



Recommender Systems in de praktijk

Brammert Ottens,
Machine Learning Engineer



Wie ben ik?

**Senior Machine Learning
Engineer@Spotify**

- Masters in AI en Logic aan de UvA
- PhD in Computer Science aan de EPFL
- Software engineer bij Quintiq
- Senior Data Scientist bij Booking.com
- Senior Machine Learning Engineer bij Spotify



Overzicht

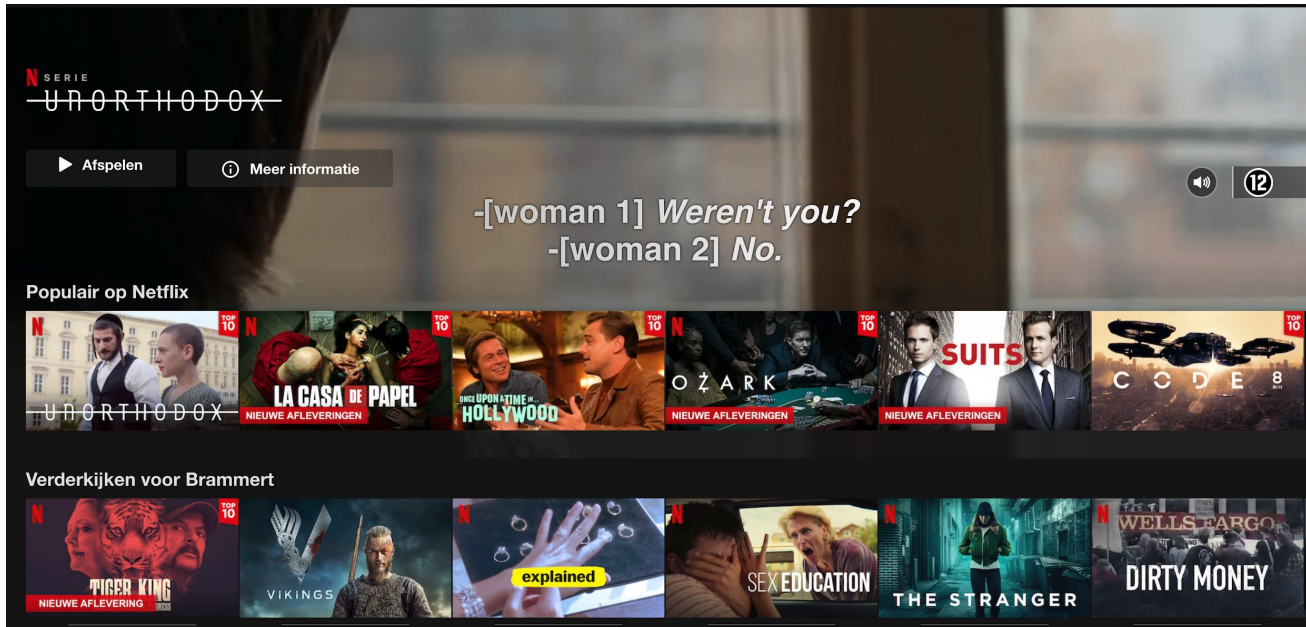
- Recommender Systems in het wild
- Waar loop je tegen aan?
 - Data Kwaliteit
 - Offline v.s. Online
 - Monitoren
 - A/B testing



