

## Werkpakket 5 / mei 2020: Social Media Monitoring, Fase 1: Exploratieve analyse van Twitter data

(10)(2e) (10)(2e), RIVM; (10)(2e), (10)(2e) LIACS  
(Universiteit Leiden); (10)(2e), CLST (Radboud  
Universiteit); (10)(2e), (10)(2e) KNAW Meertens Instituut; (10)(2e)  
(10)(2e) Netherlands eScience Center

### Fact Sheet: belangrijkste observaties

*Algemene indruk:* sociale media wordt massaal gebruikt voor het delen van ervaringen en meningen. Veel verspreiders van meningen hebben politieke of activistische motieven. Het nieuws en de overheid wordt nauwgezet gevolgd en kritisch becommentarieerd. Er is veel minder plaats voor positief gestemde tweets dan in normale tijden.

Op basis van een *algemeen corona-gerelateerde Nederlandstalige Twitter-dataset* (circa 400 duizend tweets van februari-april 2020) zijn de volgende opvallende zaken geobserveerd:

- in vergelijking met andere perioden is de algemene positiviteit in de corona-tijd, gemeten met sentimentanalyse, ruwweg gehalveerd. De inhoud van positieve berichten is meer op hoop en toekomst gericht ("ik hoop dat het snel beter gaat") dan het normale ("het gaat best goed");
- De algemene negativiteit is ongeveer even groot als in andere perioden, maar we zien binnen de negatieve berichten een inhoudelijke verschuiving naar uitingen van angst en referenties naar de dood;
- een relatief kleine groep Twitter-gebruikers genereert en verspreidt een groot deel van de berichten (42% van alle tweets zijn retweets), meer dan in andere perioden. Deze groep is grotendeels rechts-populistisch en activistisch;
- De aandacht ging in februari uit naar andere landen en naar de nieuwe hygiëne-maatregelen; in maart werden Nederland, thuisblijven en ziekenhuisopnamen de hoofdonderwerpen, en in april verschoof de aandacht naar de politiek en de effectiviteit van maatregelen;
- Er wordt veel gerefereerd naar nieuwsberichten, talkshows en persconferenties van het kabinet; dat laatste is ongewoon.

Op basis van een *specifiek mondkapjes-gerelateerde Twitter dataset* (circa 45 duizend tweets van 27 april tot 6 mei) zijn de volgende opvallende zaken geobserveerd:

- De tweets over mondkapjes zijn overwegend negatief: 80.8%;
- Het onderwerp 'mondkapjes' wordt veel in verband gebracht met de effectiviteit van de regering; veel aandacht gaat uit naar specifieke narratieven (rond kappers, zelf mondkapjes maken, en KLM die reizigers verplicht zelf mondkapjes te regelen); argumenten voor en tegen het gebruik van mondkapjes worden uitvoerig gedeeld en besproken;
- Er worden relatief veel berichten geschreven en verspreid door gebruikers die zichzelf affiliëren aan rechtse bewegingen.

### Alle observaties op een rij

In circa 400 duizend COVID-19 gerelateerde Twitterdata waarin ofwel RIVM, Hugo de Jonge of de Minister President worden genoemd, uit de periode februari tot en met april 2020 vallen de volgende zaken op:

1. De meest prominente onderwerpen (bepaald met topic modeling, O'Callaghan *et al.* 2015) zijn: het aantal patiënten dat is overleden of opgenomen; het RIVM; het aantal besmettingen; corona maatregelen; de stand van zaken; grenzen sluiten; onderzoek naar spreiding corona; en hygiënemaatregelen tegen corona.
2. De meest voorkomende hashtags relateren aan coronamaatregelen, overheidscommunicatie over corona (in andere perioden wordt over het algemeen minder gereageerd op overheidscommunicatie), solidariteit (e.g. #alleensamen) en rechtse bewegingen (e.g. #PVV).
3. Er worden ook vaak nieuwsbronnen (NOS), politici en andere gezondheidsorganisaties dan RIVM (WHO, VWS) vermeld.
4. Een relatief kleine groep twitteraars uit rechts-populistische hoek genereert relatief veel berichten, meer dan in andere perioden. 42% van alle corona-gerelateerde tweets zijn retweets.
5. Berichten gingen in februari voornamelijk over andere landen (zoals China en Italië) en handen wassen. Daarna lag de focus meer op Nederland, thuisblijven en ziekenhuisopnames. In de laatste weken wordt er meer getwitterd over de politieke besluitvorming rondom corona.
6. Veel minder positieve berichten dan in Twitterdata van voor de corona-tijd. Negatieve uitingen zijn heftiger (verwijzen naar dood of angst) terwijl positieve uitingen vaker te maken hebben met veiligheid.
7. Vergeleken met Twitterdata van voor de corona-tijd wordt er meer getweet over de dood, het menselijk lichaam (e.g. pijn), mensen, en verboden, en minder over religie, seks, tv, muziek, eten en slaap.

