

PERSONALIZING

SEARCH RESULTS IN REAL-TIME

The image displays four screenshots of e-commerce websites, each showing search results in real-time:

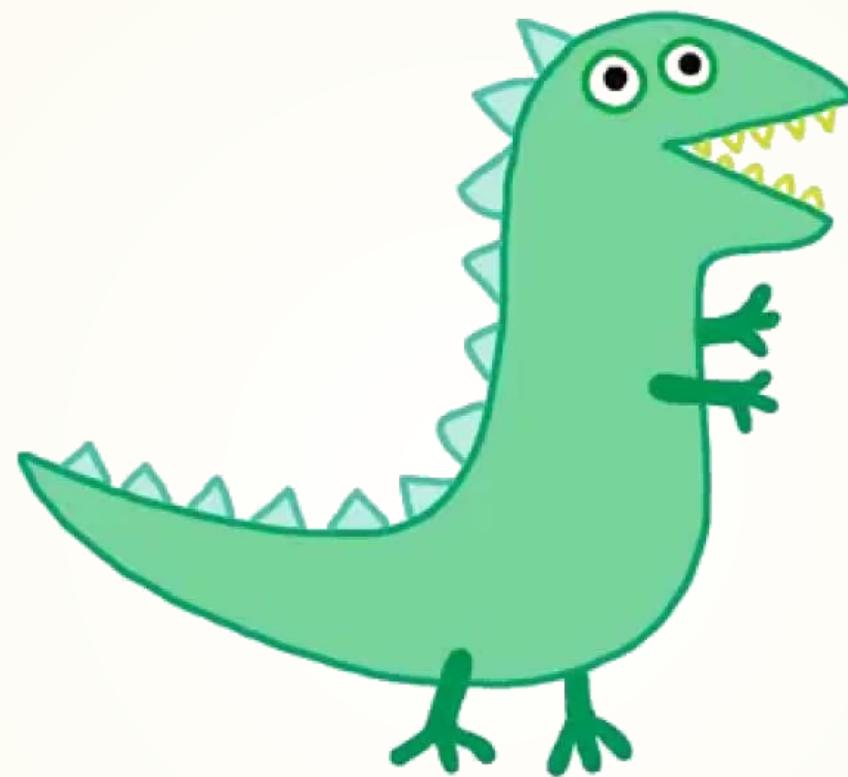
- Fountain Pen Store:** Shows search results for "pen", with filters for category (Accessories, Bottled Ink, Bottom Shelf, Brush Pens, Converters, Dip Pens), price (Under \$20.00, \$20.00 - \$50.00, \$50.00 - \$100.00, \$100.00 - \$200.00, \$200.00 - \$500.00, \$500.00 & up), brand (Aston Leather, Aurora, BENU, Cartan, Clairefontaine, Col-o-ring), and color.
- Yarn Store:** Shows search results for "yarn", with filters for category (All Categories, Knitting Kits, Knitting Leaflets, Colour Packs, Crochet Yarn, Crochet Kits, Value Pack), brand (Stylecraft, King Cole, Rowan, Sirdar, James C. Brett, Scheepjes), and color.
- Clothing Store:** Shows search results for "SEARCH RESULTS", with filters for category (NEW, CLOTHING, BLUE LIFE™, DRESSES, SWIMWEAR, SALE, DESIGNERS, STORES), color, size, brand, and price.
- Furniture Store:** Shows search results for "FLASH SALE" (June 05 SEE DEALS), with filters for category (All Categories, Bedroom Package, Chair, Dining Room Set, Sectional, End Table, Ottoman), material (Chenille, Fabric, Faux Concrete, Faux Leather, Foam, Glass), and color.

Grebennikov Roman / findify.io / @public_void_grv / grv@dfdx.me

ABOUT FINDIFY

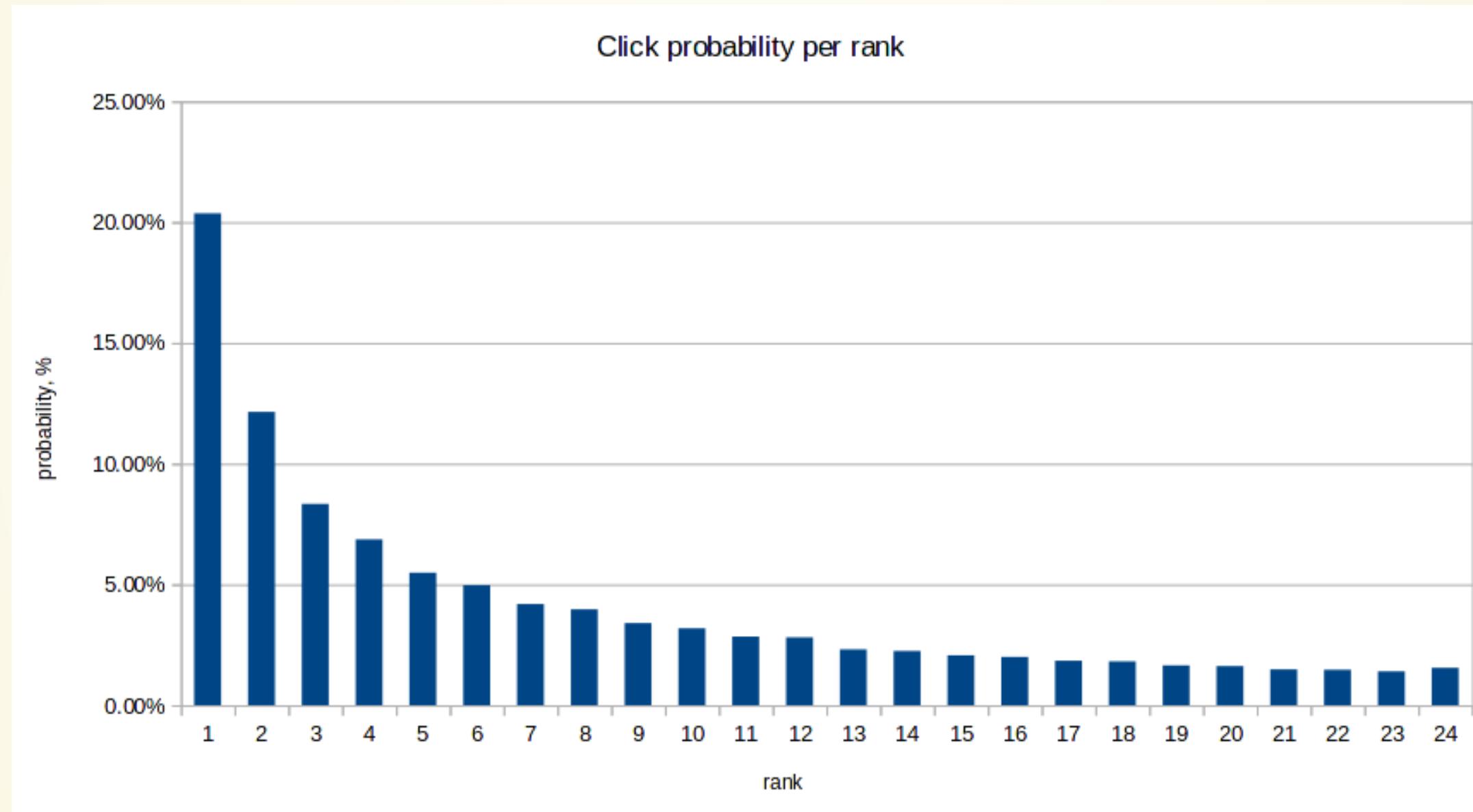
- white-label eCommerce SaaS search
- 1500 stores, 20M products
- 50M customers per month