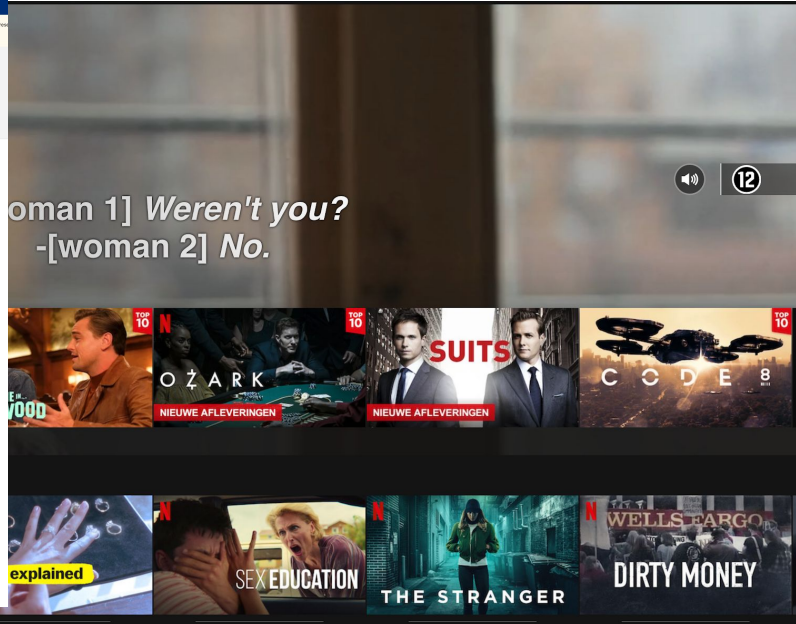
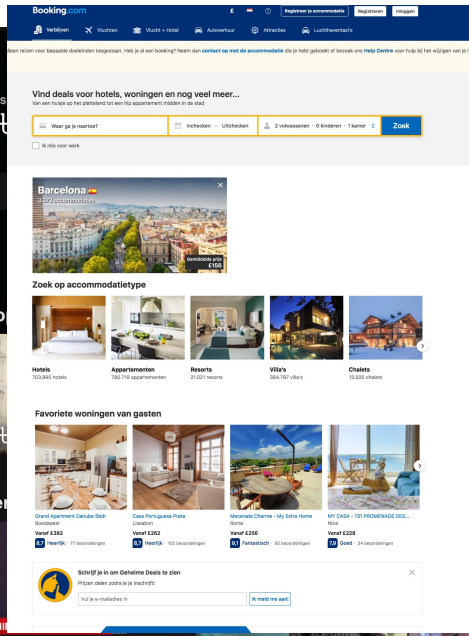
Overzicht

- **Recommender Systems in het wild**
- Waar loop je tegen aan?
 - Data Kwaliteit
 - Offline v.s. Online
 - Monitoren
 - A/B testing

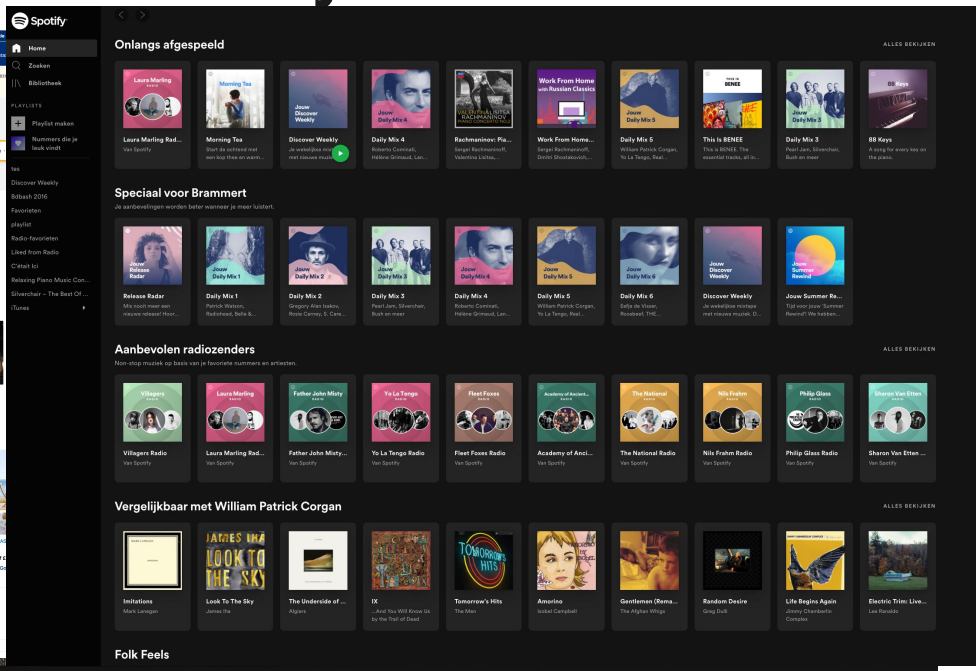
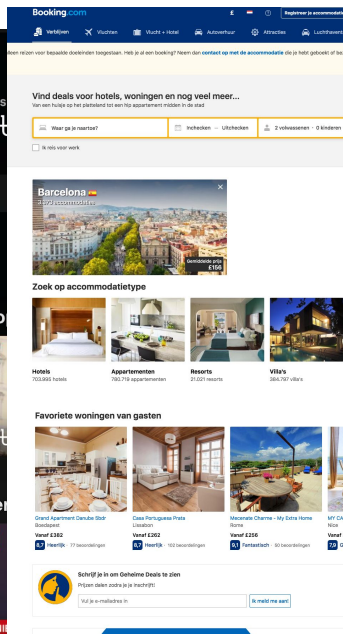
Recommender Systems in het wild



