

# Collectieve Intelligentie

Björn Out, Rabie Afqir, Frank Tamer & Gino Pennasilico

# Programma

## Cold Start Approach

- Werking
- Resultaten
- Case example

## Support Vector Machines

- Werking
- Resultaten
- Case example

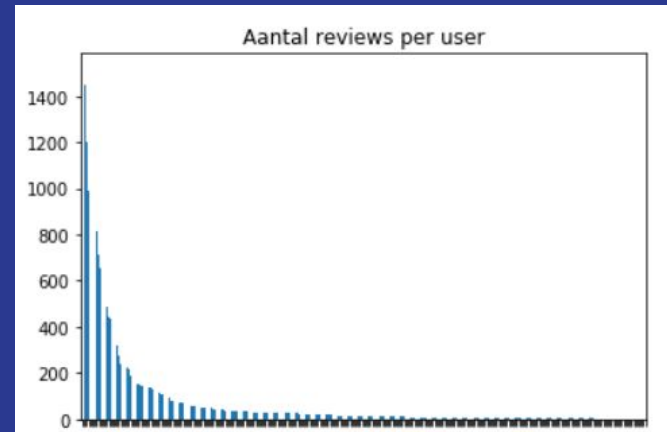
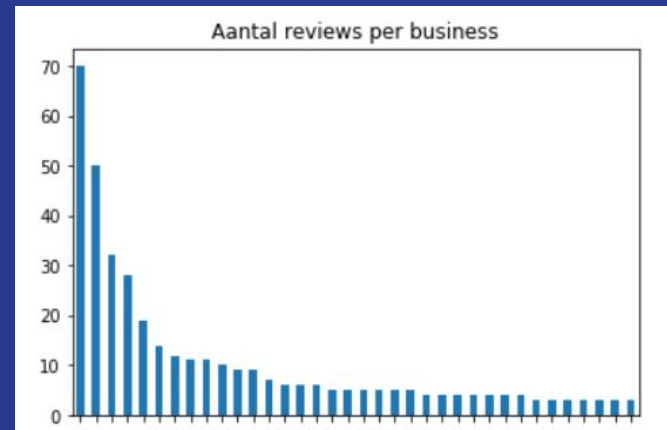
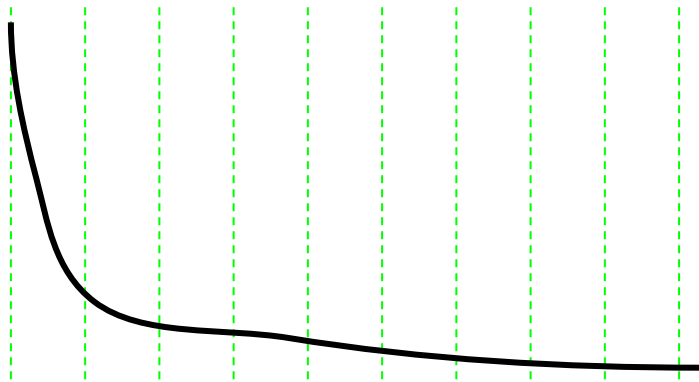
## Item-based Collaborative Filtering

- Werking
- Resultaten
- Case example



# Aanpak: Bij Cold-Start

# De Long Tail in de data



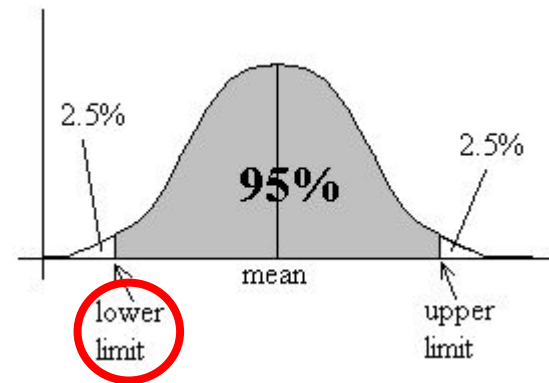
# Hoe werkt de Cold-Start Approach?

