

Algoritmes in het nieuws: appreciatie of aversie?

Glen Joris



UNIVERSITEIT
GENT

Colofon

Titel: Algoritmes in het nieuws: appreciatie of aversie?

Auteur: Glen Joris (imec-mict-UGent)

Publicatiedatum: 1 maart 2021

Contact:

Glen Joris
Glen.Joris@UGent.be

Onderzoeksgroep imec-mict-UGent
Korte Meer 11, 9000 Gent

Het gerapporteerde onderzoek verliep binnen het lopende **NewsDNA**-project, een vierjarig onderzoekproject dat zich tot doel stelt een algoritme te ontwikkelen dat nieuwsdiversiteit als uitgangspunt voor gepersonaliseerde nieuws-aanbevelingen neemt. Meer info: <https://www.ugent.be/mict/en/research/newsdna>

Inhoud

1. Inleiding	4
2. Methodologie en steekproef	7
3. Resultaten	9
3.1. Descriptieve analyse	9
3.2 Clusteranalyse	10
3.3 Hiërarchische regressieanalyse	11
4. Conclusie en aanbevelingen	15
Referenties	17
Bijlage	18

1. Inleiding

De afgelopen jaren hebben nieuwsalgoritmes een sterke intrede gemaakt in de nieuwsmediasector. Waar vroeger nieuwsalgoritmes hoofdzakelijk werden gebruikt door grote technologiebedrijven zoals Facebook en Google, maken nu ook steeds meer nieuwsorganisaties de overstap naar automatische beslissingssystemen om hun nieuwsaanbod te ordenen of gepersonaliseerde aanbevelingen te maken.¹ Door deze evolutie schakelen nieuwsorganisaties stap per stap over naar een model waarbij automatisering en personalisatie essentiële onderdelen worden van het nieuwsbelevingsproces.

Die evolutie gaat echter ook gepaard met verschillende risico's. Al sinds 2011 waarschuwen verschillende academici dat nieuwsalgoritmes te sterk gericht zijn op commerciële criteria zoals clicks en likes, waardoor burgers mogelijk te weinig worden blootgesteld aan meningen of onderwerpen die minder gangbaar zijn of minder aansluiten bij zijn of haar standpunten. In populaire termen wordt naar dit scenario ook wel verwezen met de term 'filter bubbel', of 'informatiebubbel', waarin nieuwsgebruikers zouden vastzitten.²

Hoewel empirisch onderzoek nog geen bevestiging heeft gevonden voor deze theorie³, leeft bij academici wel het idee om algoritmes in te zetten voor meer democratische doeleinden.⁴ Zo zouden nieuwsalgoritmes ook gebruikt kunnen worden om de *geconsumeerde diversiteit*, of de mate waarin iemand wordt blootgesteld aan een divers nieuwsaanbod, te stimuleren en te promoten.⁵ Nieuwsartikels zouden dan niet meer aanbevolen worden op basis van het aantal clicks of likes dat een artikel heeft, maar op basis van hoe divers het leesgedrag van een nieuwsgebruiker is. Dit idee sluit ook aan bij het idee van zogenaamde *public service algorithms* die in verschillende Europese landen hun opmars maken om maatschappelijke waarden te bewaren en te beschermen.⁶

Ondanks het potentieel van dit idee is er in het onderzoeksveld nog maar weinig geweten over hoe het publiek nieuwsalgoritmes en de principes die hieronder schuilgaan, beoordelen. Bestaand onderzoek focust zich voornamelijk op de perceptieverschillen tussen mens en algoritme⁷, terwijl er net verschillende algoritmes bestaan en deze sterk kunnen verschillen naargelang de principes waarop ze steunen. Het is bijgevolg **onduidelijk welk soort nieuwsalgoritme de voorkeur geniet én of een diversiteitsalgoritme de interesse kan opwekken van het publiek.**

¹ Bodó (2019); Thurman & Schifferes (2012)

² Pariser (2011)

³ Bruns (2019)

⁴ Helberger (2019)

⁵ Joris et al. (2019)

⁶ Jones & Jones (2019); Van den Bulck & Moe (2018)

⁷ Araujo et al. (2020); Thurman et al. (2019)

Met deze probleemstelling in het achterhoofd tracht deze studie⁸ in de eerste plaats te onderzoeken of en welk nieuwsalgoritme de voorkeur geniet van het publiek. We focussen ons daarbij op drie soorten nieuwsalgoritmes die elk andere principes hanteren om nieuwsartikelen op een automatische manier te selecteren:

- (a) op basis van het leesgedrag en de interesses van een persoon (i.e., content-based algoritme),
- (b) op basis van wat anderen lezen en interessant vinden (i.e., collaborative algoritme), of
- (c) op basis van de geconsumeerde diversiteit (i.e., diversity-based algoritme).