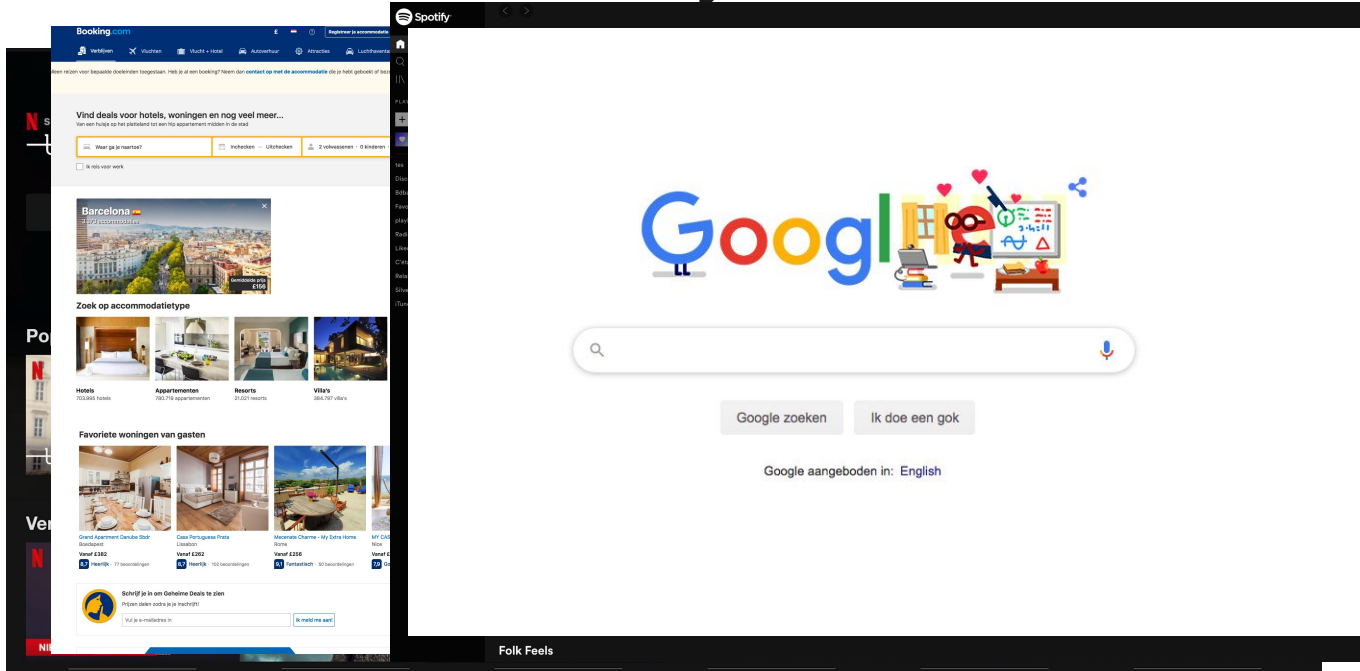
Recommender Systems in het wild



Recommender Systems in het wild



Recommender Systems in het wild





Overzicht

- Recommender Systems in het wild
- Waar loop je tegen aan?
 - **Data Kwaliteit**
 - Offline v.s. Online
 - Monitoren
 - A/B testing

Garbage in, Garbage out



- Data Kwaliteit, wat betekent dat?
- Data Kwaliteit, hoe kan je het meten?
- Data Kwaliteit, hoe kan je het bereiken?