In recente Twitterdata over mondkapjes, circa 50 duizend tweets uit de periode 27 april – 6 mei, vallen de volgende zaken op:

1. Er worden vaak nieuwsbronnen (RTL Nieuws, Telegraaf, NOS, Volkskrant) en politici vermeld: zowel ministers (naast de minister-president ook in deze periode de relevante ministers, (10)/(2e) of mjrjn) als rechtse politici (geertwilderspvv, thierrybaudet).
2. In de meest voorkomende hashtags vallen talkshows (#jinek en #op1) en KLM (#klm) op.
3. De berichten met de meeste retweets binnen dit onderwerp komen overwegend van gebruikers die zichzelf affilieren aan rechtse bewegingen.
4. Tweets over dit onderwerp zijn overwegend (80.8%) negatief volgens een sentimentanalyse.
5. Een gerichte analyse op het onderwerp mondkapjes wijst uit dat dit onderwerp veel in verband wordt gebracht met de effectiviteit van de regering; veel aandacht gaat uit naar specifieke narratieven (rond kappers, zelf mondkapjes maken, en KLM die reizigers verplicht zelf mondkapjes te regelen); argumenten voor en tegen het gebruik van mondkapjes worden uitvoerig gedeeld en besproken.

### Plannen voor onderzoek in de komende periode

Momenteel wordt er een deel van de Mondkapjes-data geannoteerd voor sentiment (Negatief, Positief en Neutraal) en voor steun voor de maatregel (Support, Neutral en Deny) door medewerkers van het RIVM. Hierdoor wordt het mogelijk om onderzoek te doen naar de steun die op Twitter wordt geuit voor deze maatregel. Deze data zal ook in de toekomst worden uitgebreid tot op heden. Het huidige onderzoek met de Relevancer tool (Hürriyetoglu et al, 2016), en de clusters die de tool oplevert, wijst uit dat het mogelijk is om verder de gebruikte argumenten te onderzoeken en onder te verdelen in onderwerpen, zoals we dat ook deden in een eerder samenwerkingsproject (Kunneman et al, 2020). Wij kijken momenteel ook naar andere bronnen zoals Reddit, Nu.nl en Youtube (met accounts als die van Robert [Jensen](#) en Janet [Ossebaard](#)) om ook de steun in die bronnen te kunnen meten en ons niet te beperken tot de groep die op Twitter zit. De invloed van rechtse bewegingen op Twitter is opvallend, en om te onderzoeken hoe het bredere publiek over maatregelen denkt zullen we hierop alert moeten zijn, en bijvoorbeeld in bepaalde analyses retweets verwijderen om zo de invloed van een aantal populaire tweets uit deze hoek op onze resultaten te verminderen.

We willen in de toekomst ook bekijken of deze invloed verschilt met Twitter data uit andere periodes.

Andere beleidsvraagstukken en maatregelen kunnen naar keuze genomen worden als startpunt voor toekomstige onderzoek; de eerste resultaten behaald rondom mondkapjes zijn illustratief voor de stappen die we vervolgens zetten, met sentimentanalyse, topic modeling, en annotatie van specifieke aspecten van gedrag en draagvlak (zoals het geven van argumenten 'voor' en 'tegen'). Met deze verdieping zal meer inzicht ontstaan in de redenen die mensen geven voor het wel of niet opvolgen en draagvlak van de maatregelen.