FINDIFY IN 2014



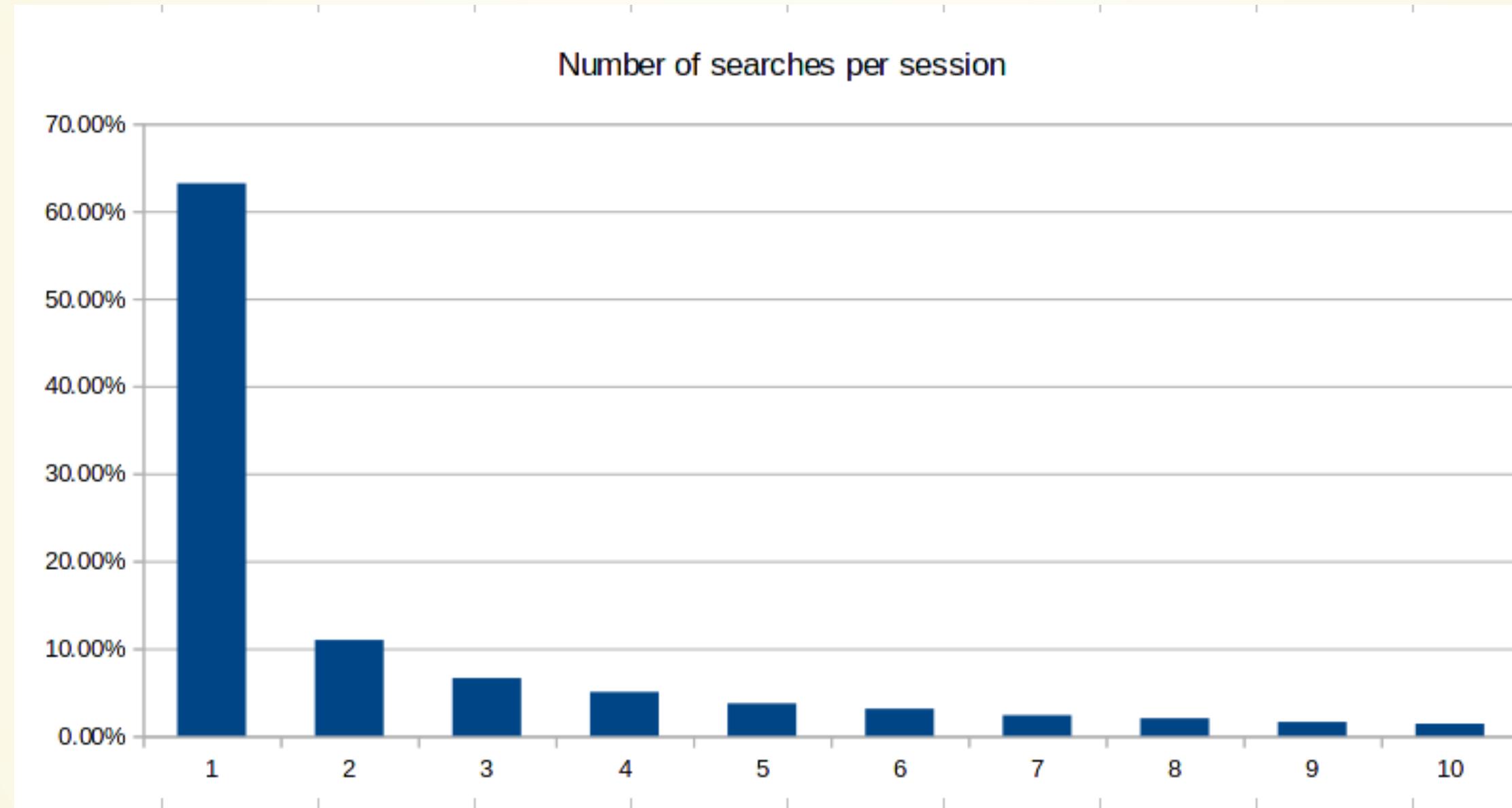
- UI-focused Shopify search addon
- Backed by ElasticSearch
- Nothing special about product ranking