Data Kwaliteit

Wat betekent het?





Data Kwaliteit

Wat betekent het?

- Betekent de data wel wat je denkt dat het betekent?
- Is het duidelijk waar de data vandaan komt?
- Is iemand “eigenaar” van de data?
- Is de data nog actueel?
- Zitten er fouten in de data?



Data Kwaliteit

Hoe meet je het?

- Betekent de data wel wat je denkt dat het betekent?
 - **Documentatie!**
- Is het duidelijk waar de data vandaan komt?
 - **Documentatie!**
- Is iemand “eigenaar” van de data?
 - **Cultuur!**
- Is de data nog actueel?
 - **Monitoren**
- Zitten er fouten in de data?
 - **Monitoren**



Data Kwaliteit

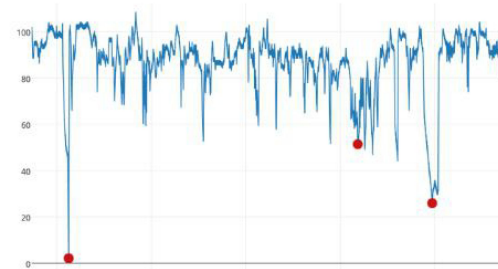
Hoe bereik je het?

- **Documentatie!**
 - Dit kan je afdwingen
- **Cultuur!**
 - Hier moet je iedereen meekrijgen
- **Monitoren**
 - Dit kan je afdwingen

Data Kwaliteit

Hoe kan je fouten in de data herkennen?

- Kijk naar de eigenschappen van je data
 - min/max
 - gemiddelde
 - aantal verschillende waarden
- Anomaly detection
 - thresholds
 - time series analysis
 - Clustering






Overzicht

- Recommender Systems in het wild
- Waar loop je tegen aan?
 - Data Kwaliteit
 - **Offline v.s. Online**
 - Monitoren
 - A/B testing

- Zijn de features wel beschikbaar online?
- Kan je ze snel genoeg berekenen?
- Zijn ze betrouwbaar genoeg?
- Zijn mijn offline en online features wel hetzelfde?