## Ondersteuning van observaties

COVID-19 gerelateerde Twitterdata waarin ofwel RIVM, Hugo de Jonge of de Minister President worden genoemd

### Data in het kort:

- circa 400k tweets tussen Feb-Apr 2020.
- Zoekterm: (rivm OR minpres OR (10)(2e)) AND (corona OR covid)

**Observatie 1: De meest genoemde onderwerpen zijn: het aantal patiënten dat is overleden of opgenomen; het RIVM; het aantal besmettingen; corona maatregelen; de stand van zaken; grenzen sluiten; onderzoek naar spreiding corona; en hygiënemaatregelen tegen corona.**

We hebben met automatische methodes (O'Callaghan *et al.* 2015) de data geclusterd in 9 onderwerpen. Deze onderwerpen zijn handmatig gelabeld aan de hand van de top 10 woorden die met het onderwerp geassocieerd worden. Deze zijn weergegeven in de onderstaande tabel. Hierbij zijn ook het relatieve gewichten van woorden in het cluster meegenomen..

**Tabel 1: Top 10 woorden geassocieerd met de 9 onderwerpen in de data**

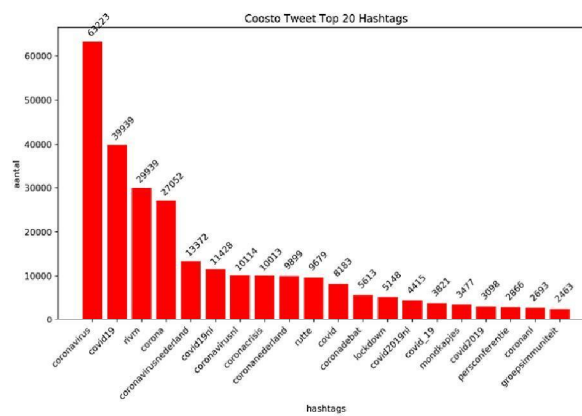
|           | Handmatig toegekend label               | Top 10 woorden in het cluster  |
|-----------|---|--|
| Cluster 1 | Aantal patiënten overleden of opgenomen | mensen totaal gisteren overleden positief gemeld getest opgenomen ziekenhuis                   |
| Cluster 2 | RIVM                                    | rivm corona mensen virus mondkapjes rutte gaat gaan besmet goed                                |
| Cluster 3 | Aantal besmettingen                     | aantal nieuwe patiënten besmettingen rivm cijfers doden coronavirus meldt ziekenhuisopnames    |
| Cluster 4 | Corona maatregelen                      | maatregelen coronavirus aanpak kabinet vandaag aangescherpt gelden overzicht covid19 maart     |
| Cluster 5 | Stand van zaken                         | stand zaken toespraak tekst lees terug coronavirus live persconferentie rutte                  |
| Cluster 6 | Corona klachten melden                  | meld klachten meedoen wekelijks help starten koorts onderzoek hoesten mail                     |
| Cluster 7 | Grenzen sluiten                         | lockdown nugrenzendicht hashtag nexit herhalen tweet ga iedere dag mee                         |
| Cluster 8 | Onderzoek naar spreiding corona         | patiënt ggd informatie contactonderzoek covid19 gestart coronavirus vanavond labonderzoek ncov |

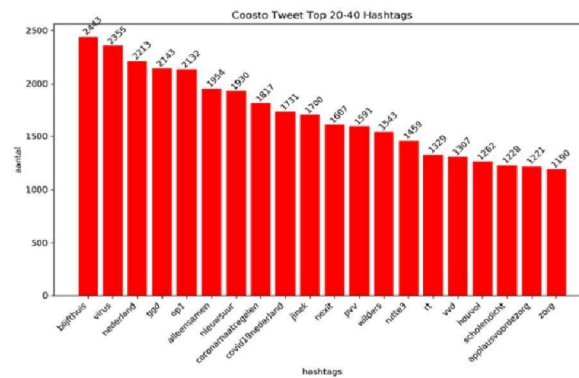
|           |                                  |  |
|-----------|----------------------------------|--|
| Cluster 9 | Hygiëne maatregelen tegen corona | handen wassen elleboog zakdoekjes binnenkant<br>papieren schudden belangrijk afstand<br>hygiënemaatregelen |
|-----------|----------------------------------|--|

**Observatie 2: De meest voorkomende hashtags relateren aan corona en corona maatregelen, overheidscommunicatie over corona, solidariteit (e.g. #alleensamen) en rechtse bewegingen (e.g. #PVV).**

Hashtags kunnen een indicatie geven van de onderwerpen waarvan de auteur van de tweet zelf denkt dat hun tweet mee te maken heeft (Bruns et al. 2013). Wij hebben hier gekozen om naar de top 40 te kijken omdat de top 20 niet informatief was: deze bestaat voornamelijk uit hashtags die deel uitmaakte van de zoekterm en verschillende hashtags voor corona (e.g. #covid19, #coronavirus, #covid, #corona, #coronavirusnl etc.).

