimentanalyse, botdetectie en peilingdata te combineren, is inzicht verkregen in de relatie tussen online politieke dynamiek en electorale ontwikkelingen voor de PVV en D66.

6.1 Correlatie

De resultaten tonen aan dat er voor zowel de PVV als D66 sprake is van een zwakke tot matige statistische samenhang tussen het gemiddelde dagelijkse sentiment op X en de trends in de peilingen, gemeten met behulp van distance correlation (dCor). De gevonden dCor-waarden (circa 0.21 voor beide partijen) wijzen erop dat online sentiment en electorale steun niet volledig onafhankelijk van elkaar zijn. Deze uitkomst ondersteunt H1 in zeer beperkte mate, waarin een statistische samenhang werd verwacht tussen online sentiment en peilingen. Tegelijkertijd dient deze samenhang met de nodige voorzichtigheid te worden geïnterpreteerd. Hoewel dCor geschikt is voor het detecteren van zowel lineaire als non-lineaire relaties en geen normaliteits of monotoniciteitsassumpties vereist, wordt niet volledig voldaan aan de assumptie van onafhankelijkheid tussen observaties. Zowel sentiment als peilingen zijn tijddreksken en vertonen duidelijke afhankelijkheid. Hierdoor kan niet worden uitgesloten dat een deel van de waargenomen samenhang het gevolg is van gedeelde trends of externe gebeurtenissen, in plaats van een directe relatie tussen online sentiment en publieke opinie. De resultaten mogen daarom niet worden opgevat als bewijs voor causaliteit, maar uitsluitend als indicatie van statistische samenhang.

6.2 Botversterking

Wat betreft botactiviteit laat het onderzoek zien dat botberichten botberichten duidelijke verschillen in sentiment ten opzichte van authentieke gebruikers. Bots zijn gemiddeld 2,525 (-0.404 / -0.160) keer zo negatief richting D66 dan richting PVV. Bots zijn ook negatief richting de PVV, maar die zijn positiever dan echte

mensen. Dit suggereert dat botactiviteit niet neutraal is, maar selectief wordt ingezet wat de literatuur ook aangeeft (Mudde, 2019). Deze bevinding ondersteunt H1 dat bots bijdragen aan een vertekend sentiment door het versterken van bepaalde politieke berichten, wat mogelijk invloed heeft op de perceptie van publieke steun. De aangepaste bereikscore, waarin sentiment werd geïntegreerd, bleek slechts beperkt in staat om electorale steun nauwkeurig te voorspellen. Eerder onderzoek bleek een betere MAE te hebben (2,71) vergeleken de MAE van dit onderzoek (6.95) (Sanders & van den Bosch, 2020). Hoewel de voorspelling voor de PVV relatief dicht bij de werkelijke peilingen lag, was de afwijking voor D66 aanzienlijk groter. Dit onderstreept dat online zichtbaarheid en sentiment slechts een deel van de electorale dynamiek verklaren en dat peilingen worden beïnvloed door een veelheid aan factoren die buiten de scope van sociale media vallen.

6.3 Trefwoordenfrequentie na gebeurtenissen

Verder laat de trefwoordenanalyse zien dat politieke gebeurtenissen, met name campagnegerelateerde momenten zoals lijsttrekkersdebatten, gepaard gaan met sterke pieken in politieke discussie op X. Dit bevestigt de rol van X als reactief platform waar politieke aandacht zich snel concentreert rondom actuele onderwerpen. De som van alle trefwoorden vertoont een toename van 347% (procentuele verandering van vóór en na, zoals berekend in Tabel 4.3) wanneer de periodes vóór en na de gebeurtenissen met elkaar worden vergeleken. Het botpercentage is bovendien ongeveer verdubbeld (van een gemiddeld botpercentage van 5,56% vóór en 10,74% ná, zoals weergegeven in Tabel 4.3). Deze veranderingen ondersteunen H1 door te laten zien dat er zowel een verandering in de frequentie van trefwoorden is als dat bots hieraan bijdragen.

