

Tekst van een jaarverslag zegt meer over fraude dan gedacht

Erik van Rein
Amsterdam

Neem de tekst van een jaarverslag, laat er een algoritme op los en aan het eind van de berekening is bekend of er mogelijk een nieuw boekhoudschandaal in staat. Kort door de bocht is dit het systeem dat Marcia Fissette ontwikkelde. Zij promoveerde afgelopen week aan de Universiteit Twente op een onderzoek naar het ontdekken van fraude via tekstanalyse.

Onderzoek naar het ontdekken van boekhoudfraude richt zich vaak op de cijfers in jaarrekenin-

gen. Logisch, omdat gesjoemel met balansen daar het zichtbaarst is.

Teksten kunnen ook aanwijzingen bevatten, vaak onbewust. 'We denken dat we onze woorden vaak heel bewust kiezen, maar dat is niet zo. Je valt toch vaak in je eigen patroon en woordenschat', zegt Fissette, die ook bij KPMG werkt.

Mensen kunnen die afwijken de woordkeuze moeilijk ontdekken, omdat ze betekenis geven aan woorden. De vraag was dus: kan een computer bepaalde jaarverslagen bestempelen als frauduleus, en andere niet?

Fissette maakte daarvoor een da-

tabase met 400 jaarverslagen van bedrijven wereldwijd waar fraude is ontdekt. De lijst werd aangevuld met ruim 1300 jaarverslagen van ondernemingen waar, voor zover bekend, geen fraude is gepleegd.

Fissette koos ervoor alleen de managementdiscussies uit de jaarverslagen mee te nemen in het onderzoek, niet alle teksten. In dat hoofdstuk licht een bestuur de resultaten toe. Analisten lezen dat het grondigst, blijkt uit ander onderzoek. Fissette: 'Daar wordt de keuze gemaakt: dit gaan we wel vertellen en dat niet.'

Die teksten zijn daarna geanaly-

seerd via 'machine learning'. Dat is een ingewikkeld proces waarbij algoritmes de patronen in een tekst kunnen herkennen. Zo kan een computer leren of het gebruik van bepaalde woorden of woordcombinaties kenmerkend is voor de aanwezigheid van fraude. Daarna maakt de computer een keuze: dit jaarverslag is frauduleus, of juist niet. Bij de test plaatste de computer de jaarverslagen in 89% van de gevallen in de goede categorie.

'Dat lijkt veelbelovend, maar de algoritmes kunnen niet aanwijzen wáár die fraude dan precies zit en welke woorden of zinnen af-

1700

In het onderzoek zijn 1700 jaarverslagen van bedrijven onderzocht, waaronder 400 van bedrijven waar fraude is ontdekt.

wijken. Een lijstje maken van verdachte teksten kan dus niet. 'Dat is het nadeel van machine learning', zegt Fissette. 'Je weet niet welke beslissingen de computer neemt.'

De uitkomst is dan ook eerder een 'red flag'. Het is een teken om nog eens extra goed naar het jaarverslag of de cijfers te kijken. Misschien valt er dan iets op. 'Bij één frauduleus jaarverslag zag ik overdreven optimistische woordkeuze. Dan hadden ze het over de fantastische en overweldigende resultaten', zegt Fissette. 'Dat soort woorden verwacht je al niet in een zakelijke tekst.'