Kijk naar de lower bound confidence interval van bedrijven die:

- Boven het minimum aantal reviews zitten
- Categorieën hebben die uniek genoeg zijn

Parameters:

- NumberOfPreds
- RepetitionsOfCategories
- MinimalAmountOfReviews

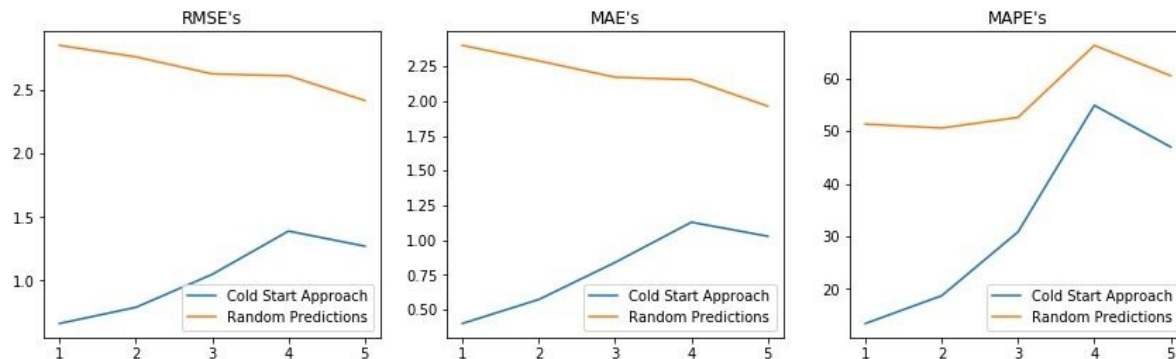


# Resultaten Cold Start Approach

	Business ID	Predicted Rating	RMSE	MAE	MAPE
1	Xg5qEQiB-7L6kGJ5F4K3bQ	4.750960	0.660154	0.401922	13.430810
2	mDR12Hafvr84ctpsV6YLag	4.605711	0.787607	0.574419	18.715428
3	mz9ltimeAly2c2qf5ctljw	4.328782	1.047926	0.840471	30.831568
4	t-o_Sraneime4DDhWrQRBA	3.932575	1.389513	1.129438	54.882167
5	WYw3Uf56DT5lwpaLNNCH5Q	3.764400	1.270705	1.028700	46.979958

Text(1, 1.1, "RMSE, MAE and MAPE's of the predictions (Jaccard Similarity = 0.00365):")

RMSE, MAE and MAPE's of the predictions (Jaccard Similarity = 0.00365):