Samenvattend vindt dit onderzoek aanwijzingen voor een zwakke tot matige statistische samenhang tussen het gemiddelde dagelijkse sentiment op X en trends in stemintenties voor PVV en D66, wat wijst op een zwakke afhankelijkheid tussen online sentiment en peilingen. Hierdoor wordt H1 niet volledig ondersteund en kunnen de resultaten, door het niet volledig voldoen aan de assumpties, niet met zekerheid worden vastgesteld. Tegelijkertijd blijkt de voorspellende waarde van sentiment-gecorrigeerde bereiksmaten voor electorale steun beperkt, al biedt dit wel gedeeltelijke ondersteuning voor H1. Ook de voorspellende waarde van dCor is laag, aangezien niet aan alle assumpties werd voldaan. Bovendien kunnen de beperkte datahoeveelheid en de kwaliteit van de annotaties de resultaten hebben verzwakt. Het temporele verband kan echter wel worden aangetoond door de sterke stijging wanneer de periodes vóór en na de gebeurtenissen met elkaar worden vergeleken, hiermee wordt H1 gedeeltelijk ondersteund. Daarnaast toont de

analyse aan dat botactiviteit systematisch en selectief is, bots verspreiden relatief negatiever sentiment richting specifieke partijen wat kan bijdragen aan vertekening van online discussies. Alle drie de onderdelen van H1 kunnen niet volledig en met volledige zekerheid worden geaccepteerd. Om die reden wordt H0 geaccepteerd, al lijkt er wel sprake te zijn van een indicatie van een mogelijk causaal verband tussen sentiment op X en peilingen.

Wat betreft de hoofdvraag: X-bots hebben nauwelijks tot geen effect gehad op de publieke opinie. Dit blijkt uit de zwakke statistische correlatie tussen sentiment en publieke opinie. Hoewel de bots negatiever zijn in sentiment richting D66 dan de PVV, botactiveit de PVV een groter bereik geven en de literatuur vermoedt dat er een verband is tussen sentiment en peilingen volgens Tumasjan e.a., 2010, wegen deze factoren weinig, omdat er geen causaal verband kan worden vastgesteld vanwege het ontbreken van een duidelijke statistische correlatie (D'Agostino McGowan e.a., 2024; Khan e.a., 2021).

A Appendix

A.1 Github pagina voor data, notebook en software-werkomgeving

<https://github.com/Niffyn/scriptie>

A.2 Taalfouten corrigeren

You are a grammar correction bot for Dutch tweets.

Given the tweet: {tweet}, correct this sentence in the most simple way possible.
If it is already written correctly, do nothing.

A.3 Annotatie voor berichten

You are a Twitter sentiment annotation bot for Dutch tweets.

You must ONLY respond with:

- One of: `highly positive`, `positive`, `neutral`, `negative` or `highly negative`
- And the Dutch political party it targets:
 - PVV (Partij voor de Vrijheid led by Geert Wilders)
 - D66 (Democraten66 led by Rob Jetten)
 - `None` if no political target is clear or it's non-political
 - `both` if both are mentioned

Given the tweet: {tweet}, what is this tweet's sentiment and which party does it target?

Remember, only respond in format: `sentiment,targeted_party`

A.4 De belangrijkste gebeurtenissen

Geef de 5 belangrijkste evenementen in de periode 1 juni 2025 t/m 27 oktober 2025 die het meeste X/Twitter-effect kunnen hebben op de Nederlandse verkiezingen 2025, specifiek voor de partijen D66 en PVV.

De output moet de volgende structuur hebben:

- **Datum**
- **Naam van het evenement**
- **Korte omschrijving** + waarom dit relevant is voor het Twitter-debat over PVV en/of D66