RANKING IS IMPORTANT



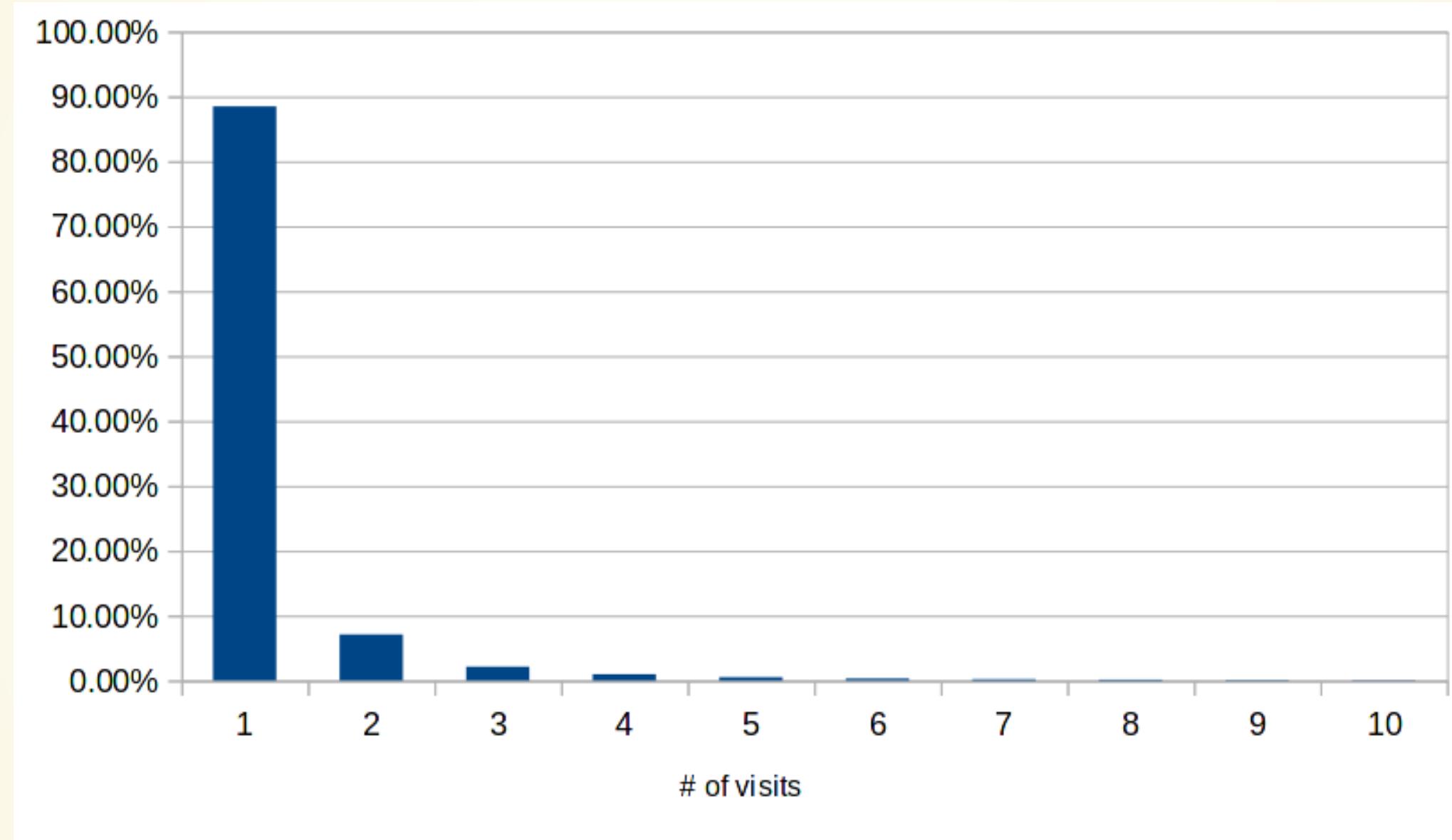
nobody is scrolling down

RANKING IS IMPORTANT



no second search

RANKING IS IMPORTANT

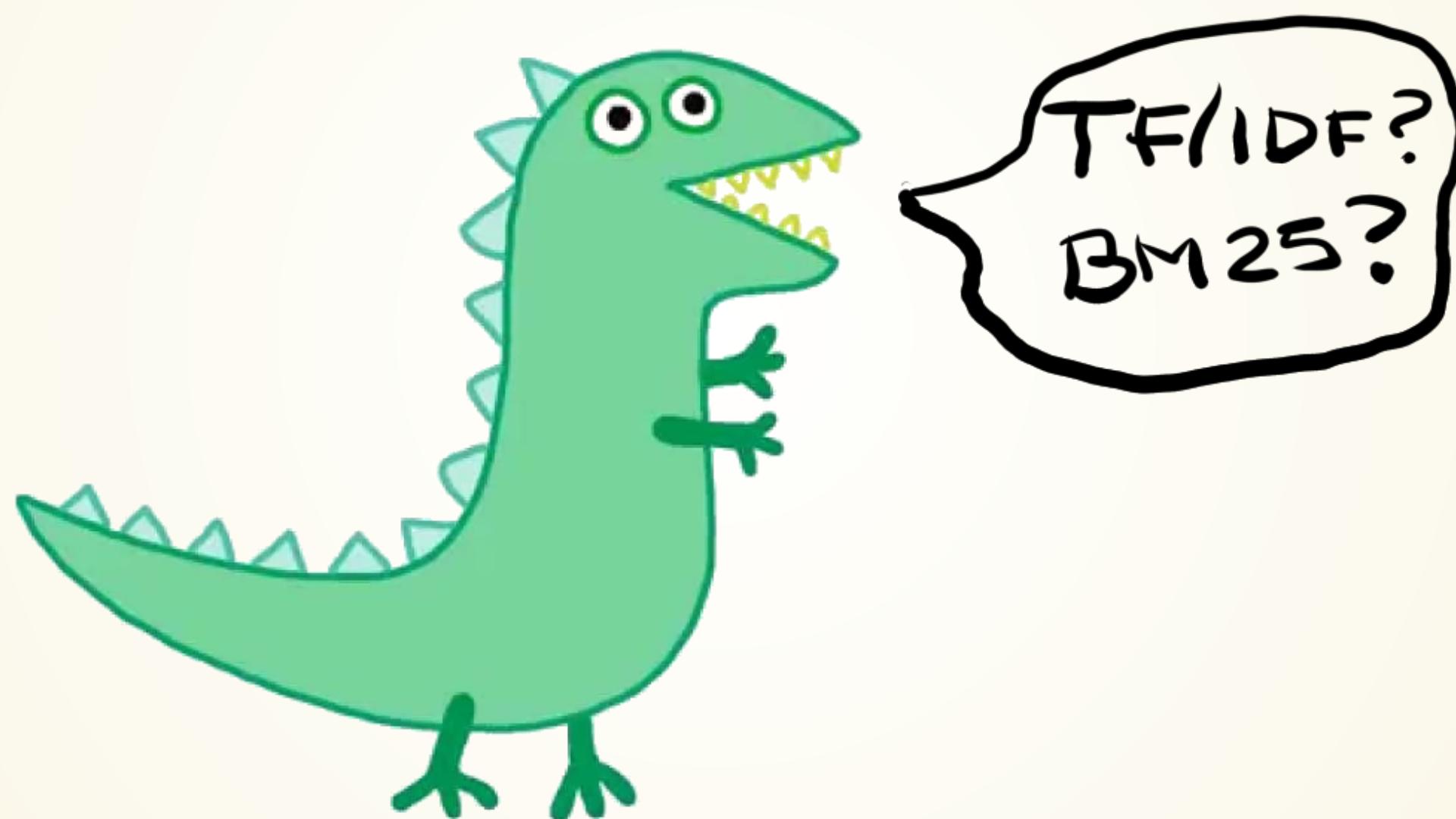


no second visit

TYPICAL CUSTOMER SESSION

1. Arrive on a landing/product page ^(0s)
2. Click on product collections ^(+10s)
3. Make a search ^(+20s)
4. Leave forever ^(+30s)

BETTER RANKING



BETTER RANKING?

$$Rel_q = score + c_1 \cdot P + c_2 \cdot P_q$$



TF/IDF?
BM25?