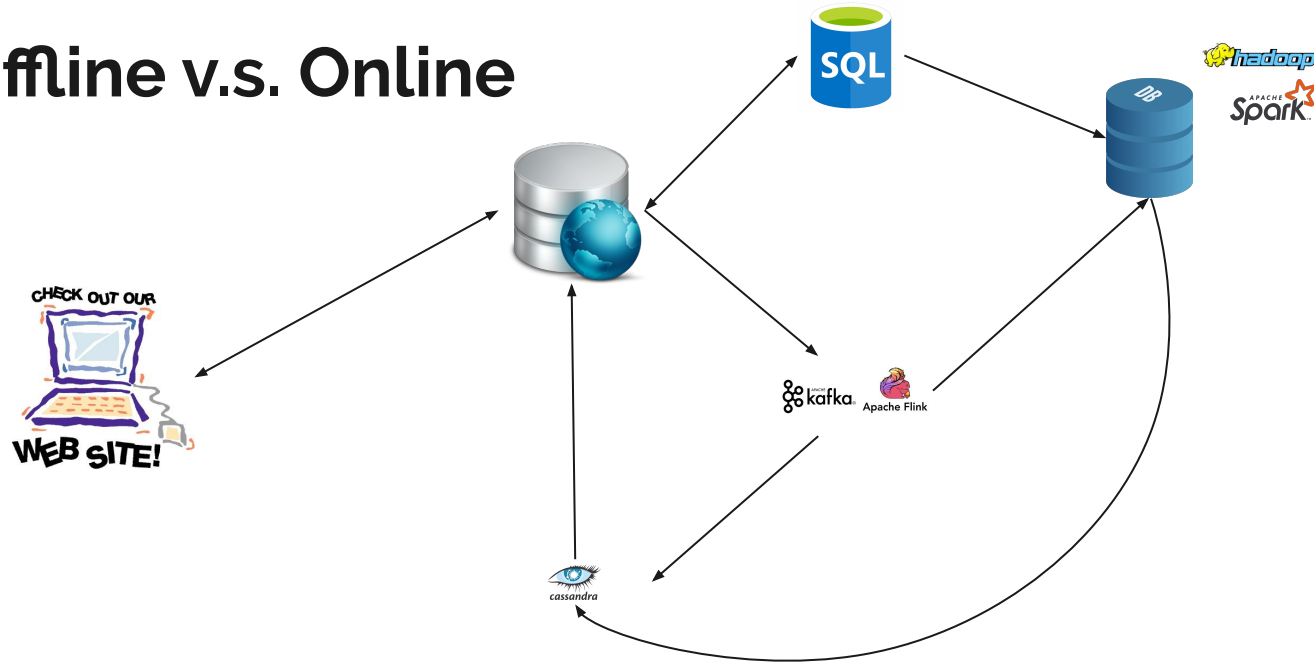
Offline
vs.
Online



Offline v.s. Online

| | | | |
|-------------------|---------------------------------------|---|--------------------------------|
| Type | SQL (Relational DB) | Distributed File System | Distributed Key Value Store |
| Examples | MySQL/PostgreSQL/SQ Lite | HDFS (Hadoop)/GCS | Cassandra/BigTable/HBase |
| Strengths | Fast/efficient querying | Able to deal with large amounts of data | Very fast lookup/Very scalable |
| drawbacks | Can't deal with large amounts of data | No fast lookup | One table per query |
| Where to use them | Power backend | Historical data store | Power backend |

Offline v.s. Online





Offline v.s. Online

| Key | Event | timestamp |
|-----|-------|-----------|
| u1 | A | t1 |
| u2 | C | t2 |
| u1 | A | t3 |
| u1 | A | t4 |

Hoe vaak komt A voor in de afgelopen 2 uur?



Offline v.s. Online

| Key | Event | timestamp |
|-----|-------|-----------|
| u1 | A | t1 |
| u2 | C | t2 |
| u1 | A | t3 |
| u1 | A | t4 |

Hoe vaak komt A voor in de afgelopen 2 uur?



Offline v.s. Online

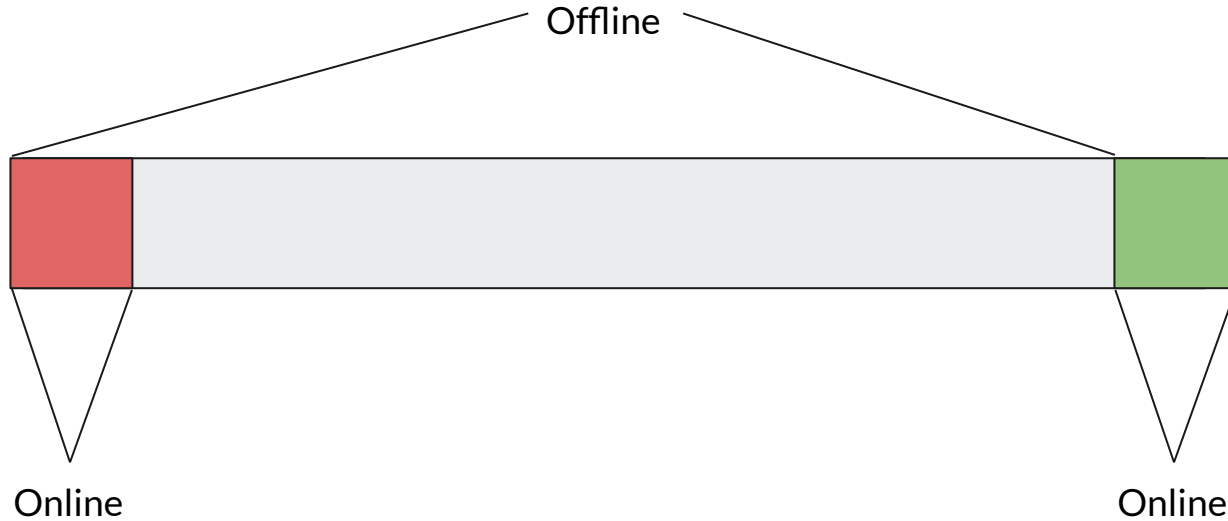
| Key | Event | timestamp |
|-----|-------|-----------|
| u1 | A | t1 |
| u2 | C | t2 |
| u1 | A | t3 |
| u1 | A | t4 |