Grafiek 1: Top 20 meest voorkomende hashtags (boven) en Top 20-40 meest voorkomende hashtags (onder)

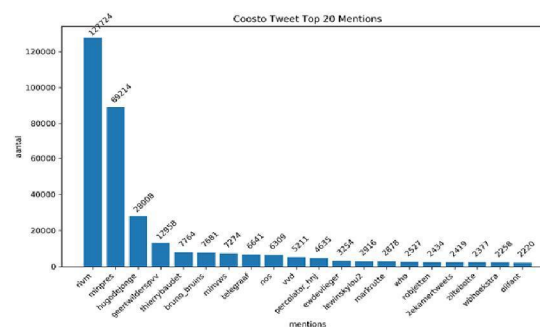




**Observatie 3: Er worden ook vaak nieuwsbronnen (NOS), politici en andere gezondheidsorganisaties dan RIVM (WHO, VWS) vermeld.**

Een vermelding is als een Twitter gebruiker met een @ wordt gelinkt aan een bericht.

Grafiek 2: Top 20 meest voorkomende vermeldingen (ofwel mentions) op Twitter.



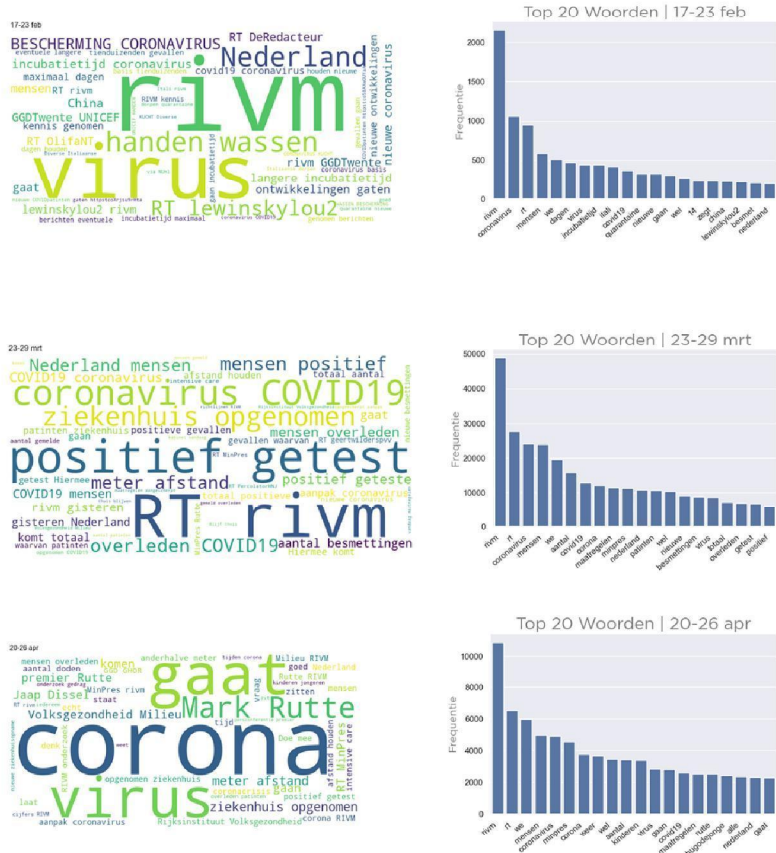
**Observatie 4: Een relatief kleine groep twitteraars uit rechts-populistische hoek genereert relatief veel berichten.**

In Grafiek 2 zien we dat meerdere rechtse politici (geertwilderspvd, thierrybaudet) genoemd worden, maar ook drie twitter accounts die zichzelf ofwel identificeren met rechtse politieke bewegingen in hun Twitter bio ofwel bekend staan om hun rechtse politieke ideeën: (10)(2e) /ergelijkbare observaties zijn al gemaakt in door reguliere media in andere landen (BBC,2020; New York Times, 2020) en ook in preprints van academische artikelen (Ferrara, 2020).