AI ML (LINEAR REGRESSION)

Algorithm	Conversion	AOV
Elasticsearch	baseline	baseline
Regression	+3.1%	+2.5%

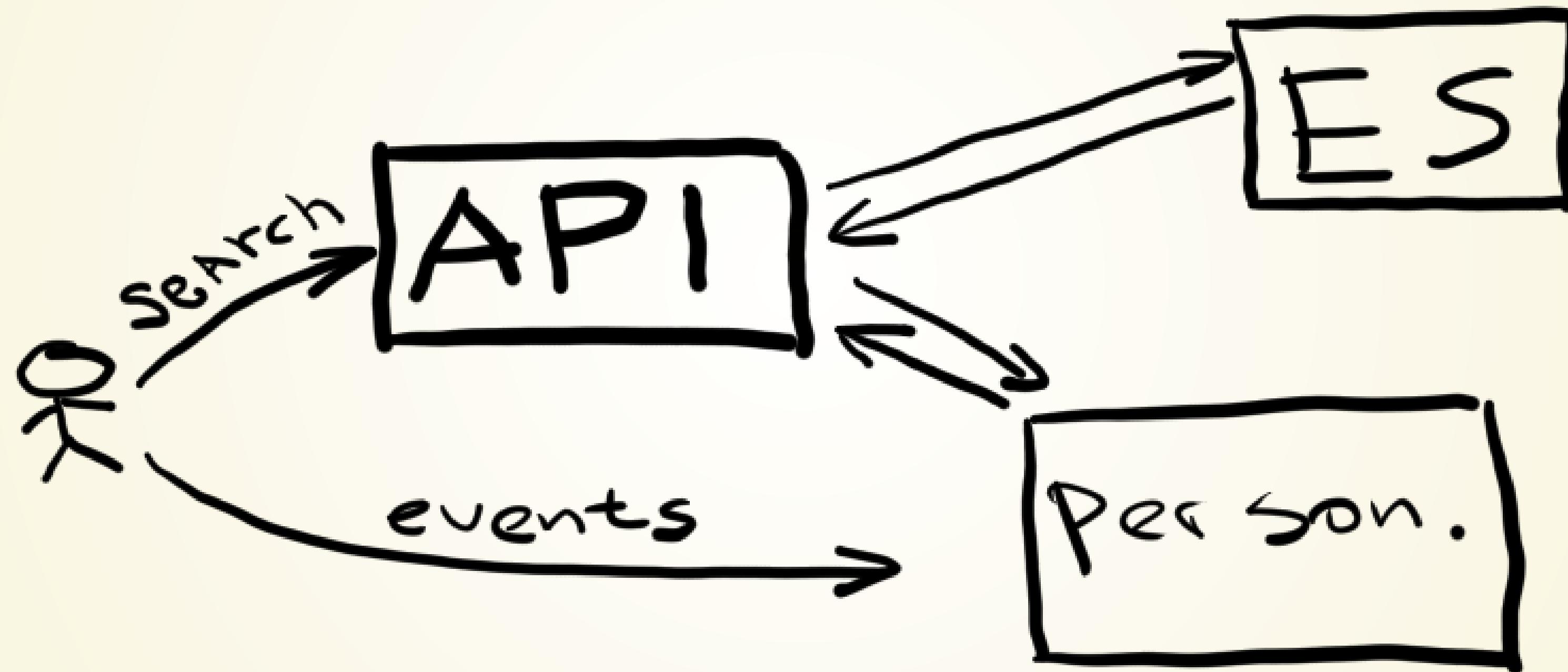
REINVENTING THE WHEEL

- Learn to Rank
- LambdaMART
- XGBoost/LightGBM/CatBoost

ELASTICSEARCH INTEGRATION



ELASTICSEARCH INTEGRATION



TRAINING

- Historical click/purchase data
- Model per merchant
- Optimize for NDCG, watch for conversion

FEATURE GROUPS

- **search**: # of terms, # of filters
- **product**: price, # of pageviews
- **variant**: color, size
- **current session**: price sensitivity, # of searches
- **historical sessions**: # of sessions
- **product and search**: # of pageviews within context
 - + different time windows

