

Recommendations at Rappi



Ariel Wolfmann
Head of Data - CPGs



About me!



DISCLAIMER:

It will NOT be a deep technical talk!

Rappi

Data Product

- 👉 A product that facilitates an end goal through the use of data .
 - Would not be possible without data
- 👉 You can't just buy good data products. - "I want to buy that data product".
- 👉 "Data is the new oil"
 - It needs to be transformed into a product before it's valuable.
 - Once you've used oil, it's gone. Data is persistent, you can use it over and over again



Product / Data Fit Challenge

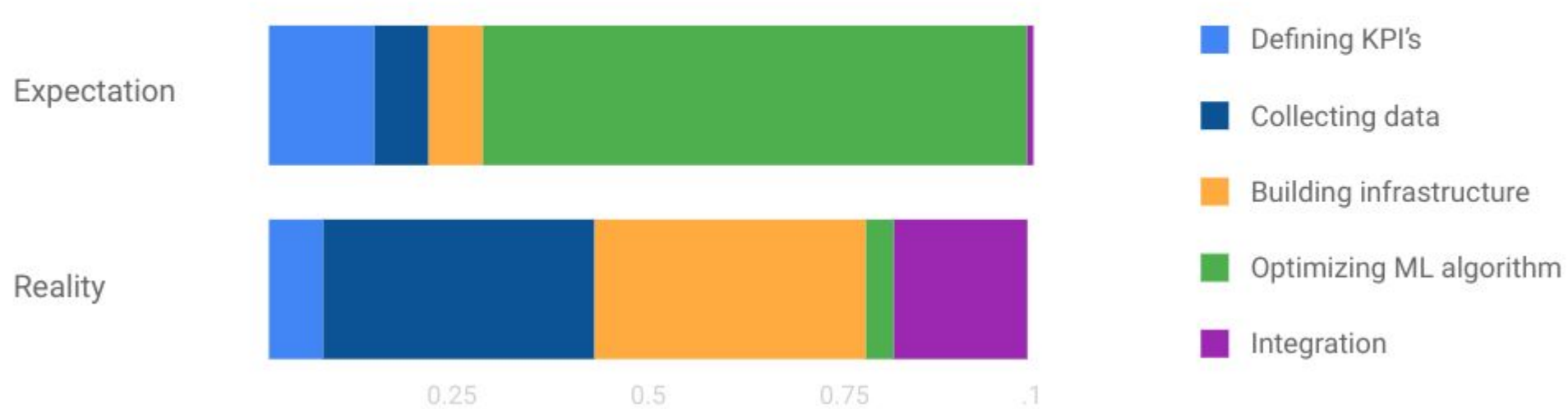
When building a data product, you are tackling at least **two sources of uncertainty** simultaneously: The market & the data

- 👉 Predictions aren't always correct.
- 👉 You can't guarantee in advance the level of performance of a model.
- 👉 Maybe you don't have enough data or that there's not enough signal, but you don't know that in advance.
- 👉 Even if the model performs well on training data, its performance in production may vary significantly.



Uncertainty

Effort Allocation



Big Data

👉 One of our main challenges is Data Collection.



Data Quality

- 👉 **Completeness** — Do we have some sort of data?
- 👉 **Validity** — Is the data that's there the right kind of data?
- 👉 **Accuracy** — Does the data reflect the real-world objects they represent?
- 👉 **Integrity** — Are the relationships between entities consistent?
- 👉 **Consistency** — Is data consistent across systems? Are there duplicates?
- 👉 **Timeliness** — Is the data there when you need it?



MVP

- 👉 It doesn't need AI or sophisticated algorithms to make products feel 'smart' — not initially



(prototype)



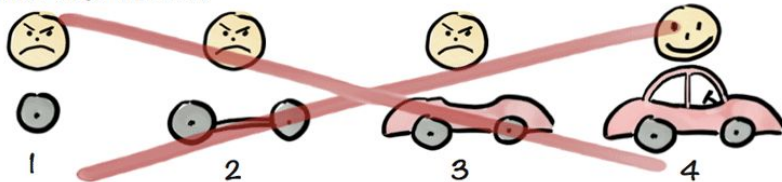
(minimum viable product)



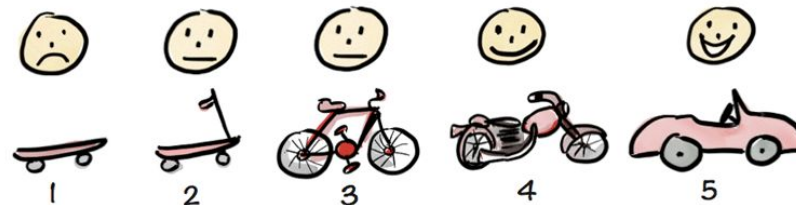
(product)



Not like this....