De eerste twee algoritmes streven naar gelijkens door respectievelijk te kijken naar het voorgaand nieuwsconsumptiegedrag van een persoon of het consumptiegedrag van iemand met een gelijkaardig profiel. Het derde algoritme, in tegenstelling, kijkt voornamelijk naar hoe nieuwsartikels verschillen van elkaar en hoe een persoon blootgesteld kan worden aan een zo divers mogelijk nieuwsaanbod.

Door te focussen op een diversiteitsalgoritme, tracht dit onderzoek in tweede instantie ook in kaart te brengen wat het adoptiepotentieel is dat achter maatschappelijke algoritmes schuilgaat. De afgelopen jaren hebben verschillende publieke omroepen in Europa onderzocht hoe maatschappelijke nieuwsalgoritmes een meerwaarde kunnen bieden voor hun activiteiten. Toch blijven maatschappelijke nieuwsalgoritmes eerder een uitzondering op de regel. Deze studie hoopt daar verandering in te brengen door waardevolle inzichten te bieden in het adoptiepotentieel van een diversiteitsalgoritme, alsook concrete suggesties aan te bieden hoe ontwikkelaars maatschappelijke nieuwsalgoritmes kunnen ontwerpen.

⁸ Dank aan Frederik De Grove, Kristin Van Damme en Lieven De Marez



2. Methodologie en steekproef

Om deze onderzoeksdoelstellingen te verwezenlijken, is het belangrijk om algemene uitspraken te kunnen doen over een grotere bevolkingsgroepen. Onderzoeksgroep imec-mict-UGent verspreidde daarom een **grootschalige survey onder 4995 bewoners in de Stad Gent**. Er werd gekozen voor de Stad Gent omwille van praktische overwegingen (i.e., de mogelijkheid om huisbezoeken te kunnen uitvoeren) en juridische overwegingen (i.e., beperkte toegang tot bevolkingsregister). De uitspraken die uit deze studie voortvloeien zijn bijgevolg niet te veralgemenen tot de Vlaamse bevolking, maar geven niettemin een indicatie van de verhoudingen in Stad Gent zelf.

Om een doorsnede van de Gentse bevolking te bevragen werd, in samenwerking met de Stad Gent, een impliciet **gestratificeerde steekproef** getrokken uit het bevolkingsregister. Er werd daarbij rekening gehouden met vier sociodemografische variabelen: leeftijd, geslacht, nationaliteit en buurt. Er werden ook drie exclusiecriteria vooropgesteld in de steekproeftrekking: personen mochten (1) niet jonger zijn dan 18 jaar, (2) niet wonen in een institutionele setting (zoals bijvoorbeeld een rusthuis of gevangenis) en (3) niet recent gecontacteerd zijn geweest door andere onderzoekers van de Universiteit Gent.

Om de inwoners van Gent aan te sporen om deel te nemen, werden er in twee gepersonaliseerde uitnodigingsbrieven verzonden. De response ratio van deze uitnodigingsronde bedroeg 20,74% (of 1.036 respondenten) voor data cleaning en 18,88% (of **943 respondenten**) na data cleaning. In Tabel 1 kan je de absolute en relatieve verdeling terugvinden van de respondenten die hebben deelgenomen.

Tabel 1. Absolute en relatieve verdeling van de respondenten die hebben deelgenomen aan de survey (n = 943)

	Buitenlandse nationaliteit			Belgische nationaliteit			Totaal			
	M	V	Total	M	V	Total	M	V	Ongekend	Totaal
18-24	7 (0.74%)	7 (0.74%)	14 (1.48%)	22 (2.33%)	41 (4.35%)	63 (6.68%)	29 (3.08%)	48 (5.09%)	0 (0.00%)	77 (8.17%)
25-34	14 (1.48%)	15 (1.59%)	29 (3.08%)	64 (6.79%)	80 (8.48%)	144 (15.27%)	78 (8.27%)	95 (10.07%)	0 (0.00%)	173 (18.35%)
35-44	7 (0.74%)	9 (0.95%)	16 (1.70%)	57 (6.04%)	55 (5.83%)	112 (11.88%)	64 (6.79%)	64 (6.79%)	1 (0.11%)	129 (13.68%)
45-54	9 (0.95%)	6 (0.64%)	15 (1.59%)	56 (5.94%)	59 (6.26%)	115 (12.20%)	65 (6.89%)	65 (6.89%)	0 (0.00%)	130 (13.79%)
55-64	5 (0.53%)	4 (0.42%)	9 (0.95%)	90 (9.54%)	92 (9.76%)	182 (19.30%)	95 (10.07%)	96 (10.18%)	0 (0.00%)	191 (20.25%)
65-74	0 (0.00%)	5 (0.53%)	5 (0.53%)	84 (8.91%)	86 (9.12%)	170 (18.03%)	84 (8.91%)	91 (9.65%)	0 (0.00%)	175 (18.56%)
75-84	2 (0.21%)	0 (0.00%)	2 (0.21%)	21 (2.23%)	29 (3.08%)	50 (5.30%)	23 (2.44%)	29 (3.08%)	1 (0.11%)	53 (5.62%)
85+	0 (0.00%)	0 (0.00%)	0 (0.00%)	9 (0.95%)	6 (0.64%)	15 (1.59%)	9 (0.95%)	6 (0.64%)	0 (0.00%)	15 (1.59%)
Totaal	44 (4.67%)	46 (4.88%)	90 (9.54%)	403 (42.74%)	448 (47.51)	851 (90.24%)	447 (47.40%)	494 (52.39%)	2 (0.21%)	943 (100%)