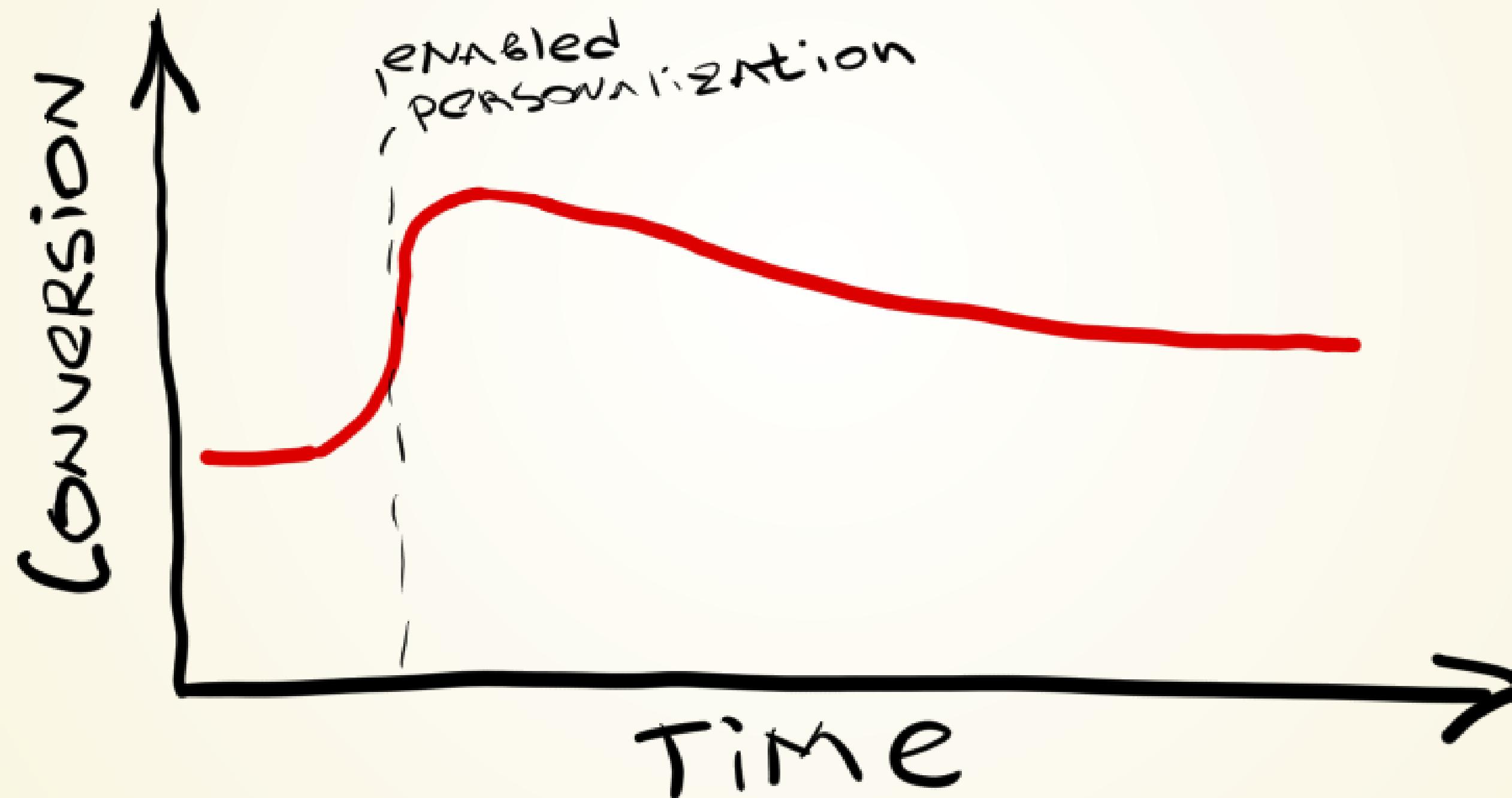


Findify

ML

Search

MIXED RESULTS



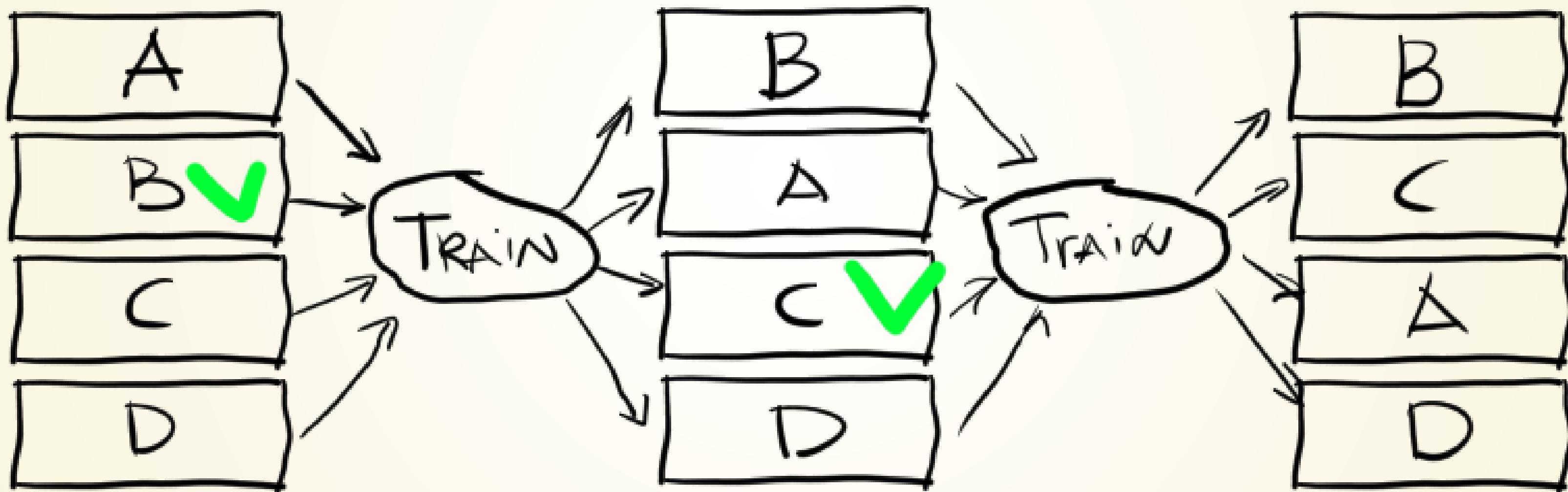
MIXED RESULTS

Algorithm	Conversion	AOV
Elasticsearch	baseline	baseline
Regression	+3.1%	+2.5%
LMART v1	+6.1% ^(+8.1%)	no data