Like this!

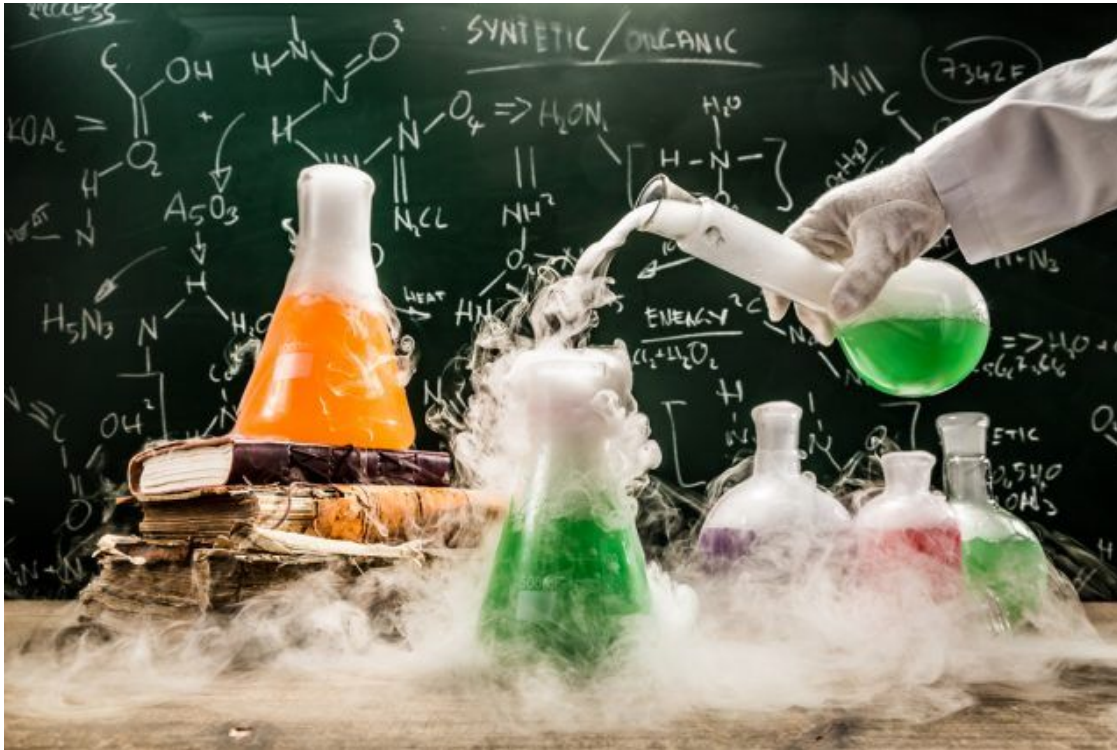


Time is critical

- 👉 Start simple and establish a **Baseline**
- 👉 The key to successful data projects is getting models into production **early**, adjusting, and pushing new models often.
 - It's not trivial to operationalize ML & AI.
- 👉 How can we estimate a research task - **Timebox**



Experimentation



- 👉 Define **GOALS**. (Product discovery?)
- 👉 Turn ideas into **hypothesis** to validate.
- 👉 Design **experiments** that let you validate each hypothesis.
- 👉 Prove your theory / intuition with science.
- 👉 A/B Testing