Om een mogelijke scheef trekking in de steekproef te beperken, werden er tijdens het dataverzamelingsproces ook huisbezoeken in verschillende wijken van de Stad Gent uitgevoerd. Er werd hierbij vooral gefocust op het bereiken van Gentenaren die ondervertegenwoordigd waren in de antwoorden. Dit ging voornamelijk om Gentenaren met een buitenlandse nationaliteit. Tabel 2 geeft een overzicht van deze huisbezoeken.

Tabel 2. Resultaten huisbezoeken survey

	Aantal respondenten
Deelgenomen	17 (11.18%)
Niet aanwezig	74 (48.68%)
Weigering	61 (40.13%)
Totaal	152 (100.00%)

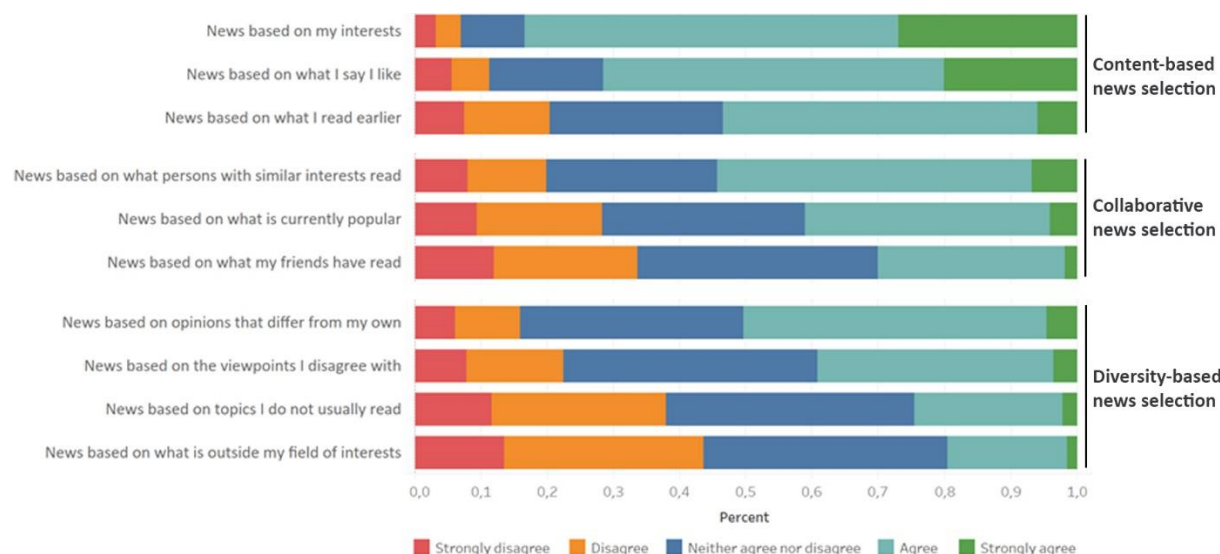
Noot: weigering omvat beperkte taalkennis, te lage bereidheid tot deelname en ongeldig/gewijzigd adres.

3. Resultaten

3.1. Descriptieve analyse

Om te begrijpen welke nieuwsalgoritme de voorkeur geniet, voerden we een descriptieve analyse uit op alle items die werden bevraagd in de survey. Figuur 1 biedt een beknopt overzicht van deze analyse en laat zien dat het publiek doorgaans een **grotere voorkeur heeft voor de nieuwsselectieprincipes die onder *content-based* nieuwsalgoritmes schuilgaan** (items 1-3) dan voor principes die behoren tot *collaborative* (items 4-6) of *diversity-based* nieuwsalgoritmes (items 7-10). Vooral het nieuwsselectieprincipes waarin de interesses van een gebruiker (item 1) of de expliciete voorkeur van een gebruiker (item 2) centraal staat, kan op veel bijval rekenen van het publiek. Respectievelijk 83,46% en 71,48% van de respondenten ging ermee (helemaal) akkoord dat deze twee nieuwsselectieprincipes goede principes zijn om nieuwsartikelen te selecteren. Een analyse van de gemiddeldes (zie Bijlage 1) bevestigde dat de gemiddelde scores van deze twee principes significant verschillen van de andere gemiddeldes.