TRAINING ISSUES

- Historical click/purchase data
- Model per merchant
- Optimize for NDCG

POSITIVE FEEDBACK LOOP



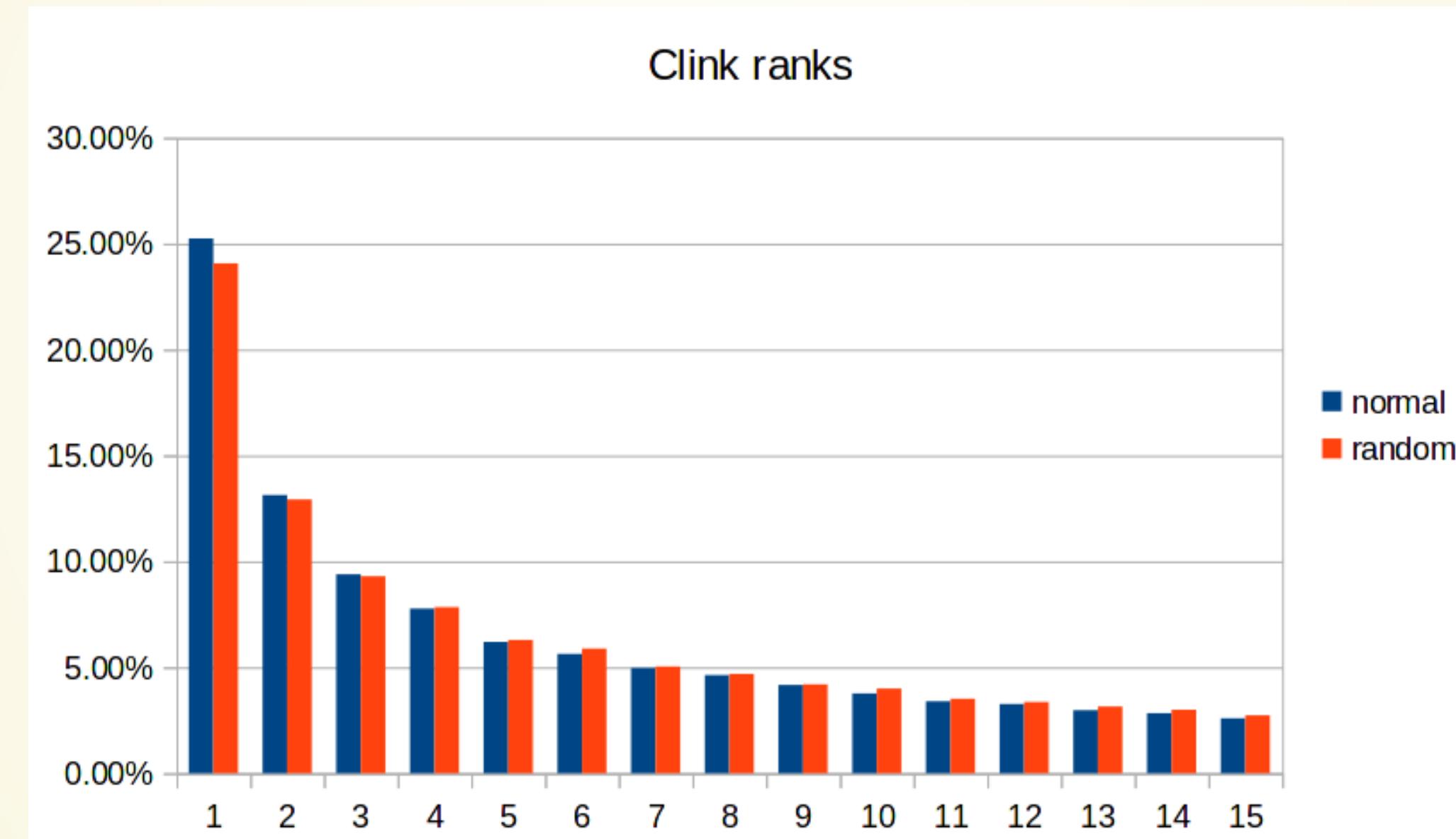
POSITIVE FEEDBACK LOOP



POSITION BIAS

CUSTOMERS ARE CLICKING ONLY ON FIRST PRODUCTS

RANDOM RANKING



RANDOM RANKING

Algorithm	Conversion	AOV
Elasticsearch	baseline	baseline
Regression	+3.1%	+2.5%
LMART v1	+6.1% <small>(+8.1%)</small>	no data
Random	-2.8%	-1.3%

POSITION BIAS

L. Li, W. Chu, J. Langford, R. Schapire. 2010.
A contextual-bandit approach to personalized news article recommendation.

- Exploration and exploitation segments
- Un-biasing the training data

EXPLORATION SEGMENT

- tiny segment, 0.1-1% of traffic
- first page is shuffled
- used for training

TRAINING ISSUES

- ~~Historical~~ Unbiased click/purchase data
- Model per merchant
- Optimize for NDCG

