

Machine Learning for a Smarter Online Review Platform

Yuan Si



Why do We Care?



Customer

TOO MANY Reviews

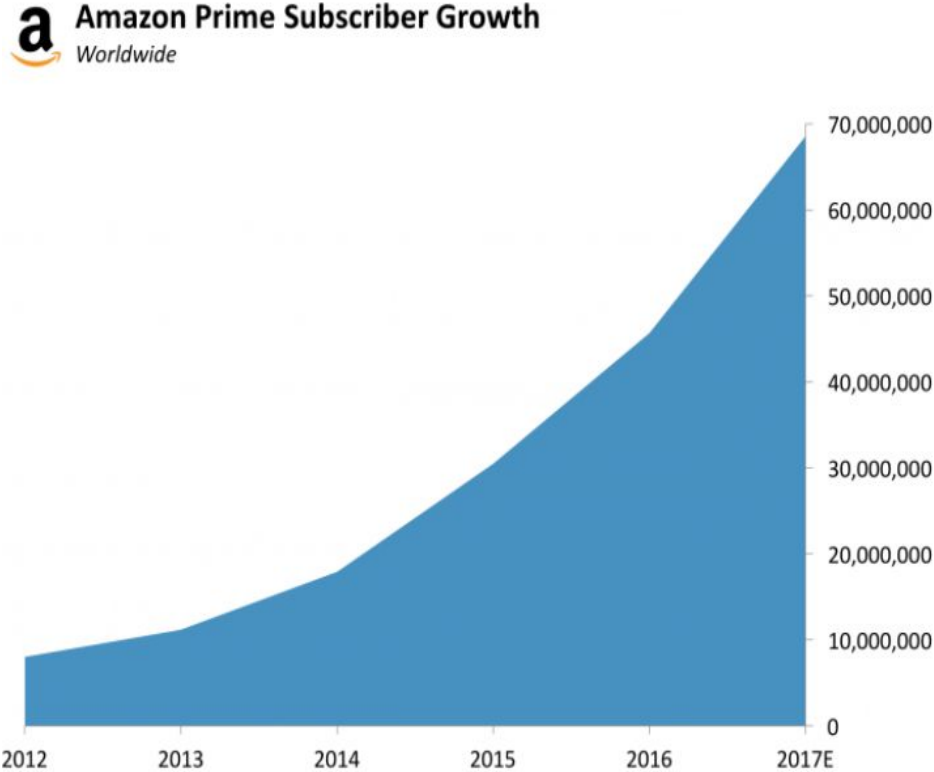
Business

HARD to Get Reviews
($<1\%$)

A smarter review platform !