Hoe vaak komt A voor in de afgelopen 2 uur?

Dit werkt goed, als er niet te veel events zijn, en het timeframe niet te lang is. Dus, hoeveel bookings heeft iemand gemaakt werkt niet met deze methode.



Offline v.s. Online





Overzicht

- Recommender Systems in het wild
- Waar loop je tegen aan?
 - Data Kwaliteit
 - Offline v.s. Online
 - **Monitoren**
 - A/B testing



Monitoren

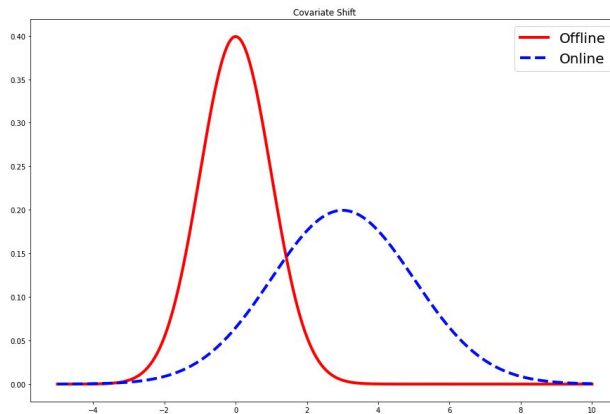
Monitoring van je features

- Covariate shift
 - De distributie van je features is veranderd
- Label set shift
 - De distributie van de labels is veranderd
- Concept drift
 - De relatie tussen de input en output variabelen is verandert

Als je dit detecteert, is het enige dat je kan doen opnieuw je model trainen!

Monitoren

Covariate Shift
 $P(x,y) = P(x)P(y|x)$



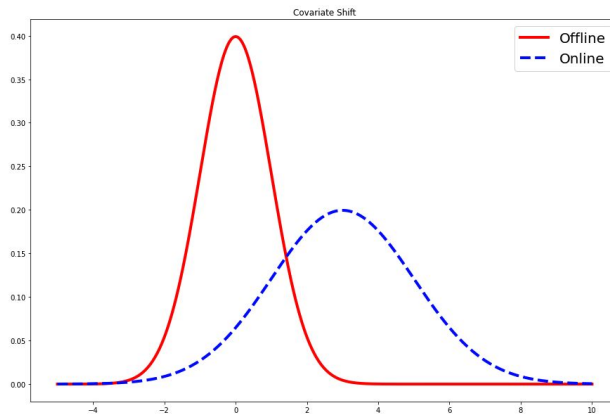
- De functie die je geleerd hebt is niet meer geldig
- Population Stability Index (PSI)

$$\sum_{i=1}^B (\hat{p}_i - \hat{q}_i)(\ln(\hat{p}_i) - \ln(\hat{q}_i))$$

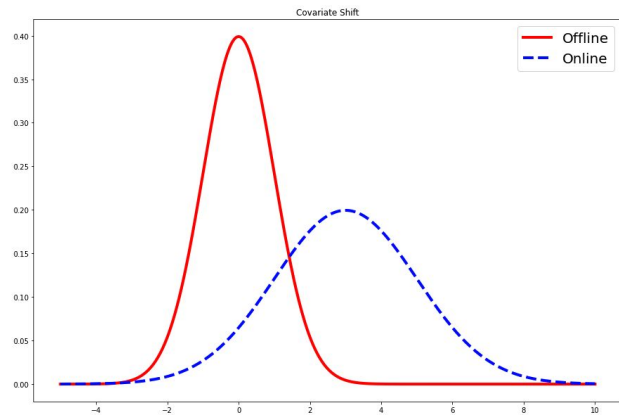
- Gerelateerd aan Kullback-Leibner Divergence