# Use case Cold Start Approach



# Use Case Cold Start Approach







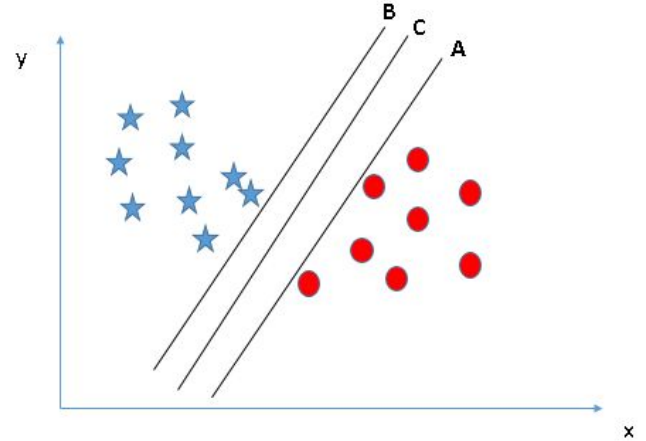
Niet slecht, maar niet  
persoonlijk



# Aanpak: SVM

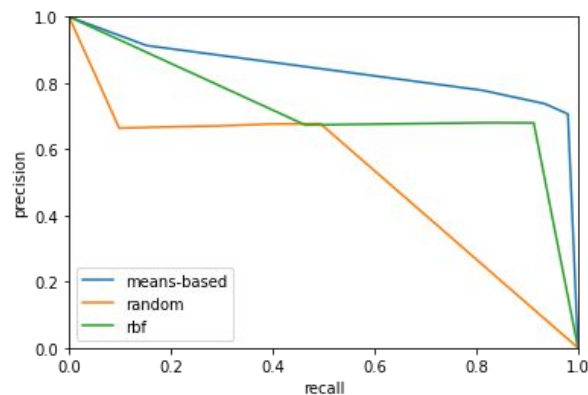
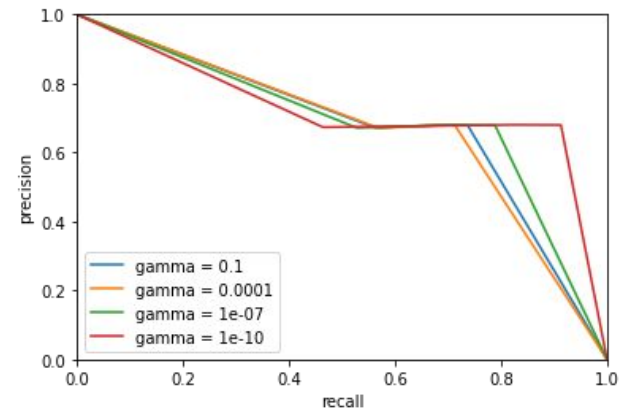
# Hoe werkt een SVM?

- Machine learning
- Classificatie (threshold)
- Hyperplane
- Trainen op goed gereviewde businesses
- Selectie via Jaccard



# Evaluatie

- Precision-recall (classificatie)
- Optimale parameters
- Baselines
- Variatie in bedrijven



Jaccard: 0.0, baseline 0.0596

# Carter

Cleveland, OH




# Goede reviews



# Suggesties





Gevarieerd, maar  
mean item-based  
werkt beter



# Aanpak: Item-Based CF

# Hoe gebruiken wij Item-Based CF?

- Een bepaalde stad gekozen
- Adjusted cosine similarity tussen businesses & utility matrix
- Input: user\_id
  - -> predicted rating



# Samantha

Brooklyn, NY



# Use Case Item based

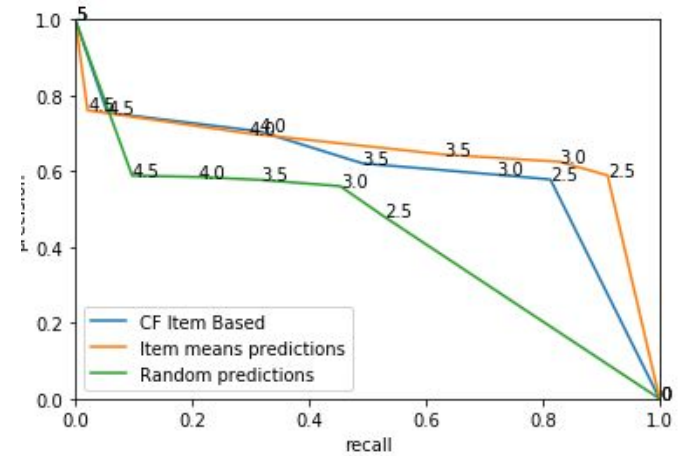


- Samantha bekijkt Chinees restaurant 'Hong Kong King Buffet' in Brooklyn
- Hoe krijgt zij haar aanraders?
  - Lijstje van business id's met 10 hoogste predicted ratings



# Evaluatie

- Evaluatie: MSE
  - Item based
  - Item mean
  - Random



Methode	MSE
Collaborative Filtering item based	3.26
Random predictions (Baseline)	4.36
Item means predictions (Baseline)	1.98

Prima, maar Item  
Means predictions  
relatief beter (en  
simpeler)

# Afweging & conclusie

- Precision het meest belangrijk
- Recall ondersteunt in variatie
- Jaccard
- Item-based means
- Cold-start