Figuur 1. Algemene voorkeur nieuwsalgoritmes (n = 943)



Wanneer we kijken naar de nieuwsselectieprincipes die schuilgaan achter het collaborative nieuwsalgoritme en het diversity-based nieuwsalgoritme, zien we een licht hogere voorkeur voor de principes die behoren tot het collaborative nieuwsalgoritmes zoals, bijvoorbeeld, het principe waarbij nieuwsartikelen worden geselecteerd op basis van populariteit of op basis van personen met gelijkaardige interesses.

Toch tekent deze voorkeur voor collaborative nieuwsalgoritmes zich niet duidelijk af tegenover alle principes die vallen onder het diversity-based nieuwsalgoritme. Items 7 en 8, bijvoorbeeld, worden door het publiek op eenzelfde manier beoordeeld dan de items die behoren tot het collaborative nieuwsalgoritme. Ook in de analyse van de

gemiddelde scores (zie Bijlage 1) zien we dat de gemiddeldes van deze principes niet significant van elkaar verschillen. We kunnen dit verschil verklaren doordat de eerste twee principes van het diversity-based algoritme focussen op diversiteit in standpunten en opinies, terwijl de andere items meer samenhangen met diversiteit in onderwerpen en interesses. Dit resultaat toont aan dat een deel van het publiek dit verschil opmerkt en het ook anders beoordeelt.

3.2 Clusteranalyse

In dit rapport spreken we al even over 'het publiek', terwijl er verschillende groepen binnen deze allesomvattende term bestaan die elk op een andere manier kijken naar nieuwsalgoritmes en de verschillende principes voor nieuwsselectie. Om deze verschillende groepen te onderscheiden, voerden we een clusteranalyse uit en onderzochten we hoe deze groepen de verschillende nieuwsselectieprincipes gemiddeld beoordelen. In Tabel 3 geven we een overzicht van deze analyse waaruit vier profielen werden gedestilleerd.

Tabel 3. Gemiddelde scores (vijfpuntenschaal) van het vier clustermodel: beoordeling van verschillende nieuwsselectieprincipes

	Cluster 1: optimistische nieuwskiezer (32,47%)	Cluster 2: kieskeurige nieuwskiezer (31,02%)	Cluster 3: breeddenkende nieuwskiezer (27,91%)	Cluster 4: ongeïnteresseerde nieuwskiezer (8,60%)
Nieuws op basis van mijn interesses	3,67	4,45	4,33	2,57
Nieuws op basis van wat ik zelf aangeef leuk te vinden	3,28	4,32	4,14	2,21
Nieuws op basis van wat ik eerder heb gelezen	3,05	3,67	3,70	1,81
Nieuws op basis van wat personen met gelijkaardige interesses lezen	2,91	3,79	3,84	1,65
Nieuws op basis van wat momenteel populair is	2,87	3,34	3,44	1,72
Nieuws op basis van wat mijn vrienden hebben gelezen	2,63	3,04	3,37	1,46
Nieuws op basis van meningen die verschillen met mijn eigen mening	3,40	3,16	3,96	1,66
Nieuws op basis van standpunten waar ik het niet eens mee ben	3,14	2,92	3,81	1,60
Nieuws op basis van onderwerpen die ik normaal niet lees	3,06	2,21	3,47	1,46
Nieuws op basis van wat buiten mijn interesseveld valt	3,01	2,02	3,28	1,40

Het eerste profiel is de 'optimistische nieuwskiezer' (32,47%) die over het algemeen vrij tot zeer positief staat tegenover alle nieuwsselectieprincipes. Personen met dit profiel verkiezen om zowel nieuws op basis van hun interesses als op basis van diversiteit in opinies te ontvangen. Nieuws selecteren op basis van wat hun vrienden hebben gelezen, is het minst geliefkoosde principe, maar geniet ook nog van een zekere acceptatie.

Het tweede profiel is de 'kieskeurige nieuwskiezer' (31,02%). Personen in dit profiel zijn vooral voorstander van het nieuwsselectieprincipe waarin rekening wordt gehouden met diens interesses. Zij zijn geen voorstander om nieuwsonderwerpen te ontvangen die ze normaal gezien niet lezen of niet aansluiten bij hun interesses. Als ze diversiteit in hun nieuwsdieet wensen, dan verkiezen ze nieuwsartikelen met verschillende opinies.

Het derde profiel is de 'breeddenkende nieuwskiezer' (27,91%) die open staat voor elk nieuwsselectieprincipe maar, in tegenstelling tot de voorgaande profielen, vooral open staat voor alle principes die behoren tot het *diversity-based* nieuwsalgoritme. In het bijzonder ontvangen ze graag nieuwsartikelen waarin opinies staan die verschillen met hun eigen mening.