Monitoren

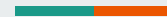
Covariate Shift
 $P(x,y) = P(x)P(y|x)$



- De functie die je geleerd hebt is niet meer geldig
- Population Stability Index (PSI)
- Classificeer
 - Probeer het verschil tussen trainings data en de huidige data te modelleren.



- leidt tot onder of over predictie



Monitoren

Label Set Shift
 $P(x,y)=p(y)p(x|y)$



Monitoren

Concept shift

- De relatie tussen X en Y is verandert
- Monitor de kwaliteit van je online voorspellingen
- Dit is niet altijd makkelijk, want het kan soms best een tijdje duren voordat je het label weet, i.e. iemand annuleert wel of niet.



Monitoren

- Gebruik streams om data te verzamelen
 - voor iedere voorspelling de features, de output, en iets van een identifier
 - Zorg ervoor dat je de features and labels kan reconstrueren
- Automatiseer alerting



Overzicht

- Recommender Systems in het wild
- Waar loop je tegen aan?
 - Data Kwaliteit
 - Offline v.s. Online
 - Monitoren
 - **A/B testing**



A/B Testing

- Waarom wil je A/B testen?
 - Een model werkt nooit in isolatie, i.e. design, copy, aanbod helpen allemaal ook mee bij de vraag of een nieuw aanpassing aan de website succesvol is
 - Vaak is het label dat je leert maar een proxy voor wat je echt wil doen

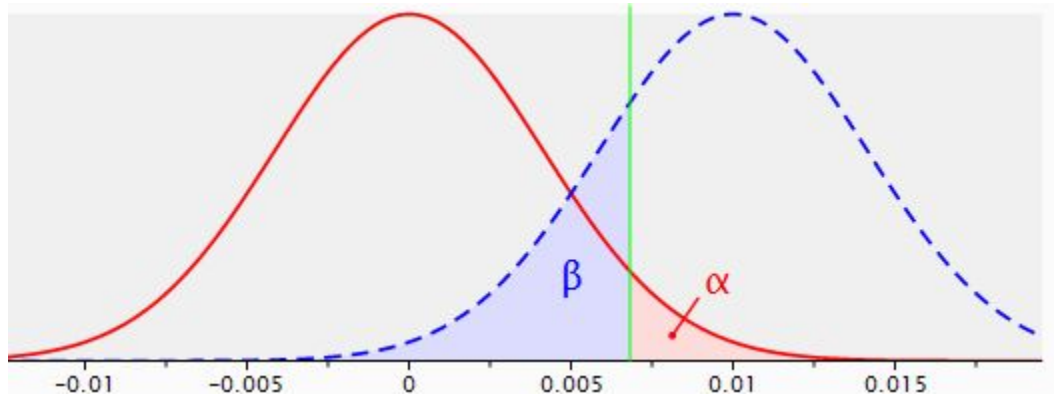


Hoe werkt A/B Testing

Wat wil je testen?