- Summarize the reviews
- Easy to collect reviews from customers

Data : Amazon Customer Reviews on Yoga Pants



★ **1,659 Full Customer Reviews**

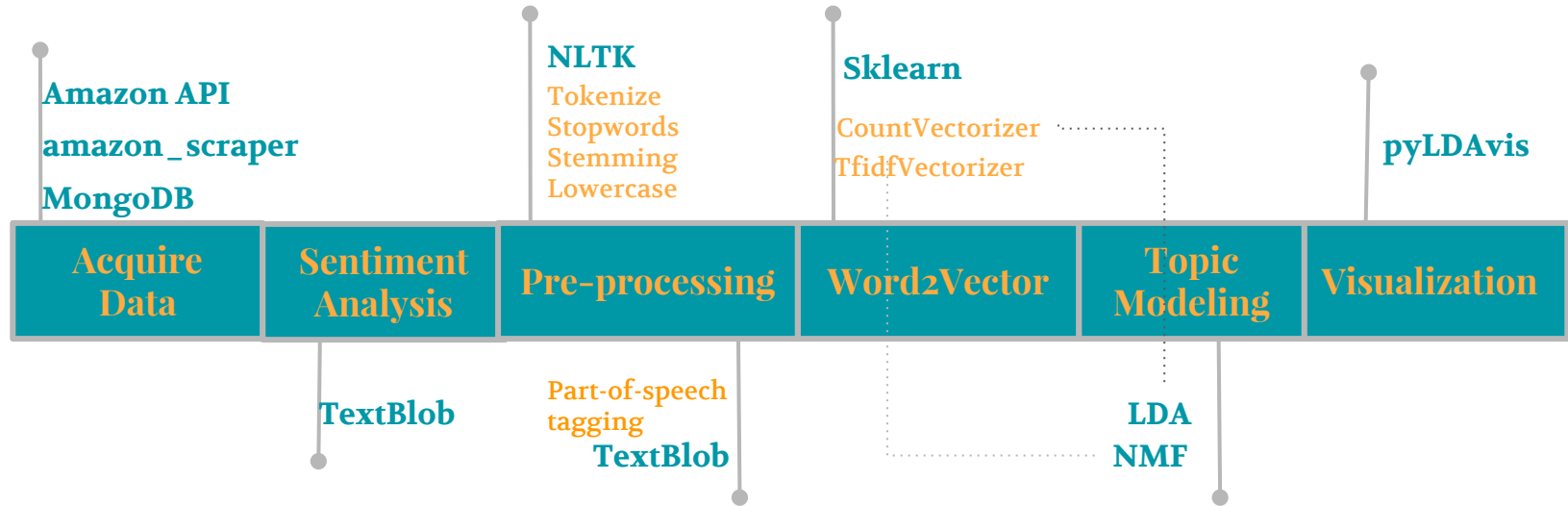
★ **Top 4 Best Sellers**

★ **Different styles**



* Amazon accounts for 43% of US online retail sales

Project Workflow



Complications & Solutions

Frequent Words

- ★ Stopwords
- ★ Vectorize (max_df)

