

Most experiments fail.
(at Microsoft it's 5000 to 1)

you have to avoid assumptions and test early and test often.

If a statistic looks interesting or unusual, it's probably wrong.



CONVERSION HOTEL

@timhengeveld

Ronny Kohavi, general manager van het analyse- en experimententeam van Microsoft, begon in 2001 al met het uitvoeren van online experimenten. Hij is nu verantwoordelijk voor het gehele experimentenplatform van Microsoft – dat voert gemiddeld 1.200 experimenten per maand uit. Vijf vragen uit het interview met Kohavi tijdens Conversion Hotel 2017.

Experimenteren op grote schaal

Wat is de winnaar-ratio bij Microsoft?

'Bij Microsoft zijn er verschillende teams die zelfstandig experimenteren. Er zijn teams die pas net begonnen zijn met experimenteren waar ongeveer een derde van de experimenten een positief significant resultaat heeft (de aanpassing zorgt voor verbetering), een derde geen effect heeft en een derde een significant negatief effect heeft (de aanpassing maakt de resultaten slechter). Maar voor de meer geoptimaliseerde domeinen, zoals Bing, dat al heel lang experimenten uitvoert, is het succes moeilijker

te vinden. Hier is de winnaar-ratio ongeveer 10 tot 20 procent van alle testideeën.'

Tijdens je presentaties en workshops heb je het veel over de OEC. Kun je daar wat meer over vertellen?

'Elke organisatie die experimenteert moet een duidelijk doel formuleren waar ze naartoe werkt. De Overall Evaluation Criteria is idealiter één statistiek of soms een combinatie van belangrijke statistieken die je vertellen of een idee positief is voor de organisatie. Het moeilijkst hierbij is het vertalen van kortetermijnmetingen - een experiment dat een week of twee duurt - naar iets dat voorspellend is voor het succes op de lange termijn.'

Denk je dat er een correlatie is tussen het aantal experimenten op jullie platform dat het afgelopen jaar is gestegen en het toenemende succes van Bing?

'Nee, ik geloof dat de groei van Bing te

danken is aan de juiste beslissing om te investeren in de infrastructuur en de volledige focus op de OEC. Eén van de dingen die je je realiseert als je op zo'n grote schaal experimenteert, is dat je door het uitvoeren van experimenten niet alleen de "goede" functies van een product identificeert, maar juist de slechte eruit filtert. Als je niet experimenteert dan heeft je product altijd een paar goede en een paar slechte elementen. De waarde van experimenteren is dat je die twee kunt onderscheiden.'

Wat is volgens jou de toekomst van experimenteren?

'Elk bedrijf dat geavanceerde technieken toepast als machine learning en het bouwen van classifiers, zal moeten experimenteren. Je test niet meer elementen, maar complete modellen. En geen enkel mens kan testen of een dergelijk model werkt. Je zult dus alleen maar meer moeten A/B-testen als je over gaat op machine learning.'

Dus de markt voor A/B-testen zal blijven groeien?

'Ik geloof absoluut dat het alleen maar meer zal worden.'

Webanalisten.nl

Dit artikel is geschreven door Tom van den Berg, analytics & optimization expert bij Online Dialogue, voor het online analyse- en optimalisatieplatform Webanalisten.nl. Het volledige artikel lees je hier: <https://anali.st/interviewronnykohavi>.