Referenties

- Apify. (z.d.). Tweet Scraper V2. Verkregen december 1, 2025, van <https://apify.com/apidojo/tweet-scraper>
- AVROTROS. (z.d.). Zes lijsttrekkers EenVandaag Verkiezingsdebat bekend. <https:////pers.avrotros.nl/zes-lijsttrekkers-eenvandaag-verkiezingsdebat-bekend/>
- Bennett, W. L., & Pfetsch, B. (2018). Rethinking political communication in a time of disrupted public spheres. *Journal of communication*, 68(2), 243–253.
- Bijleveld, C., Engbersen, G., Schrijvers, E., Damstra, A., Askari, L., & Hoogendoorn, K. (2024). *Aandacht voor media: Naar nieuwe waarborgen voor hun democratische functies* (tech. rap. Nr. 111) (Geraadpleegd op 1 januari 2026). Wetenschappelijke Raad voor het Regeringsbeleid. Den Haag. <https://www.wrr.nl/publicaties/rapporten/2024/10/03/aandacht-voor-media>
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Buccoliero, L., Bellio, E., Crestini, G., & Arkoudas, A. (2020). Twitter and politics: Evidence from the US presidential elections 2016. *Journal of Marketing Communications*, 26(1), 88–114.
- Chadwick, A. (2017). *The hybrid media system: Politics and power*. Oxford University Press.
- Chen, T. (2016). XGBoost: A Scalable Tree Boosting System. *Cornell University*.
- Congge, U., Guillamón, M.-D., Nurmandi, A., Salahudin & Sihidi, I. T. (2023). Digital democracy: A systematic literature review. *Frontiers in Political Science*, 5, 972802.
- Cresci, S., Yang, K.-C., Spognardi, A., Di Pietro, R., Menczer, F., & Petrocchi, M. (2023). Demystifying misconceptions in social bots research. *Social Science Computer Review*, 08944393251376707.
- D'Agostino McGowan, L., Gerke, T., & Barrett, M. (2024). Causal inference is not just a statistics problem. *Journal of Statistics and Data Science Education*, 32(2), 150–155.
- Dalton, R. J. (2018). *Political realignment: Economics, culture, and electoral change*. Oxford University Press.
- Delobelle, P., Winters, T., & Berendt, B. (2020). Robbert: a dutch roberta-based language model. *arXiv preprint arXiv:2001.06286*.

- Farjam, M. (2021). The bandwagon effect in an online voting experiment with real political organizations. *International Journal of Public Opinion Research*, 33(2), 412–421.
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104.
- Fujiwara, T., Müller, K., & Schwarz, C. (2024). The effect of social media on elections: Evidence from the united states. *Journal of the European Economic Association*, 22(3), 1495–1539.
- Garreta, R., & Moncecchi, G. (2013). *Learning scikit-learn: machine learning in python* (Deel 2013). Packt Publishing Birmingham.
- Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. *Proceedings of the National Academy of Sciences*, 120(30), e2305016120.
- Granger, C. W., Newbold, P., & Econom, J. (2001). Spurious regressions in econometrics. *A Companion to theoretical Econometrics*, 557–561.
- Hauke, J., & Kossowski, T. (2011). Comparison of values of Pearson's and Spearman's correlation coefficients. *Comparison Of Values Of Pearson's And Spearman's Correlation Coefficients*.
- Hickey, D., Fessler, D. M., Lerman, K., & Burghardt, K. (2025). X under Musk's leadership: Substantial hate and no reduction in inauthentic activity. *PLoS One*, 20(2), e0313293.
- Hoo, Z. H., Candlish, J., & Teare, D. (2017). What is an ROC curve?
- Huang, Q., Mao, J., & Liu, Y. (2012). An improved grid search algorithm of SVR parameters optimization. *2012 IEEE 14th International Conference on Communication Technology*, 1022–1026.
- Irwin, G. A., e.a. (2021). Impediments to Foreign Interference in Elections in the Netherlands. *Election Law Journal*. <https://doi.org/10.1089/elj.2020.0654>
- Jennings, W., & Wlezien, C. (2018). Election polling errors across time and space. *Nature Human Behaviour*, 2(4), 276–283.
- Khan, A., Zhang, H., Boudjellal, N., Ahmad, A., Shang, J., Dai, L., & Hayat, B. (2021). Election prediction on twitter: a systematic mapping study. *Complexity*, 2021(1), 5565434.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology*.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Louwerse, T. (2025). Trends sinds de vorige verkiezingen. *Peilingwijzer*. https://peilingwijzer.tomlouwerse.nl/grafiek_trends.html