**Observatie 5: Berichten gingen in februari voornamelijk over andere landen (zoals China en Italië) en handen wassen. Daarna lag de focus meer op Nederland, thuisblijven en ziekenhuisopnames. In de laatste weken wordt er meer getwitterd over de politieke besluitvorming rondom corona.**

Om te bepalen over tijd welke termen het meest besproken hebben, hebben we voor elke week gekeken naar de meest voorkomende termen. In Grafiek 3 presenteren we voorbeelden van een aantal weken in het begin, midden en richting het einde van de data.

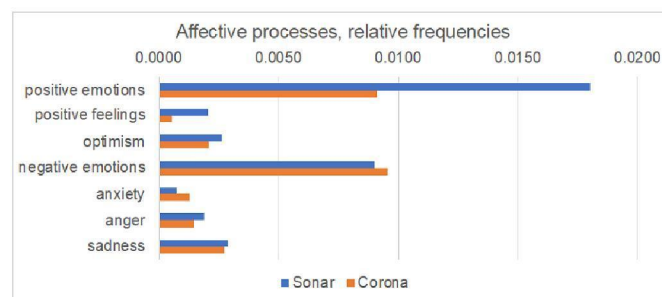
*Grafiek 3: Meest voorkomende termen in de Twitter data in (top) week 8 van 17-23 Feb, (midden) week 13 van 23-29 Mrt en (onder) week 17 van 20-26 April.*



**Observatie 6:** Er is veel minder positieve Twitterdata dan in voorgaande periodes. Negatieve uitingen zijn heftiger (verwijzen naar dood of paniek) terwijl positieve uitingen vaker te maken hebben met veiligheid.

Om de vergelijking te maken met voorgaande periodes, hebben we data gebruikt van het SoNaR nieuwe media corpus (Versie 1), een corpus van ongeveer 1.5M tweets. Om de positieve en negatieve emoties te meten maakte we gebruik van de Nederlandse versie van de Linguistic Inquiry and Word Count (LIWC) (Boot *et al.*, 2017). Eerdere onderzoek heeft uitgewezen dat de LIWC goed correleert met uitdrukkingen van gevoelens rondom de corona crisis in korte teksten (Kleinberg *et al.*, 2020).

Grafiek 4: Relatieve frequentie van emoties in Corona Corpus vergeleken met reguliere Twitter data (het Sonar Corpus, 2013)



Tabel 2: Meeste frequente woorden die hoorde bij positieve en negatieve emoties in reguliere Twitter data (Sonar Corpus) en onze data

|                          | Meeste frequente woorden in deze categorie in het Sonar Corpus | Meeste frequente woorden in deze categorie in ons corpus                       |
|--------------------------|--|--|
| <b>Positieve emoties</b> | goed leuk lekker mooi dank beter gelukkig hoop lol heerlijk    | goed positief beter hoop verwacht controle vertrouwen lekker veilig veiligheid |
| <b>Negatieve emoties</b> | alleen erg jammer helaas druk geleden last mis gek slecht      | alleen doden geleden erg druk helaas paniek bang leden gevaar                  |

**Observatie 7:** Vergeleken met Twitterdata uit voorafgaande periodes wordt er meer getweet over de dood, het menselijk lichaam (e.g. pijn), mensen, en verboden, en minder over religie, seks, tv, muziek, eten en slaap.

De LIWC is ontwikkeld door sociaal psychologen met het idee dat taalgebruik psychologische factoren (bv. motivatoren) reflecteert. De LIWC bevat naast emoties ook woordenlijsten voor andere psychologische factoren (voor een uitgebreide lijst van alle categorieën zie: Boot *et al.*, 2017). We tellen een categorie als meer prevalent dan normaal als de relatieve frequentie van woorden in die categorie minstens 2 keer zo hoog

is als in het Sonar Corpus. We tellen een categorie als minder prevalent dan normaal als de relatieve frequentie van woorden in die categorie minstens 2 keer zo klein is als in het Sonar Corpus.

Het verschil in de verzamelingen van meest frequente woorden die een positieve emotie uitdrukken in de twee corpora kan geïnterpreteerd worden als een contrast tussen “het gaat best goed” in het Sonar Corpus en een combinatie van “hoop” en “vertrouwen” op een goede afloop en veiligheid nu. Een soortgelijk verschil is te vinden in de woorden die negatieve emoties uitdrukken; van “jammer” is verschoven naar veel acutere negatieve emoties.

