

PERSBERICHT 23 SEPTEMBER 2021

Statistiek Vlaanderen start samenwerking met KU Leuven rond het gebruik van data science voor nieuwe en betere statistieken over Vlaanderen

Statistiek Vlaanderen zet via de uitbouw van een eigen Data Science Hub sinds enige tijd volop in op het toepassen van data science technieken voor het creëren van nieuwe en betere statistieken over Vlaanderen. Daarbij wordt nu een versnelling hoger geschakeld: de Data Science Hub van Statistiek Vlaanderen werkt de komende 4 jaar samen met het Research Center for Information Systems Engineering (LIRIS) van de KU Leuven aan state-of-the-art onderzoek om de bruikbaarheid van data science na te gaan voor nieuwe en betere openbare statistieken.

Statistiek Vlaanderen is in 2019 gestart met het uitbouwen van een eigen Data Science Hub om in te spelen op het snel veranderende data landschap waar volop wordt ingezet op de nieuwe mogelijkheden van 'data science' en 'big data'. De hub zorgt voor het uitwerken van innovatieve data science projecten, het aanbieden van data science opleidingen binnen de Vlaamse overheid en het samenbrengen van data science initiatieven binnen en buiten de Vlaamse overheid.

Zo werkt de Data Science Hub bijvoorbeeld aan het ontwikkelen van accuratere cijfers over innovatie in Vlaanderen door middel van het 'scrapen' van webpagina's van bedrijven. In combinatie met een machine learning algoritme worden bedrijven zo gecategoriseerd als 'innovatief' of 'niet innovatief'. Ook sociale media staan op de radar van Statistiek Vlaanderen. Een machine learning model maakt het mogelijk om automatisch het sentiment van tweets te bepalen. Dat wordt dan vertaald naar een sentiment score die een inschatting geeft van hoe het met het algemene welbevinden van de Vlamingen gesteld is.

De eerste resultaten van deze projecten worden momenteel (nog) niet gepubliceerd als openbare statistieken maar beschouwd als 'experimentele statistieken'. Als bron van betrouwbare data en statistieken wil Statistiek Vlaanderen deze resultaten eerst in de diepte analyseren en verder inzetten op het opbouwen van kennis over de bruikbaarheid en betrouwbaarheid van deze nieuwe methoden. De gebruikte technieken zijn veelbelovend, maar er wachten tegelijk ook nog veel uitdagingen. Die uitdagingen variëren van het efficiënt gebruiken van nieuwe databronnen tot het correct inzetten van complexe machine learning algoritmen en innovatieve data-visualisatie technieken.

De samenwerking met het Research Center for Information Systems Engineering (LIRIS) van de KU Leuven die nu opgestart wordt en loopt tot september 2025, kadert in deze kennisopbouw. LIRIS staat bekend om state-of-the-art data science onderzoek met sterke praktische relevantie. Een doctoraatsstudent zal onderzoek verrichten over de bruikbaarheid van data science voor de opmaak van openbare statistieken. Ook zullen er elk jaar verschillende master thesis topics in dit domein uitgewerkt worden. Het onderzoek zal steeds focussen op het gebruik van data science technieken om informatie over Vlaanderen te ontginnen uit nieuwe ('big') databronnen. Tijdens de loop van de onderzoekssamenwerking zullen er geregeld seminars georganiseerd worden waarop de opgedane kennis gedeeld wordt met de buitenwereld en zullen resultaten van het onderzoek als 'experimentele statistieken' gepubliceerd worden op de websites van beide partners.

Voor meer informatie

Contact:

Michael Reusens | T 0473 84 75 55 | michael.reusens@vlaanderen.be

Statistiek Vlaanderen is het netwerk van Vlaamse overheidsinstanties die openbare statistieken ontwikkelen, produceren en publiceren. Statistiek Vlaanderen wil deze Vlaamse openbare statistieken via één portaalsite voor iedereen vlot toegankelijk maken zodat burgers, organisaties en bedrijven beschikken over correcte en gedetailleerde informatie over Vlaanderen. Statistiek Vlaanderen heeft als opdracht om alle openbare statistieken te ontwikkelen in overeenstemming met de wettelijke bepalingen en internationale principes rond kwaliteit, relevantie en onafhankelijkheid.