Het vierde profiel is de 'ongeïnteresseerde nieuwskiezer' (8,60%). Personen met dit profiel zijn over het algemeen niet te vinden voor de automatische selectie van nieuwsartikelen. Bijgevolg staan ze ook weigerachtig tegenover elk nieuwsselectieprincipe. Als ze één nieuwsselectie principe moeten kiezen, kiezen ze voor het principe waarin de interesses van diens persoon centraal staan.

3.3 Hiërarchische regressieanalyse

Om te verklaren waarom een bepaald nieuwsselectieprincipe slechter of beter scoort in de beoordeling, voerden we ook een hiërarchische regressieanalyse uit. Deze analyse laat toe om de relatie tussen de beoordeling van verschillende nieuwsselectieprincipes enerzijds, en verschillende individuele kenmerken anderzijds, te bestuderen. Voor deze analyse hebben we gekozen om 20 individuele kenmerken mee te nemen die, op basis van een uitgebreide literatuurstudie, mogelijke verschillen in de beoordeling kunnen verklaren. We kunnen deze kenmerken opdelen in drie groepen: (1) sociodemografische kenmerken, (2) kenmerken die gerelateerd zijn aan het nieuwsgebruik en de nieuwsinteresses van een persoon, en (3) kenmerken die gerelateerd zijn aan de cognitieve en technologische context waarin iemand nieuws consumeert.

In Tabel 4 geven we een overzicht van de resultaten van deze hiërarchische regressieanalyse voor de drie verschillende nieuwsselectieprincipes: content-based, collaborative en diversity-based. Om de resultaten voldoende gedetailleerd te kunnen beschrijven, focussen we ons in deze bespreking enkel op het diversity-based nieuwsselectieprincipe waarin de respondenten werden gevraagd hoe ze stonden tegenover een nieuwsalgoritme dat diversiteit in onderwerpen en opinies wil promoten. We konden uit de regressieanalyse de volgende zaken afleiden:

- Van de sociodemografische variabelen, bleken enkel de variabelen 'geslacht' ($\beta = -.085$, $p < .05$, referentiecategorie = 'man') en 'opleiding' ($\beta = -.076$, $p < .05$, referentiecategorie = 'lager dan hoger onderwijs') een significante relatie te hebben met de beoordeling van het diversity-based nieuwsselectieprincipe. Andere variabelen zoals leeftijd, nationaliteit (ouders) of arbeidsactiviteit kunnen geen verschillen in de beoordeling verklaren. Concreet wil dat zeggen dat **vrouwen en mensen met een hoger opleidingsniveau het nieuwsselectieprincipe lager beoordelen dan mannen en mensen met een lager opleidingsniveau.**

- Van de variabelen gerelateerd aan nieuwsgebruik, bleken enkel de variabelen 'nieuwsinteresse' ($\beta = -.113$, $p < .05$; $\beta = -.103$, $p < .05$, met referentiecategorie = 'breed geïnteresseerd'), nieuwsgebruik in aantal minuten ($\beta = .075$, $p < .05$) en het gebruik van buitenlandse nieuwsbronnen ($\beta = .076$, $p < .05$, referentiecategorie = 'nee') een significante relatie te hebben met de beoordeling van het diversity-based nieuwsselectieprincipe. Als we deze relaties van dichtbij bekijken, dan zien we voor de eerste variabele 'nieuwsinteresse' dat personen die behoren tot de clusters 'gemiddeld geïnteresseerd' en 'laag geïnteresseerd' het nieuwsselectieprincipe lager beoordelen dan personen die behoren tot de cluster 'breed geïnteresseerd'. Daarnaast zien we dat **hoe meer tijd een persoon spendeert aan nieuws, hoe positiever hij staat tegenover dit nieuwsselectieprincipe**. Tenslotte zien we dat mensen die buitenlandse nieuwsbronnen raadplegen positiever staan tegenover dit nieuwsselectieprincipe dan personen die geen buitenlandse nieuwsbronnen gebruiken.
- Van de variabelen gerelateerd aan de technologische en cognitieve context waarin nieuwsgebruik plaatsvindt, bleken enkel de variabelen 'overbelasting nieuwsinformatie' ($\beta = .127$, $p < .05$) en 'bezorgdheden over ontvangen van verschillende invalshoeken' ($\beta = .198$, $p < .05$) een significante relatie te hebben met de beoordeling. Andere variabelen zoals vertrouwen in nieuwsmedia of bezorgdheden over privacy of het missen van belangrijke informatie heeft geen significante relatie. Concreet houdt dat het volgende in: **hoe meer je je overladen voelt door de snelheid en hoeveelheid van het nieuws, hoe positiever je staat tegenover het diversity-based nieuwsselectieprincipe**. Tenslotte, hoe bezorgder je bent dat nieuwspersonalisatie leidt tot het missen van uitdagende standpunten, hoe beter je het diversity-based nieuwsselectieprincipe beoordeelt.

