

### Bart Verbeeren

Tijdens zijn studie Burgerlijk Ingenieur hebben zowel doorgedreven wiskunde als programmeren de interesse voor data science opgewekt bij Bart. Later vulde hij zijn technische kennis aan met de nodige business kennis dankzij een cursus 'Postgraduate Studies in Business Administration' aan de KU Leuven. Door het ontwikkelen van innovatieve business modellen en enkele R&D projecten heeft hij deze kennisdomeinen ten volle kunnen omzetten naar de praktijk. Na enkele specifieke opleidingen en projecten rond Machine Learning en Artificialle Intelligentie, besloot Bart om van zijn passie zijn beroep te maken en als Data Scientist aan de slag te gaan bij Canguru.

### CANGURU

Canguru is een jong en innovatief consulting bedrijf uit Mechelen. Het data science team helpt bedrijven om betere beslissingen te nemen aan de hand van "slimme data". We implementeren de juiste tools om kwalitatieve data te vergaren en te stockeren. Vervolgens brengen we die informatie naar boven dewelke het bedrijf moet helpen in het nemen van de juiste beslissingen

### Meer weten?

[www.canguru.be](http://www.canguru.be)

[Jan.Roels@canguru.be](mailto:Jan.Roels@canguru.be)



### Wat doet een Data Scientist ?

Data scientists zijn bezig met het verzamelen, modeleren en interpreteren van bedrijfsgegevens. Ze helpen organisaties om betere beslissingen te nemen op basis van deze data.

### Niko Home Control

Niko, een toonaangevende leverancier van huisautomatisering, zette eerder dit jaar een pilootproject op in een beperkt aantal studentenhuizen. Het doel: exploreren of Niko Home Control een meerwaarde kan brengen aan de uitbaters en de studenten. Eerst en vooral dienden de gegevens van alle sensoren en meters centraal en gestructureerd opgeslagen te worden in een zodanig formaat dat verdere verwerking toelaat. Daarnaast werden deze gestructureerde data verder geëxploreerd en gevisualiseerd om zo na te gaan of bestaande hypothesen konden getest worden en nieuwe inzichten konden gegenereerd worden.

CANGURU heeft Niko bijgestaan in dit gehele proces. Er werd een geautomatiseerde 'data pipeline' opgesteld om de ongestructureerde gegevens, afkomstig van verschillende bronnen, centraal op te slaan in een cloud hosted SQL database. De bekomen datasets waren zo direct bruikbaar voor verdere verwerking. Deze data werd grondig geëxploreerd om duidelijkheid te krijgen over welke informatie er allemaal beschikbaar was en - minstens even belangrijk – welke data ontbrak.

Na analyse kon de grote hoeveelheid inkomende gegevens met de helft worden gereduceerd. Dit is belangrijk om enerzijds de nuttige informatie te scheiden van de redundante en is anderzijds interessant om schaalbaarheid en langdurige monitoring toe te laten.

Tot slot werden een groot aantal verschillende perspectieven aangeleverd om deze gegevens te visualiseren, en dit zowel op statische als dynamische wijze.