## Twitterdata over mondkapjes, 27 april – 6 mei 2020

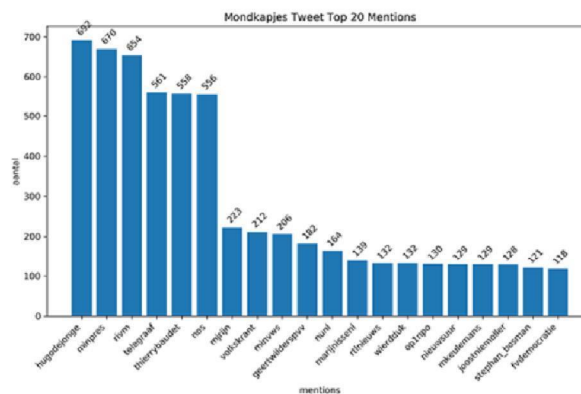
**Data in het kort:**

- circa 45k tweets tussen 27 Apr - 6 Mei.
- circa 18k unieke tweets
- Zoekterm: mondkapjes OR mondkapje

**Observatie 1:** Er worden vaak nieuwsbronnen (RTL Nieuws, Telegraaf, NOS, Volkskrant) en politici vermeld: zowel ministers (e.g. [\(10\)\(2e\)](#)) of mijrijn) als rechtse politici (geertwilderspvv, thierrybaudet).

Een vermelding is als een Twitter gebruiker met een @ wordt gelinkt aan een bericht.

Grafiek 5: De Twitter gebruikers met de meeste vermeldingen (@) in de Twitter data over mondkapjes.

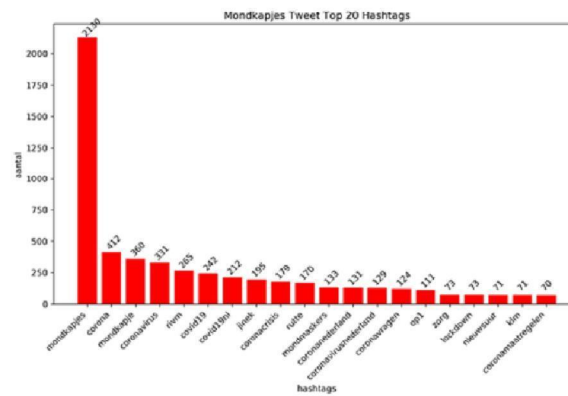


**Observatie 2:** In de meest voorkomende hashtags vallen talkshows (#jinek en #op1) en KLM (#klm) op

Van de 20 meest voorkomende hashtags refereren een aantal direct aan corona (#corona, #coronavirus, #coronaland etc.) of coronamaatregelen (#mondkapje, #lockdown, #coronamaatregelen #mondmaskers). Verder wordt het RIVM (#rivm), de minister president (#rutte), de zorg (#zorg) en nieuwsuur

(#nieuwsuur) nog genoemd. De talkshows (Jinek en Op1) en het bedrijf KLM vielen op omdat ze niet evenveel voorkomend waren in onze andere corona-gerelateerde Twitter data.

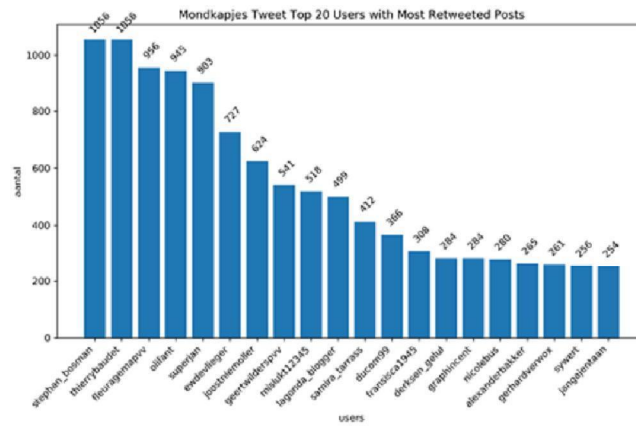
Grafiek 6: De 20 meest voorkomende hashtags in de Twitter data over mondkapjes



**Observatie 3: De berichten met de meeste retweets binnen dit onderwerp komen overwegend van gebruikers die zichzelf affiliëren aan rechtse bewegingen.**

In de grafiek hieronder staan de 20 gebruikers met de berichten met de meeste retweets. Wij observeren hier veel gebruikers die zich affiliëren met rechtse politieke bewegingen, namelijk: (10)(2e) Het Twitter bio diende als basis voor onze handmatige categorisering..