Tabel 4: Linear and hierarchical regression analysis: standardized regression coefficients

	Afhankelijke variabele	<i>Content-based</i> nieuwsselectie			<i>Collaborative</i> nieuwsselectie			<i>Diversity-based</i> nieuwsselectie		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
Sociodemografische variabelen	Regressiemodellen									
	Leeftijd	-.227*	-.222*	-.246*	-.243*	-.223*	-.232*	-.068	-.059	-.016
	Geslacht (man)	-.026	-.050	-.043	.047	.008	.021	-.061	-.056	-.085*
	Nationaliteit (buitenlands)	.107*	.185*	.206*	.027	.065	.071	.055	.056	.024
	Nationaliteit ouders (buitenlands)	-.003	.004	.007	-.031	-.024	-.026	.033	.024	.026
	Hoger opleidingsniveau (nee)	.064	.076*	.072*	-.019	.021	.009	-.027	-.068	-.076*
Variabelen gerelateerd aan nieuwsgebruik en nieuwsinteresses	Arbeidsactiviteit (niet actief)	-.055	-.068	-.053	-.111*	-.118*	-.098*	-.009	-.014	-.022
	Interesse 2: politiek geïnteresseerd		-.163*	-.136*		-.212*	-.189*		-.039	-.067
	Interesse 3: gemiddeld geïnteresseerd		-.078*	-.082*		-.072*	-.073*		-.126*	-.113*
	Interesse 4: laag geïnteresseerd		-.137*	-.116*		-.203*	-.181*		-.112*	-.103*
	Nieuwsgebruik 2: getemperde omnivoor		.017	.016		.038	.037		.064	.065
	Nieuwsgebruik 3: traditionele gebruiker		-.010	.033		.043	.085		-.128	-.183
	Nieuwsgebruik 4: getemperde traditionele		-.008	.021		.006	.044		-.061	-.083
	Nieuwsgebruik (in minutes)		.010	.015		.025	.038		.066	.075*
	Gebruik buitenlandse bronnen (nee)		-.037	-.004		-.023	.010		.081*	.076*
	Interactie-effect: nationaliteit		.101*	.103*		.041	.029		.012	-.005
	Betalen voor nieuws (nee)		.069*	.067*		.006	-.004		.054	.041
	Mobiel nieuwsgebruik (nee)		.006	-.020		.040	.007		-.006	.019
	Online nieuwsgebruik (nee)		.038	.051		.129	.132		-.122	-.158
	Online nieuwsparticipatie (nee)		.045	.021		.185*	.154*		.046	.043
	Variabelen gerelateerd aan de technologische en cognitieve context	Overbelasting nieuwsinformatie			.111*			.166*		
Optimisme in technologie				.057			.131*			-.049
Vertrouwen in nieuwsmedia				.112*			.087*			.029
Bezorgdheden over privacy				-.073*			-.050			-.060
Bezorgdheden: belangrijke informatie missen				.011			.006			.016
Bezorgdheden: uitdagende standpunten				-.134*			-.076			.198*
Observaties	937	937	937	937	937	937	937	937	937	
R ²	.052	.099	.145	.047	.151	.201	.009	.059	.111	
Gewijzigde R ²	.046	.081	.121	.041	.133	.179	.003	.038	.086	
Standaardfout residuele	.976	.958	.937	.976	.928	.904	.994	.976	.951	
F Statistiek	8.388*	5.268*	6.085*	7.588*	8.456*	9.052*	1.449*	2.910*	4.494*	



4. Conclusie en aanbevelingen

Dit rapport presenteert de resultaten van een surveyonderzoek dat onderzocht hoe nieuwsconsumenten nieuwsalgoritmes en de nieuwsselectieprincipes die eronder schuilgaan, beoordelen. Door deze percepties in kaart te brengen, tracht deze studie te achterhalen welk nieuwsalgoritme en nieuwsselectieprincipe de voorkeur uitdraagt van het publiek en wat het adoptiepotentieel is van een maatschappelijk nieuwsalgoritme dat nieuwsdiversiteit wil stimuleren en promoten. Deze studie tracht in het bijzonder ook de verscheidenheid van algoritmes onder de aandacht te brengen. In de literatuur worden (nieuws)algoritmes te vaak over één kam gegooid, zonder aandacht te schenken aan de **diversiteit aan onderliggende principes die onder algoritmes schuilgaan** en bepalen hoe artikels worden geselecteerd en geordend.

Het surveyonderzoek dat we in Gent hebben uitgevoerd, toont aan dat het publiek een grotere voorkeur heeft voor nieuwsselectieprincipes die behoren tot het *content-based* nieuwsalgoritme dan voor nieuwsselectieprincipes die behoren tot het *collaborative* of *diversity-based* nieuwsalgoritme. Het bevestigt daarmee de dominante redenering die onder de huidige nieuwsalgoritmes schuilgaan: deze focussen doorgaans op de interesses van een bepaald persoon. Door op dit principe te focussen, geven nieuwsorganisaties dus wat de nieuwsgebruiker graag wil.