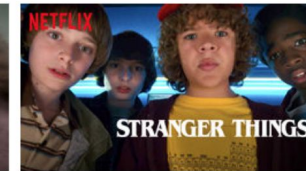
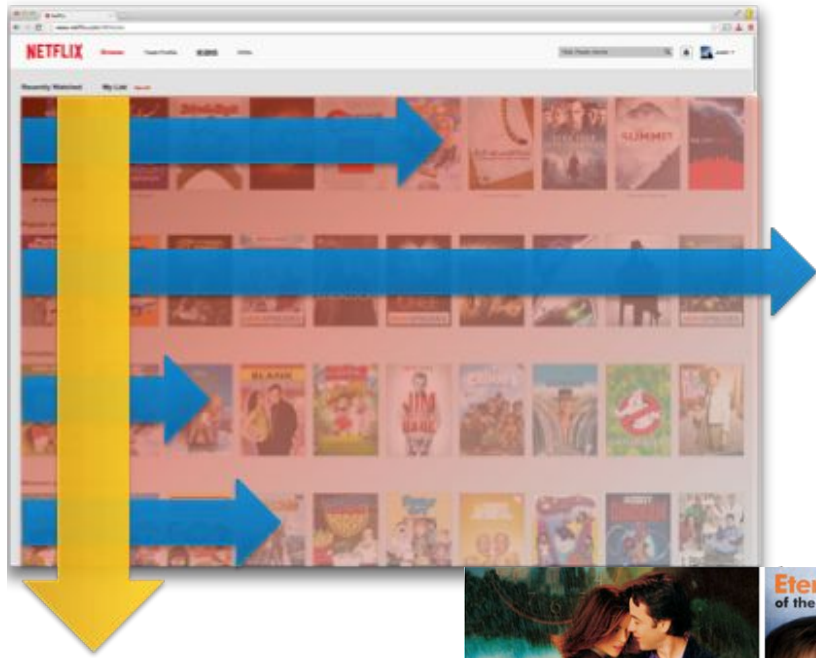


Metrics

- 👉 Analytics Infrastructure is a must!
- 👉 Online Metrics - KPIs - Biz Metrics
- 👉 Offline Metrics - Modeling Metrics
 - Accuracy, Precision, Recall.
- 👉 Biz metrics should drive modeling metrics



Recommendations at Netflix



Netflix Prize



- 👉 U\$S 1 Million prize championship.
- 👉 Recommendations system.
- 👉 Predict accurately movies customers would like.
- 👉 Netflix never use the winner solution because due to engineering costs.



Humans behind ML

EL PAÍS | ECONOMÍA | SOCIEDAD | CULTURA Y ESPECTÁCULOS | EL MUNDO | DEPORTES | PSICOLOGÍA | CONTRATAPA | SECCIONES

Página 12 Edición Impresa | 07 de junio de 2018
Hoy: **UNIVERSIDAD NO**

CULTURA Y ESPECTÁCULOS
10 de abril de 2018

Un trabajo soñado para los fanáticos del streaming

Netflix paga por ver y etiquetar series y películas

Twitter Facebook Email Print



Arts and Entertainment

Netflix tagging: Yes, it's a real job

By **Jhaan Elker** June 11, 2015 [Email the author](#)



Josh Garrell is a Netflix tagger. His job? He is paid to watch television shows and movies for hours a day. (Jhaan Elker/The Washington Post)

Starting in the wee hours of June 12, hundreds of thousands of fans are downloading services from a new Netflix original production "Orange Is



Requirements

- 👉 **Relevance:** Items should be relevant for the user.
- 👉 **Diversity:** Dissimilar items, increases the chance to convert.
- 👉 **Item Fatigue:** Don't show the same item always.
- 👉 **Novelty / Freshness:** There is a time dependency for the recomms.
- 👉 **Serendipity:** Recomms should be both attractive and surprising to the user.



Diverse Goals

Depending on the step of the funnel you can give different kind of recommendations

- 👉 Increase avg ticket size and value
- 👉 Increase buy frequency
- 👉 Stock clearance
- 👉 Discoverability for long tail items
- 👉 Promote brands or products
- 👉 Customer Loyalty



Learn to Rank

Recommendations is a Learn To Rank problem. So we use LTR metrics:

- 👉 Precision @K
- 👉 Recall @K
- 👉 F-measure @K
- 👉 NDCG Normalized Discounted Cumulative Gain
- 👉 Coverage



Embeddings

- 👉 Since we have “Big Data”, with information provided from different sources we can train different kind of embeddings.
- 👉 Similar items: trained with content
- 👉 Complementary items: trained with context
- 👉 Users related items: trained with consumer behavior
- 👉 You can use Word2Vec, Fasttext, Bert, etc. Or even combine multiple embeddings



Personalization



- 👉 Real Time predictions.
- 👉 Bigger computational cost.
- 👉 Feature engineering -> You need a "feature store".
- 👉 Depending on the goal it could have diminishing returns.