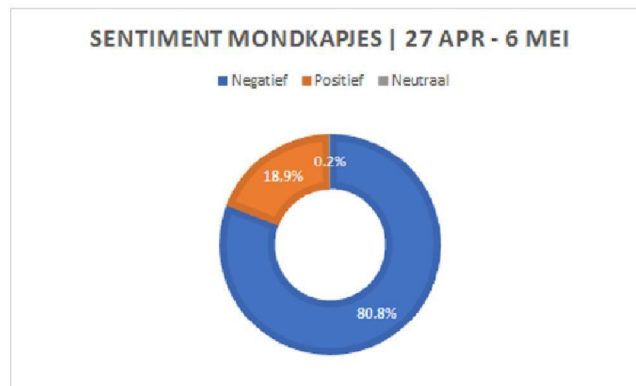
Grafiek 7: De top 20 gebruikers met berichten in de Twitter data over mondkapjes met de meeste retweets



**Observatie 4: Tweets over dit onderwerp zijn overwegend (80.8%) negatief volgens een sentimentanalyse.**

Sentimentanalyse classificeert tweets in drie categorieën: negatief, neutraal en positief. We hebben een classificatiemodel getraind met BERTje (de Vries *et al.* 2019) op 1318 restaurantreviews en 1018 smartphonereviews van de Sem-Eval 2016 Task 5 dataset (Pontiki *et al.* 2016). Deze analyse is gedaan op alleen de unieke tweets.

Grafiek 8: Sentiment analyse van Twitter data over mondkapjes



**Observatie 5: Een gerichte analyse op het onderwerp mondkapjes wijst uit dat dit onderwerp veel in verband wordt gebracht met de effectiviteit van de regering; veel aandacht gaat uit naar specifieke narratieven (rond kappers, zelf mondkapjes maken, en KLM die reizigers verplicht zelf mondkapjes te regelen); argumenten voor en tegen het gebruik van mondkapjes worden uitvoerig gedeeld en besproken.**

De gerichte analyse van narratieven en argumenten werd gedaan met *Relevancer* (Hürriyetoğlu et al. 2016), een data-gedreven 'bottom-up' tool voor het annoteren van tweets door experts, met hulp van machine-learningmethoden. De eerste stap wordt gezet door een unsupervised learning-methode, k-means clustering. Daarna labelen menselijke experts de clusters die door de clusteringmethode zijn gevonden in de gewenste categoriën (zoals 'voor' of 'tegen'). Relevancer werkt vervolgens iteratief; na labeling worden automatische machine learning methoden getraind om meer data automatisch te labelen op basis van de door mensen gelabelde voorbeelden.

Voor de huidige analyse zijn alleen de unieke tweets gebruikt; retweets en (bijna-) duplicate tweets zijn verwijderd.

Tabel 3 toont een tweetal clusters die automatisch zijn gevonden (gekozen uit een verzameling van enkele honderden clusters met een vergelijkbare coherentie), waarin links een bekend argument terugkomt voor mondkapjes (door het buurland Duitsland erbij te halen), en waarbij de tweets over KLM ingaan op een narratief dat veel stof doet opwaaien. De clusters zijn gekozen als illustratieve voorbeelden; in het onderzoek van de komende perioden worden alle clusters geannoteerd op onder meer relevantie, relatie tot een maatregel, en houding ten opzichte van de maatregel die uit de tweets in het cluster spreekt.