Dit resultaat toont echter ook aan dat wanneer het publiek de keuze heeft om te bepalen hoe ze het nieuws willen ontvangen, ze de neiging hebben om nieuwsartikelen te selecteren die hen interesseren en dus mogelijks ook **nieuwsartikelen te mijden waarmee ze niet vertrouwd zijn**. Dit impliceert dat selectieve blootstelling ook in een algoritmische context een gevaar vormt, zeker wanneer onze algoritmes hier geen weerstand tegen bieden. Om tegemoet te komen aan dit risico, kunnen nieuwsorganisaties overwegen om bepaalde aanpassingen uit te voeren. Een voorbeeld van zo'n aanpassing is het gebruik van het zogenaamde *serendipity*-concept waarin, naast de artikels die focussen op de interesses van een bepaald persoon, ook onverwacht een willekeurig artikel wordt geselecteerd en aangereikt. **Het toevallig botsen op een artikel dat een gebruiker niet verwacht te zien, leidt bovendien ook tot een hoger engagement.**⁹

Om dit risico op een meer structurele manier in te dijken, raden we de **implementatie van een diversity-based nieuwsalgoritme aan**. Hoewel het niet de voorkeur uitdraagt van het publiek, zien we dat de nieuwsselectieprincipes die focussen op diversiteit in opinies en ideeën ook kunnen rekenen op een groot draagvlak. Vanuit de clusteranalyse zien we bijvoorbeeld dat de acceptatiegraad van dit soort algoritme zeer hoog is bij twee groepen – optimistische nieuwskiezer (32,47%) en de breeddenkende nieuwskiezer (27,91%) – die samen circa 60% vertegenwoordigen. Het is echter een klein groep – de ongeïnteresseerde nieuwskiezer (8,60%) – die niet warmloopt voor dit soort algoritme.

Om een nog groter draagvlak te vinden, raden we bij de ontwikkeling van een diversiteitsalgoritme aan om steeds **rekening te houden met het leesgedrag en -interesses van een nieuwsgebruiker**. Dit betekent concreet dat een persoon die enkel sportnieuwsartikelen leest niet meteen nieuwsartikelen moet aanbevolen krijgen die thematisch gezien ver vanaf staan (bv. buitenlandse zaken). Nieuwsorganisaties kunnen best op zoek gaan naar 'het ideale pad' die iemand in die richting kan leiden. Dat kan zijn door artikels te zoeken die de aansluiting vinden bij zowel sport

⁹ De Pessemier et al. (2014)

als buitenlandse zaken, maar kan ook door kleine tussenstapjes te nemen waarbij de persoon ook nieuwsartikelen leest die gaan over meer maatschappelijke thema's zoals onderwijs, milieu of wetenschappen. We noemen deze techniek ook wel '**gepersonaliseerde diversiteit**' waarbij het uiteindelijke doel van het diversiteitsalgoritme behouden blijft, maar waarbij er wel gebruik wordt gemaakt van de personalisatie-technieken die onder *content-based* nieuwsalgoritmes schuilgaan en hen ook zo aantrekkelijk maken.

Referenties

- Araujo. T., Helberger. N., Kruijkemeier. S., & de Vreese. C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & SOCIETY*. doi:10.1007/s00146-019-00931-w
- Bodó. B. (2019). Selling News to Audiences—A Qualitative Inquiry into the Emerging Logics of Algorithmic News Personalization in European Quality News Media. *Digital Journalism*. 7(8). 1054-1075. doi:10.1080/21670811.2019.1624185
- Bruns. A. (2019). Filter bubble. *Internet Policy Review*. 8(4).
- De Pessemier, T., Dooms, S. & Martens, L. (2014). Context-aware recommendations through context and activity recognition in a mobile environment. *Multimedia Tools and Applications*, 72(3), 2925–2948. doi: 10.1007/s11042-013-1582-x
- Helberger. N. (2019). On the democratic role of news recommenders. *Digital Journalism*. 1-20. doi:10.1080/21670811.2019.1623700
- Jones. B., & Jones. R. (2019). Public Service Chatbots: Automating Conversation with BBC News. *Digital Journalism*. 7(8). 1032-1053. doi:10.1080/21670811.2019.1609371
- Joris. G., Colruyt. C., Vermeulen. J., Vercoutere. S., De Grove. F., Van Damme. K., . . . Hoste. V. (2019). *News diversity and recommendation systems: setting the interdisciplinary scene*. Paper presented at the IFIP Summer School on Privacy and Identity Management.
- Pariser. E. (2011). *The filter bubble: What the Internet is hiding from you*. Penguin, UK.
- Thurman. N., Möller. J., Helberger. N., & Trilling. D. (2019). My friends, editors, algorithms, and I: Examining audience attitudes to news selection. *Digital Journalism*. 1-23. doi:10.1080/21670811.2018.1493936
- Van den Bulck. H., & Moe. H. (2018). Public service media, universality and personalisation through algorithms: mapping strategies and exploring dilemmas. *Media, Culture & Society*. 40(6). 875-892. doi:10.1177/0163443717734407