- 👉 **Data Collection:**—We use it to make our **ETLs** to have transactional and analytics data available to train our models
- 👉 **Re Train** — We have a DAG scheduled defining each step and how to continue to performs ETLs and re train the models after a defined period.

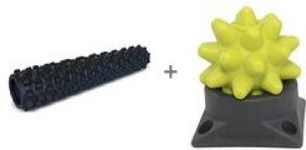


Recommendations at Amazon

- 👉 35% of Amazon.com's revenue is generated by its recommendation engine.
- 👉 Not as pure science.

Explainability

Frequently Bought Together



Total price: **\$94.90**

[Add both to Cart](#)

[Add both to List](#)

- ✓ **This item:** Rumble Roller - Textured Muscle Foam Roller Manipulates Soft Tissue
- ✓ Rumble Roller X-Firm Beastie and Base - Extra Firm Spiky Massage Ball - Come

Recommended for you, Thomas

<p>Literature & Fiction 62 ITEMS</p>	<p>Exercise & Fitness Equipment 8 ITEMS</p>	<p>Health, Fitness & Dieting Books 37 ITEMS</p>	<p>Tableware 12 ITEMS</p>
<p>Prime Video - Unlimited Streaming for Prime Members 12 ITEMS</p>	<p>Coffee, Tea & Espresso 98 ITEMS</p>	<p>Biographies & Memoirs 17 ITEMS</p>	<p>Engineering Books 7 ITEMS</p>



Peppa Pig: Fairy Tale Little Library
by Lauren Holowaty (2 September 2010)
Average Customer Review: ★★★★★ (156)

In stock

M.R.P.: —250.00

Price: **149.00**

[27 offers](#) from 149.00

☐ I own it ☐ Not interested ☒ ★★★★★ Rate this item

Recommended because you purchased **Peppa Pig: Bedtime Little Library** and more [\(Fix this \)](#)



UX Matters!

89% of properties are reserved in Amsterdam

There is also availability in these nearby destinations

Hoofddorp Schiphol Badhoevedorp

Amsterdam is 88% unavailable for your dates on our site. Have you checked these nearby destinations?

88% reserved

Schiphol 43% cheaper 11.4 km away, 8 properties	Badhoevedorp 64% cheaper 8.3 km away, 4 properties
Lijnden 60% cheaper 9.3 km away, 1 property	Amstelveen 53% cheaper 8.1 km away, 5 properties
Schiphol Airport 54% cheaper 11.4 km away, 21 properties	

Vind deals voor elk seizoen

Van een huisje op het platteland tot een hip appartement midden in de stad

Bestemming, accommodatienaam of adres:

Paris

	Le Touquet-Paris-Plage, Noord - Nauw van Calais, Frankrijk
	Paris, Ile de France, France Populaire
	Paris City Centre, Paris, Ile de France, France

Paris

Visitors love: Eiffel Tower, Musee D'orsay, Notre Dame

Paris, a magnet for lovers and their beloved; a sleepless and urbane world's art and culture connoisseurs. No trip is complete without a visit to the city of light. [Show more](#)

Aparthotel Adagio Paris XV ★★★★★ **Great Value Today**

Family friendly

20, rue d'Oradour-sur-Glane, 15th arr., 75015 Paris, France – [Show map](#) – Metro access

Filter by:

☒ Remove all filters

Your budget

<input type="checkbox"/> € 100 - € 150 per night	9
<input type="checkbox"/> € 150 - € 200 per night	15
<input type="checkbox"/> € 200 + per night	27

Popular filters

<input type="checkbox"/> Whole apartments	45
<input type="checkbox"/> Parking	36
<input checked="" type="checkbox"/> Hotels	6
<input type="checkbox"/> Very good: 8+	29
<input type="checkbox"/> Family rooms	51
<input type="checkbox"/> Paris City Centre	18
<input type="checkbox"/> Good: 7+	49
<input type="checkbox"/> Guest favourite area	35
<input checked="" type="checkbox"/> Family friendly properties	

Paris

Top reasons to visit:
museums, sightseeing, culture

Recommended destinations for you:

Amsterdam

43 properties available

Compared to the past 28 days - prices in Amsterdam are now lower for your dates.