Tabel 3: Voorbeelden van tweets in twee van de gevonden clusters

| Mondkapjes in Duitsland  | Mondkapjes en KLM  |
|--|--|
| ready voor een (voetbal)ronde Duitsland #mondkapjes  | haha miljardensteun maar wel je eigen mondkapje meebrengen   |
| Duitsland. het land van mooie auto's en mondkapjes   | hoeveel heeft die #klm ook al weer gekregen van ons belastinggeld? ze durven wel te eisen dat je eigen mondkapje mee neemt als je met ze meevliegt |
| in Duitsland afstand en verplicht mondkapje!!!!!! werkt perfect  | neem de trein of ga met je eigen auto zonder mondkapje. #klm   |
| aan de andere kant lijken de mondkapjes wel iets te doen, kijk naar Tsjechië en Jena, Duitsland  | verwijder eerst uw eigen mondkapje voordat u het mondkapje van uw kind verwijdert  |
| gister ook in Duitsland boodschappen gedaan. iedereen (wij natuurlijk ook) had een mondkapje op. voelde een stuk veiliger dan hier in de supermarkten. heb het idee dat het in Duitsland een stuk serieuzer gaat |  |
| er moeten gewoon meer doden vallen. hij kan niet aan mondkapjes komen terwijl België, Duitsland, Frankrijk en Noorwegen maar op wel. in ons eigen land worden ze ook gemaakt, maar                               |  |

|  |  |
|--|--|
| die nigeert hij  |  |
| in duitsland loopt heftige discussie over bewezen toegevoegde waarde van mondkapjes. ook veel experts daar die nut in twijfel trekken  |  |
| dus kabinet raadt ons af mondkapjes te dragen, omdat we er niet mee om zouden kunnen gaan.. #coronavirusnl #mondkapje * terwijl duitsland en belgie, *op basis van hetzelfde rivm advies* juist wel mondkapjes adviseren.. * |  |

## Referenties

- BBC (2020). Coronavirus: Far-right spreads Covid-19 'infodemic' on Facebook. <https://www.bbc.com/news/technology-52490430>
- Boot, P., Zijlstra, H., & Geenen, R. (2017). The Dutch translation of the Linguistic Inquiry and Word Count (LIWC) 2007 dictionary. *Dutch Journal of Applied Linguistics*, 6(1), 65-76
- Bruns, A., & Stieglitz, S. (2013). Towards more systematic Twitter analysis: metrics for tweeting activities. *International Journal of Social Research Methodology*, 16(2), 91-108. <https://doi.org/10.1080/13645579.2012.756095>
- Ferrara, E. (2020). #COVID-19 ON TWITTER: BOTS, CONSPIRACIES, AND SOCIAL MEDIA ACTIVISM. Preprint on ArXiv. <https://arxiv.org/abs/2004.09531>
- Hürriyetoğlu, A., Gudehus, C., van den Bosch, A., & Oostdijk, N. (2016). Relevancer: Finding and labeling relevant information in tweet collections. In *Social Informatics: Proceedings of the 8th International Conference on Social Informatics*, p. 210-224. Lecture Notes in Computer Science, Vol. 10047. Berlin: Springer.
- Kleinberg, B., van der Vegt, I., & Mozes, M. (2020). *Measuring Emotions in the COVID-19 Real World Worry Dataset*. Retrieved from <http://arxiv.org/abs/2004.04225>
- Kunneman, F., Lambooi, M., Wong, (10)(2e), A., & Mollema, L. (2020). Monitoring stance towards vaccination in twitter messages. *BMC Medical Informatics and Decision Making*, 20(1), 1-14. <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1046-y>
- New York Times (2020). How Right-Wing Pundits are Covering Coronavirus. <https://www.nytimes.com/2020/03/11/us/politics/coronavirus-conservative-media.html>
- O'Callaghan, D., Greene, D., Carthy, J., & Cunningham, P. (2015). An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, 42(13), 5645-5657. <https://doi.org/10.1016/j.eswa.2015.02.055>
- Pontiki, M., Galanis, D., Papageorgiou, H., Androutsopoulos, I., Manandhar, S., Al-Smadi, M., ... & Hoste, V. (2016, June). Semeval-2016 task 5: Aspect based sentiment analysis. In *10th International Workshop on Semantic Evaluation (SemEval 2016)*. Data available from: <http://alt.qcri.org/semeval2016/task5/index.php?id=data-and-tools>
- SoNaR Nieuwe Media Corpus (Version 1.0) (2013) [Data set]. Available at the Dutch Language Institute: <http://hdl.handle.net/10032/tm-a2-k3>
- de Vries, W., van Cranenburgh, A., Bisazza, A., Caselli, T., van Noord, G., & Nissim, M. (2019). BERTje: A Dutch BERT Model. *arXiv preprint arXiv:1912.09582*.