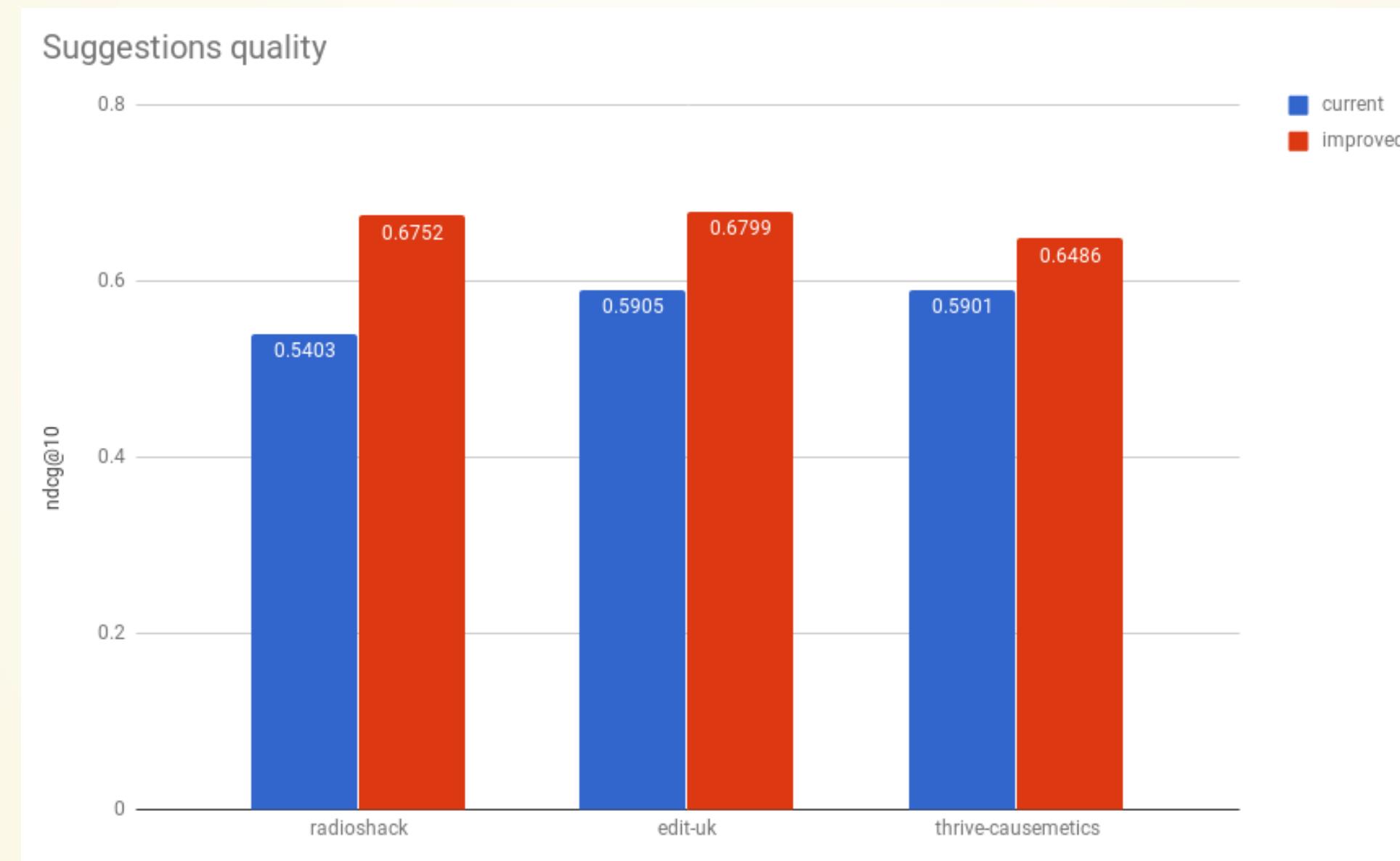
MODEL PER MERCHANT

- Low-traffic merchants
- Onboarding and data collection time
- Sacrificing ranking for "Exploration segment"

SUGGESTIONS HACKATHON

- Replace heuristics with ML
- Simpler problem than search
- All features are language-specific
- small, medium, large merchant

BETTER SUGGESTIONS?



MODEL TRANSPLANT

from large-traffic merchant to small-traffic:



GENERIC MODEL

- More training samples
- More diverse dataset
- No need for per-merchant data collection
- All features need to be scaled

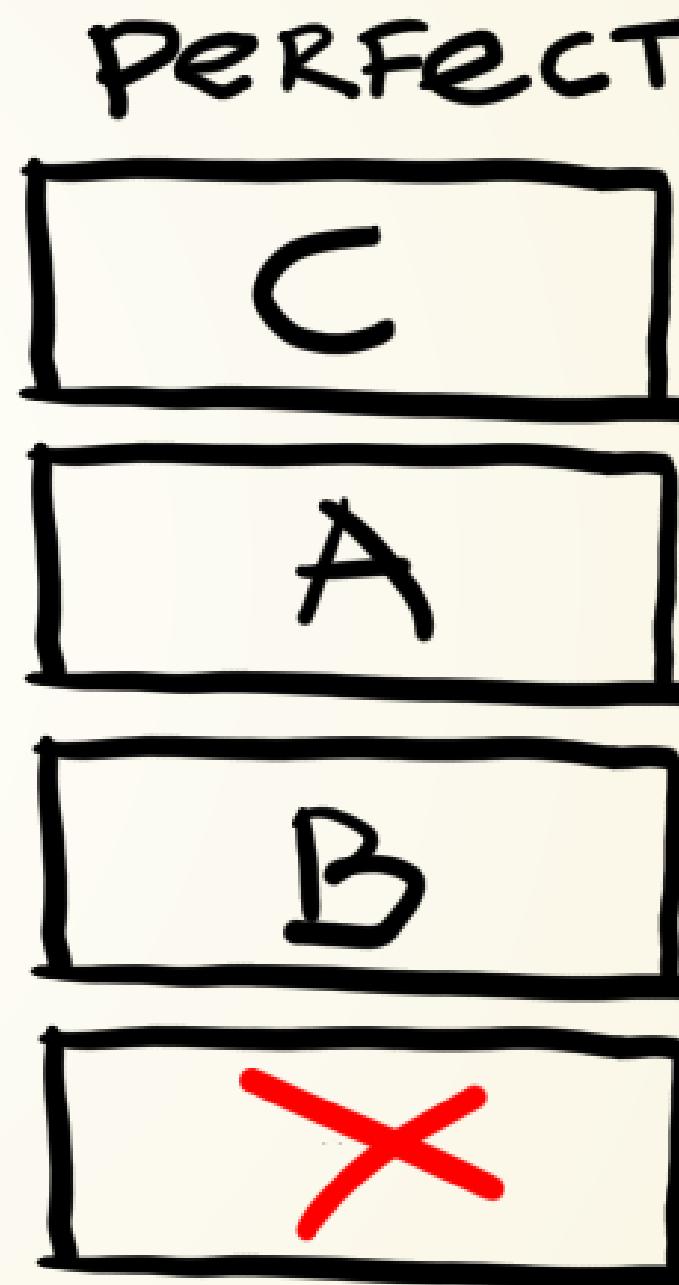
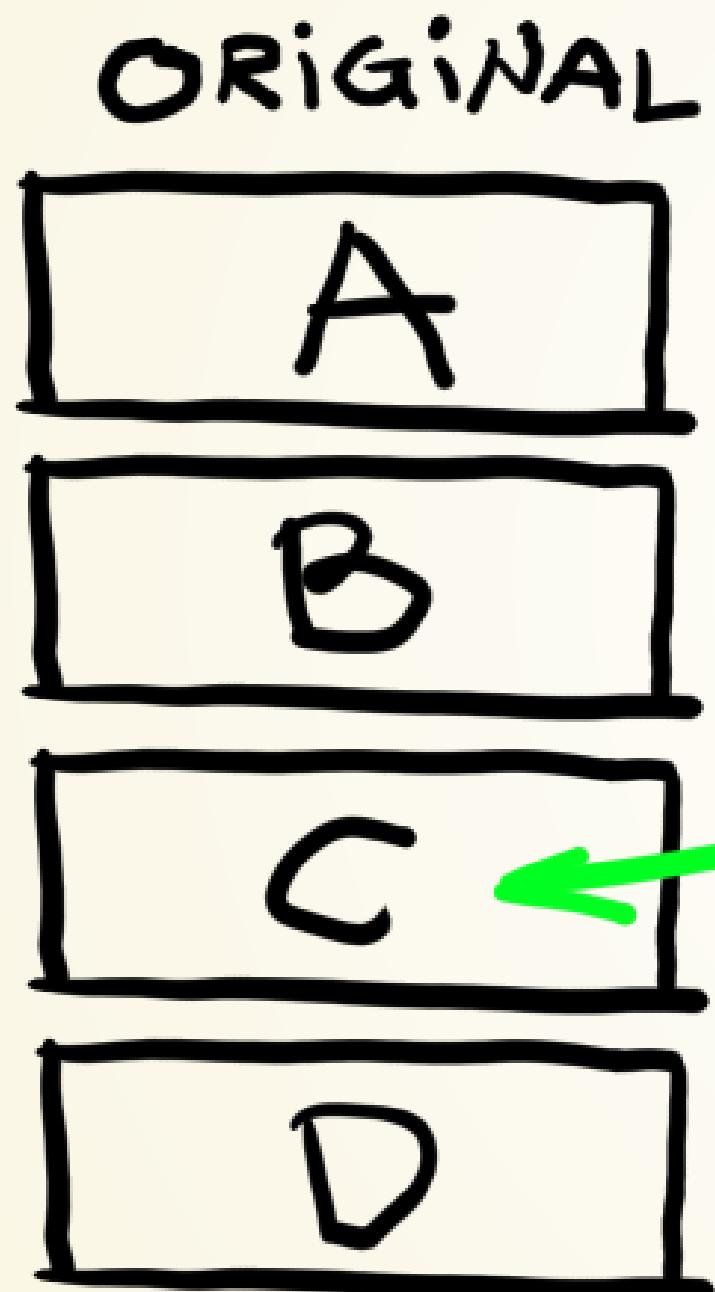
TRAINING ISSUES