- Luceri, L., Deb, A., Badawy, A., & Ferrara, E. (2019). Red bots do it better: Comparative analysis of social bot partisan behavior. *Companion proceedings of the 2019 world wide web conference*, 1007–1012.
- Manser, A. (2016). The facebook effect. <https://www1.udel.edu/udaily/2016/sep/politics-social-media-092315.html>
- Middendorp, C. P. (1992). Left-Right Self-identification and (Post) materialism in the Ideological Space; their Effect on the Vote in the Netherlands. *Electoral Studies*, 11(3), 249–260.
- Mudde, C. (2019). *The far right today*. John Wiley & Sons.
- Naranjo-Vinueza, A., Casillas-Martín, S., Cabezas-González, M., & Nevado-Batalla Moreno, P. T. (2025). From tweets to power: an integrative thematic review of political communication and platform governance on Twitter/X (2009–2024). *Frontiers in Political Science*, 7, 1666104.
- Narayanan, A. (2023). Understanding social media recommendation algorithms.
- Noelle-Neumann, E. (1974). The spiral of silence a theory of public opinion. *Journal of communication*, 24(2), 43–51.
- Ollama. (z.d.-a). Mistral: Latest generation large language models (Ollama library).
- Ollama. (z.d.-b). Qwen3: Latest generation large language models (Ollama library).
- Poell, T., Nieborg, D., & Van Dijck, J. (2019). Platformisation. *Internet policy review*, 8(4), 1–13.
- Python Software Foundation. (z.d.). re — Regular expression operations. Verkregen januari 1, 2026, van <https://docs.python.org/3/library/re.html>
- Ruess, C., Hoffmann, C. P., Boulian, S., & Heger, K. (2023). Online political participation: the evolution of a concept. *Information, Communication & Society*, 26(8), 1495–1512.
- Sanders, E., & van den Bosch, A. (2020). Optimising twitter-based political election prediction with relevance and sentiment filters.
- Schmidt, C. (2025). Definitief verschil tussen D66 en PVV bijna een halve zetel, maar die zien we niet. *Trouw*.
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia*, 126(5), 1763–1768.
- Seethaler, J., & Melischek, G. (2019). Twitter as a tool for agenda building in election campaigns? The case of Austria. *Journalism*, 20(8), 1087–1107.
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature communications*, 9(1), 4787.
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591–611.

- Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances.
- Theocharis, Y., Barberá, P., Fazekas, Z., Popa, S. A., & Parnet, O. (2016). A bad workman blames his tweets: The consequences of citizens' uncivil Twitter use when interacting with party candidates. *Journal of communication*, 66(6), 1007–1031.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal About Political Sentiment. *Proceedings of the International AAAI Conference on Web and Social Media*, 4(1), 178–185.
- Vadlapati, P. (2023). Investigating the impact of linguistic errors of prompts on llm accuracy. *ESP Journal of Engineering & Technology Advancements*, 3(2), 144–147.
- Valentino, N. A., Brader, T., Groenendyk, E. W., Gregorowicz, K., & Hutchings, V. L. (2011). Election night's alright for fighting: The role of emotions in political participation. *The journal of politics*, 73(1), 156–170.
- Varol, O., Ferrara, E., Davis, C., Menczer, F., & Flammini, A. (2017). Online human-bot interactions: Detection, estimation, and characterization. *Proceedings of the international AAAI conference on web and social media*, 11(1), 280–289.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is All You Need. *Advances in Neural Information Processing Systems*, 5998–6008. <https://arxiv.org/abs/1706.03762>
- White, C., Dooley, S., Roberts, M., Pal, A., Feuer, B., Jain, S., Shwartz-Ziv, R., Jain, N., Saifullah, K., Dey, S., Shubh-Agrawal, Sandha, S. S., Naidu, S. V., Hegde, C., LeCun, Y., Goldstein, T., Neiswanger, W., & Goldblum, M. (2025). LiveBench: A Challenging, Contamination-Free LLM Benchmark. *The Thirteenth International Conference on Learning Representations*.
- Wun, L.-M., & Pearn, W. L. (1991). Assessing the statistical characteristics of the mean absolute error or forecasting. *International Journal of Forecasting*, 7(3), 335–337.
- Zerback, T., Koch, T., & Krämer, B. (2015). Thinking of others: Effects of implicit and explicit media cues on climate of opinion perceptions. *Journalism & Mass Communication Quarterly*, 92(2), 421–443.