- Meestal is het een binaire variabele (clicks, bookings, streams)
- In een AB test heb je twee **onafhankelijke** groepen
- Je wilt weten of de kans op een click in B hoger is dan in A

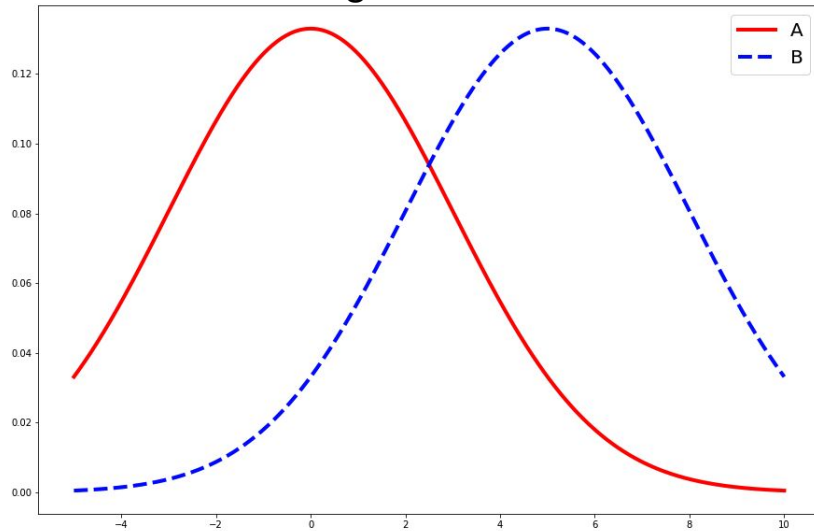
Hoe werkt A/B Testing



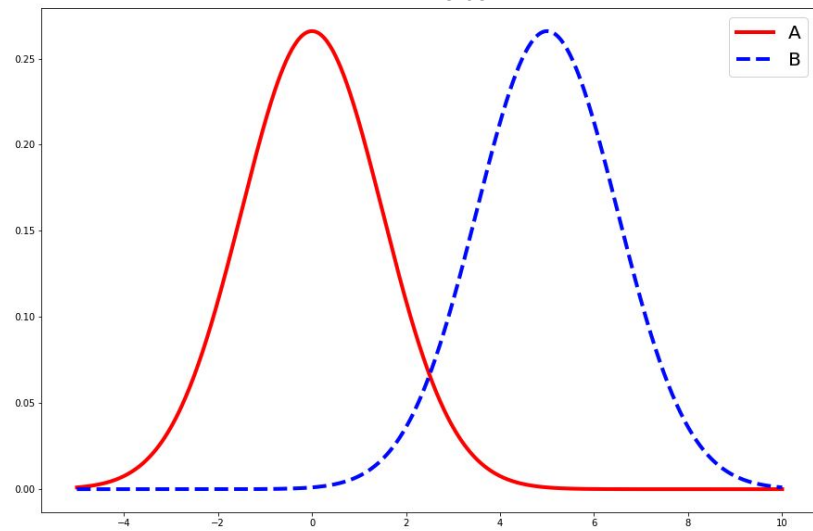


Hoe werkt A/B Testing

Weinig data



Veel data





Hoe werkt A/B Testing

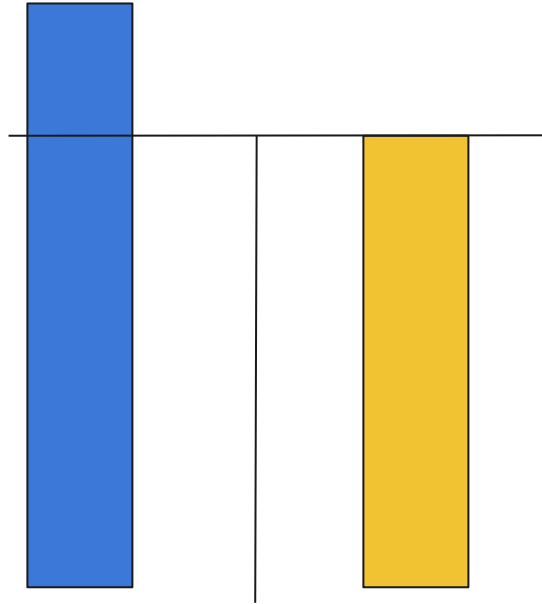
Waar moet je op letten?

- Bereken van te voren hoeveel visitors je nodig hebt
- stop je experiment op het moment dat je dat aantal hebt bereikt

Wat als je een ranker moet testen?



AB testen van een ranker



Zijn er nog vragen?

—