- ~~Historical~~ Unbiased click/purchase data
- ~~Model per merchant~~ Generic model
- Optimize for NDCG

NDCG

- 1.0 - good, 0.0 - bad, 0.4-0.7 - normal
- compares perfect ranking with real
- what is a perfect ranking?

PERFECT RANKING



STANLEY BONG ISSUE

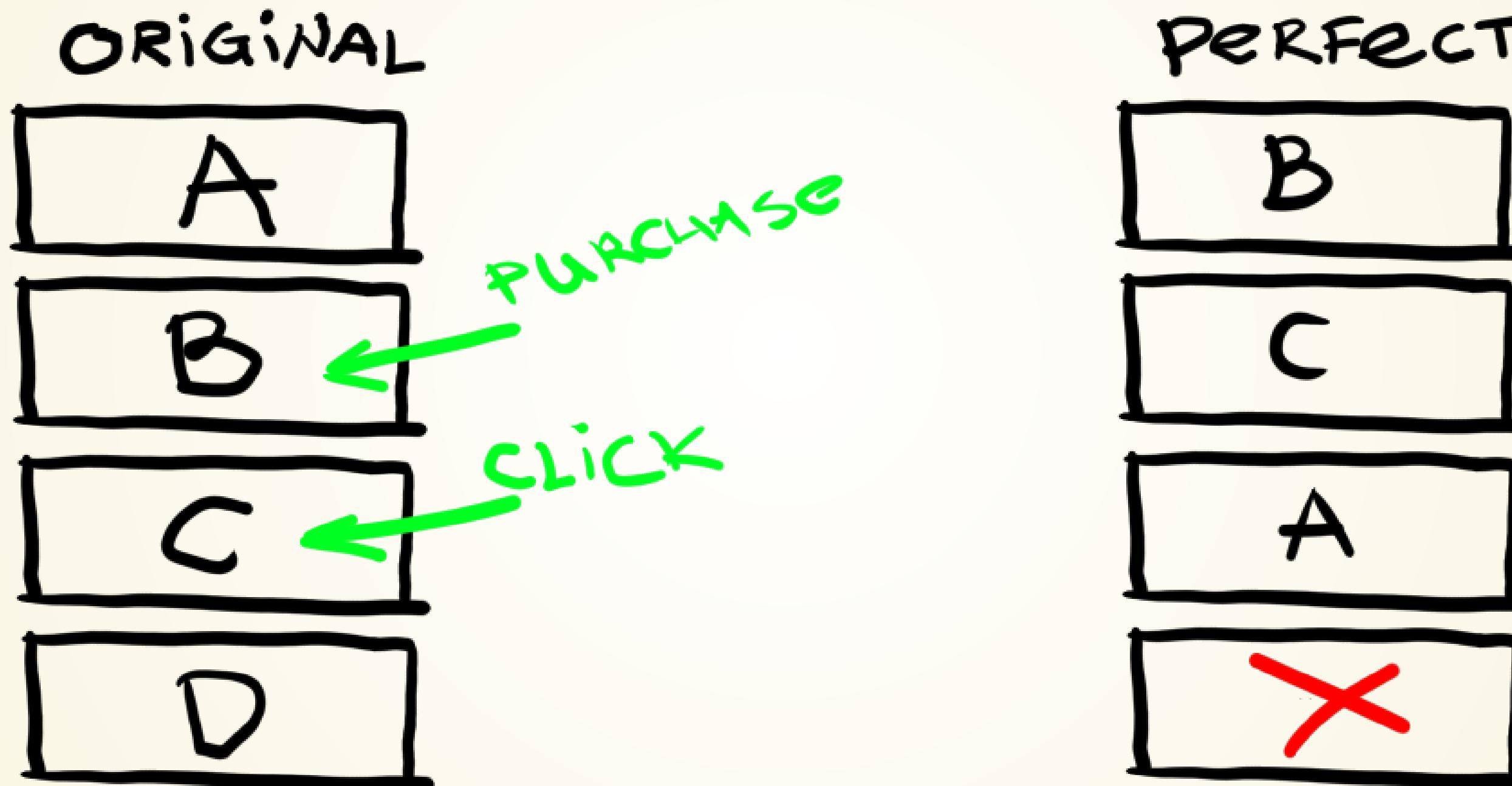
- Rank improved from #20 to #1
- Never bought
- Costs 3500\$

STANLEY BONG ISSUE



over-optimized for clicks

PERFECT RANKING



TRAINING ISSUES

- ~~Historical~~ Unbiased click/purchase data
- ~~Model per merchant~~ Generic model
- Optimize for NDCG (with proper weights)

RESULTS

NDCG WITH PERSONALIZATION

Algorithm	NDCG <small>(offline)</small>
Random	0.544
Popularity	0.578
Elasticsearch	0.601
Regression	0.615
LMART v1	~0.621
LMART unbiased	0.635

NDCG AND BUSINESS METRICS

Algorithm	NDCG	CTR	Conversion	AOV
Elasticsearch	0.601	baseline	baseline	baseline
Random	0.544	-7.1%	-2.8%	-1.3%
Regression	0.615	-1.1%	+3.1%	+2.5%
LMART v1	~0.621	no data	+6.1%	no data
LMART unbiased	0.635	no data	+8.1% ^(est)	no data

CONCLUSION

- Better ranking = more \$\$\$
- A lot of pitfalls



- Multiply development estimates by π



That's all folks!