Brussels

146 properties available

Prices in Brussels have been increasing over the past 32 days.



MVP & Iterations for Recomms

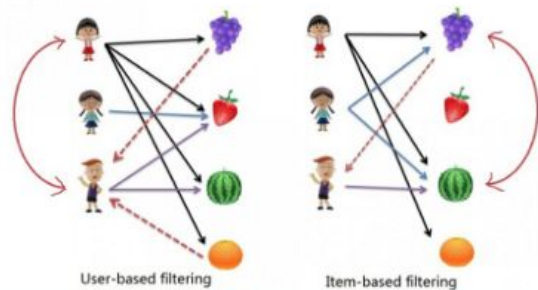
- 👉 Random, top sold, trending
 - Validate infrastructure, analytics tools, metrics,
 - A/B testing - baseline
- 👉 Cooccurrences (statistics)
 - Informed Baseline - Pareto Law
- 👉 Machine Learning
 - Simple Models - Linear Regression
 - Define Hypothesis and experiments
 - Complex models, Add other data sources
 - Offline predictions vs real time predictions
 - Deep Learning / Neural networks

- 👉 Personalization
- 👉 Validate user needs
 - Define main UX Flow,
 - UI views, interactions
 - Evaluate KPIs
- 👉 Improve UX, User Research.
- 👉 ETC!!

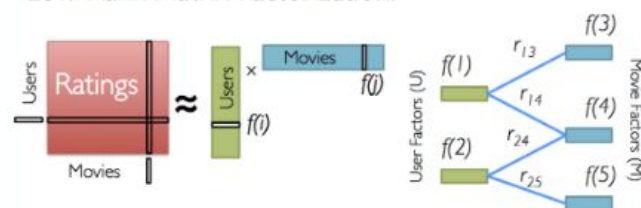


MVP & Iterations for Recommends

Not like this...

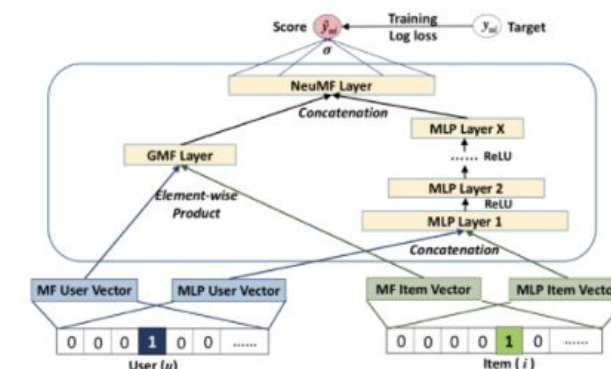


Low-Rank Matrix Factorization:



Iterate:

$$f[i] = \arg \min_{w \in \mathbb{R}^d} \sum_{j \in \text{Nbrs}(i)} (r_{ij} - w^T f[j])^2 + \lambda \|w\|_2^2$$

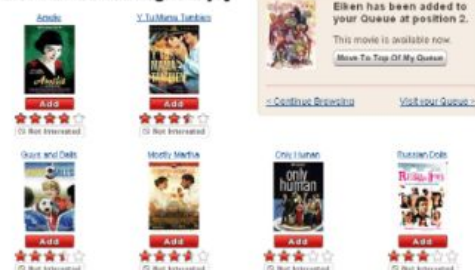


Like this!

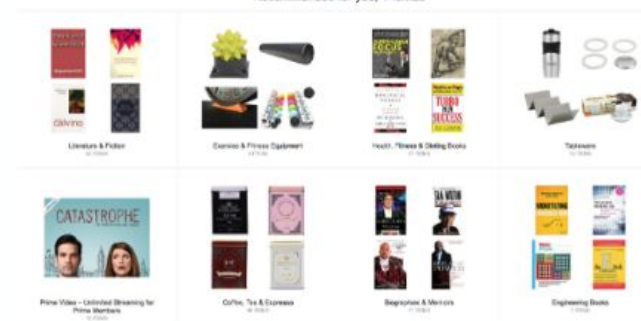
Top Products



Other Movies You Might Enjoy



Recommended for you, Thomas



¿PREGUNTAS?

👉 ¡We are hiring!

Contact me at **ariel.wolfmann@rappi.com**

Rappi