Unrelated Topics

- ★ Nouns
- ★ Adjs

Stopwords

{ re, 's, 'm, n't, 're..., little, would, leggings, yoga, pants, really, pair }

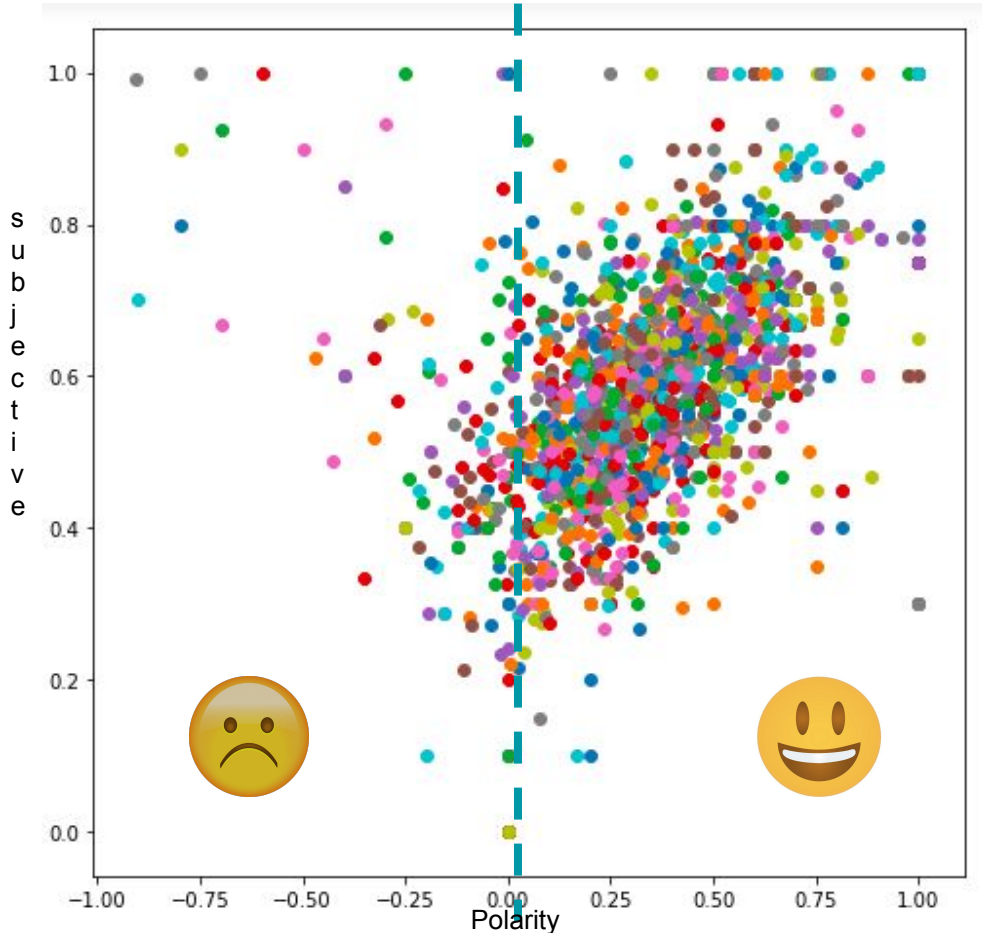
Nouns

{ fit, color, quality, legs, giraffe }

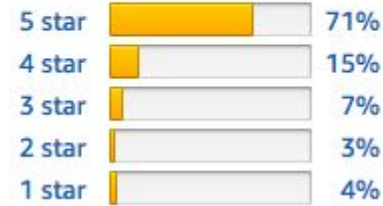
Adjs

{ comfortable, perfect, soft }

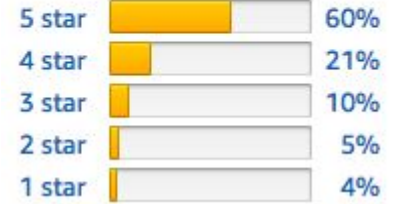
Sentiment Analysis: How do customers feel?



★★★★☆ 5,647
4.4 out of 5 stars ▾



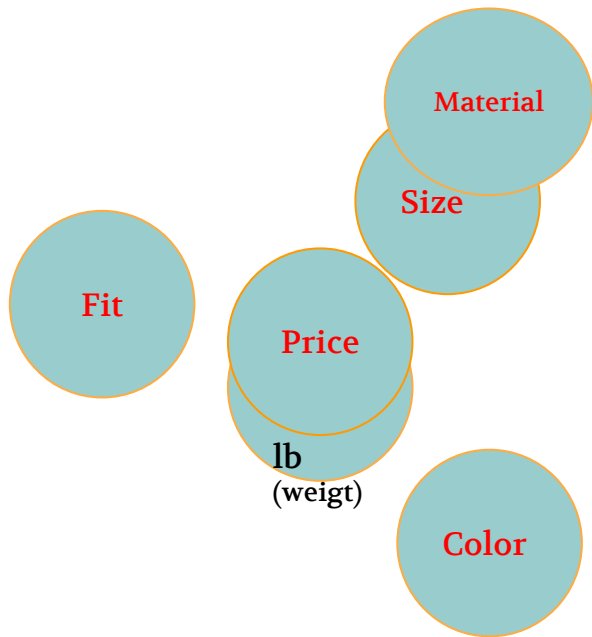
★★★★☆ 3,171
4.3 out of 5 stars ▾



- *Customer reviews are generally positive and subjective.*
- *Review score does effectively reflect how customers feel.*

What customers really care about: Topics in nouns

LDA



NMF



Topic #0:
material waist price brand fabric

Topic #1:
fit sizing good fit sizing

Topic #2:
quality **price** time workout capris

Topic #3:
color fabrics colors fabric workout

Topic #4:
size price waist pairs bit

* words in red show up in both models

How they describe the products: Topics in adjs

LDA



NMF

Topic #0:
Good small big small big box

Topic #1:
Great soft soft fast great fast

Topic #2:
Comfortable ultra comfortable

Topic #3:
Fit soft black great soft favorite

Topic #4:
Perfect small big small big box



* words in red show up in both models

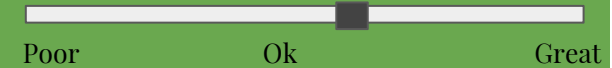
Business Application: A smarter review platform



Overall fit: *based on user review*



Overall quality: *based on user review*



Material: *based on user review*



Color: *based on user review*



What our customers like about this product?

---soft, comfortable

Namaste!