Bijlage

Bijlage 1. Descriptieve gegevens: beoordeling nieuwsselectieprincipes

Survey items	Gemiddelde	Standaard-afwijking	Variantie	Score-bereik	N
(1) Nieuws op basis van mijn interesses	4,00	0,89	0,80	1-5	943
(2) Nieuws op basis van wat ik zelf aangeef leuk te vinden	3,75	1,02	1,04	1-5	943
(3) Nieuws op basis van wat ik eerder heb gelezen	3,32	1,02	1,04	1-5	943
(4) Nieuws op basis van wat personen met gelijkaardige interesses lezen	3,33	1,04	1,07	1-5	943
(5) Nieuws op basis van wat momenteel populair is	3,08	1,04	1,09	1-5	943
(6) Nieuws op basis van wat mijn vrienden hebben gelezen	2,86	1,02	1,03	1-5	943
(7) Nieuws op basis van meningen die verschillen met mijn eigen mening	3,33	0,94	0,88	1-5	943
(8) Nieuws op basis van standpunten waar ik het niet eens mee ben	3,13	0,97	0,94	1-5	943
(9) Nieuws op basis van onderwerpen die ik normaal niet lees	2,77	1,00	0,99	1-5	943
(10) Nieuws op basis van wat buiten mijn interesseveld valt	2,64	0,98	0,95	1-5	943

Bijlage 2. Analyse van de gemiddelde scores

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1)	-	-.255***	-.685***	-.670***	-.928***	-1.140***	-.673***	-.878***	-1.232***	-
(2)		-	-.431***	-.416***	-.673***	-.885***	-.419***	-.624***	-.978***	-1.108***
(3)			-	.015	-.243***	-.455***	.012	-.193***	-.547***	-.678***
(4)				-	-.258***	-.470***	-.003	-.208***	-.562***	-.692***
(5)					-	-.212***	.467***	.050	-.304***	-.435***
(6)						-	.255***	.262***	-.092*	-.223***
(7)							-	-.205***	-.559***	-.689***
(8)								-	-.354***	-.485***
(9)									-	-.130***
(10)										-

Noot: een verschil tussen twee gemiddelde scores is significant wanneer p-value kleiner is dan .05 (*). Om het probleem van kanskapitalisatie te vermijden, werden de volgende significantieniveaus berekend: Holm-correctie: p-value < 0.01138 (**); Bonferroni-correctie: p-value < .001111 (***)

Bijlage 3. Gemiddelde scores van de 4 geïdentificeerde clusters uit de variabele 'nieuwsinteresse'

News interest topics	Cluster 1: breed geïnteresseerd (43.38%)	Cluster 2: politiek geïnteresseerd (32.25%)	Cluster 3: gemiddeld geïnteresseerd (18.93%)	Cluster 4: laag geïnteresseerd (5.44%)
Maatschappelijke kwesties	8.16	8.58	5.71	5.73
Bekende personen	5.68	1.97	3.68	0.33
Politiek	6.70	7.78	3.65	3.36
Sport	5.03	3.67	4.25	0.64
Verhalen van mensen en het alledaagse leven	6.89	4.60	4.79	1.71
Misdaad en accidenten	7.04	4.79	5.42	2.35
Cultuur	7.07	5.81	4.78	2.41
Lifestyle	6.60	3.31	4.19	0.54

Bijlage 4. Gemiddelde scores van de 4 geïdentificeerde clusters uit de variabele 'nieuwsgebruik'

News use channels	Cluster 1: nieuwsomnivoor (36.30%)	Cluster 2: Getemperde omnivoor (29.97%)	Cluster 3: traditionele gebruiker (20.30%)	Cluster 4: Getemperde traditionele gebruiker (13.43%)
Nationale televisie	3.83	2.32	3.98	2.56
Regionale televisie	2.45	1.19	2.90	1.36
Radio	3.63	2.86	3.47	3.00
Fysieke krant	2.79	2.02	3.25	2.12
Nieuwswebsite	3.39	3.25	1.01	1.01
Digitale nieuwsbrief	2.48	1.72	1.00	1.00
Nieuws via sociale media	2.20	2.72	1.00	1.01
Nieuws via zoekmachines	2.36	2.20	1.01	1.02

Meer weten over NewsDNA?

Contacteer:

Prof. dr. Lieven De Marez

Lieven.DeMarez@UGent.be