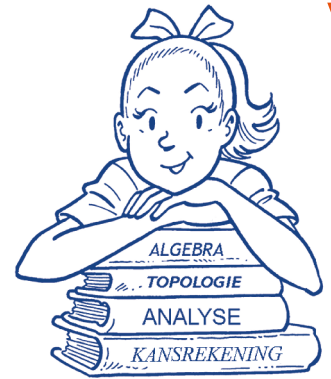
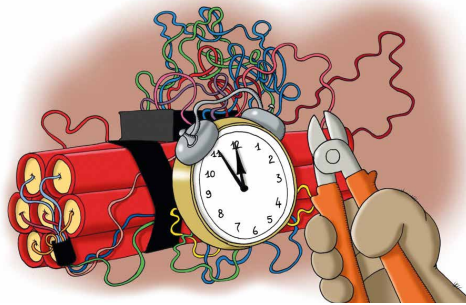


# WISKUNNEND WISKE

## DE TIKKENDE TIJDBOM



© 2019, Standaard Uitgeverij, Antwerpen, België



### OPGAVE 2

Krimson heeft in Wiskes slaap een tijdbom aan haar been bevestigd. Wanneer Wiske wakker wordt, staat de aftelklok op 1 uur. Krimson vertelt dat de bom 1000 draadjes heeft, waarvan er exact één onzichtbaar verbonden is met de klok. Wanneer ze dat draadje kan doorknippen, is de bom onschadelijk gemaakt en zal Krimson haar bevrijden. Echter, wanneer ze het verkeerde draadje knipt, zal de bom onmiddellijk ontploffen. Wiske kan de bom niet zelf van haar been losmaken.

Krimson is niet geheel slecht en geeft Wiske wel een kans om te ontsnappen. Ze krijgt de kans om 10 verschillende combinaties van draadjes op papier neer te schrijven. Krimson zal na 50 minuten terugkomen, pen en papier afnemen en van elke combinatie zeggen of ze het bewuste draadje al dan niet bevat.

Hoe kan Wiske te weten komen welk het juiste draadje is en zo de bom op tijd onschadelijk maken?

1. Los het vraagstuk eerst op met een bom met 15 draadjes, waarbij Wiske de kans krijgt om 4 combinaties op te schrijven.
2. Los het nu op met 1000 draadjes en 10 combinaties. Geef een volledige wiskundige argumentatie.

### WISKUNDIG WEETJE

Wat Wiske moet doen, is op een gerichte manier data verzamelen, en uit die data het juiste antwoord afleiden.

Daar waar het vroeger heel moeilijk was om gegevens te verzamelen over een zeker probleem is er nu eerder een overvloed aan gegevens. Je kan denken aan zoekmachines op het internet maar ook metingen bij wetenschappelijke experimenten worden steeds makkelijker en goedkoper. Van een bepaald staal bloed krijg je tegenwoordig meerdere bladzijden met allerlei parameters en waarden terug. Voor de interpretatie van deze massa aan gegevens (in combinatie met elkaar) moet de arts weten wat de relevante gegevens zijn voor een bepaalde patiënt en dan hiermee tot een diagnose komen. Met de *big data* tak van de wiskunde kunnen we dokters helpen met het verwerken van deze massa aan data tot een interpreteerbaar resultaat. De voornaamste uitdaging hier is om uit de massa gegevens die we uit een klein aantal experimenten krijgen te beslissen welke parameters al dan niet relevant zijn. Een experiment (denk aan 1 bloedstaal) geeft aanleiding tot honderden, soms duizenden, waarden die gemeten worden. Meetkundig komt dit overeen met 1 punt in een ruimte met honderden of duizenden dimensies. Omdat er in deze hoge dimensies zoveel plaats (of volume) is, wordt het moeilijk om te beslissen wanneer twee datapunten dicht bij elkaar liggen. Het is best mogelijk dat ze voor de (onbekende) relevante parameters heel dicht bij elkaar liggen maar in Euclidische afstand (waar alle coördinaten in rekening gebracht worden) heel ver. Vele statistische methoden die steunen op de nabijheid van datapunten worden onbruikbaar of moeten aangepast worden voor zeer hoge dimensies. Je kan immers nagaan dat het aantal datapunten dat je nodig hebt om statistisch relevante conclusies te vatten, exponentieel groeit met de dimensie. De wetenschap is de laatste jaren zo geëvolueerd dat we voor vele hoog-dimensionale problemen over relatief weinig datapunten beschikken. Hieruit de nuttige en relevante dimensies halen is de uitdaging! Dit fenomeen kreeg de naam *The curse of dimensionality*.

Dit is een van de vele aspecten die aan bod komen in de nieuwe opleiding *Wiskunde & Data Science* van de VUB.